HPC at INFN: Astroparticle

Matteo Tenti
(INFN - CNAF)

04/05/2018

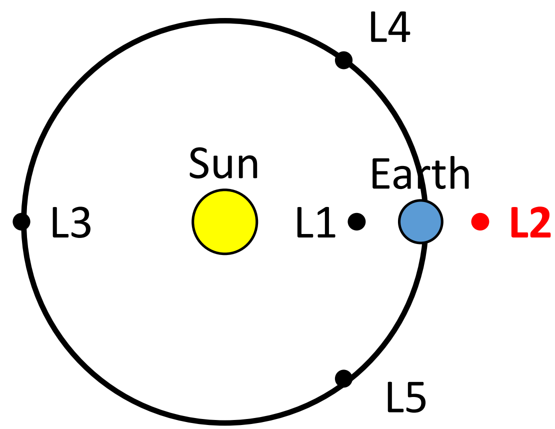
INFN and The Future of Scientific Computing – Episode I: The HPC Opportunity
Torino

Euclid *(<u>Matteo Tenti</u> & Alessandro Renzi)*

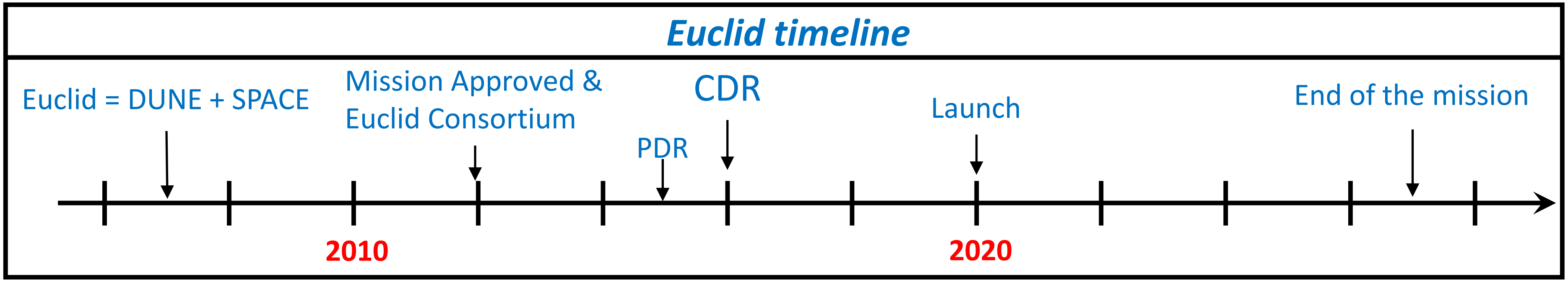
LSPE *(Francesco Piacentini)*

VIRGO *(Sebastiano Bernuzzi & Cristiano Palomba)*

Euclid mission



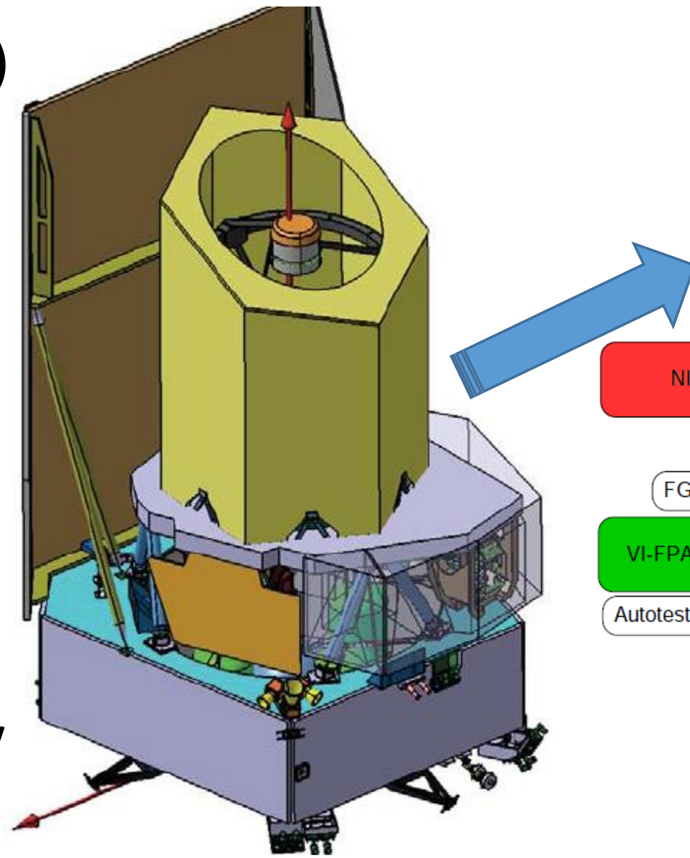
- Euclid is **medium class** mission of ESA **Cosmic Vision** program (2015 - 2025)
- **Launch: 2020** from ESA spaceport in Kourou (French Guyana) with Soyuz ST-2.1B rocket
- Orbit around the Lagrangian point **L2** of Sun-Earth system
- Nominal mission: **6 years**

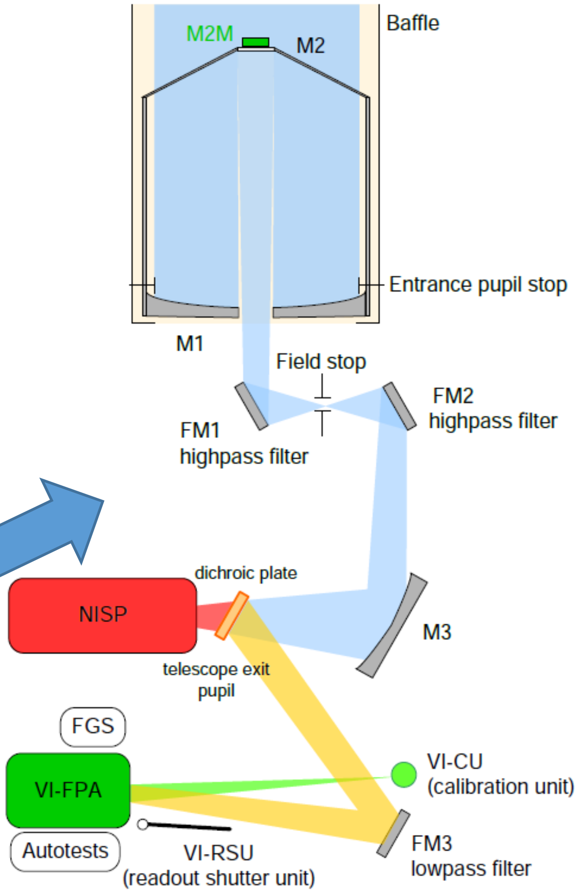


Euclid Payload

1.2m diameter telescope

- **VIS: Visual Imager (36 CCDs)**
- **NISP: Near-Infrared Spectrometer and Photometer (16 H2RG)**
- **Two wheels:**
  - **filters for photometry**
  - **grisms for spectroscopy**





Euclid Physics

- *Why*: understand the physical **origin of the accelerated expansion** of the Universe

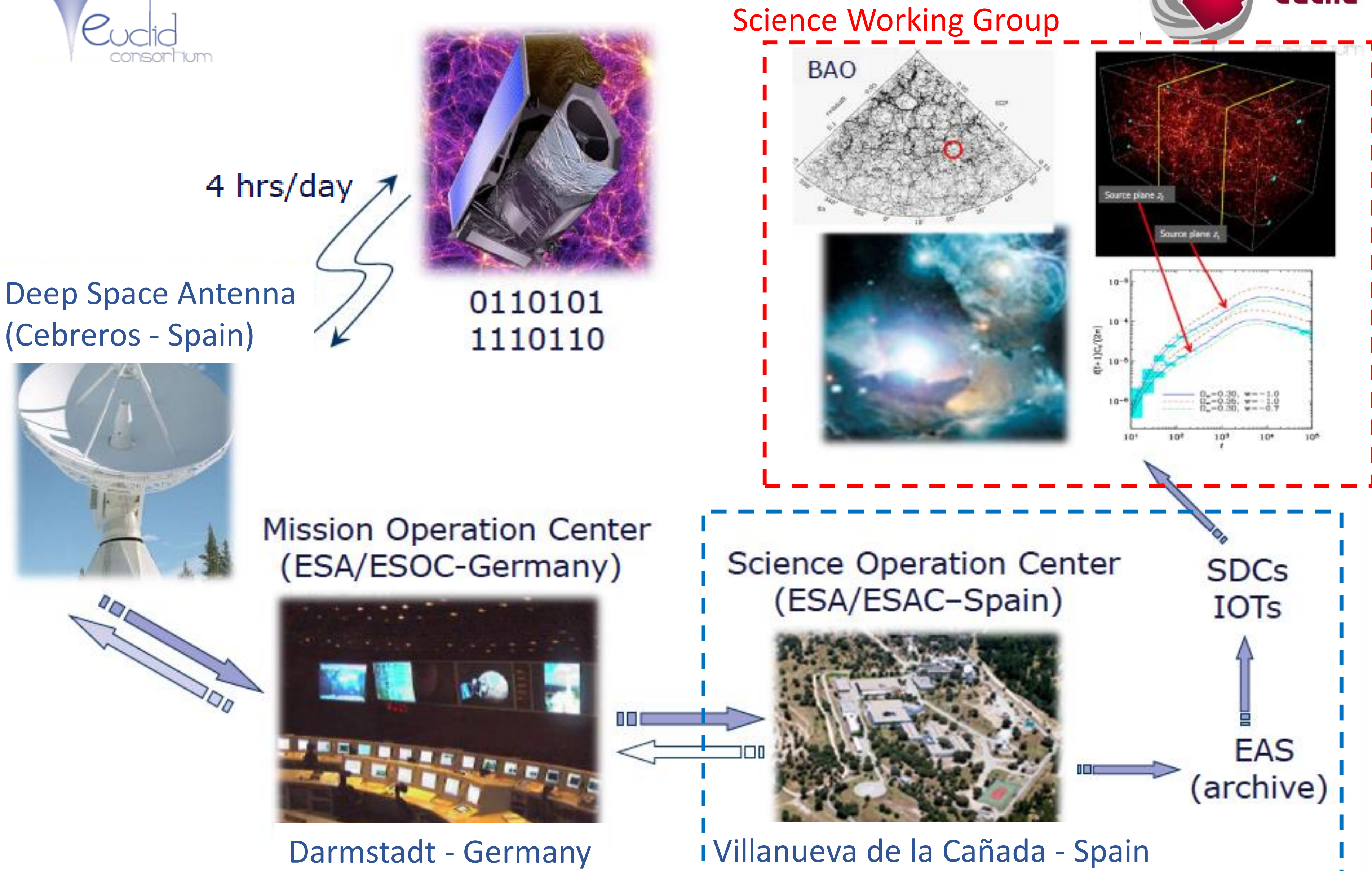
- *How*: **mapping matter** distribution at different redshifts (i.e. looking ~ 10 Gy back in time)

  - **Dark Energy equation of state** $[w = p/\rho]$:
    cosmological costant $(w = -1)$ or **scalar field?**

  - **Structures' growth rate** $[f(z) = \Omega_m(z)^\gamma]$:
    general relativity $(\gamma = 0.55)$ or **modified gravity?**

  - **Neutrino mass & effective numbers**

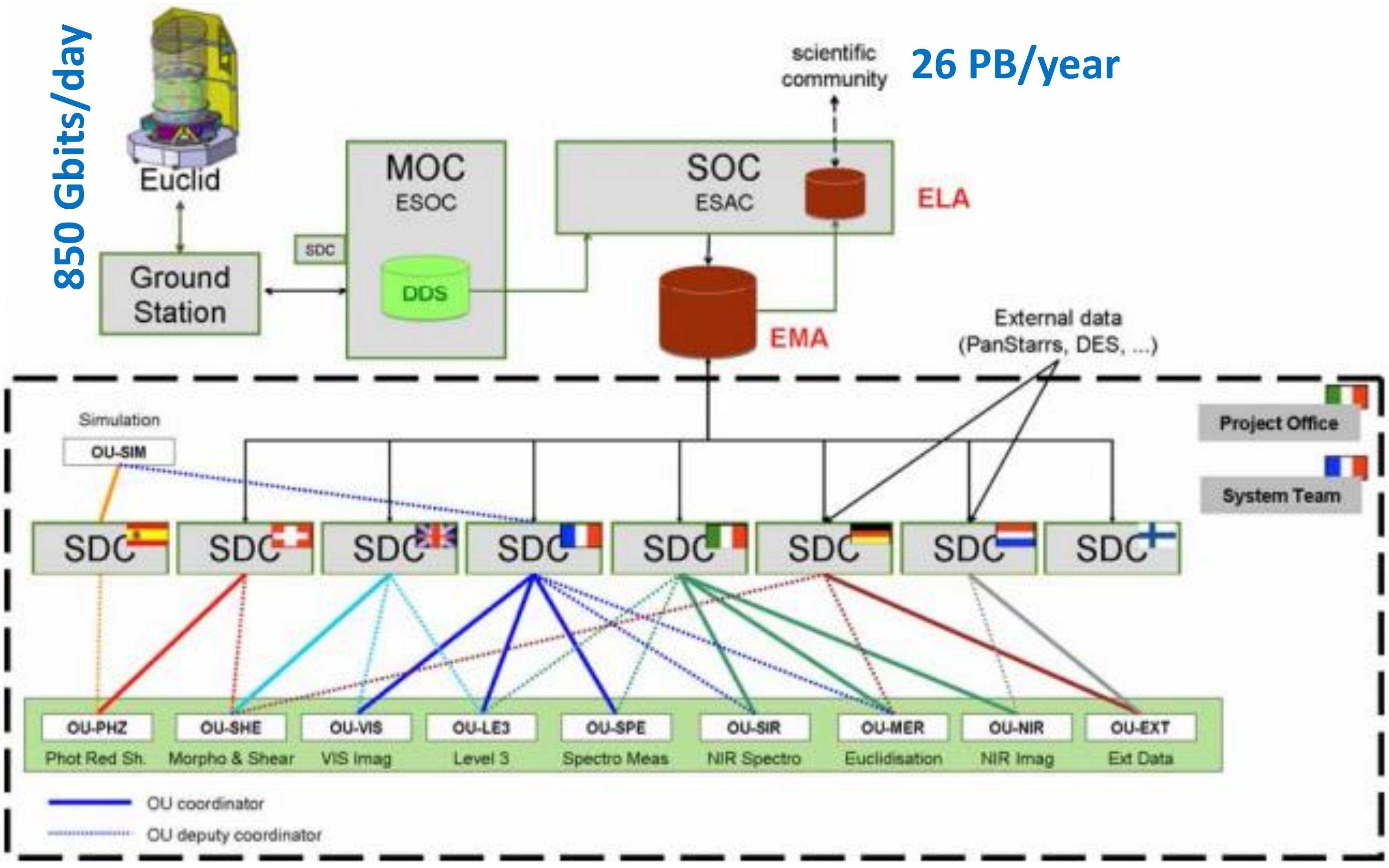
Euclid Data Flow



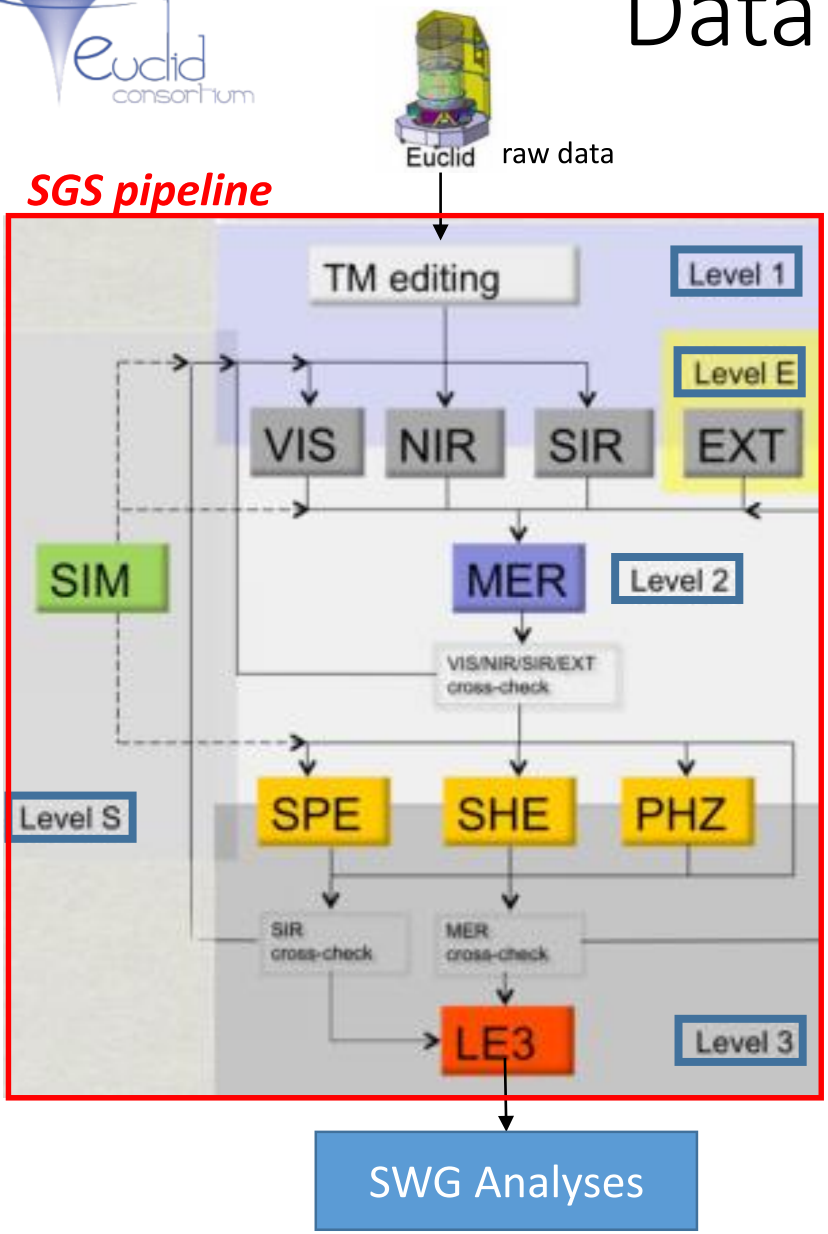
Science Ground Segment

Science Ground Segment

- Euclid data are managed by the **Science Ground Segment (SGS)**



Data Pipeline



Data are organized in different **levels**

- **Raw data** from satellite
- **Level 1**: edited telemetry
- **Level 2**: calibrated signal from instrument
- **Level 3**: physical observables (redshift, shear) reconstruction
- **Level Q**: quick release
- **Level E**: external data from ground-based experiment
- **Level S**: simulation

Alessandro Renzi

Computing resources in Cosmology: *some numbers*

	[1603.09303] PLANCK	[1603.09303] 	[1701.08158] euclid
HPC Computation (cores x hours / year)	~ 10M – 100M	**~ 1G – 10G**	**~ 100M – 1G**
Storage for Products	~ 1 PB	**~ 10P – 100P**	**> 100 PB**

With respect to Planck, next generation cosmology experiments will require:

order of magnitude in CPU-hour needs
orders of magnitude in Disk Space needs

Take-home messages

Science Working Group

Cosmology:

- Weak Lensing
- Galaxy Clustering
- Theory
- Clusters of galaxies
- CMB Cross-correlations
- Strong Lensing

Legacy:

- Primerval Galaxies
- Galaxy and AGN evol.
- Nearby Galaxies
- Milky Way
- Planets
- SNs & Transient

Cosmological Simulations

Science Working Group (SWG)

- The science activities performed by **Science Working Groups** *[using Level 3 data]*
- **INFN groups** are mainly interested in
  - Dark Energy
  - Neutrinos

Alessandro Renzi

Cosmological simulation: an example

- **Simulations** are the main tool to investigate evolution of the universe at small scale where **non linear effects become important**
- C. Carbone, M. Petkova, K. Dolag [1605.02024]
- 4 simulations with different Σm_{ν} [0, 0.17, 0.3, 0.53 eV]
- 62 sampling in time
- 3D grid dimensions: $L_{box} = 2h^{-1}$Gpc and a mesh of 4096^3 cells
- For every time sampling save a snapshot of:
  - CDM particles
  - Neutrinos
  - 3D grid of the gravitational potential
  - 3D grid of the derived gravitational potential
- Resources for **<u>one</u>** simulation:
  - **~ 1M CPU-hours**
  - **~ 90 TB**

Alessandro Renzi

Cross-correlations between CMB and large scale structure

euclid

The CMB fluctuation originates at last scattering surface when radiation and baryon matter decouple at z ~ 1000

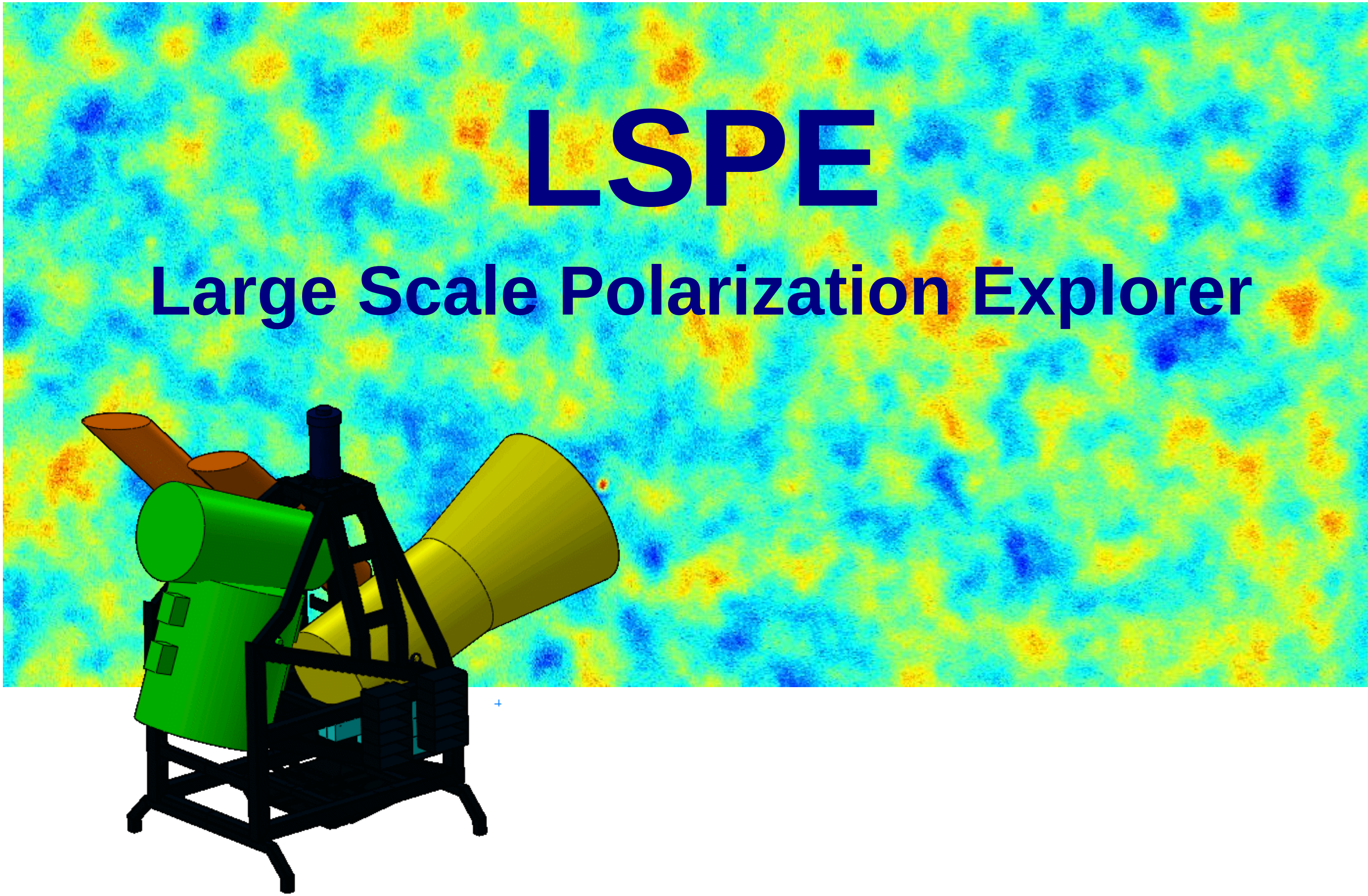
Structures formed at low redshift (z ~ 2) leave imprints on the CMB through different effects (tSZ, kSZ, ISW, gravitational lensing) which give information on cosmological parameters

- Estimated Requirement for **developing and optimization** of data analysis codes:
  - Dedicated HPC/MPI with high-speed low-latency communication
  - 1k cores (8 GB RAM/core)
  - ~ 1 PB Storage
- The use of **large HPC facilities** (like CINECA) will be soon a must

Large Scale Polarization Explorer

LSPE

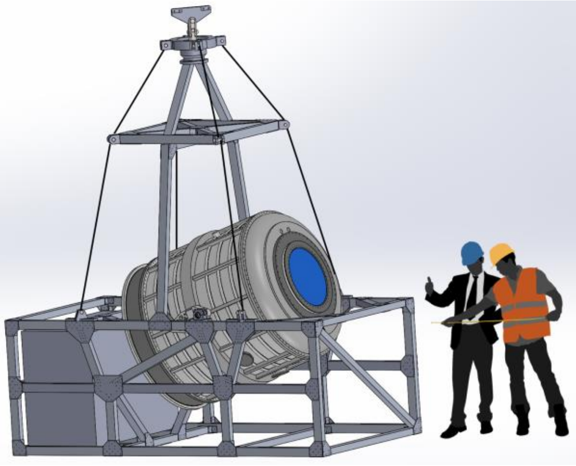
Large Scale Polarization Explorer

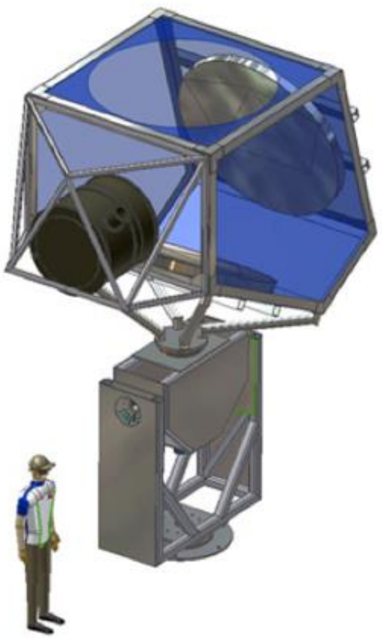


LSPE: measurement of CMB polarization

LSPE: Large Scale Polarization Explorer

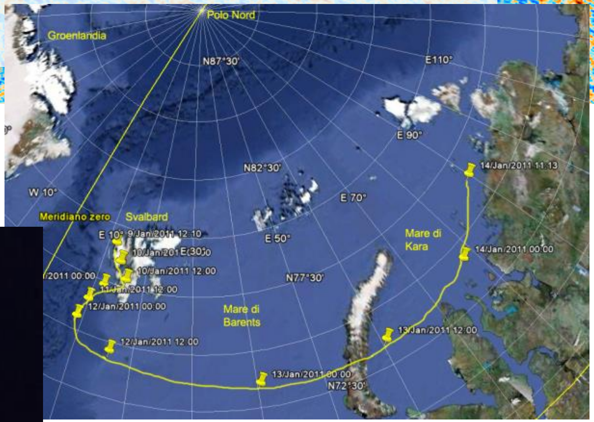
- ***Goal***:
  - Measure the **polarization of the CMB**, at **large angular scales** (30% of the sky, ~1 deg angular resolution)
  - To detect the signature of **primordial gravitational waves**, which imprint the polarization of the CMB with a particular pattern, named B-mode
  - To improve knowledge of inflationary mechanism in the early Universe
- Two ***instruments***:
  - **SWIPE**: stratospheric balloon based, launched in the polar night from Svalbard (Norway)
  - **STRIP**: ground based, in Tenerife





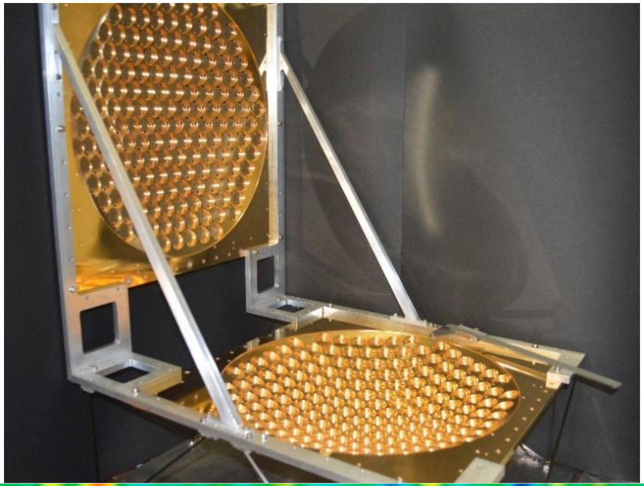
Francesco Piacentini

LSPE instrumentation / 1



- LSPE SWIPE: Winter Arctic stratospheric balloon:
  - Svalbard, Dec 2019
  - Full dark for 15 days
  - Powered by batteries
  - Microwave: 150 - 220 - 240 GHz
  - 300 mK cryogenic system
  - Spinning telescope
  - 500 mm aperture
  - 1.5 deg angular resolution
  - Covering 30% of the sky

  - INFN contribution:
    - Multimode TES bolometers
    - readout for TES bolometers
    - Cryogenic facility
    - Data analysis

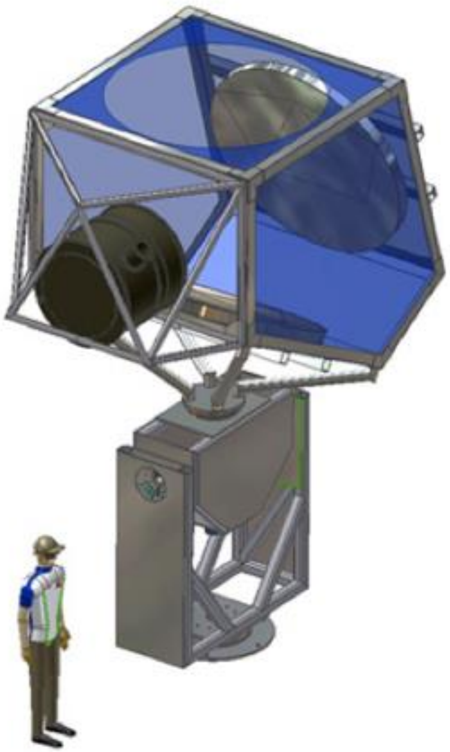


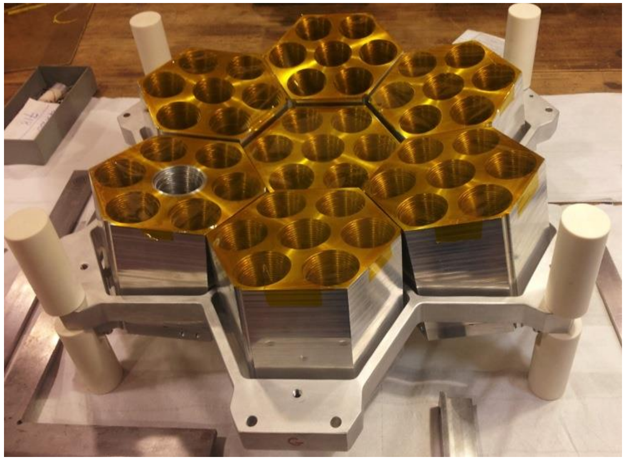
Francesco Piacentini

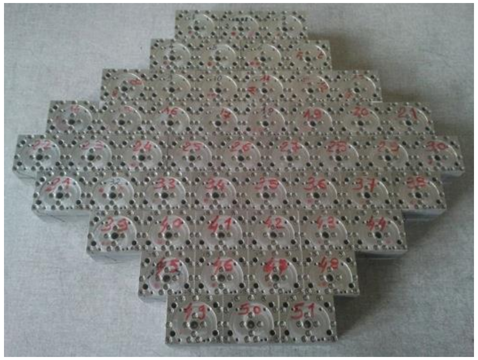
LSPE instrumentation / 2

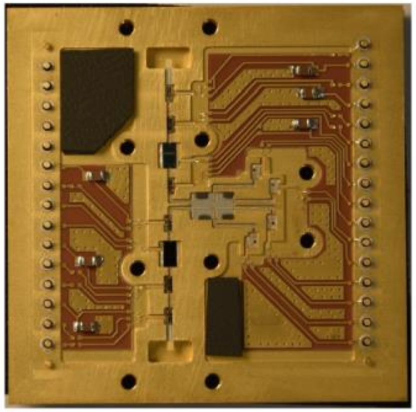
- LSPE STRIP: **Ground based**
  - Tenerife, Mar 2018
  - 2 years campaign
  - Microwave: 43 - 90 GHz
  - 20K cryogenic system
  - Spinning telescope
  - 1500 mm aperture
  - 0.3 deg angular resolution
  - Covering 30% of the sky (same as SWIPE)

  - INFN contribution:
    - Telescope mount
    - Test instrumentation
    - Data analysis







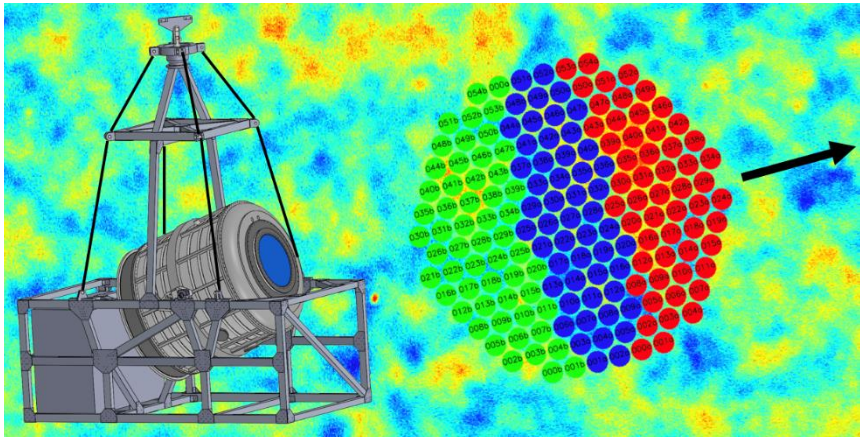


Francesco Piacentini

LSPE HPC requirements: instruments simulators

- Simulation of:
  - 326 detectors SWIPE + 49 STRIP (HPC required to run efficiently)
  - Each one sampled at 100 Hz for (~10^{10} samples)
  - Each sample is the convolution of the sky with the angular response (on a 5 degrees radius)
  - Simulation of signal, noise and systematic effects
  - Time domain Filtering effects
  - Projection to sky maps

- These are very **heavy HPC simulations**, in particular in case of Montecarlo realizations



Francesco Piacentini

LSPE Computational model and resources

- Heaviest tasks are:
  - Instrument simulator:
    - Parallel Fortran code
    - The more complex version requires **1.3 k core hr** for one simulation
    - ~100 detectors per band; 16 GB RAM for each detector simulation
    - **1000 simulations require 1.3 M core hr**
  - Mapmaking, with correlated noise
    - Parallel Fortran code
    - Can produce pixel-pixel inverse noise covariance matrix
    - 320 cores x 10 hr with 10 GB RAM per core, for one map
    - **300 maps require 1 M core hr**

Francesco Piacentini

LSPE Computational model and resources

- Heaviest tasks are:
  - CMB component separation
    - R&D to select adequate methods for LSPE
    - Not computationally demanding (~ 32 core node)
  - Power spectrum estimation (Quadratic Maximum Likelihood method)
    - MPI/OpenMP
    - Flexible (32 to 1000+ cores)
  - Cosmological parameter estimation
    - Critical piece is likelihood code
      - Starting from map pixels (32-256 cores adequate)
      - power spectrum based: minimal computational burden at likelihood level, but heavily relies on simulations and QML

Francesco Piacentini


VIRGO

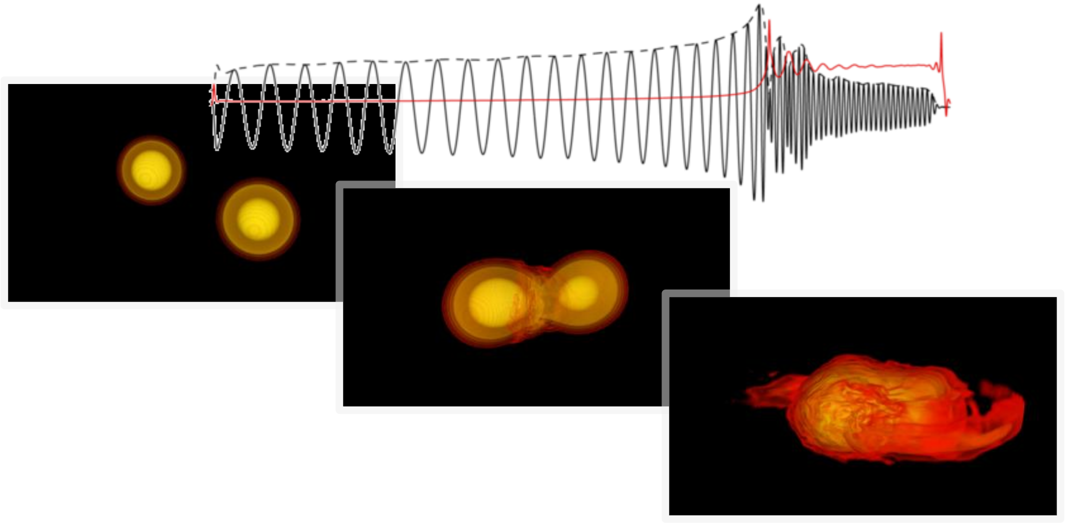
Theoretical modeling for gravitational-wave and multi-messenger astrophysics

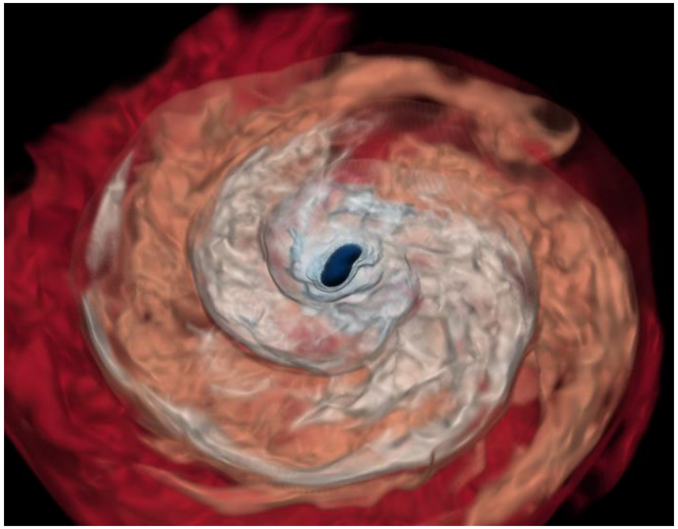
Sebastiano Bernuzzi

Extreme phenomena: **neutron stars collision**, black hole collision and supernovae explosions

- Strong and dynamical gravity, matter at supranuclear density, interactions matter / radiation at extreme conditions

Multi-physics and multi-scale 4D simulations from *first principles* (all fundamental theories)



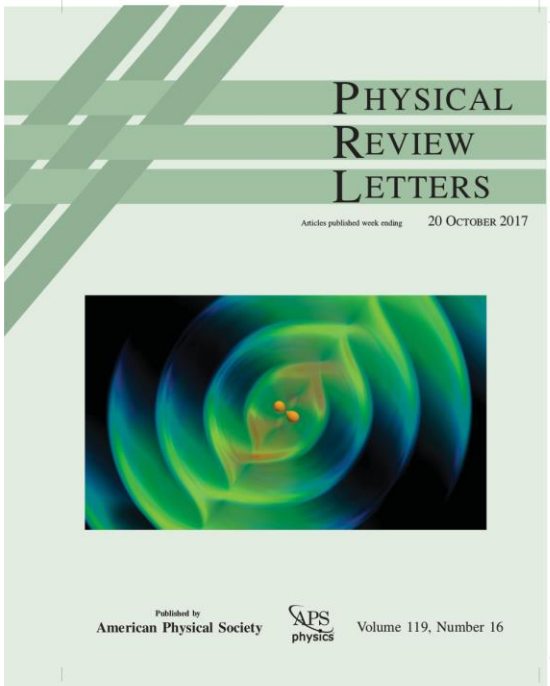


Neutron star merger simulations of the ***Computational-relativity collaboration*** *(CoRe).*
[S.Bernuzzi, B.Bruegmann, T.Dietrich, A.Nagar, D.Radice, A.Perego, W.Tichy, M.Ujevic, and others]

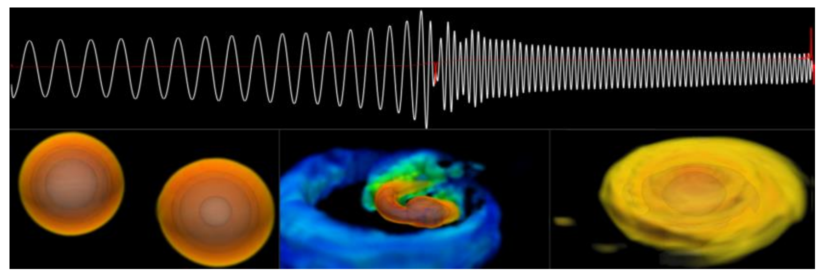
FAU FLORIDA ATLANTIC UNIVERSITY · FRIEDRICH-SCHILLER-UNIVERSITÄT JENA · INFN · Nikhef · UFABC · VIRGO

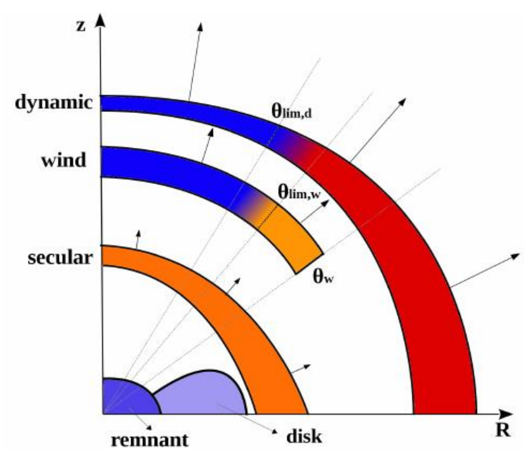
Simulations are key for multi-messenger science

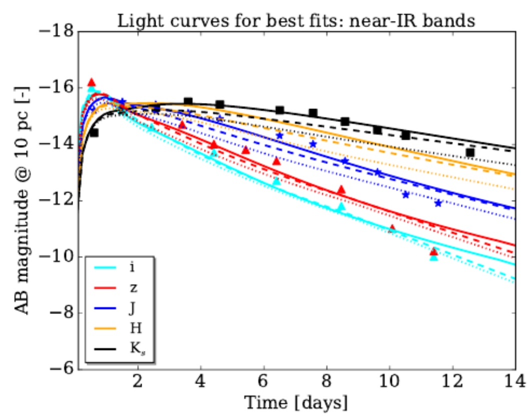
Sebastiano Bernuzzi

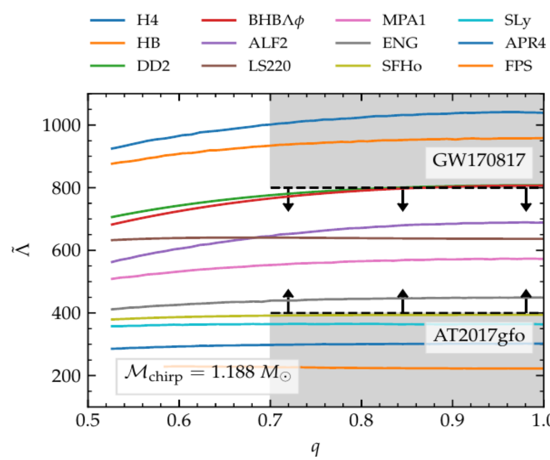


- Example: **GW170817**
- State of art gravitational-wave models
  [S.Bernuzzi, A.Nagar et al PRL 114 (2015) 161103]
- Support LIGO-Virgo data analysis with models and interpretation
- Large database of waveforms
- Understand origin of electromagnetic transient
- Develop models of counterparts





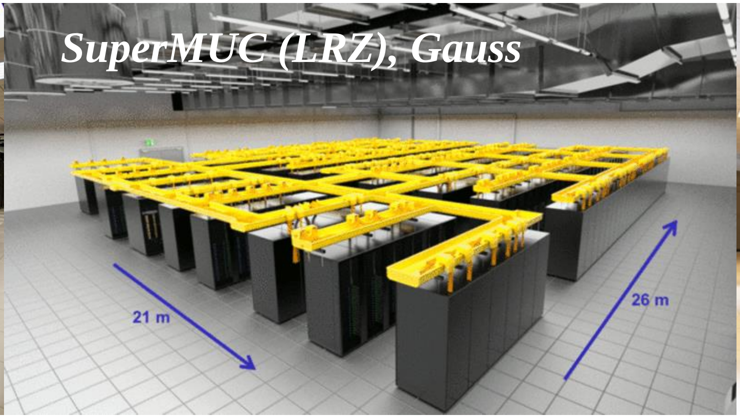




[A.Perego et al ApJL (2017) Multicomponent model of the AT2017gfo kinolova]
[D.Radice at al ApJL (2018) Joint constraint on the neutron star equation of state from GW170817 and AT2017gfo]

Simulation details & Resources

Sebastiano Bernuzzi

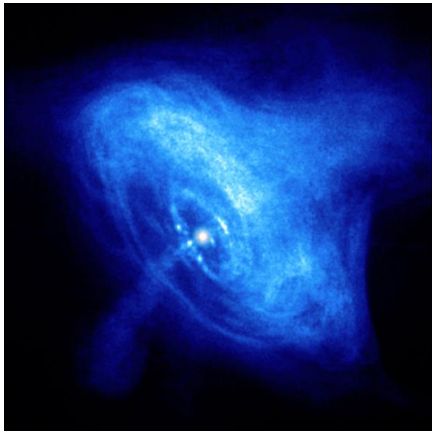


- **Large-scale HPC simulations**
- 3D + time, explicit evolution of spacetime and general relativistic hydrodynamics
- Microphysics and radiation hydrodynamics (in some cases)
- Resolve multiple scales consistently, from strong field to radiation zone in a single simulation
- Parallel adaptive mesh-based codes with hybrid **MPI/OpenMP techniques**
- **Typical simulation: 100 - 3000 CPUs** (depending on mesh resolution)
- **Runtime: weeks to months** (depending on mesh resolution)
- Explore different physical configurations
- Control results with multiple resolution simulations
- Mature production codes but also active developments of new computational strategies

Last 2 years usage: > 160 M CPU-hours on largest machines world-wide
Peer-reviewed applications through different agencies

Cristiano Palomba

Data analysis for the search of persistent GW signals from spinning neutron stars

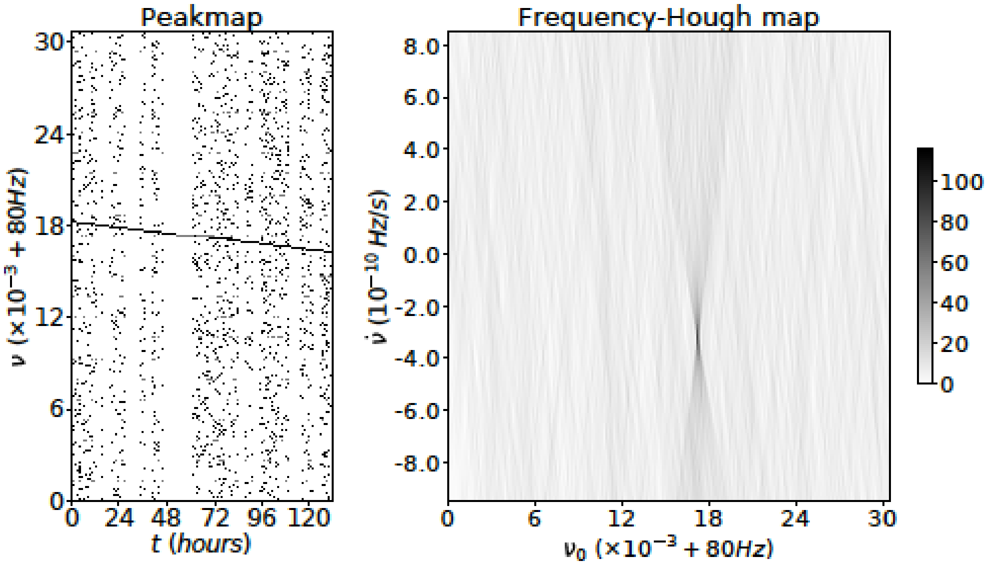


- Blind analyses explore a huge parameter space searching for signals from neutron stars with no electro-magnetic counterpart

- O(15) M CPU-hours for a typical blind search over three detectors (2 LIGOs + Virgo) and 6 months of data

- Most of the load in computing several independent Hough transforms (HT): ***"embarrassing" parallelism***

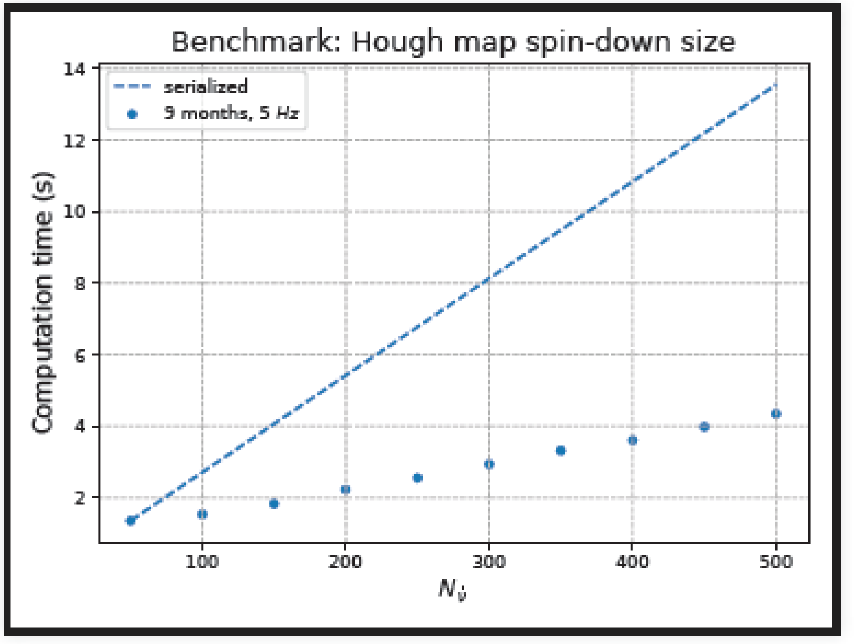
- Production analyses currently on GRID



- Porting of the HT code on **GPU** has been recently done (Iuri La Rosa's Master Thesis)

Cristiano Palomba

- Based on TensorFlow framework
- **Speed gain: ~20**
- Further improvements by better exploiting GPU parallelism
- At fixed available computing resources it will allow to make deeper searches

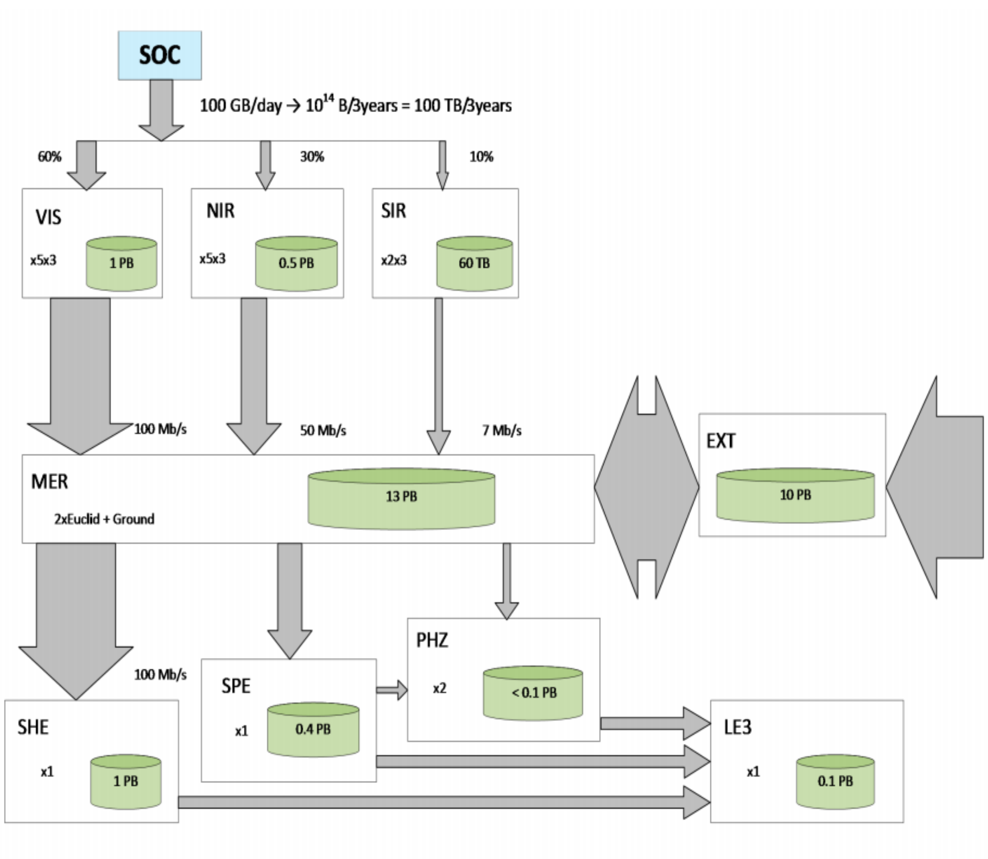


- We aim at having the new GPU code in production for O3 run (beginning of 2019). Internal LIGO-Virgo review starting today!
- Other analysis pipelines will be ported on GPU

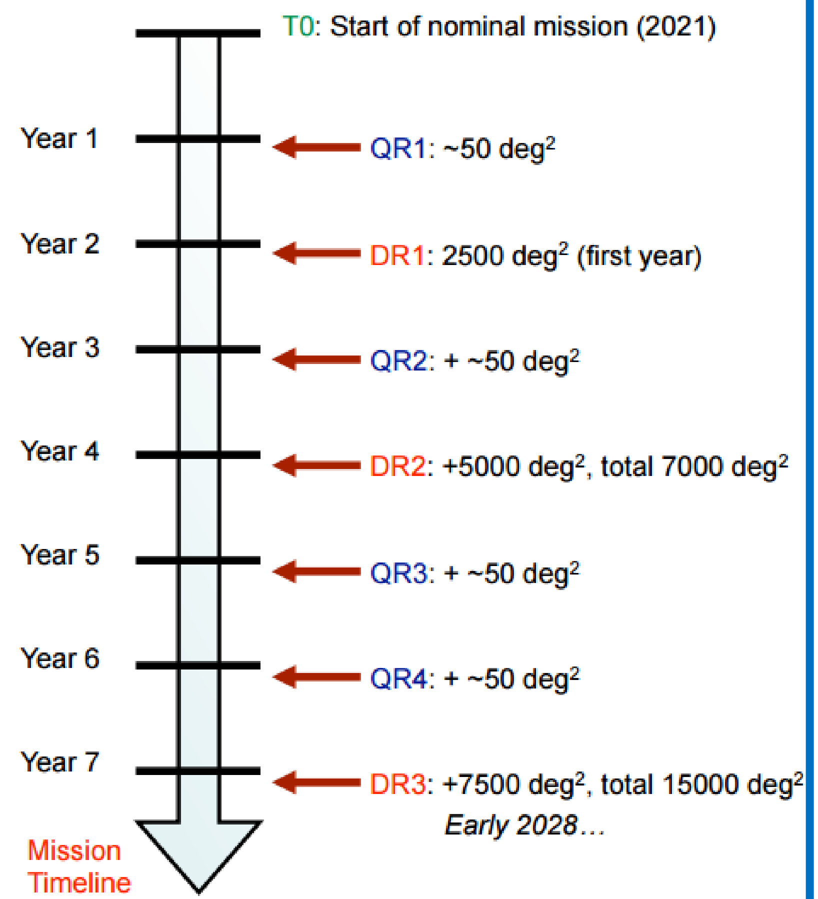
Spares

Data

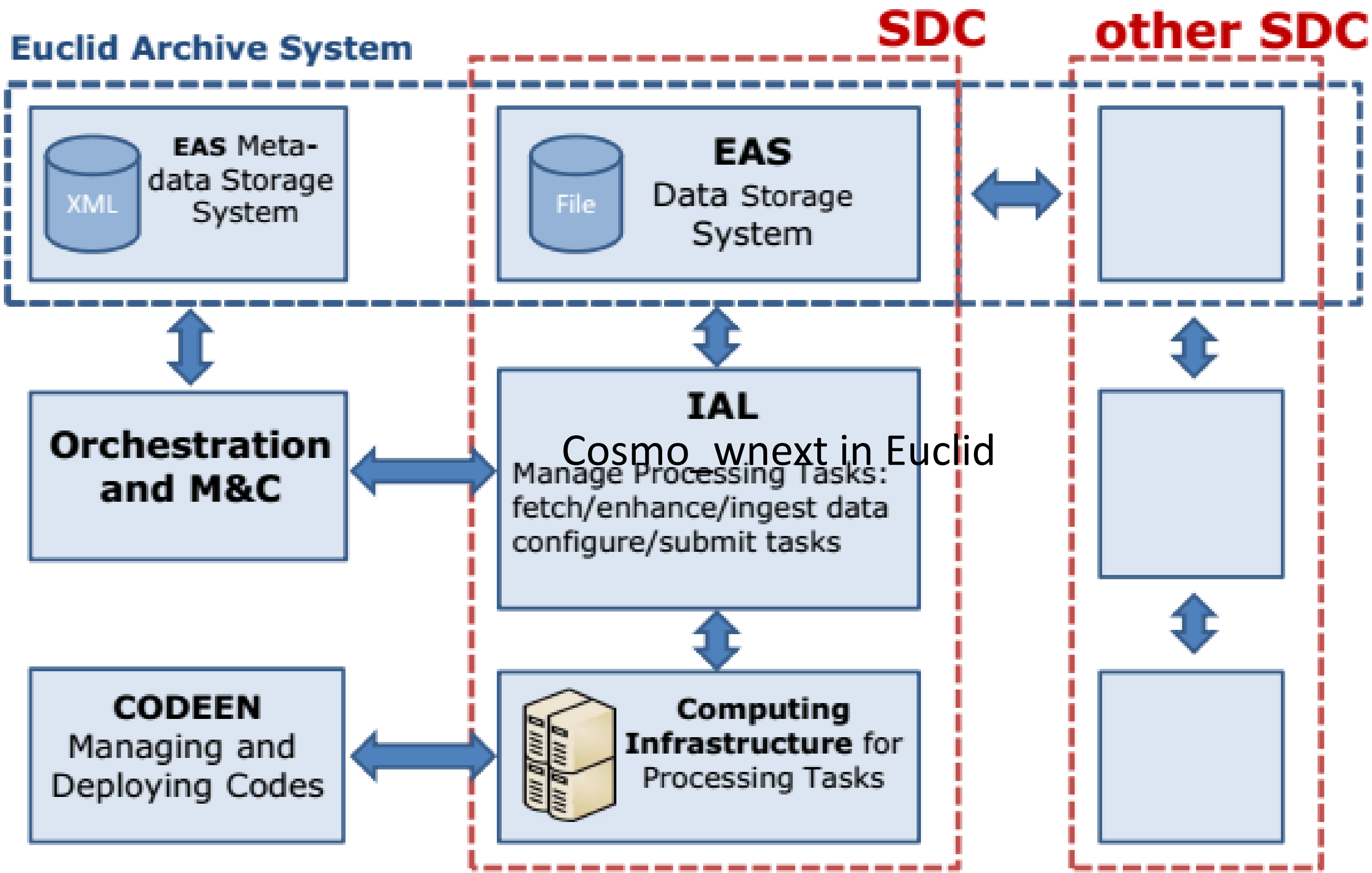
Expected data flow among PF



Planned data release



Cosmo_wnext in Euclid



Development Environment

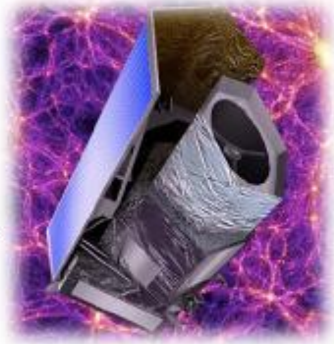
Common tools

Different domains

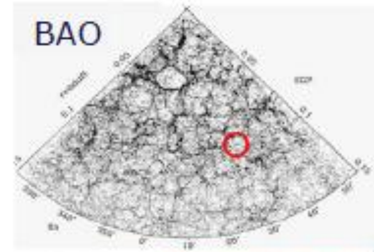
- Development environment
  - ✓ CODEEN (high-level centralized for dev/build/doc/packaging etc…)
  - ✓ EDEN (encompasses all common standards and tool for developers)
  - ✓ LODEEN (Virtual machine pre-installed with EDEN tools)
- Infrastructure environment : deployment/installation
- Scientific environment :
  - ✓ common (third parties) scientific libs
  - ✓ common scientific packages (developed once and shared among different processing functions)
  - ✓ Mission Data Base : data to be shared (physical constants, instrumental parameters, environment parameters,…)

Euclid Operations: In Flight

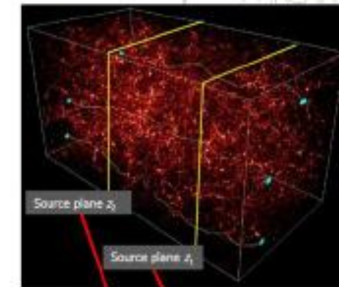
4 hrs/day

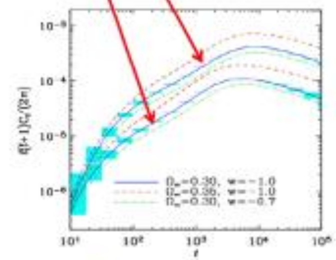


0110101
1110110



1.5 billion sources with shapes, 10 slices





Deep Space Antenna (Spain

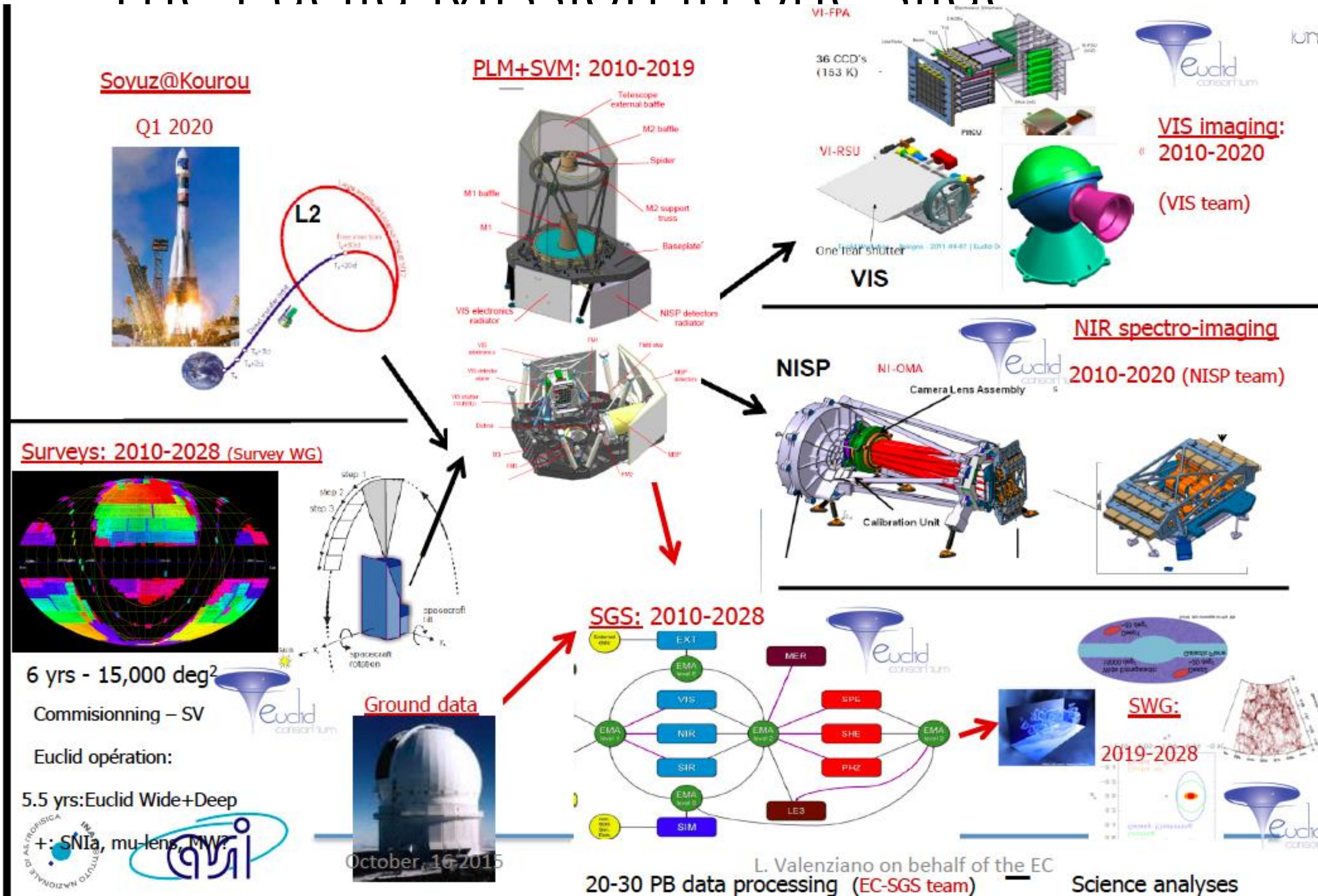
Mission Operation Center (ESA/ESOC-Germany)

Science Operation Center (ESA/ESAC–Spain)

SDCs
IOTs

EAS (archive)

The Euclid Mission in one slide



SDC - Italia

La SDC italiana si basa sul cluster DHTCS a INAF-OA_TS.
Si tratta di un cluster HPC basato su Linux con circa 800 core e 8 GB di RAM per core.
I nodi di elaborazione sono collegati fra loro e al GPFS (400 TB di storage parallelo) con 40 Gbps InfiniBand.

Secondo le stime SGS esistenti, la prevista rampa in tempo di risorse fornite è tale da avere 7 kcores e 35 PB di storage entro il 2027.

Data size

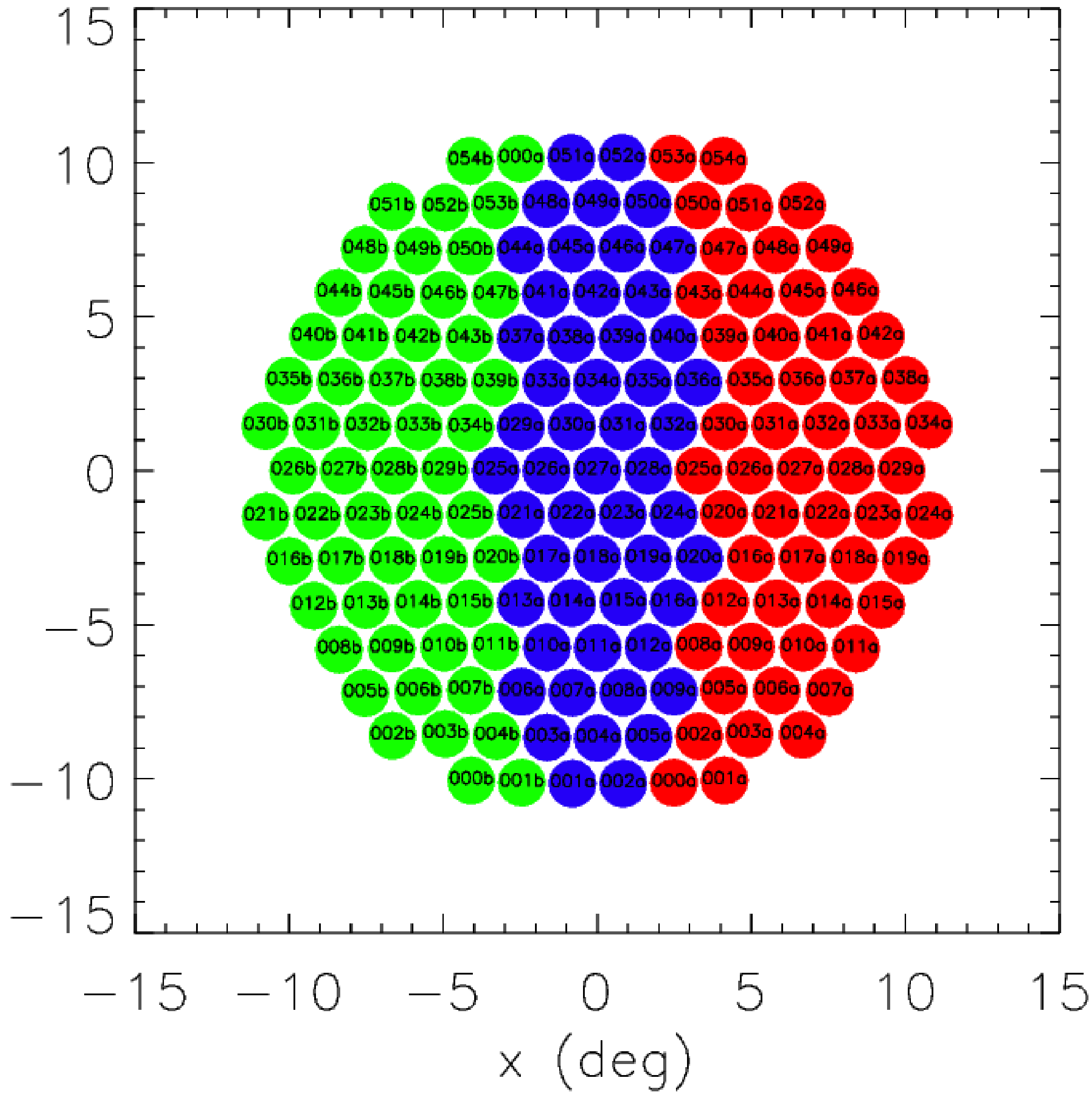
Data type	Parameter	Value	Unit
TOD	Sampling	100	Hz
	detectors	330	
	duration	15	gg
	data	$\sim 4.3 \cdot 10^{10}$	
	byte per data	4	
	Total	117	GB (INT)
ATTITUDE	sensors	4	
	data	$5.2 \cdot 10^{8}$	
	byte per data	4	
	Total	2	GB (INT)
MAPS	nside	256	
	Map pixels	786432	
	pixel_TQU	2359296	
	coverage	0.3	
	data_map	707788.8	
	Total	5.66	MB (double precision)
Covariance†	nside	64	
	Map pixels	49152	
	pixel_TQU	147456	
	coverage	0.2	Useful sky fraction
	Covariance matrix	$\sim 8.7 \cdot 10^{8}$	
	Total	6.957	GB (double precision)
Monte Carlo	realizations	1000	
	tod signal	171	TB
	tod noise	171	TB
	maps signal	5.66	GB (double precision)
	maps noise	5.66	GB (double precision)

† This is the noise covariance matrix of the map, expressed in pixel space.

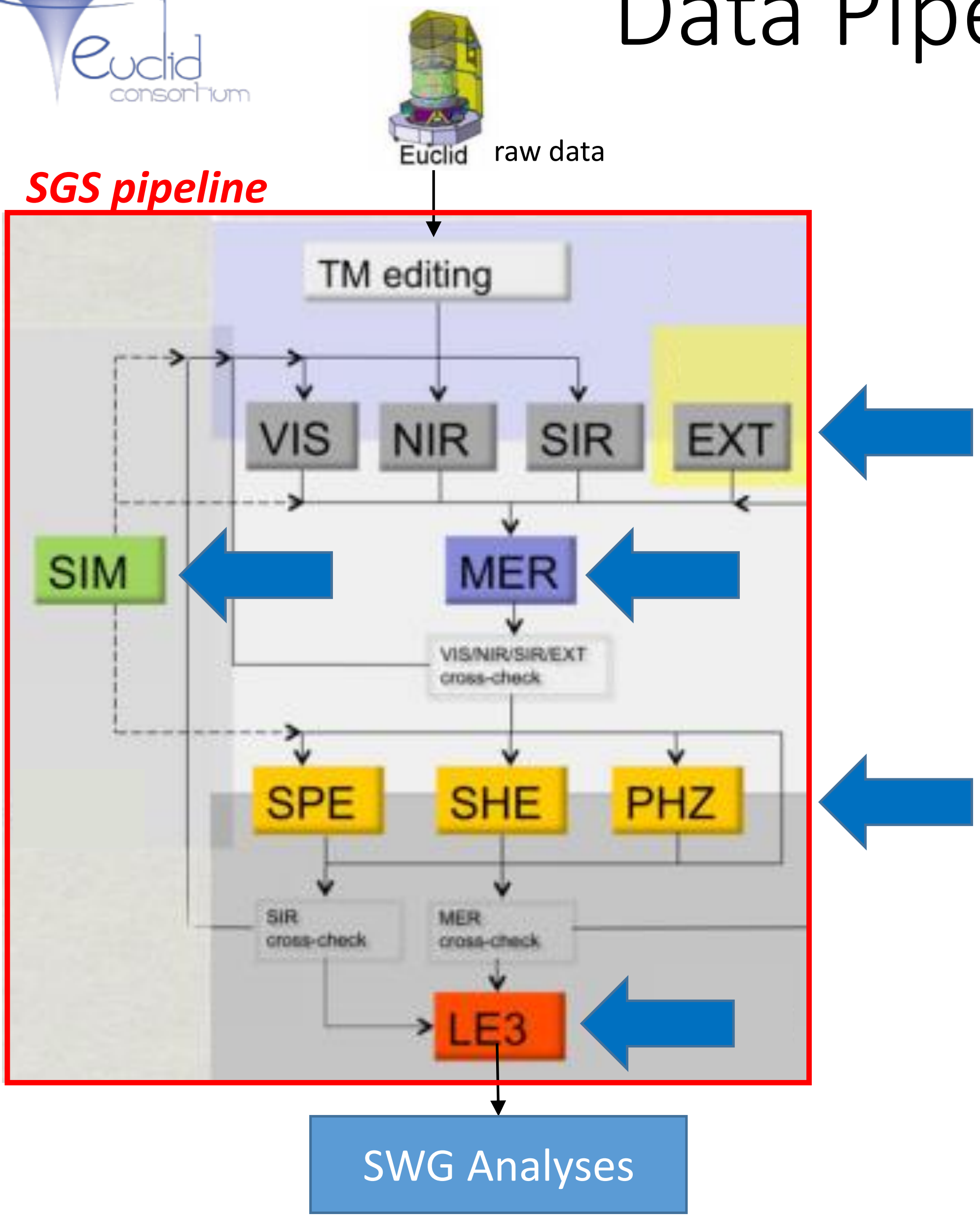
SWIPE - Focal plane

- **2 focal planes**
- **Total: 163*2 = 326 detectors**

- **140GHz: 55*2 = 110 detectors**
- **220GHz: 56*2 = 112 detectors**
- **240GHz: 52*2 = 104 detectors**



Data Pipeline



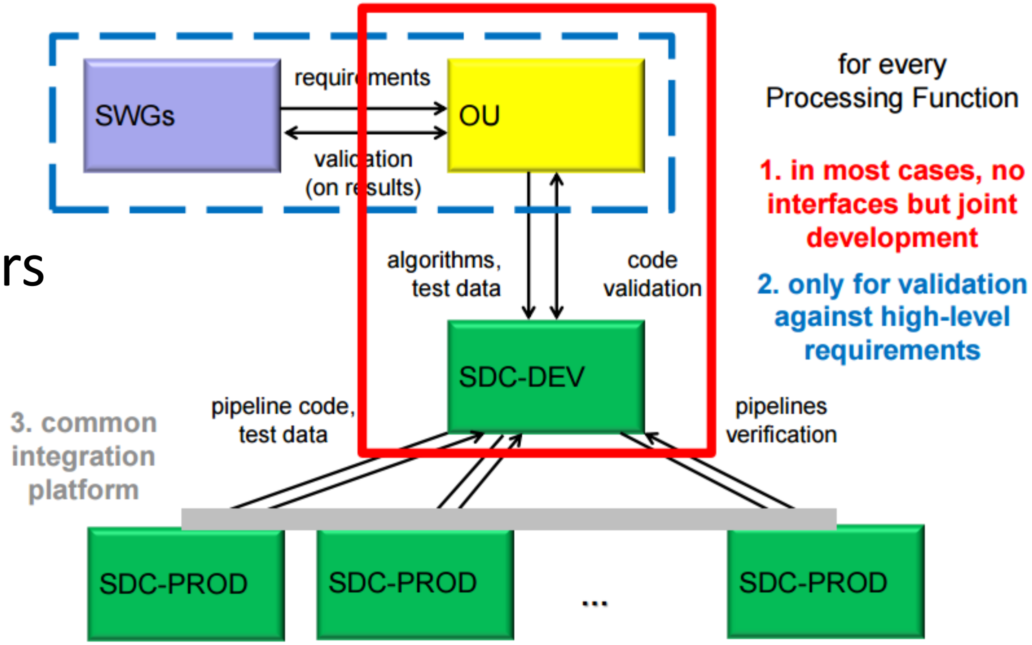
- Pipeline elements are the **Processing Functions** (PF), each having a specific task

- Each PF is mainly (*but not exclusively*) related to one Science Data Centers (SDC)

Processing Functions

- **Designed** by Organization Units according to the requirements from Science Working Groups

- **Developed** in collaboration between the Organization Units and Science Data Centers



- **Integrated** and **Run** by the Science Data Centers

- **Continuous development and implementation** during the entire mission

Architecture Concepts

- **Data-centric** information system:
  - ***Move the code***, *not the data*
- **Centralised** information repository
  - *separation of metadata from data*
- **Distributed** data and processing
  - *each SDC is both a processing and a storage node*
- **No dedicated** SDC
  - *each SDC runs the same code through* ***virtualization*** (CernVM, Docker, …)
- **MapReduce** model: **µpipelines** lower level of processing
  - *operates on the minimal processable set of data covering a given sky area*
- Requirements: **Robust**, **Reliable**, **Scalable**, **Maintainable**

Logical Architecture

- A **Euclid Archive System** (EAS)
  - *Central Metadata Repository*
    inventories, indexes and localizes distributed data
  - *Distributed Storage over the SDCs*
    balance between data availability, data transfers and redundancy
- A set of services (middleware) **decoupling** SGS components
  - *e.g. metadata query and access, data localization and transfer, data processing M&C, …*
- An **Infrastructure Abstraction Layer** (IAL)
  - *decouple data processing software and underlying IT infrastructure*
- **Monitoring&Control** (M&C) **and Orchestration** layers
  - *responsible for distributing data and processing among the SDCs.*

Work Packages

Cosmology WGs:

- Weak Lensing
- Galaxy Clustering
- Theory
- Clusters of galaxies
- CMB Cross-correlations
- Strong Lensing

Legacy WGs:

- Primerval Galaxies
- Galaxy and AGN evol.
- Nearby Galaxies
- Milky Way
- Planets
- SNs & Transient

Cosmological Simulations WGs

Science Working Group (SWG)

Each WG task is divided in several **Work Packages**.

INFN groups, beside instrument commitments, are currently involved in the **CMB cross-correlations WG** *(more if other groups, hopefully, join in)*

CMB Cross-correlations

Work Packages

- WP1: Integrated Sachs-Wolfe (ISW)
- WP2: Rees-Sciama
- WP3: Lensing Extraction
- WP4: CMB lensing likelihood
- WP5: Cross-Correlation estimators
- WP6 SZ maps and Euclid Cluster Cross-Correlation
- WP7: Validation

COSMO_WNEXT

- **COSMO_WNEXT** is an INFN GRII research program, which includes activities in two ESA space missions (Euclid & Planck)
  from Bologna and Padova Divisions : ~ 20 people

- In Euclid
  - responsibility at instrument level (AIV Warm Electronics)
  - Data analysis & simulation:
    - *mildly non-linear large scale structure simulations*
    - *estimation of the tomographic cross-correlation power spectra between CMB and density fields derived from Euclid galaxy*

- Necessity of HPC resources (possibly at **INFN-CNAF**):
  - ***O(1k) cores (physical)***
  - ***Filesystem GPFS ~400 TB***
  - ***Infiniband 40 Gbps (nodes and storage)***

Cosmology HPC needs forecast

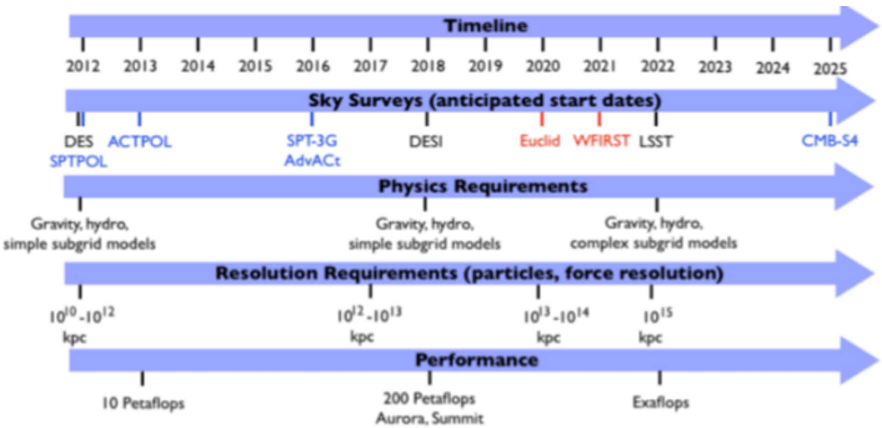


Figure 7: *Timelines for cosmological surveys (blue: CMB, black: optical, ground-based, red: optical/NIR satellite) and supercomputing resources as well as simulation requirements. DOE HEP plays a major role in the ground based optical surveys and in some of the CMB surveys.*

Code: TOAST	Column 1: Current Usage	Future Usage: 2020 (As a factor of column 1)	Future Usage: 2025 (As a factor of column 1)
Computational core hours (Conventional)	O(100M)	30x	1000x
Memory per node	0.1 GB	30x	1000x
Aggregate memory	1 TB	30x	1000x
Data read and written per run	Read: 10 TB Write: 500 TB	Read: 30x Write: 3x	Read: 1000x Write: 10x
Maximum I/O bandwidth needed	0.1 GB/sec	10x	10x
Percent of runtime for I/O	25%	0.1x	0.1x
Scratch file system space needed	500 TB	3x	10x
Permanent online data storage	5 PB	3x	10x
Archival data storage needed	50 PB	3x	10x

Cosmology HPC needs

	Planck (see 1603.09303)	CMB-S4 (see 1603.09303)	Euclid (see 1701.08158)
Core-Hours	~ 10-100M/year	~ 1-10G/year	~ 100M-1G/year
Storage Disk Required (Products)	~ 1PB	~ 10-100 PB	< 100 PB
Products	# MC Maps ~ 10^6 ~ 400MB/Map	#MC Maps ~ 10^8 (considering ~ 400MB/Map)	1.5G galaxies (estimated >33GB); 30M spectra (?dim)

Take-home messages

With respect to Planck satellite mission (at least):

1 order of magnitude in CPU-hour needs
2 orders of magnitude in Disk Space needs

Moreover:
Scientific analysis of data will require many medium-size N-body simulations

An example of Medium-size N-body for Euclid

See the paper: C. Carbone, M. Petkova, K. Dolag https://arxiv.org/abs/1605.02024

- Four different simulation with Σm_ν = (0, 0.17, 0.3, 0.53) eV
- Have been run on the Fermi supercomputer at CINECA employing ~ **1M CPU-hours per simulation**
- 62 “points” in time
- For every time point are saved snapshot of:
    - CDM particles,
    - neutrino particles,
    - 3D grid of the gravitational potential,
    - 3D grid of the derived gravitational potential

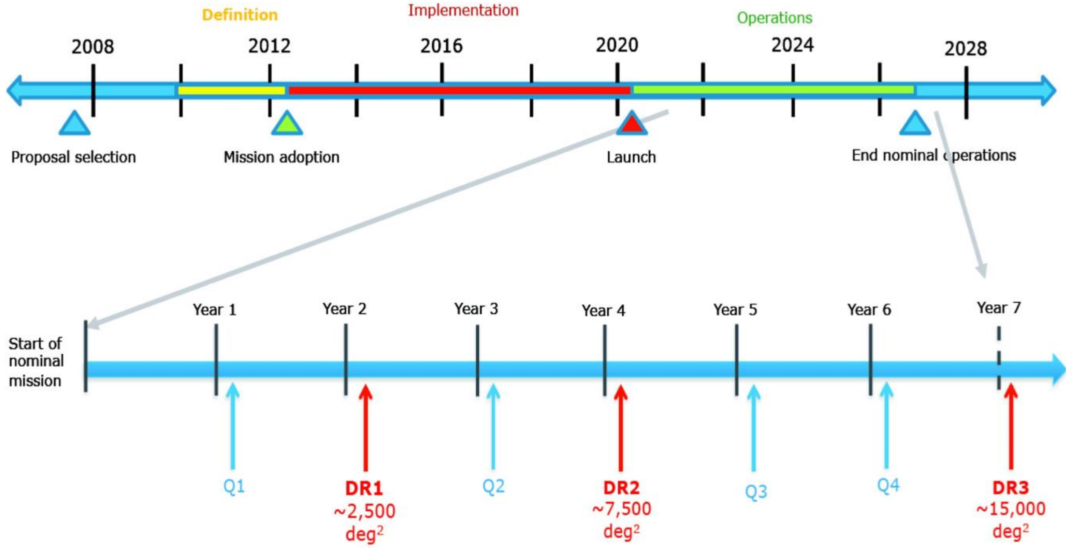
[the 3D grid have dimensions: Lbox = $2h^{-1}$Gpc and a mesh of 4096^3 cells]
with a total size of **~ 90 TB per simulation**

Total (only to produce the simulations):

- 4M CPU-hours
- 360 TB of space

NOTE: those are **Medium-size** simulations, to study neutrino

Euclid timeline



→ Present requirement (**for developing and [parzial] optimization** of the data analysis codes):

Dedicated HPC/MPI with high-speed low-latency communication
1000 cores
8 GB RAM/core
~ 1 PB Storage

→ It is critical a reliable **long term large Disk Storage space**

→ High bandwidth connection with Euclid HPC facilites (only part of data are needed at a time)

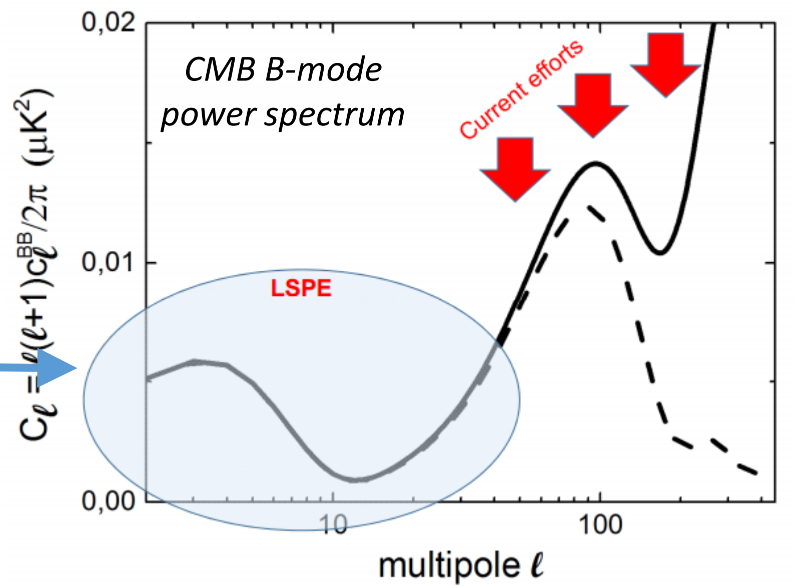
→ The use of **large HPC facilities** (like CINECA) it will be soon a priority!

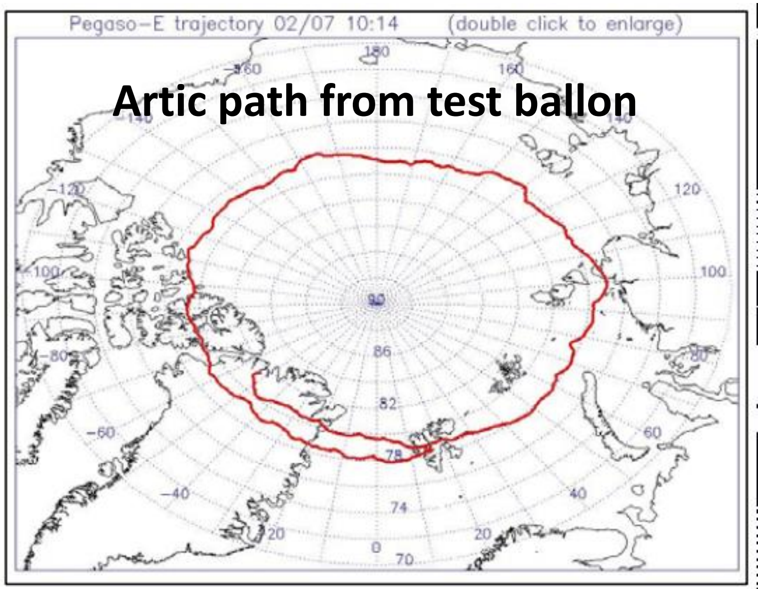
Groups involved in data analysis

WP	Task	Participants
0	Coordination	RM1
1	Instrument simulator SWIPE	RM1, FE, RM2, PI
1	Instrument model SWIPE	RM1, PI, GE, RM2
1	Focal plane definition SWIPE	RM1
1	Mission planning SWIPE	RM1
2	Calibration SWIPE	RM1, PI, GE
2	Attitude and pointing SWIPE	RM1
2	Calibration SWIPE	RM1, FE, RM2, PI, GE
2	Polarimetry SWIPE	RM1, GE
2	Data cleaning SWIPE	PI, RM1, RM2, GE
1b	Instrument simulator STRIP	UniMI
1b	Instrument model STRIP	UniMI, UniMIB
1b	Focal plane definition STRIP	INAF-BO, UniMI
1b	Mission planning STRIP	UniMI, SISSA
2b	Calibration STRIP	INAF-BO, UniMI, UniMIB
2b	Attitude and pointing STRIP	INAF-BO, OATs, UniMI
2b	Calibration STRIP	INAF-BO, UniMI, UniMIB
2b	Polarimetry STRIP	INAF-BO, UniMI, UniMIB
2b	Data cleaning STRIP	UniMI, SISSA
3	Noise estimation	RM1, RM2, FE, OATs, MI, MIB
3	Map-making	RM2, FE, PI, RM1, OATs
3	Monte-Carlo	RM2, FE, PI, OATs
4	Component separation	FE, RM2, SISSA
4	Power spectra estimation	FE, RM2
4	Likelihood	FE, RM1, RM2, PI
4	Consistency tests	RM1, FE, GE
5	Cosmological parameters	RM1, FE, RM2, PI
6	Foreground characterization	SISSA, RM2, UNIPD
6	Secondary science	RM1, GE, SISSA, UNIPD

LSPE

- Large Scale Polarization Explorer:
  - measure the **polarization of CMB** at large angular scales
  - targeting the **reionization peak of B-mode** power spectrum
  - spinning (3 rpm) **stratospheric balloon** payload
  - flying long duration (15 days) in polar night
- Frequency coverage: **40 – 250 GHz**
  (5 channels, 2 instruments: STRIP & SWIPE)
- Sky coverage: **25% of the sky**





LSPE Data flow

Large Scale Polarization Explorer

Inputs

Algorithms

Products

Data rate: ~ 400kb/s stored on solid state disk on board

Sky signal

Systematic effects

Instrument

After payload rescue

Data (timelines)

Calibration, noise estimation, data cleaning

Map-making

Maps

Components separation

Monte-Carlo: uncertainties and bias

CMB, likelihood, cosmological parameters

LSPE
Large Scale Polarization Explorer

Simulation & Analysis

- **Instrument simulator** is essential:
  - in the pre-launch phase to control of systematic effects
  - in post flight analysis to de-bias the cosmological estimators from systematic contributions
- The simulator is a parallel (**MPI**) code. One detector (326 in SWIPE) per task. **16 GB RAM** needed per detector.
- Data size produced by simulation are dominated by the signal and noise time ordered data (TOD) simulation, equal to **342 TB**
- **Data Analysis**: Map-making and power spectrum estimation
- Needed HPC facility, preferably at CNAF:
  - ~ 650 physical cores
  - ~ 350 TB GPFS
  - Infiniband 40 Gbps (node and storage)
  - Currently simulator runs in «opportunistic mode» at NERSC HPC facility

Science Working Group

Cosmology:	Legacy:	
Weak Lensing Galaxy Clustering Theory Clusters of galaxies CMB Cross-correlations Strong Lensing	Primerval Galaxies Galaxy and AGN evol. Nearby Galaxies Milky Way Planets SNs & Transient	**Cosmological Simulations**

Science Working Group (SWG)

- The science activities performed by **Science Working Groups** *[using Level 3 data]*