Experimentalist’s affairs

Calibrations, statistics, data interpretation

Tommaso Chiarusi - INFN Sezione di Bologna

KM3NeT

INFN

Summary

Part 1: general concepts useful for data taking: **from signals to event**

Part 2: mandatory operations to adjust the data: **calibrations**

Part 3: looking at the sky: **celestial coordinates**

Part 4: interpreting data: **a touch of statistics**

PART 1
Approaching data analysis

Data-analysis means “squeezing” any record taken from our experiment to extract “juicy” information on several Physics cases.

In principle, neutrino telescopes are made for running for decades, 24h/day, detecting optical and acoustic signals coming from every direction (4π).

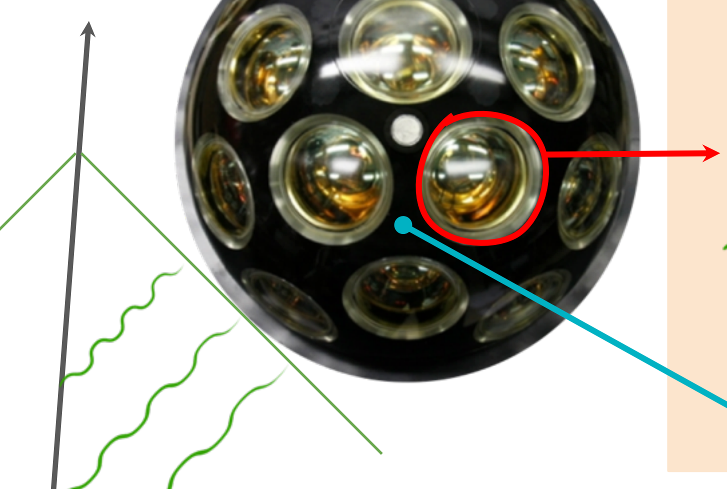
The observer (i.e. we!) sets his own point-of-view to define the search.

This determines the playground where to build the data-analysis context:
defining the analysis by mapping the expected ***signal patterns*** onto the data.

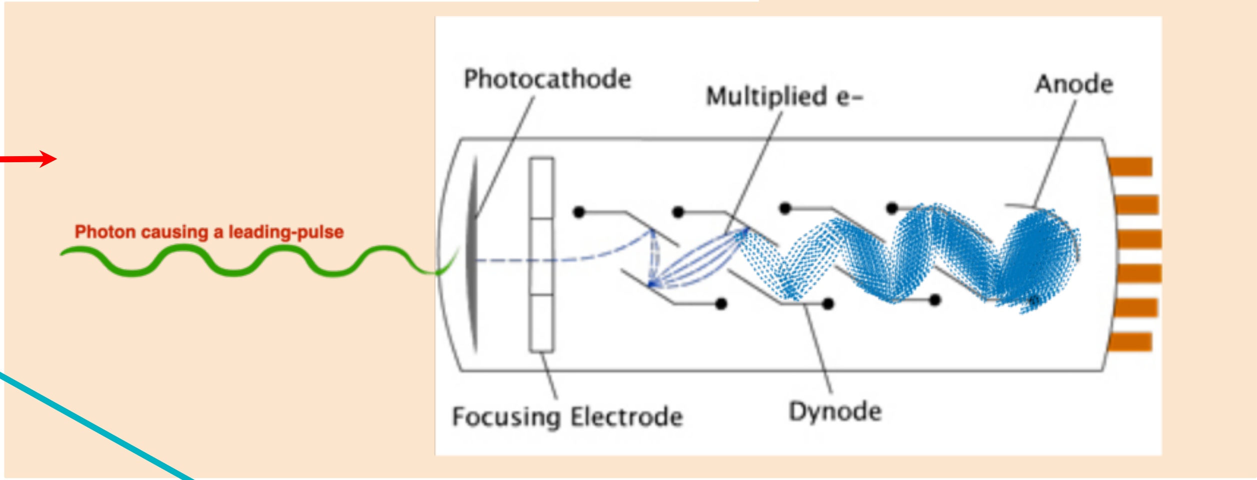
KNOW YOUR DATA !

From the signal to the hit

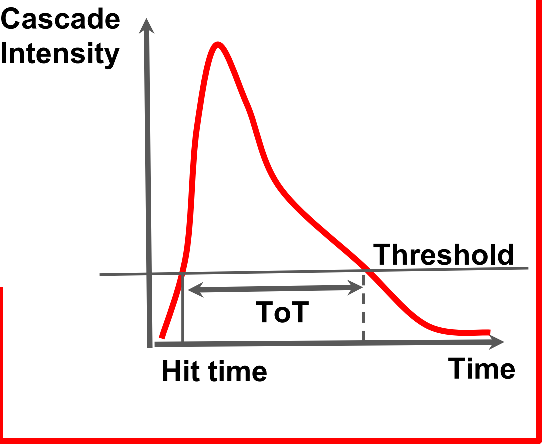
THE DOM



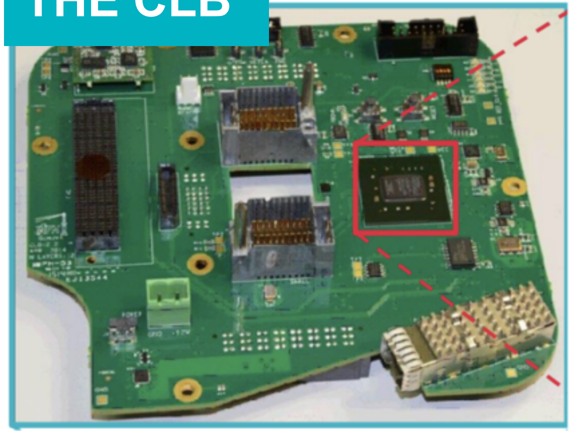
The PhotoMultiplier



THE PULSE



THE CLB



THE HIT

- *Hit Time*
- *PMT id*
- *Time over Threshold (ToT)*

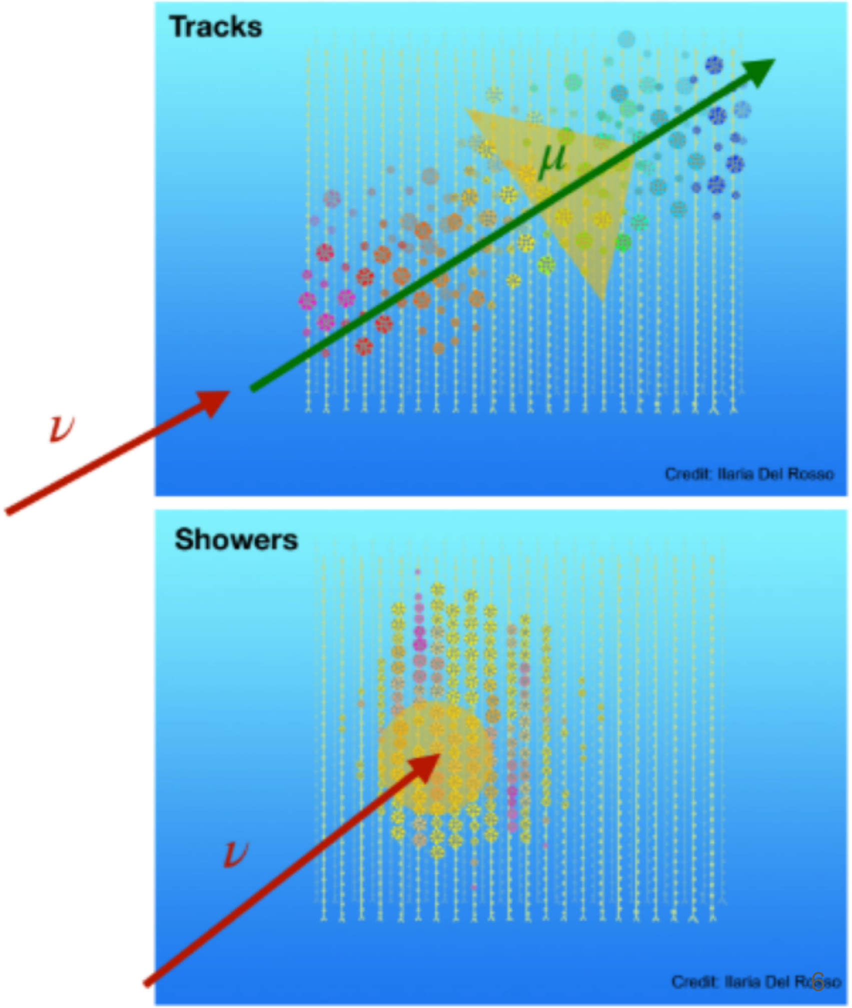
Hits aggregation: the *EVENT* concept

Tracks

- Mainly from charge current (CC) interactions of a ν_μ
- Golden channel for neutrino astronomy
- Clear long track-like signature in the detector (~1 TeV muon in water travel > 5 km!)
- Very good angular resolution while energy resolution is poorer

Showers/cascades

- From all flavours neutral current (NC) interactions and CC interactions of ν_e and ν_τ
- “blobby” light emission (almost symmetrical, but…)
- Good energy resolution while poorer angular resolution



Question to be answered

All directions are good:

- Downward-going (from above)
- Horizontal (from side)
- Upward-going (from below)

Crucial questions:

What particles get to the detector?

What their energies?

What probability to be stopped in the Earth?

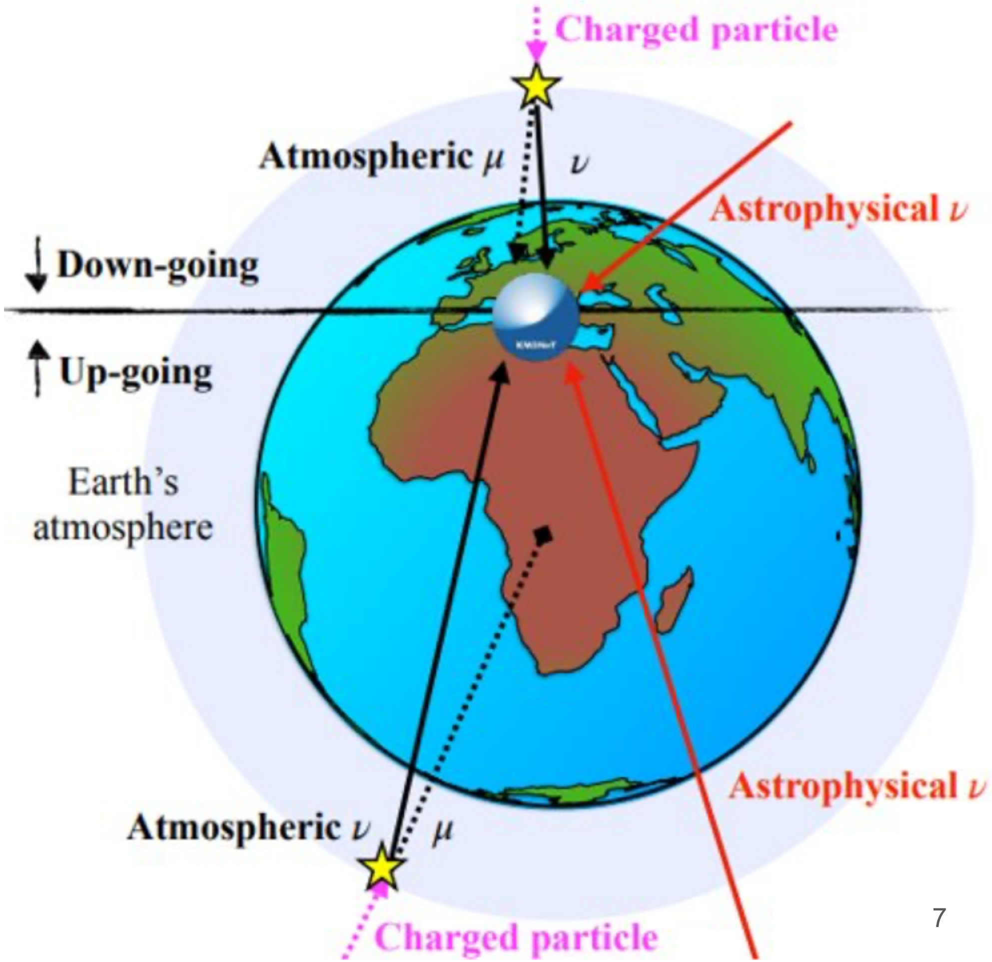
What is signal and background?

What/How background can be eliminated?

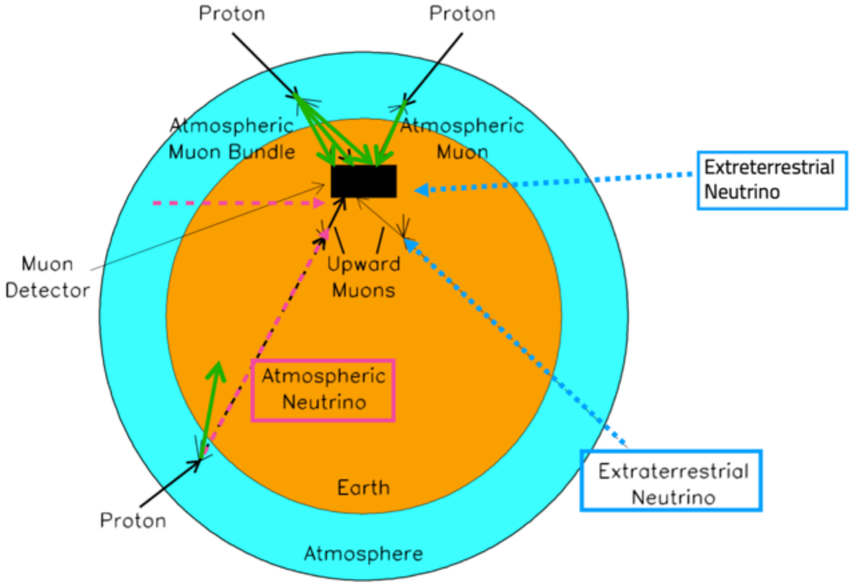
Most basic ones:

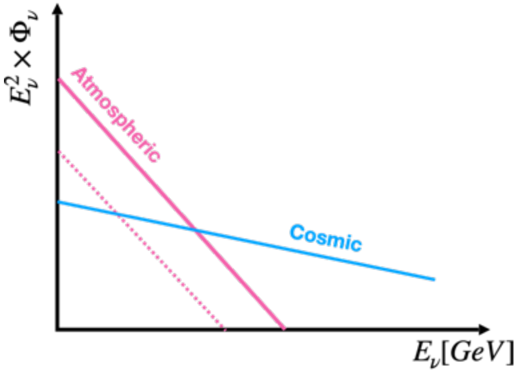
How would an event look like in the data snapshot ?

Am I able to see the difference from an interesting event from bad ones?

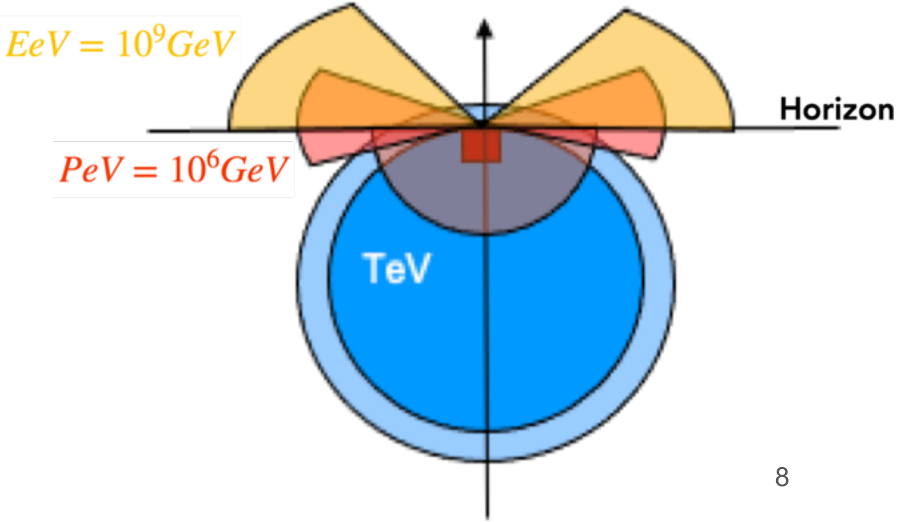


Signal vs. Background





ATMOSPHERIC NEUTRINOS vs COSMIC NEUTRINOS

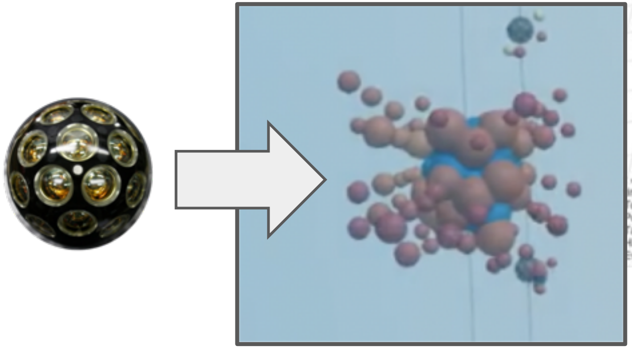


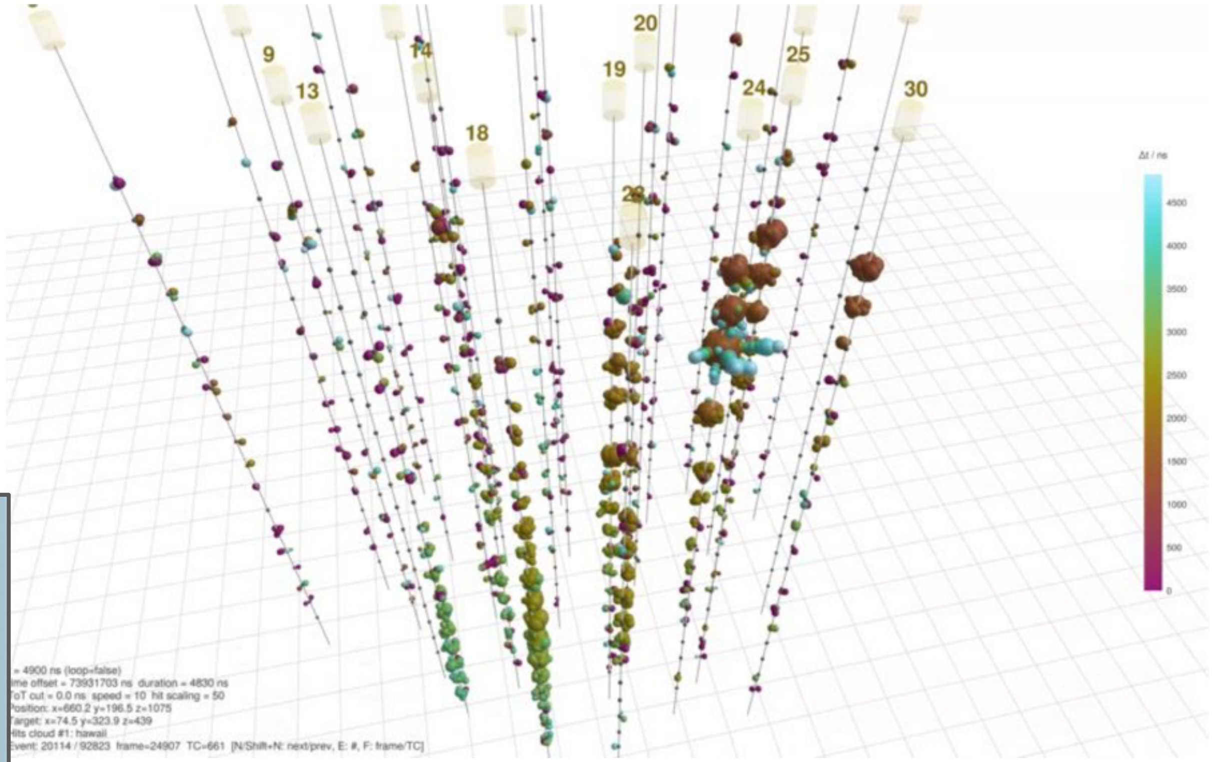
Introduction to event display exercise

That’s why the event display!

- Event displays are a powerful means to visually identify and infer the main features of the events

- They are also valuable tools to monitor and validate the performance of the detector





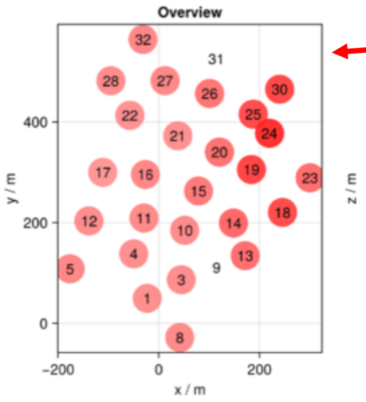
What you will find for this exercise:

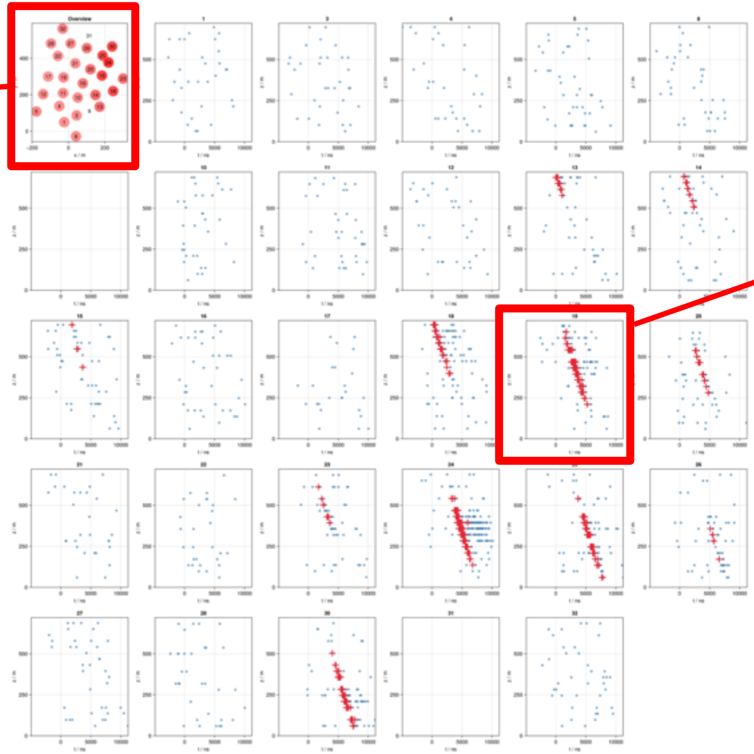
- 10 event displays (movies)
- 10 sets of plots for the same events

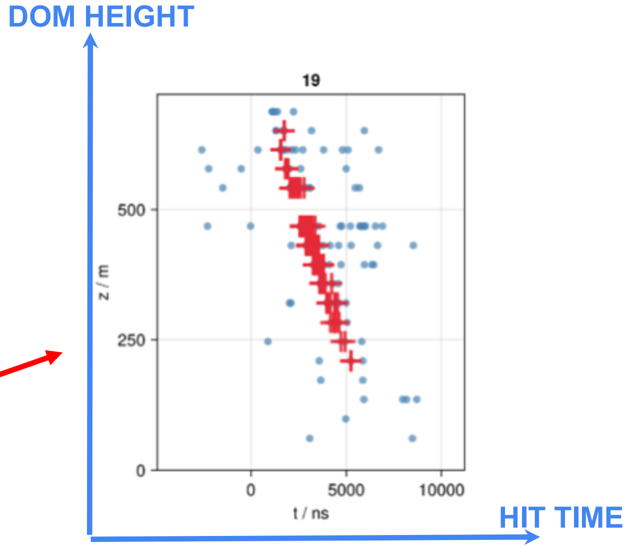
What you are requested to do:

- Find the association between movies [1,..,5] and plots [A,..,E] (if you have time, you can play with events 6-10 and the corresponding F-L plots)
- Determine for each event the category that you consider more appropriate, choosing from:
  - Downwardgoing / horizontal / upwardgoing tracks
  - Cascades

How to interpret the plots







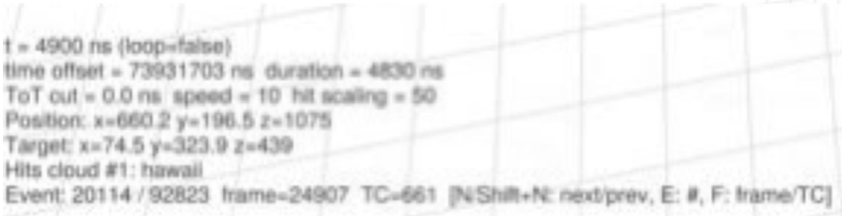
Top left: layout of the detector, showing the positions of the detection units and the amount of signals they detected (increasing with increasing level of colour)

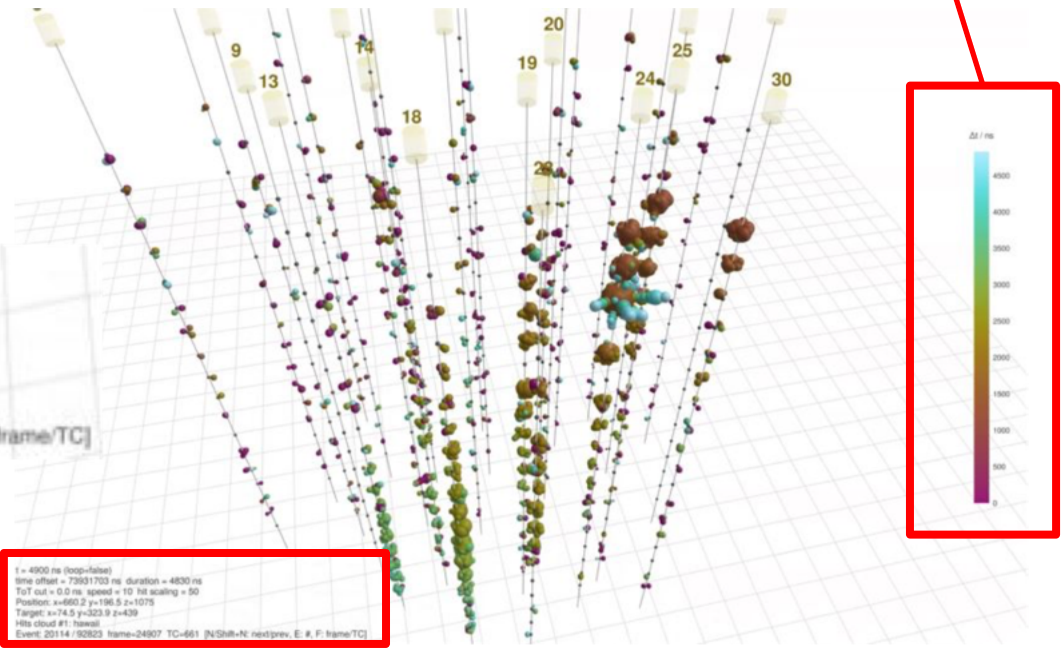
Rest of the page: one plot per detection unit, showing the vertical level and the time of detection of each signal (Remark: red signals are those contributing to the event trigger, which occurs at $t = 0$)

How to interpret the movies

The colours indicate the detection time of the signals (with the count starting at the time of the first triggering hits)

In the bottom left you find key information about the event (not really relevant for your exercise)





PART 2
Calibrations

Why do we need ***calibration***?

Calibration is meant to ensure that what we measure corresponds to the quantity that we want to measure:

- Detectors are not perfect.
- Each detector has intrinsic offsets.
- Environmental conditions can affect the measurements.
- Calibrating means to compensate for these effects.
- Without calibration → <u>biased physics results</u>.
- Calibrations can be “on-line” (when running) or “off-line” (when analysing the data)
- Calibration information must be coherent with the run conditions

The study of these features is referred to as **study of systematics**…
… and how to account for them (when it is possible!)

Systematics

Timing

Positioning

Orientation

Pointing

Detector conditions

Sea conditions

Systematics

Timing

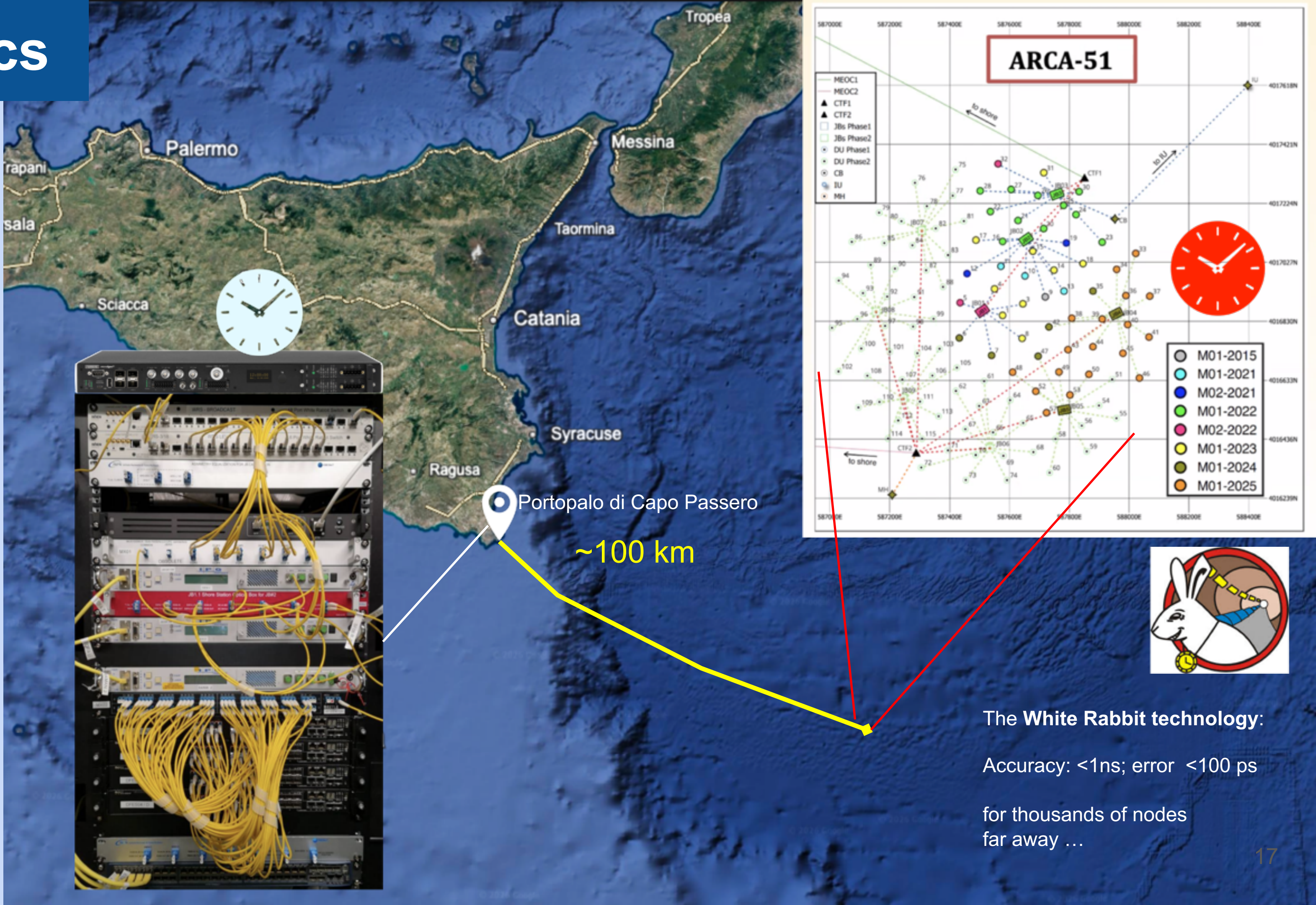
Positioning

Orientation

Pointing

Detector conditions

Sea conditions



The **White Rabbit technology**:

Accuracy: <1ns; error <100 ps

for thousands of nodes far away …

Systematics

Timing

Positioning

Orientation

Pointing

Detector conditions

Sea conditions

Systematics

Timing

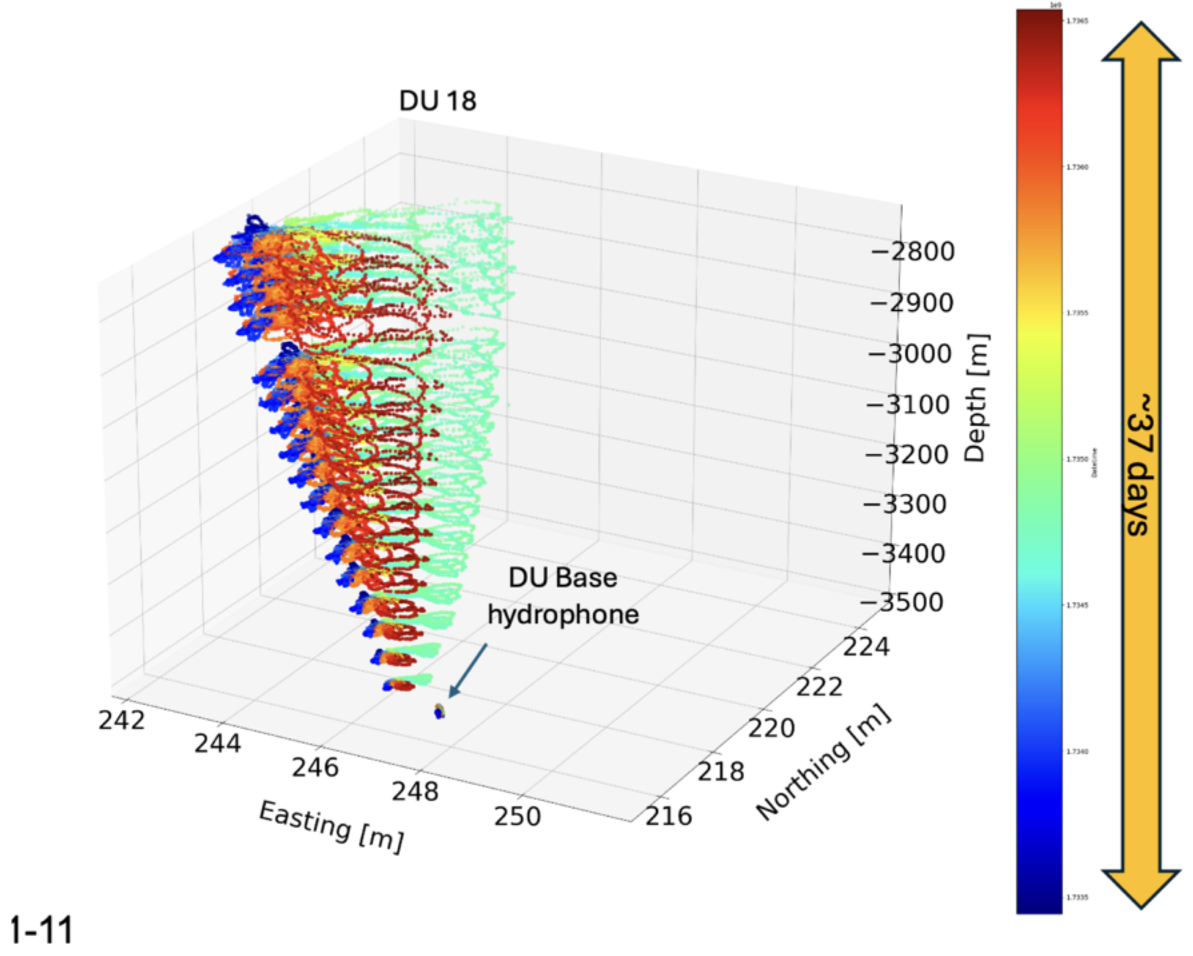
Positioning

Orientation

Pointing

Detector conditions

Sea conditions



Systematics

Timing

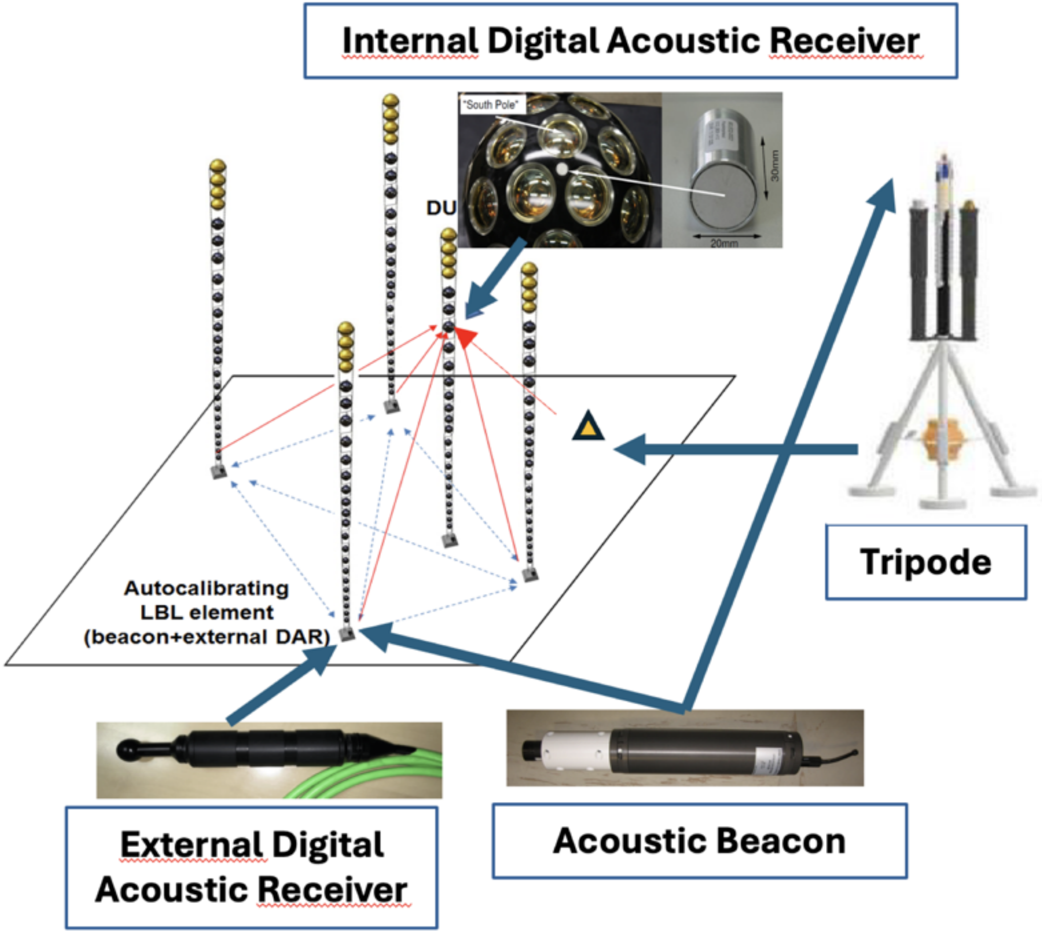
Positioning

Orientation

Pointing

Detector conditions

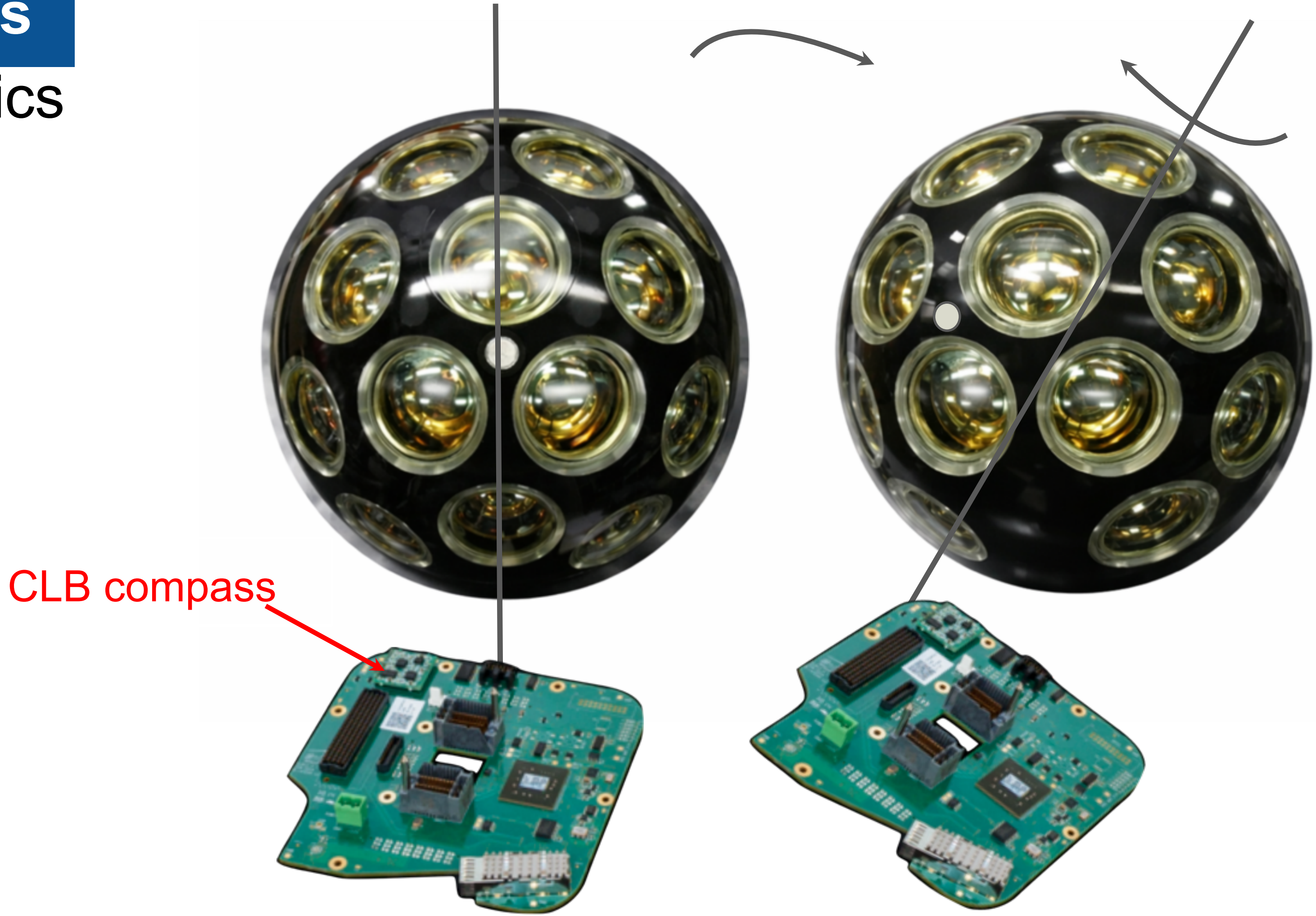
Sea conditions



Systematics

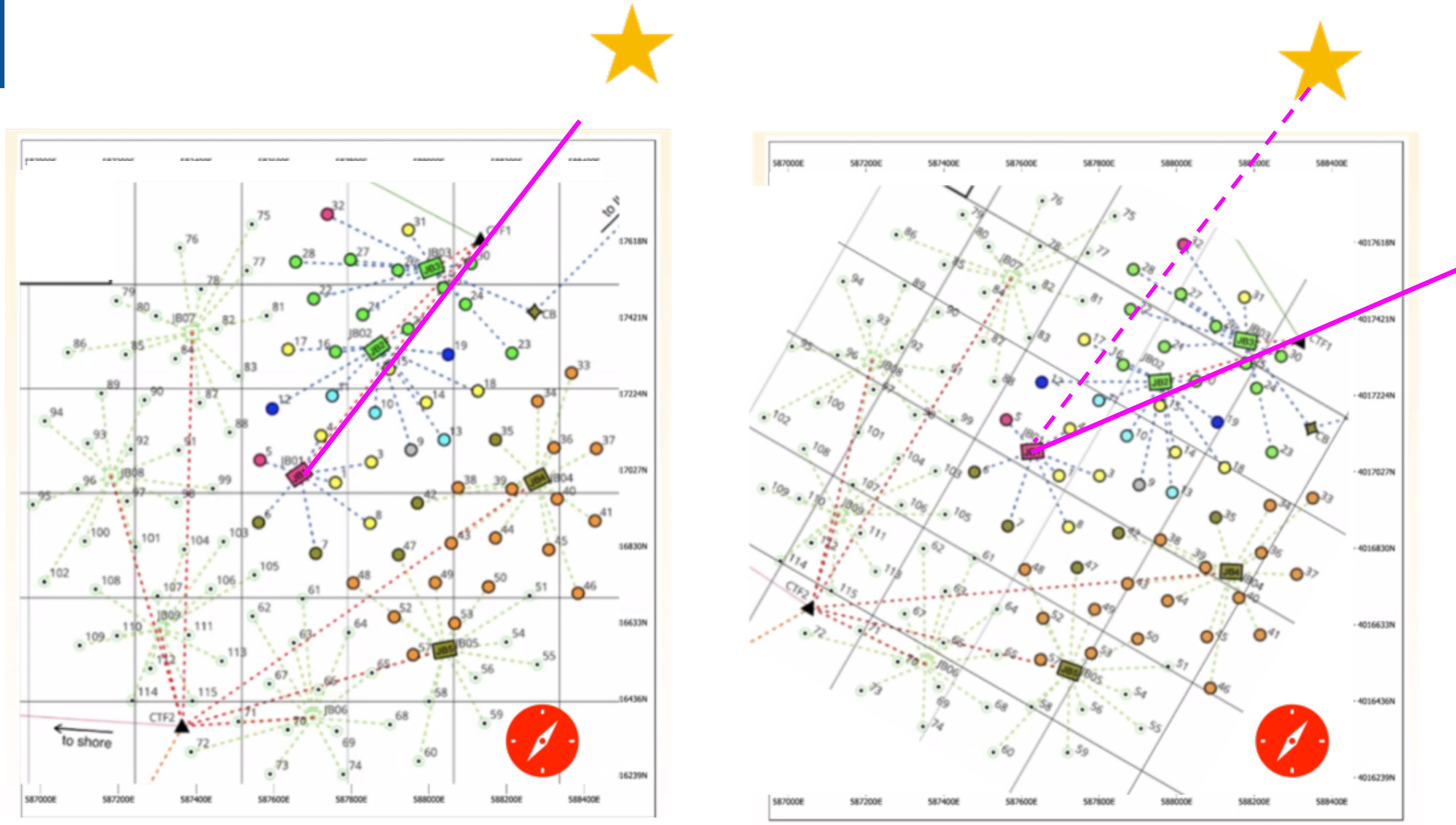
Systematics

- Timing
- Positioning
- ***Orientation***
- Pointing
- Detector conditions
- Sea conditions



Systematics

- Timing
- Positioning
- Orientation
- ***Pointing***
- Detector conditions
- Sea conditions



A bad knowledge of the detector position can fake the pointing in the sky

Systematics

Timing

Positioning

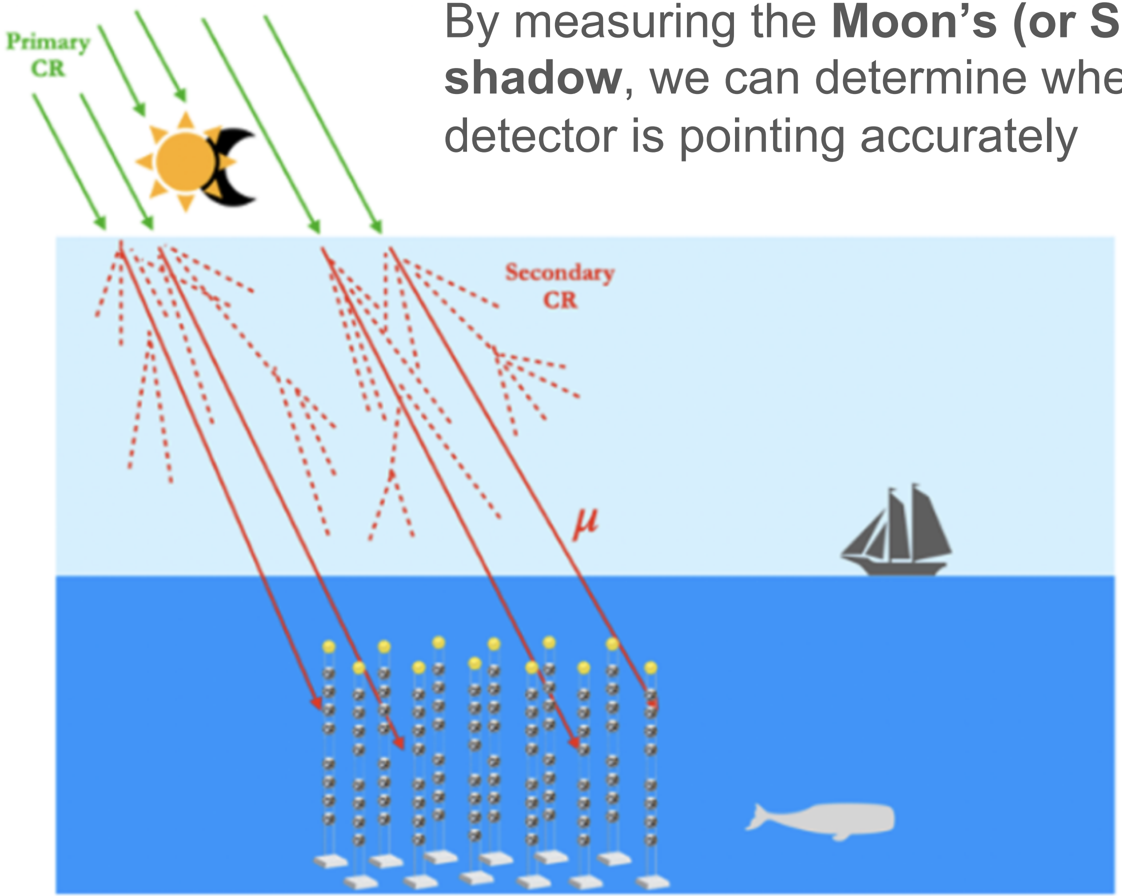
Orientation

Pointing

Detector conditions

Sea conditions

By measuring the **Moon's (or Sun's) shadow**, we can determine whether our detector is pointing accurately



Systematics

Time to time: erratic losses of some DOMs

Timing

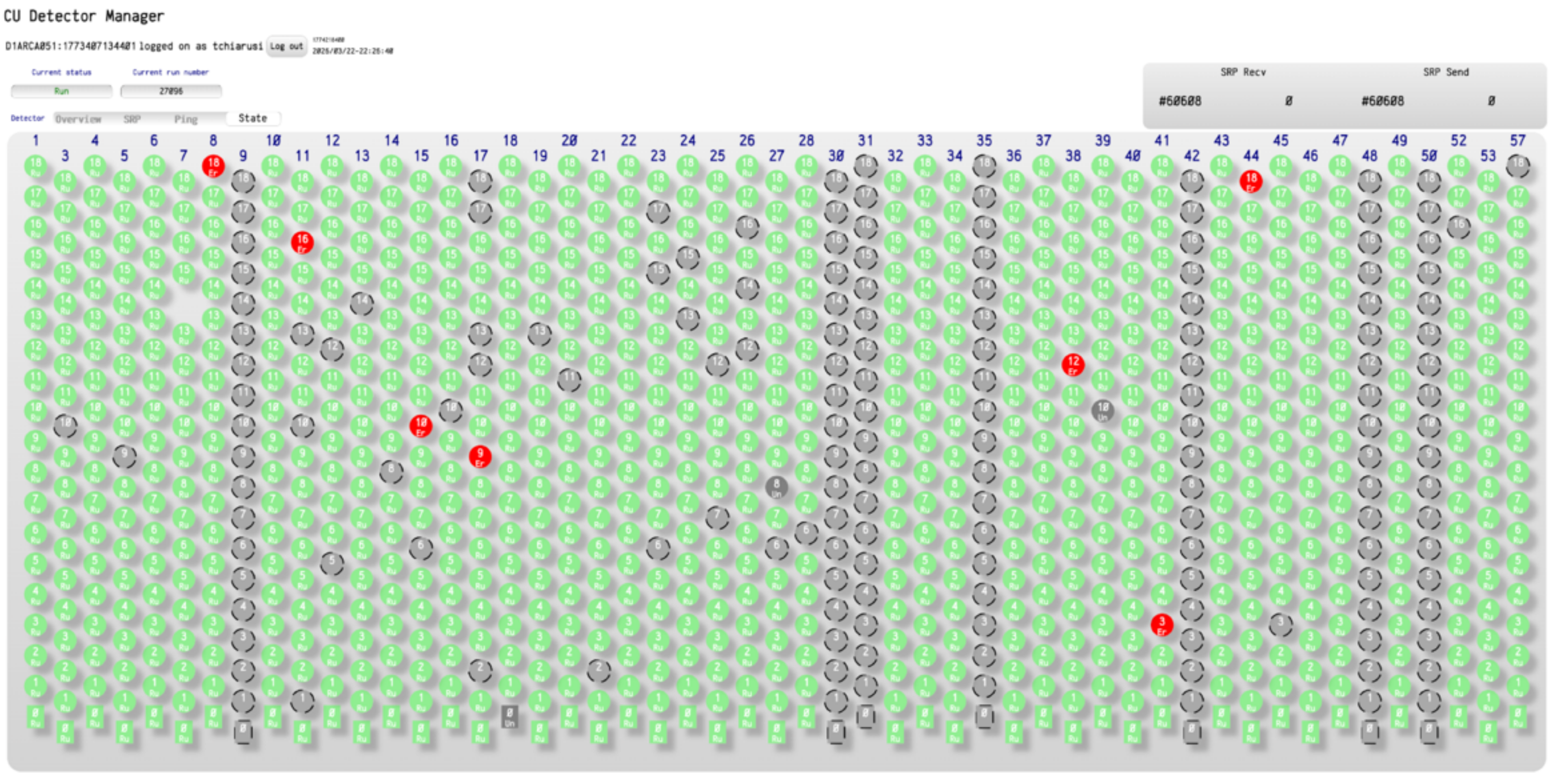
Positioning

Orientation

Pointing

Detector conditions

Sea conditions

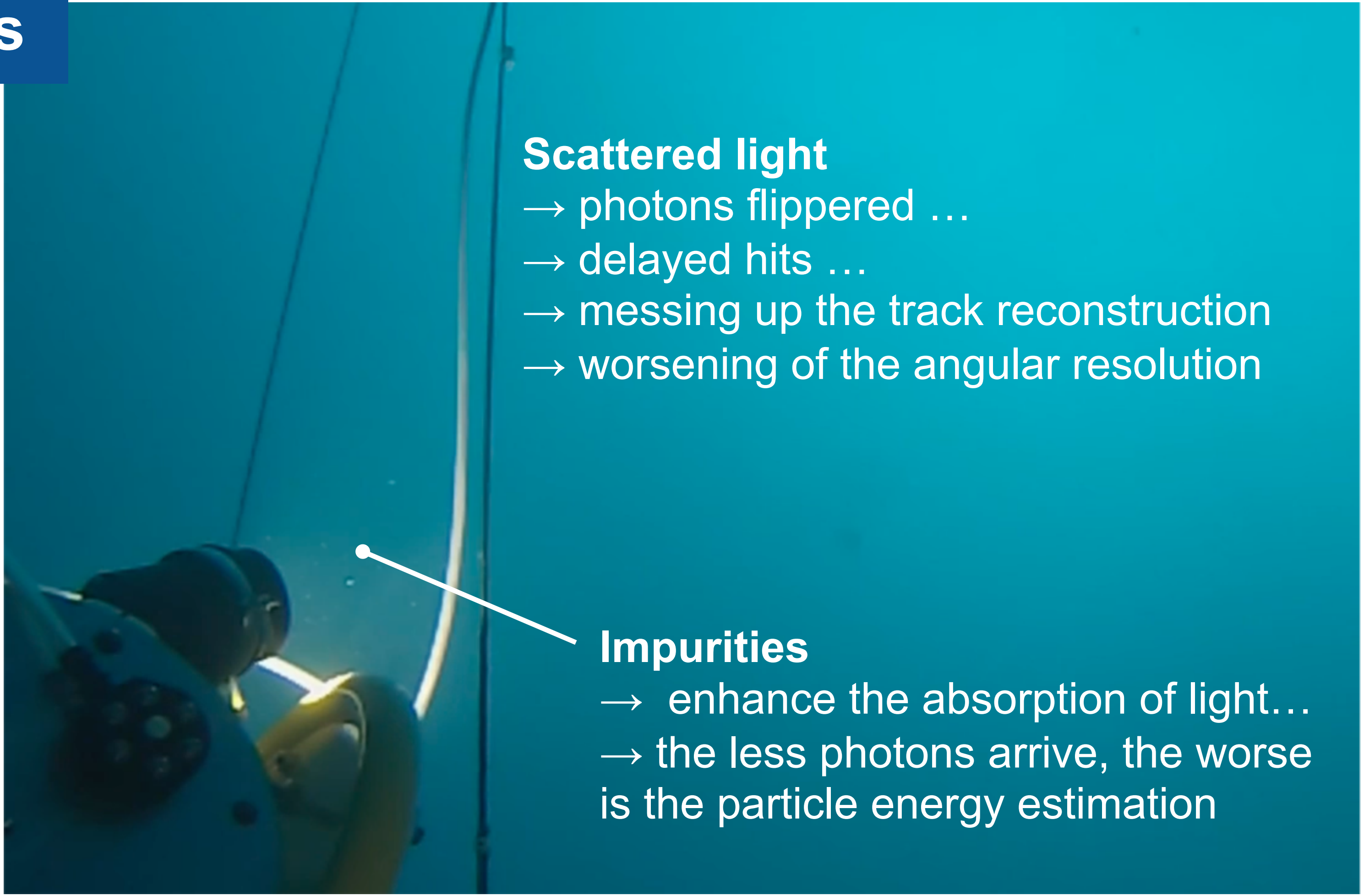


Systematics

- Timing
- Positioning
- Orientation
- Pointing
- Detector conditions
- ***Sea conditions***

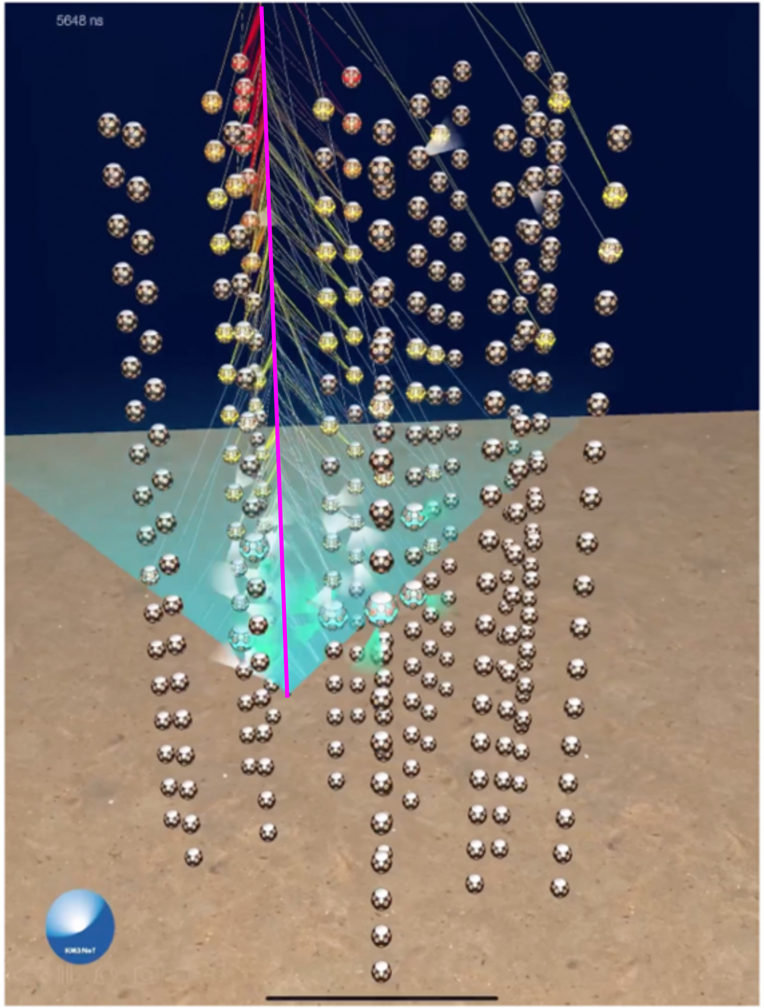
Systematics

- Timing
- Positioning
- Orientation
- Pointing
- Detector conditions
- ***Sea conditions***



Scattered light
→ photons flippered …
→ delayed hits …
→ messing up the track reconstruction
→ worsening of the angular resolution

Impurities
→ enhance the absorption of light…
→ the less photons arrive, the worse is the particle energy estimation



TIME RESIDUALS

In a given time-interval of data-taking,
=> a **set of measured hits**
(i.e.: *time-stamps and positions for each hit*)
that “*should be due*” to the passage of a muon

Due to geometrical assumptions:

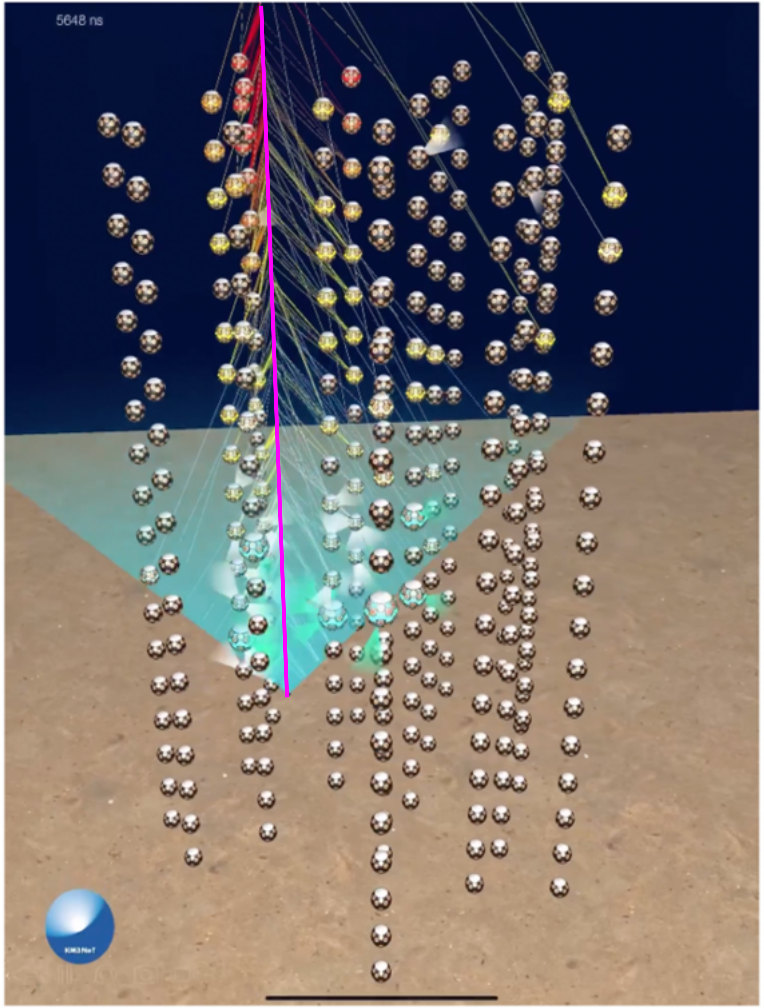
- space-time constraints
- Cherenkov angle

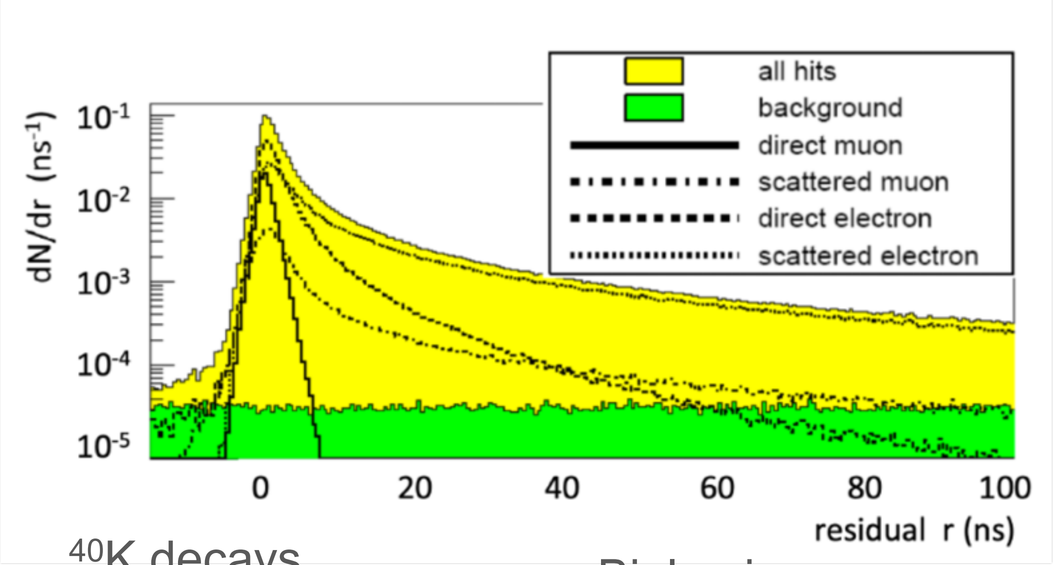
We can reconstruct ***the best muon track fitting the data***.
=> ***infer “what-if” time-stamps and positions***
… and calculate the ***time-residuals:***

$$r_i = t_i^{meas} - t_i^{reco}$$

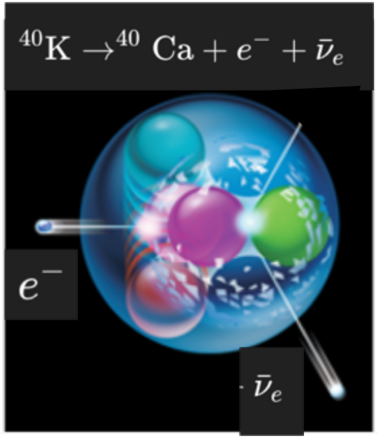
With i the index of i-th PMT

The best reco track is when r is the least one





^{40}K decays



Bioluminescence

What is Calibration?

Calibration is the process of correcting systematic deviations in detector measurements.

True signal → Detector → Offset → Measured signal

↓

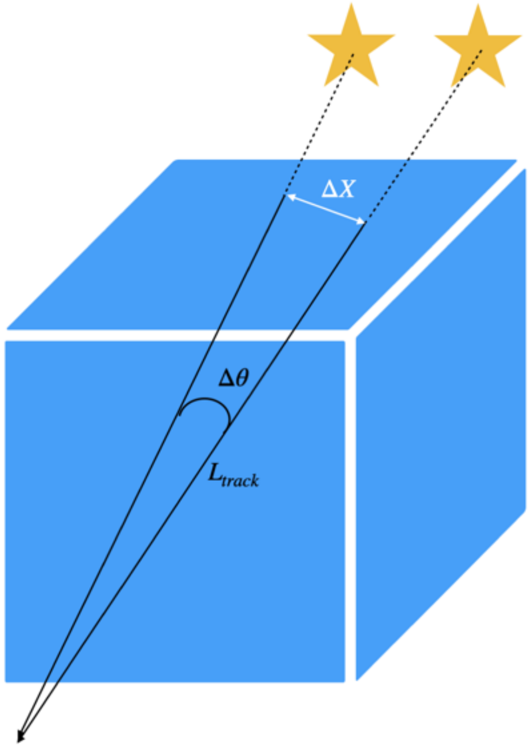
Calibration → Corrected signal

Example: Timing Offsets in Detection Units

- Each Detection Unit (DU) may have a non-negligible time offset.
- Even a few nanoseconds matter!
- Timing affects:
  - Direction reconstruction
  - Event classification
  - Angular resolution

<u>Small timing shifts can translate into large angular errors.</u>

One last thing about a **good timing…**



Assumptions

$L_{track} = 1km$ — Due to detector dimensions

$\Delta\theta = 0.1^{\circ}$ — Our **requirement** to do Astronomy

$v_{light}^{-1} = 5\frac{ns}{m}$ **in water** — Due to sea water conditions: temperature, salinity, density

Calculations

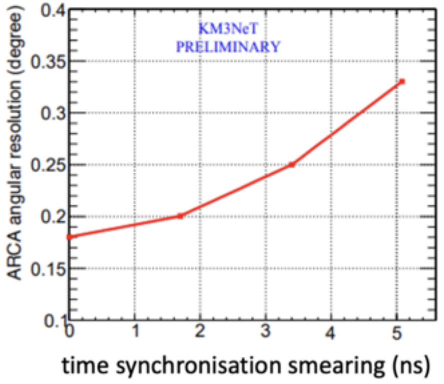
$\Delta X = L_{track} \times \tan \Delta\theta \sim 1.7m$

$\Delta X = 1.7m \rightarrow \Delta t < 10\ ns$

Acoustic positioning can push to $O(20)\ cm$

$\rightarrow \Delta t \sim 1\ ns$

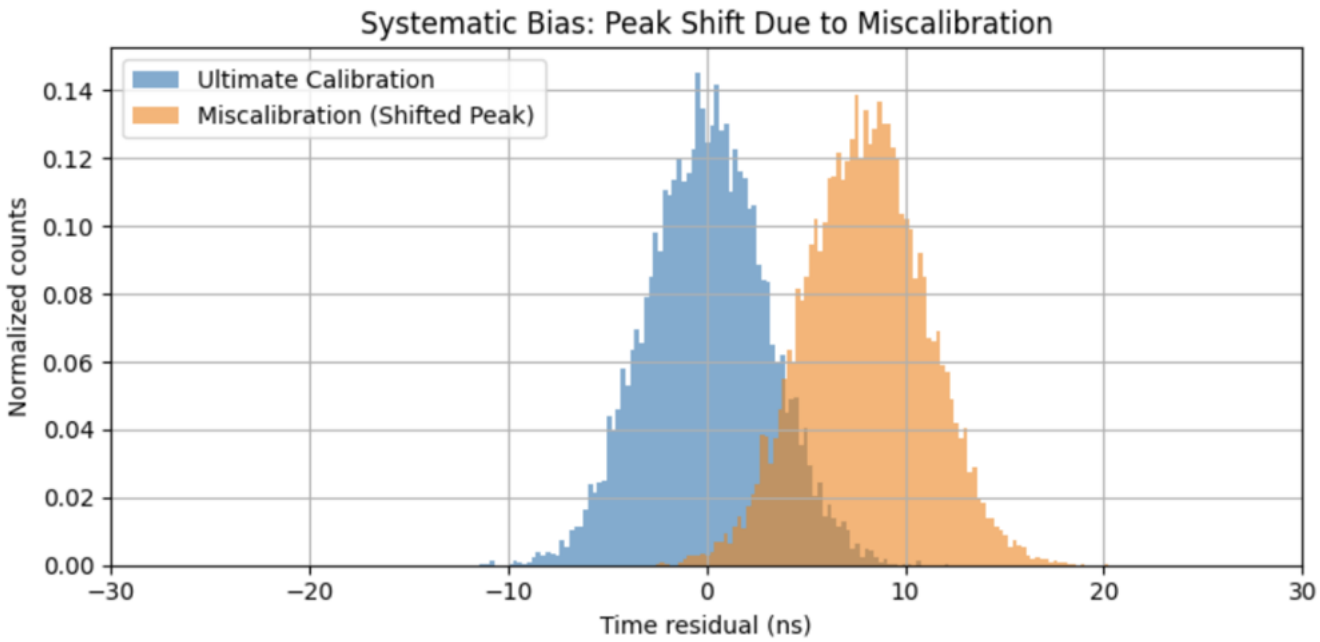
Time Synchronisation and DOM position calibration is the key parameter to optimise angular resolution



(DOM position uncertainty 20 cm≈ 1ns)

Miscalibration Effects

Miscalibration increases systematic uncertainty.



Time residual = the difference, for each PMT hit detected in a detection unit, between detection time and the estimated time (inferred from the reconstruction of a muon track based on the signals collected from the other detection units)

Introduction to the calibration exercise

Our exercise

In this exercise, **we reconstruct the direction of particles using timing**.
If one detection unit is even slightly miscalibrated, the reconstruction algorithm may think the particle came from a **slightly different direction**.

You have a set of ***15***, miscalibrated detection units that you need to calibrate based on their time residual plots:

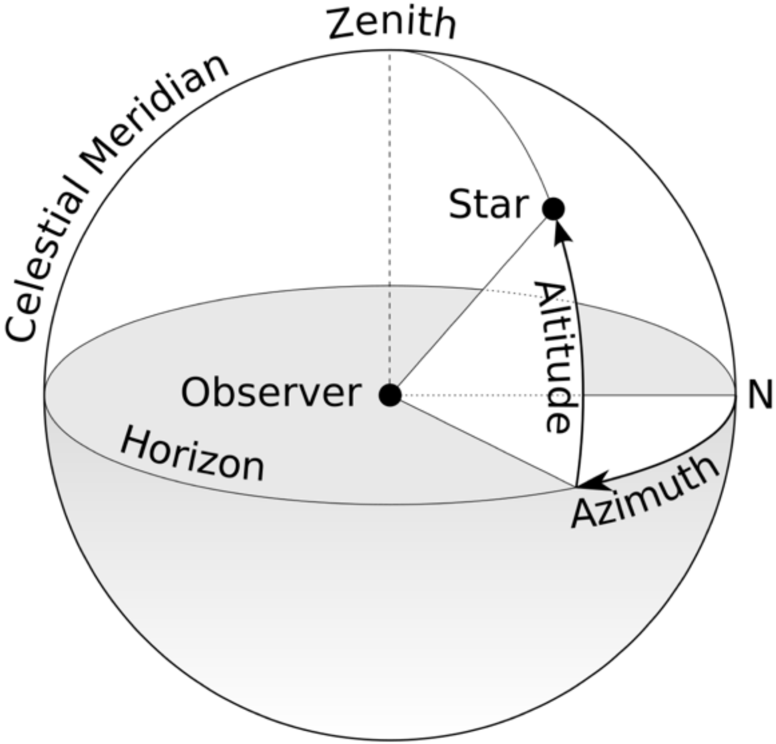
-> look at the plots using a null-calibration set + look at the corresponding sky map
-> recursively apply a correction offset to each DU (in an attempt to place the peaks as close as to zero time residual), then look at the changes in the sky map
-> set up the best calibration you can reach and finalize the analysis

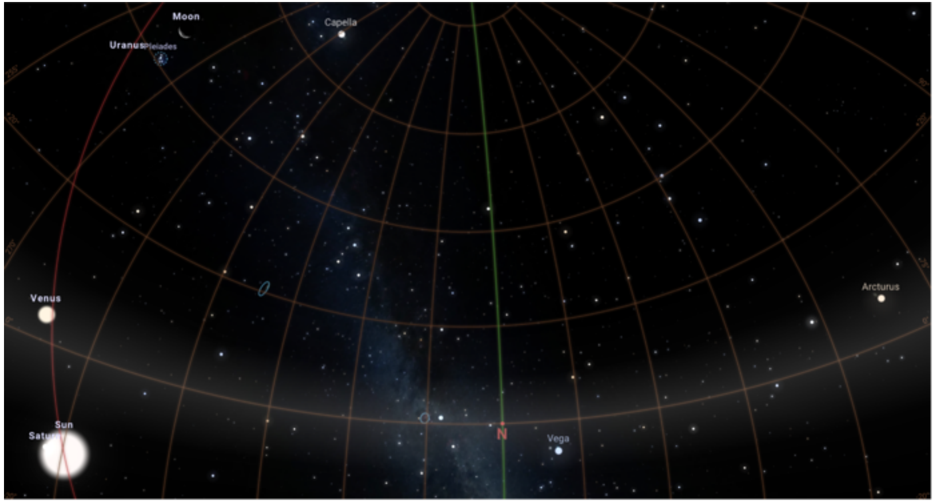
PART 3
Approaching data analysis

How we locate a point in the sky

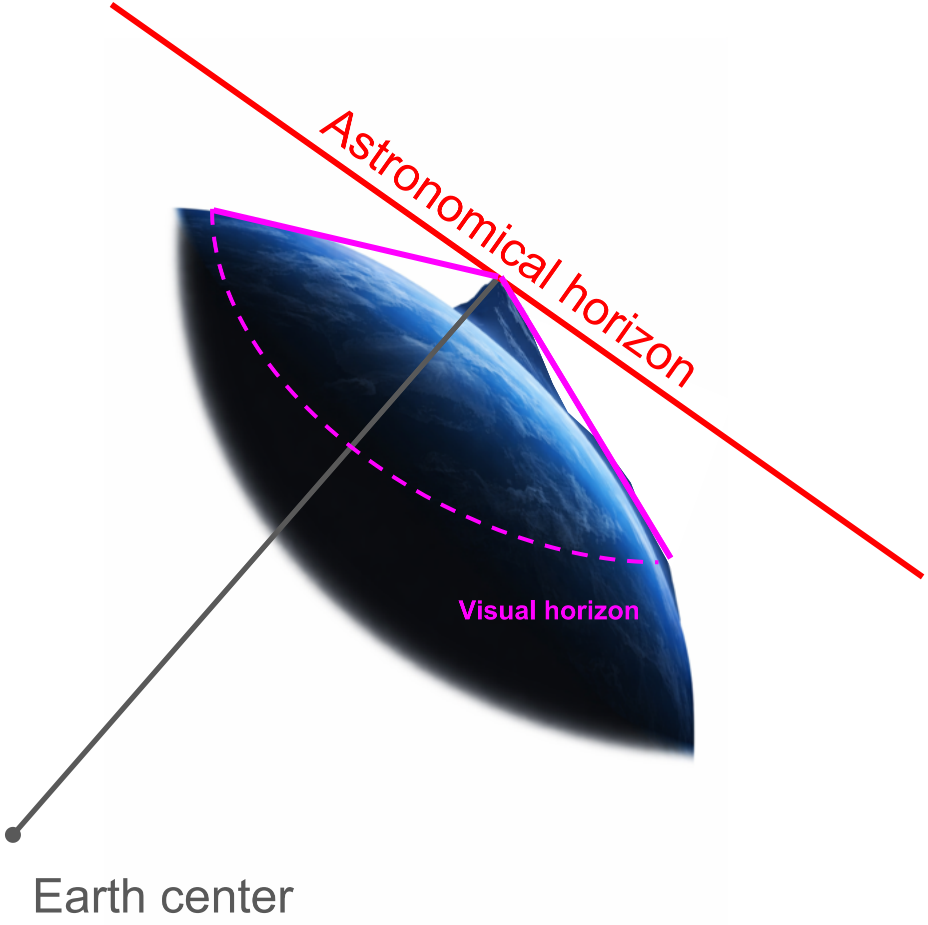
The Observer and the Sky

- Imagine the sky as a giant dome surrounding the observer.
- The observer stands at the center.
- The point straight above your head is the zenith.
- In astronomy, the ***astronomical horizon*** is defined as the intersection between the celestial sphere and a plane perpendicular to the zenith axis of the observer.
  - This is a geometrical reference plane used to define celestial coordinates.
  - It does not coincide with the place where the sky appears to meet the Earth. That would only be true if the Earth were flat.
  - The line where the sky seems to touch the ground is instead called the ***visual horizon***, which depends on the observer's position and on the curvature of the Earth.




Moon
Capella
Uranus
Pleiades
Venus
Arcturus
Sun
N
Vega


Moon
Capella
Uranus
Pleiades
Venus
Arcturus
N


Astronomical horizon
Visual horizon
Earth center

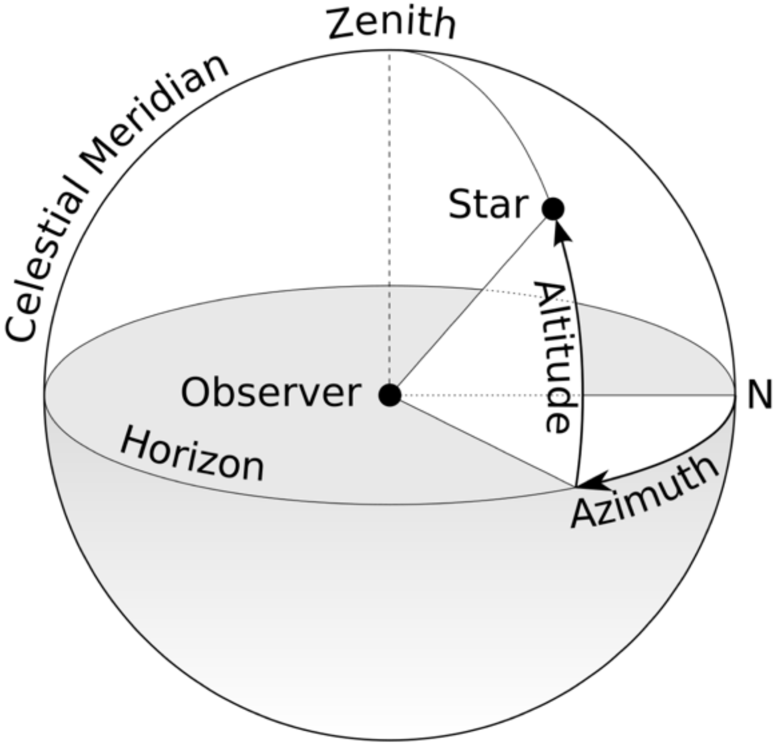
How we locate a point in the sky

Altitude = “Sky Latitude”

- Altitude tells us how high an object is in the sky.
- It is measured upward from the horizon.
  - 0° → object is on the horizon
  - 90° → object is at the zenith (straight overhead)

Analogy:

Like latitude on Earth, which tells how far North or South a location is.



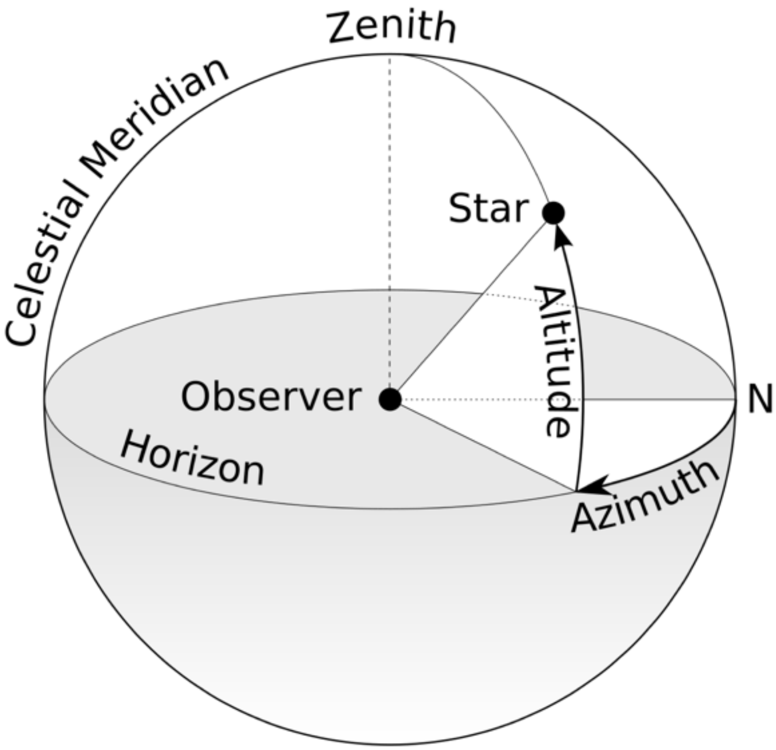
How we locate a point in the sky

Azimuth = “Sky Longitude”

- Azimuth tells us which direction to look along the horizon.
- It is measured around the horizon, starting from North.
    - North = 0°
    - East = 90°
    - South = 180°
    - West = 270°

Analogy:

Like longitude on Earth, which tells how far East or West a location is.



Measuring with KM3NeT in a LOC COORD sky map

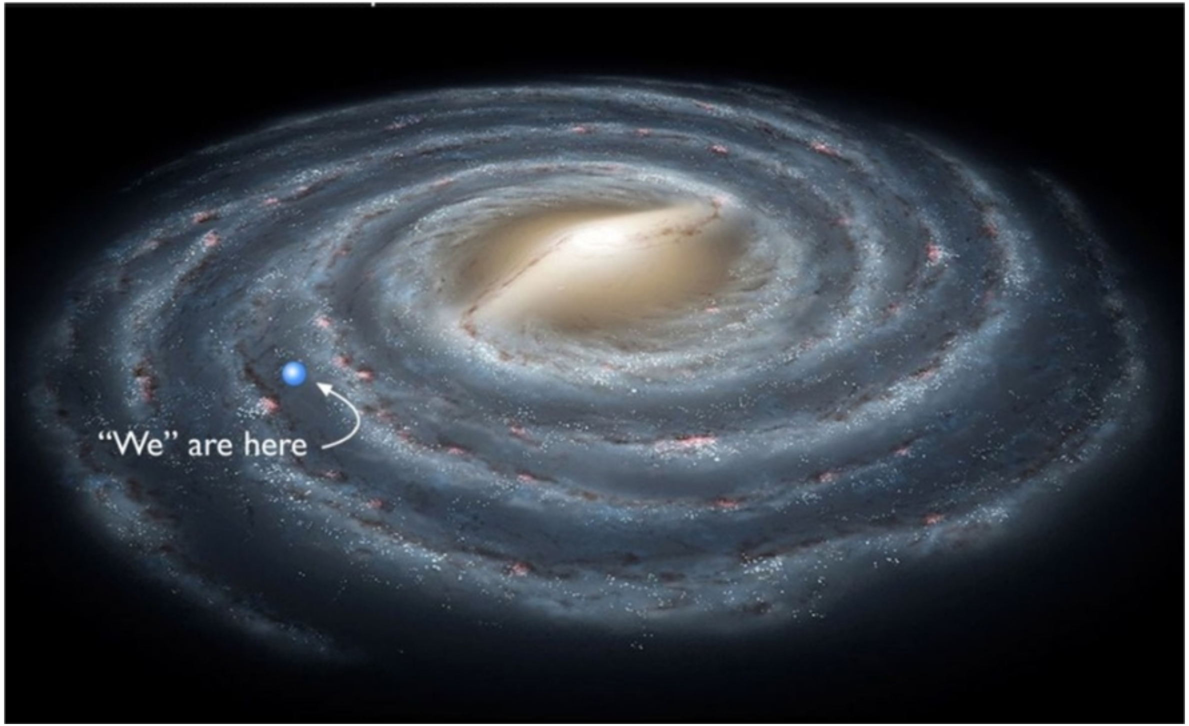
- **Above the horizon (altitude > 0°):**
  - mainly the huge contribution of atmospheric muons (background)
  - atmospheric and cosmic neutrinos are completely overwhelmed by muons

- **Slightly below the horizon (-10° < altitude < 0°)**
  - many poorly reconstructed downgoing muons that decrease more and more as we move to more negative altitude (background)
  - atmospheric and cosmic neutrinos are still significantly overwhelmed by muons

- **Below the horizon (altitude < -10°)**
  - No atmospheric muons (muons cannot cross the Earth to reach the detector from below)
  - some atmospheric neutrinos (background)
  - a few cosmic neutrinos (signal!)

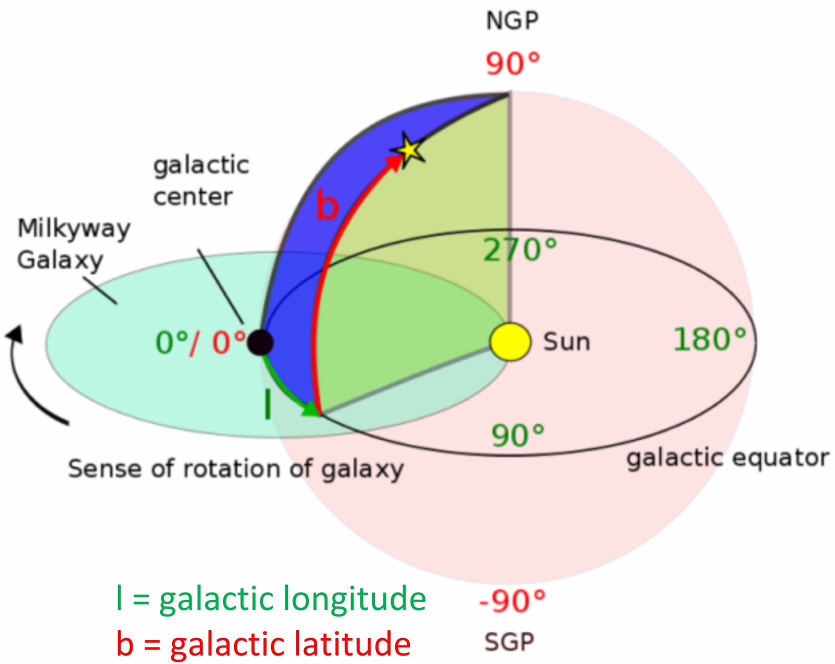
Galactic coordinates

- The Milky Way is shaped like a flat disk (like a giant pancake).
- We live inside this disk, not outside it.



Galactic coordinates

- The Sun is our reference point (yellow dot).
- The galactic centre is the middle of the galaxy (black dot).



l = galactic longitude
b = galactic latitude

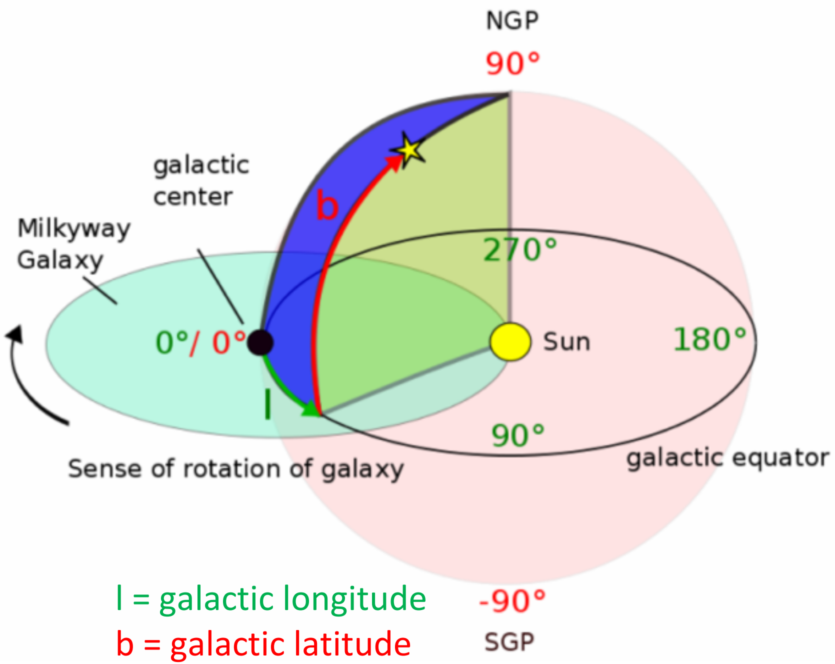
Galactic coordinates

Galactic Longitude (Gal Lon)

- Galactic longitude tells us which direction we look within the galaxy disk.
- It is measured around the galactic disk:
  - 0°→ toward the galactic center
  - 90°, 180°, 270° → other directions around the galaxy

Analogy:

Like longitude on Earth, measuring East–West position.



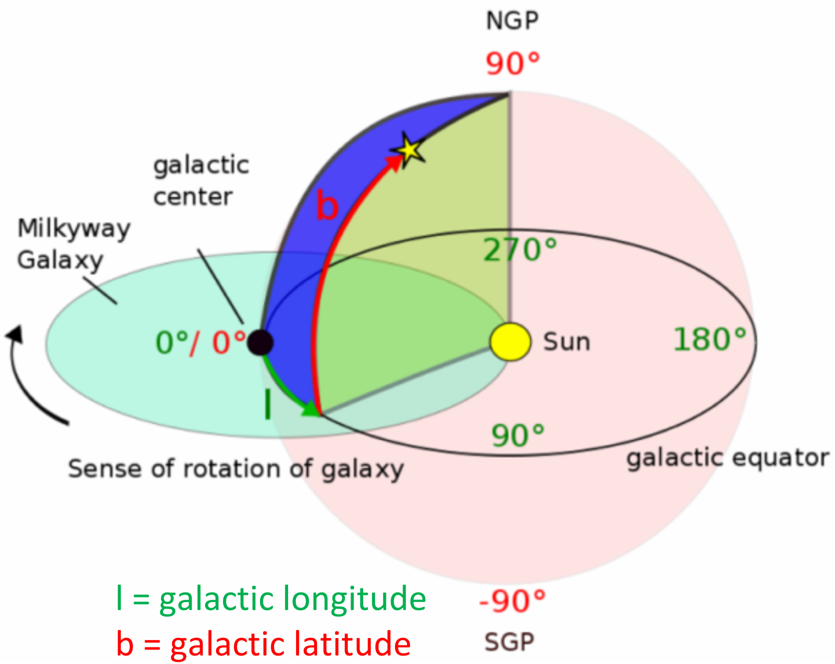
Galactic coordinates

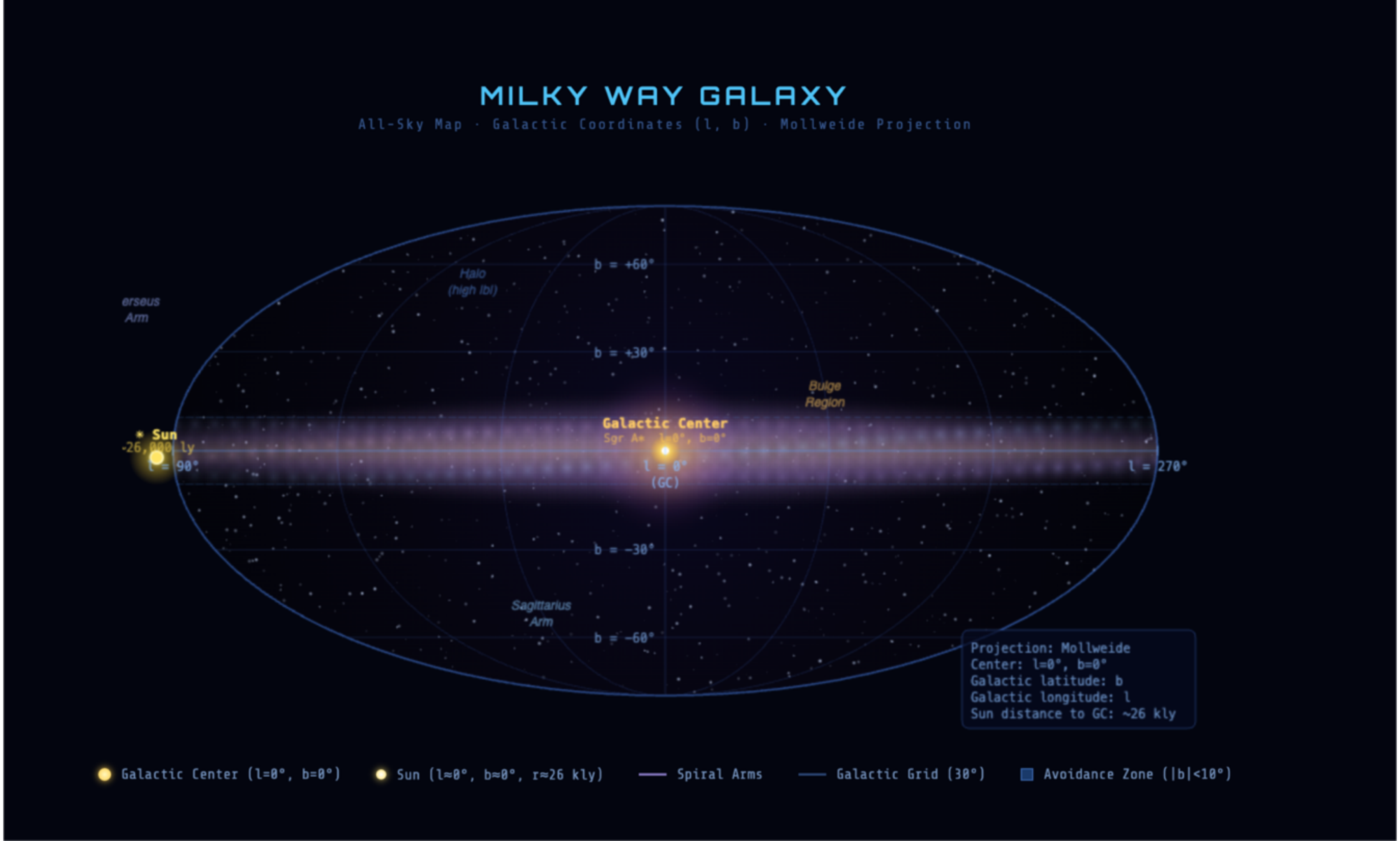
Galactic Latitude (Gal Lat)

- Galactic latitude tells us how far above or below the galactic disk an object is.
- It is measured up or down from the galactic equator:
  - 0°→ inside the galactic disk
  - +90° → North Galactic Pole (NGP)
  - −90° → South Galactic Pole (SGP)

Analogy:

Like latitude on Earth, measuring North–South position.

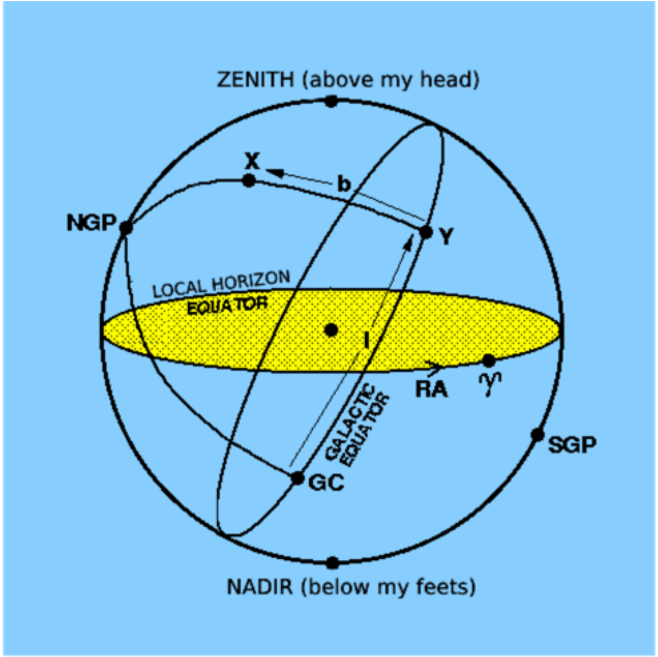



MILKY WAY GALAXY
All-Sky Map · Galactic Coordinates (l, b) · Mollweide Projection
Halo
(high |b|)
erseus
Arm
b = +60°
b = +30°
b = −30°
b = −60°
Bulge
Region
Galactic Center
Sgr A* · l=0°, b=0°
Sun
-26,000 ly
l = 90°
l = 0°
(GC)
l = 270°
Sagittarius
Arm
Projection: Mollweide
Center: l=0°, b=0°
Galactic latitude: b
Galactic longitude: l
Sun distance to GC: ~26 kly
Galactic Center (l=0°, b=0°)
Sun (l≈0°, b≈0°, r≈26 kly)
Spiral Arms
Galactic Grid (30°)
Avoidance Zone (|b|<10°)

Local coordinates vs Galactic coordinates

The celestial sphere does not change!

- All stars are projected onto the same celestial sphere
- What changes is how we draw the reference plane on that sphere
- The picture shows different planes crossing the same sphere

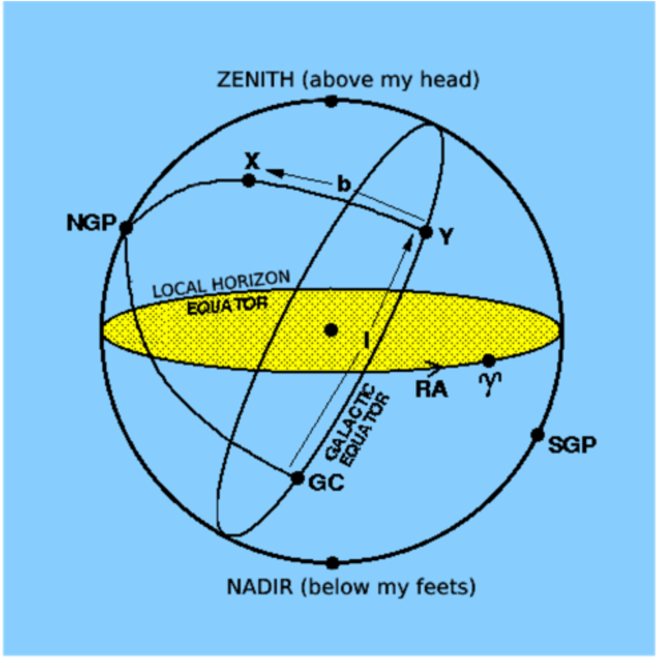


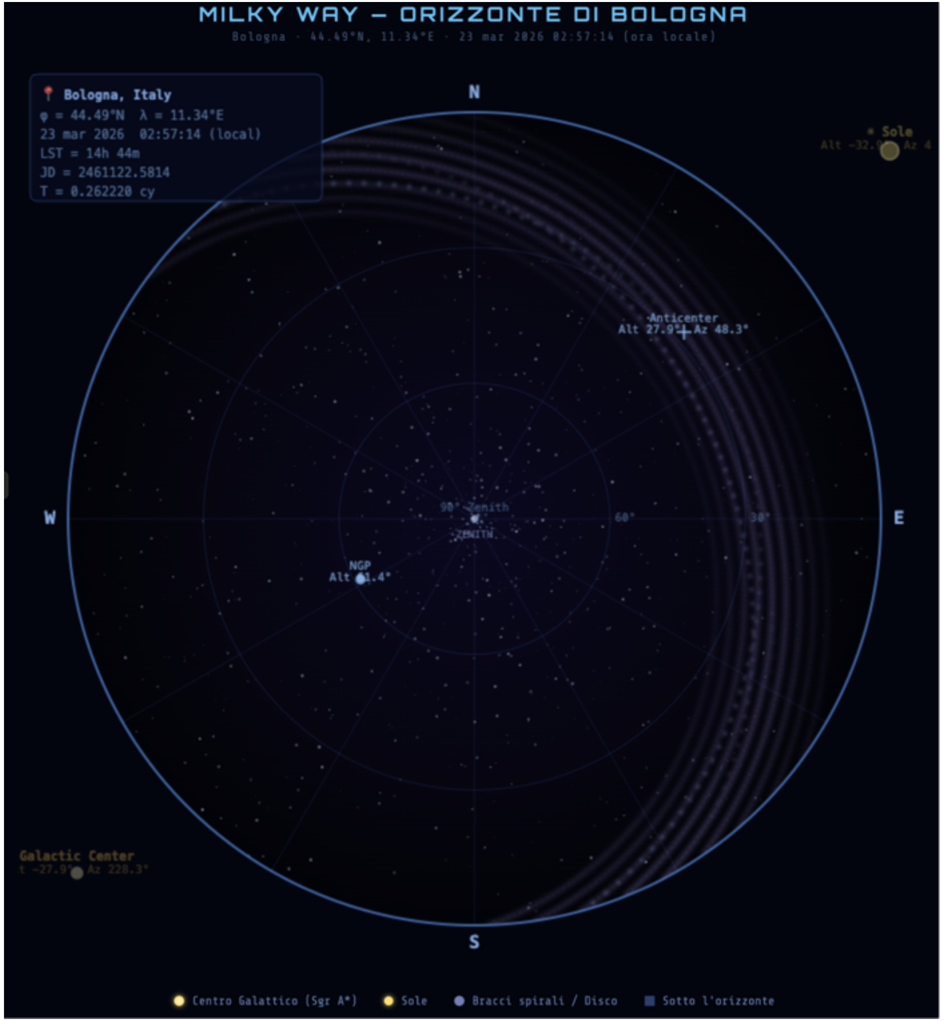
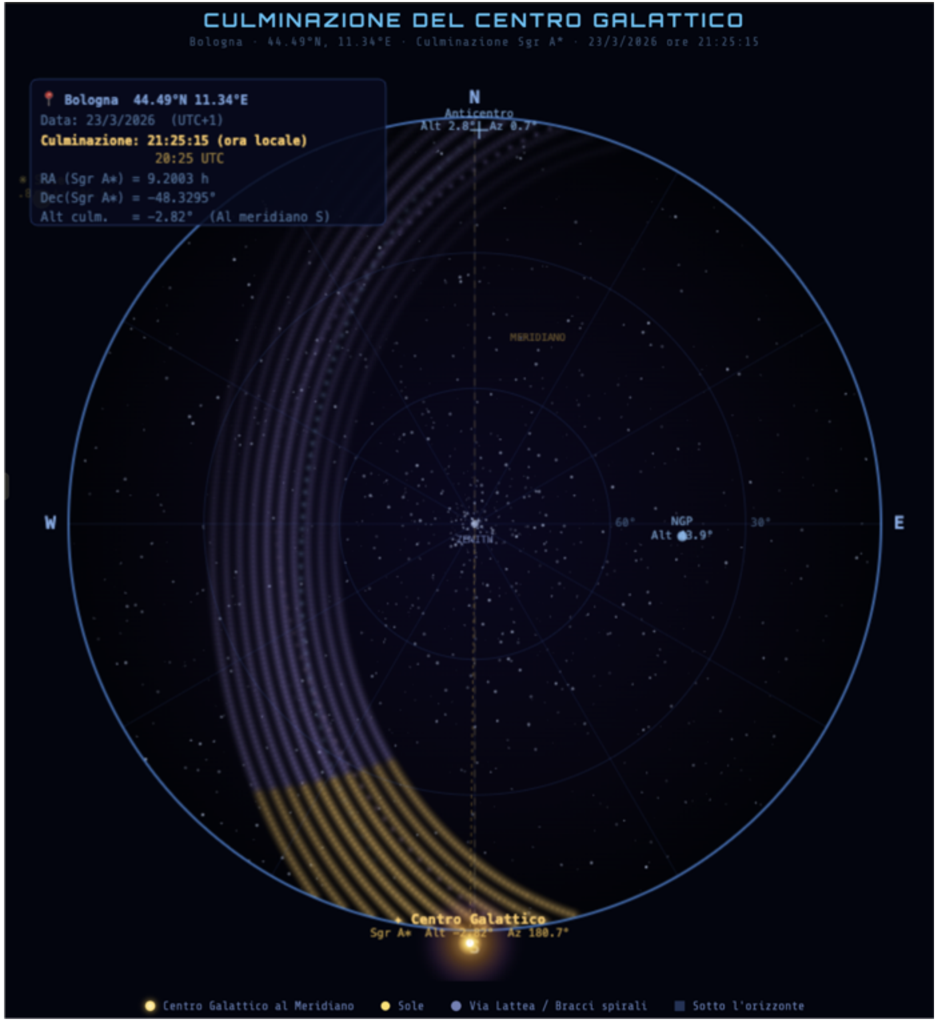
Local coordinates vs Galactic coordinates

A star stays in the same position in space

But its coordinates change because:

- We rotate the reference plane (Earth rotates!)
- Mathematically, this is like rotating the coordinate system, not moving the star



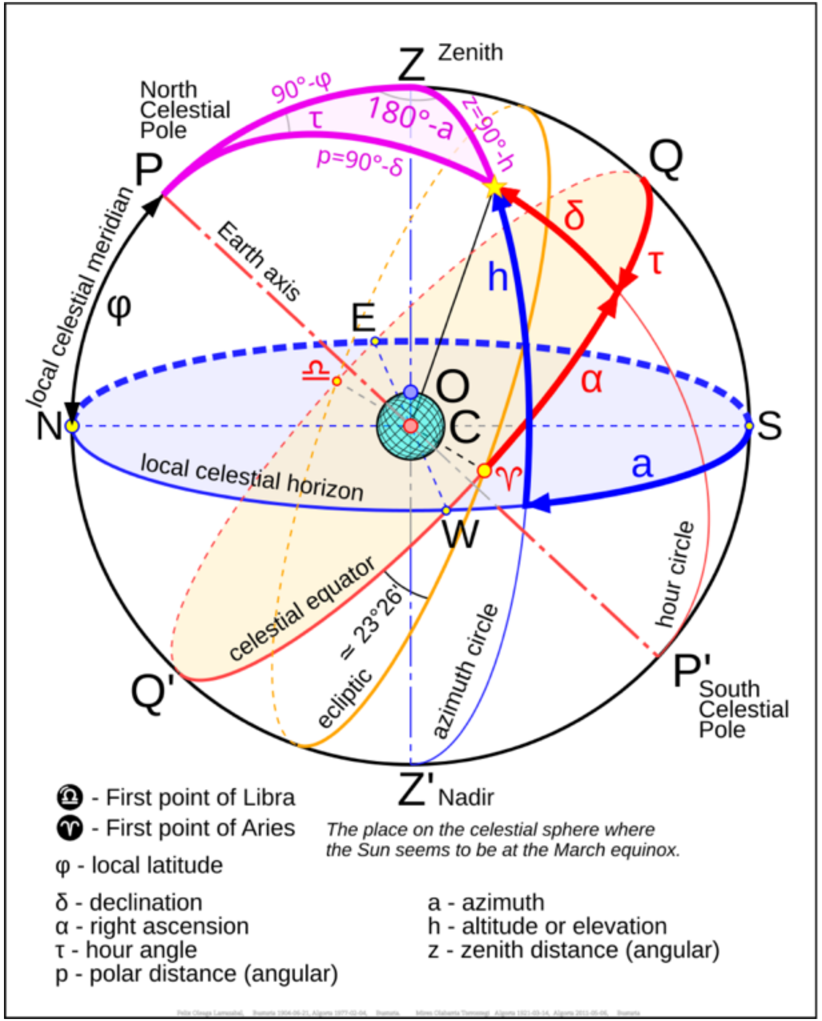

MILKY WAY – ORIZZONTE DI BOLOGNA
Bologna · 44.49°N, 11.34°E · 23 mar 2026 02:57:14 (ora locale)
Bologna, Italy
φ = 44.49°N λ = 11.34°E
23 mar 2026 02:57:14 (local)
LST = 14h 44m
JD = 2461122.5814
T = 0.262220 cy
N
S
W
E
Sole
Anticenter
Alt 27.9° Az 48.3°
90° Zenith
ZENITH
60°
30°
NGP
Galactic Center
Az 228.3°
Centro Galattico (Sgr A*)
Sole
Bracci spirali / Disco
Sotto l'orizzonte

CULMINAZIONE DEL CENTRO GALATTICO
Bologna · 44.49°N, 11.34°E · Culminazione Sgr A* · 23/3/2026 ore 21:25:15
Bologna 44.49°N 11.34°E
Data: 23/3/2026 (UTC+1)
Culminazione: 21:25:15 (ora locale)
20:25 UTC
RA (Sgr A*) = 9.2003 h
Dec(Sgr A*) = -48.3295°
Alt culm. = -2.82° (Al meridiano S)
N
W
E
Anticentro
MERIDIANO
ZENITH
60°
30°
NGP
Centro Galattico
Az 180.7°
Centro Galattico al Meridiano
Sole
Via Lattea / Bracci spirali
Sotto l'orizzonte

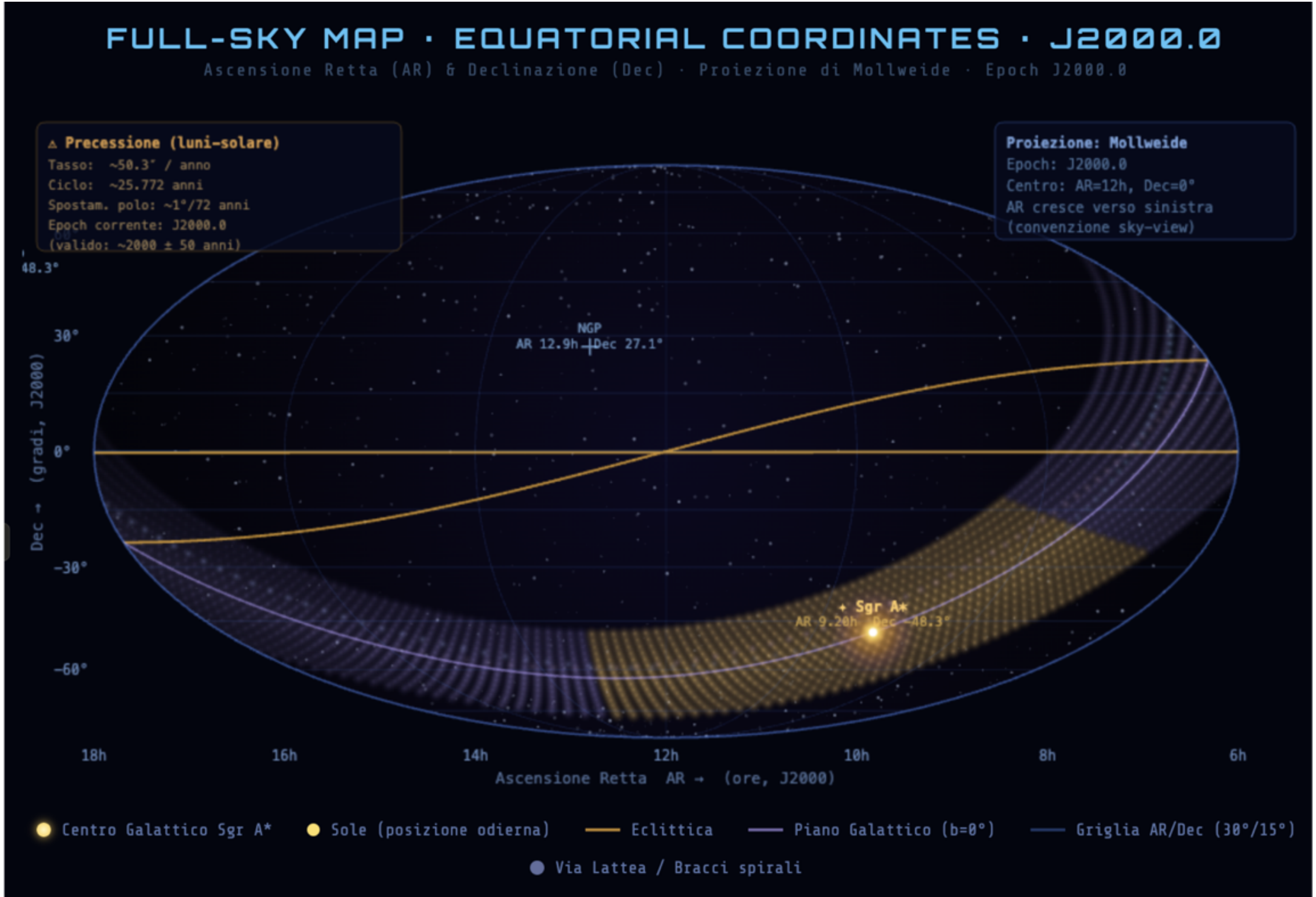
EQUATORIAL SYSTEM

δ = Declination (-90,+90): rises from the equatorial plane to the North (geographical)

α = Right Ascension (0-24h): from Gamma-Aries (ɣ) eastward

It depends on Earth precession of the equinox… so not stable like the Galactic Coordinate System





FULL-SKY MAP · EQUATORIAL COORDINATES · J2000.0
Ascensione Retta (AR) & Declinazione (Dec) · Proiezione di Mollweide · Epoch J2000.0
⚠ Precessione (luni-solare)
Tasso: ~50.3″ / anno
Ciclo: ~25.772 anni
Spostam. polo: ~1°/72 anni
Epoch corrente: J2000.0
(valido: ~2000 ± 50 anni)
Proiezione: Mollweide
Epoch: J2000.0
Centro: AR=12h, Dec=0°
AR cresce verso sinistra
(convenzione sky-view)
NGP
AR 12.9h Dec 27.1°
Sgr A*
AR 9.20h Dec -48.3°
48.3°
30°
0°
-30°
-60°
Dec → (gradi, J2000)
18h
16h
14h
12h
10h
8h
6h
Ascensione Retta AR → (ore, J2000)
Centro Galattico Sgr A*
Sole (posizione odierna)
Eclittica
Piano Galattico (b=0°)
Griglia AR/Dec (30°/15°)
Via Lattea / Bracci spirali

Local coordinates vs. Galactic coordinates

Local Coordinates (Altitude & Azimuth)

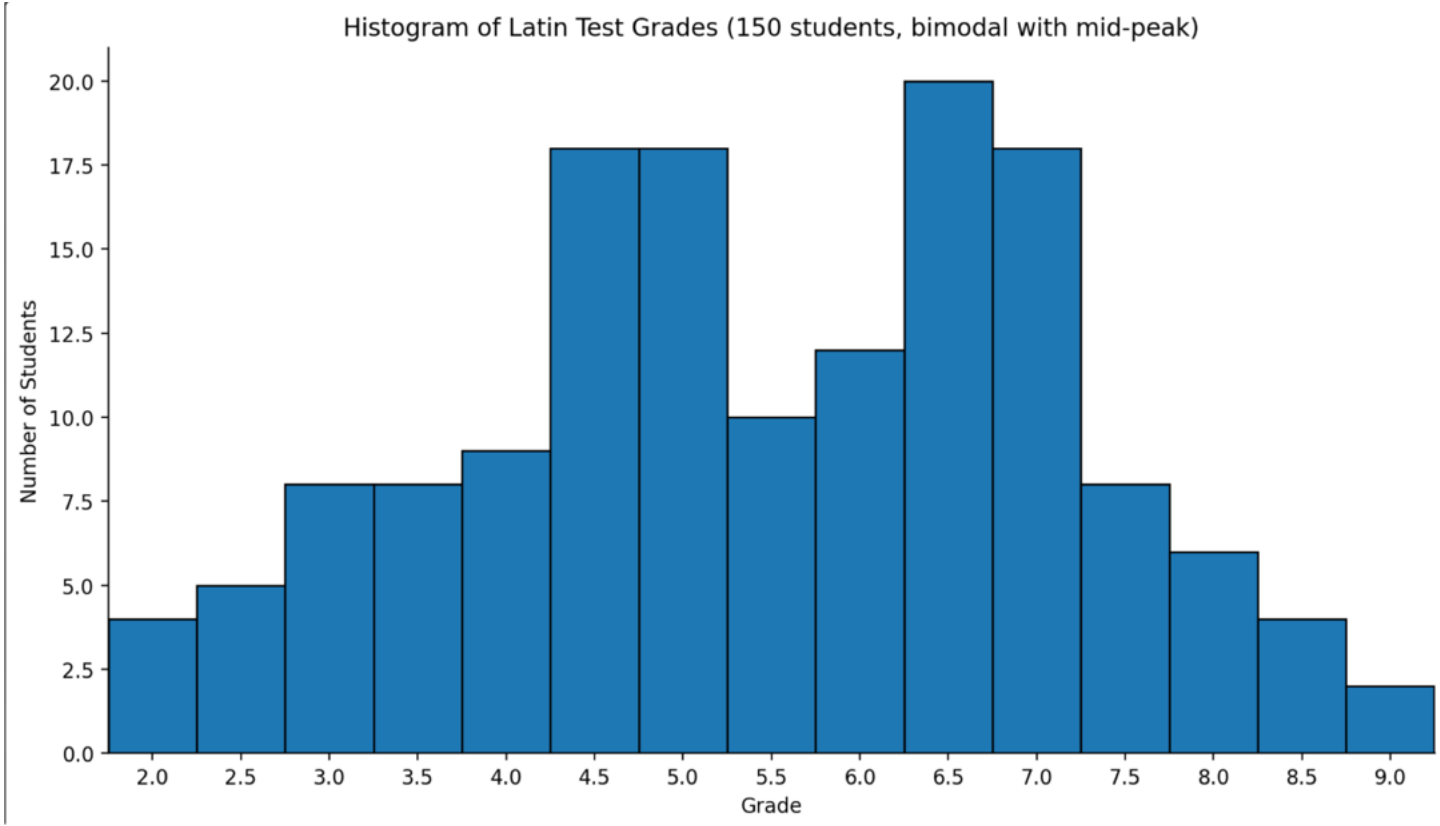
- Based on where you are standing on Earth
- **Dependant on:**
  - **Your location**
  - **The time of day**
- Suitable for:
  - Pointing a telescope
  - Saying "Look there, above the horizon!"
- Problem:
  - The same star has different coordinates for different observers/time
- They answer the question:
  - "Where is the object in my view right now?"

Galactic Coordinates (Galactic Lat/Lon)

- Based on the structure of the Milky Way
- **Independent of:**
  - **The observer**
  - **The time**
- Ideal for:
  - Making sky maps
  - Sharing scientific data
- Advantage:
  - Everyone measures the same coordinates
- They answer the question:
  - "Where is the object in the universe?"

PART 4
Statistics

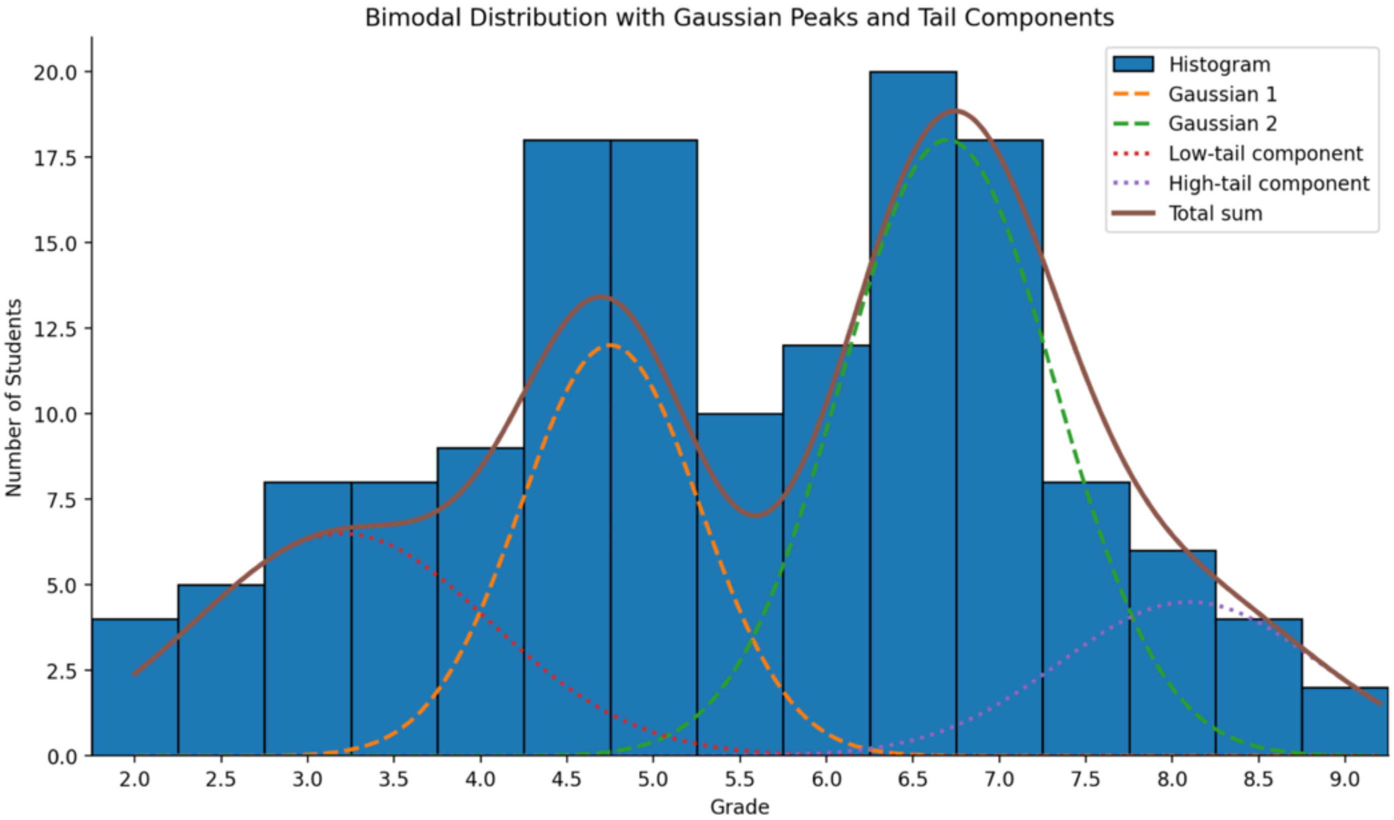
What we call a **distribution**: a histogram with the frequencies of a given "**observable**" (in this case the ***Grade*** *of a Latin Test* for 150 students)

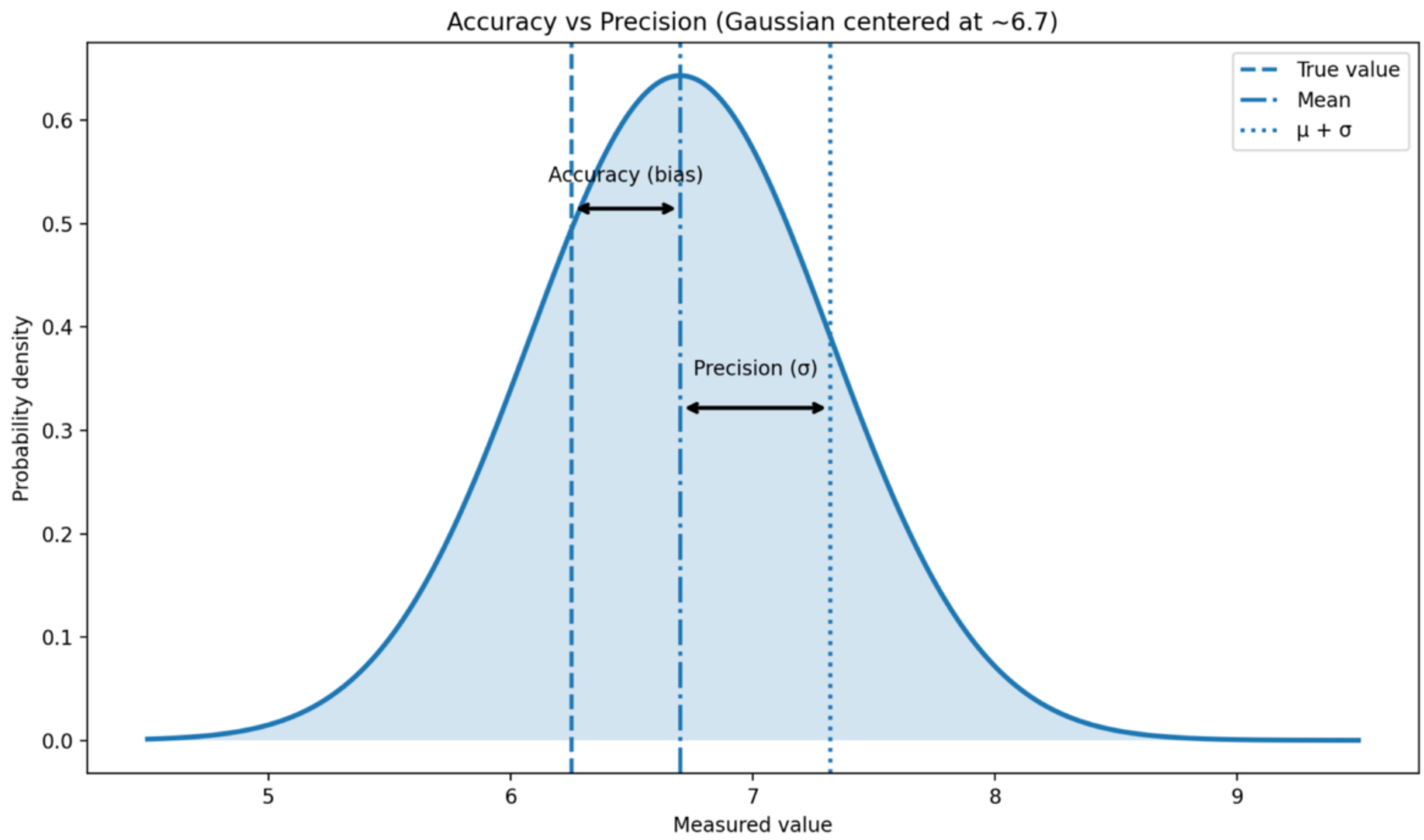


We can make assumptions about the laws (continuous lines) behind our measurements.
In this case we interpret that the students are split in 4 groups:

- Low-performing
- Average
- Satisfactory
- High-performing

Each group has its own **Probability function: the Gaussians** (in this case!)

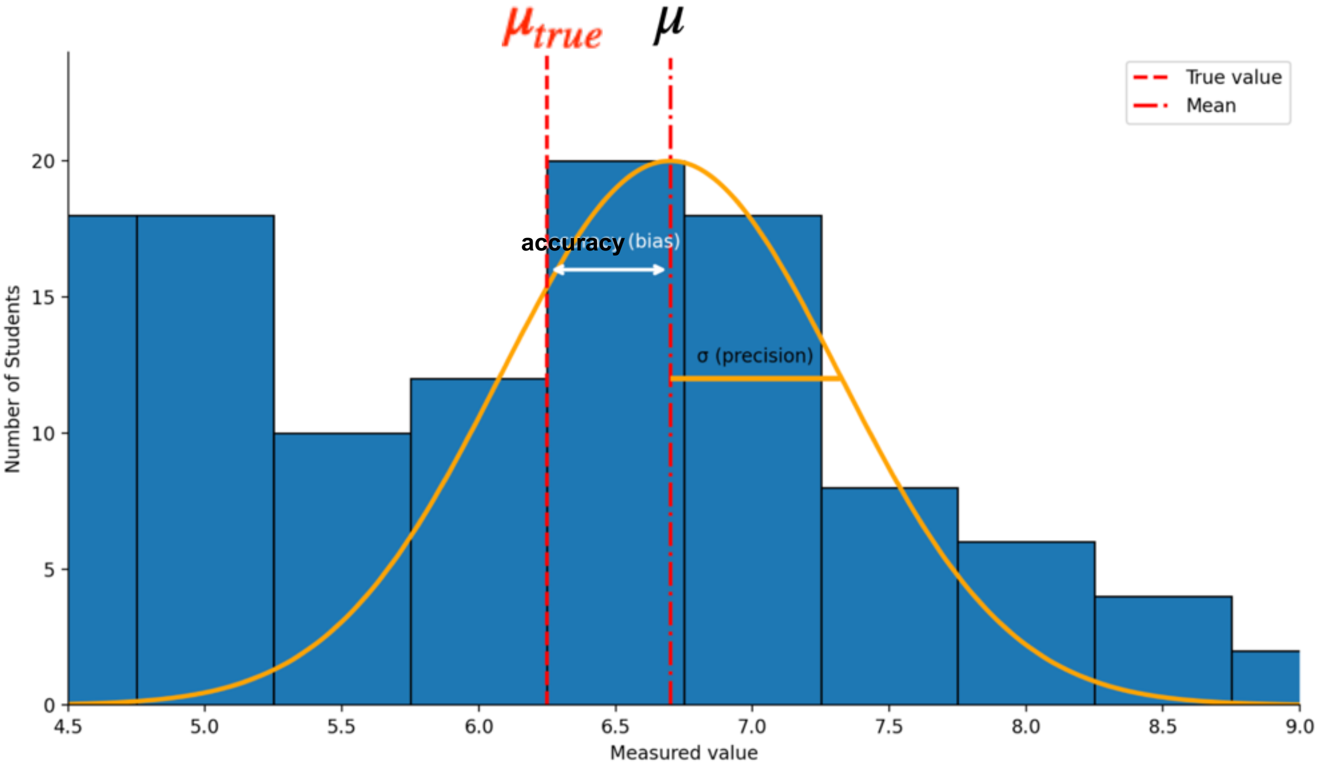



Accuracy vs Precision (Gaussian centered at ~6.7)
True value
Mean
μ + σ
Accuracy (bias)
Precision (σ)
Probability density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
5
6
7
8
9
Measured value

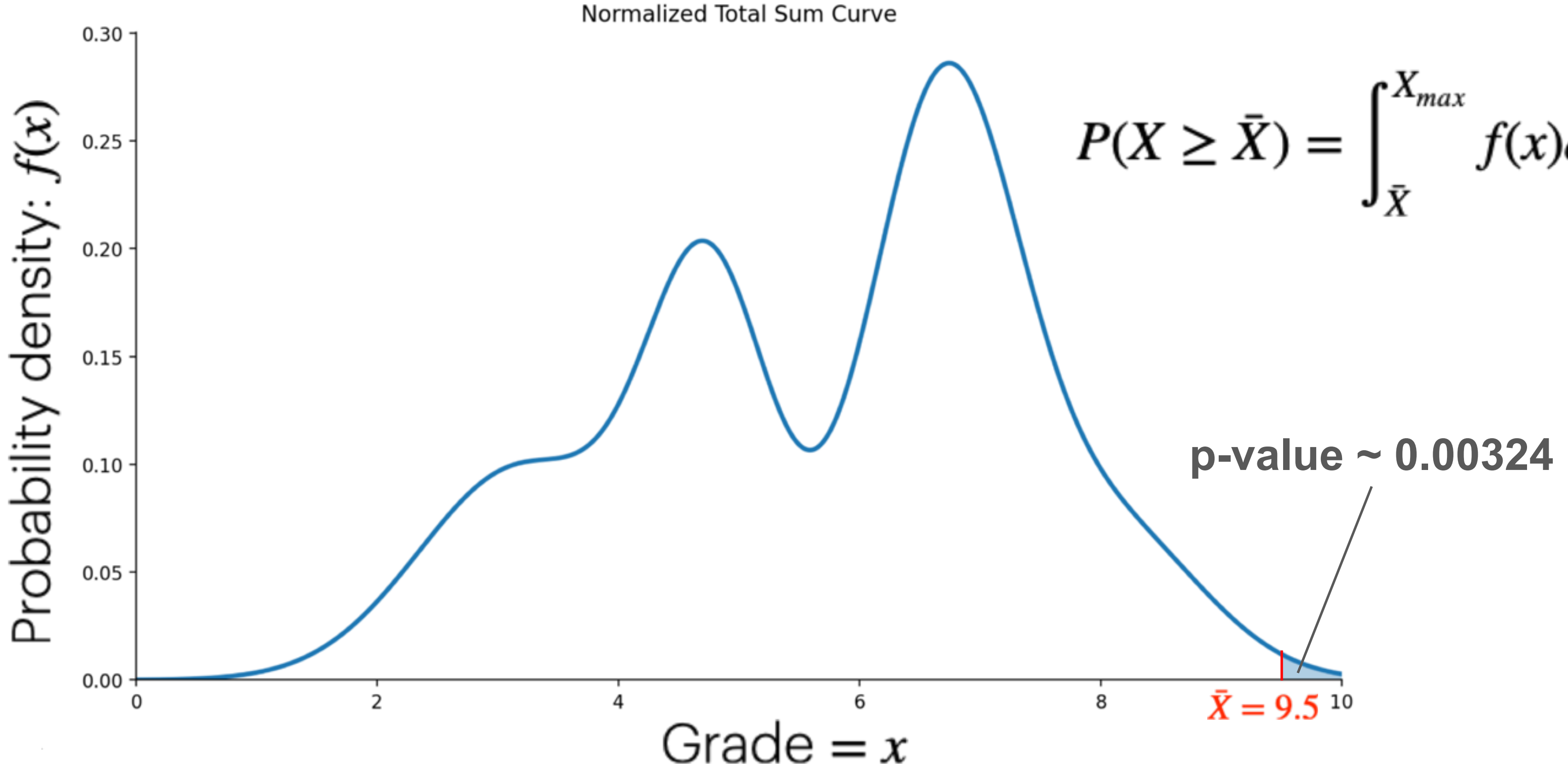
N = 150

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}$$

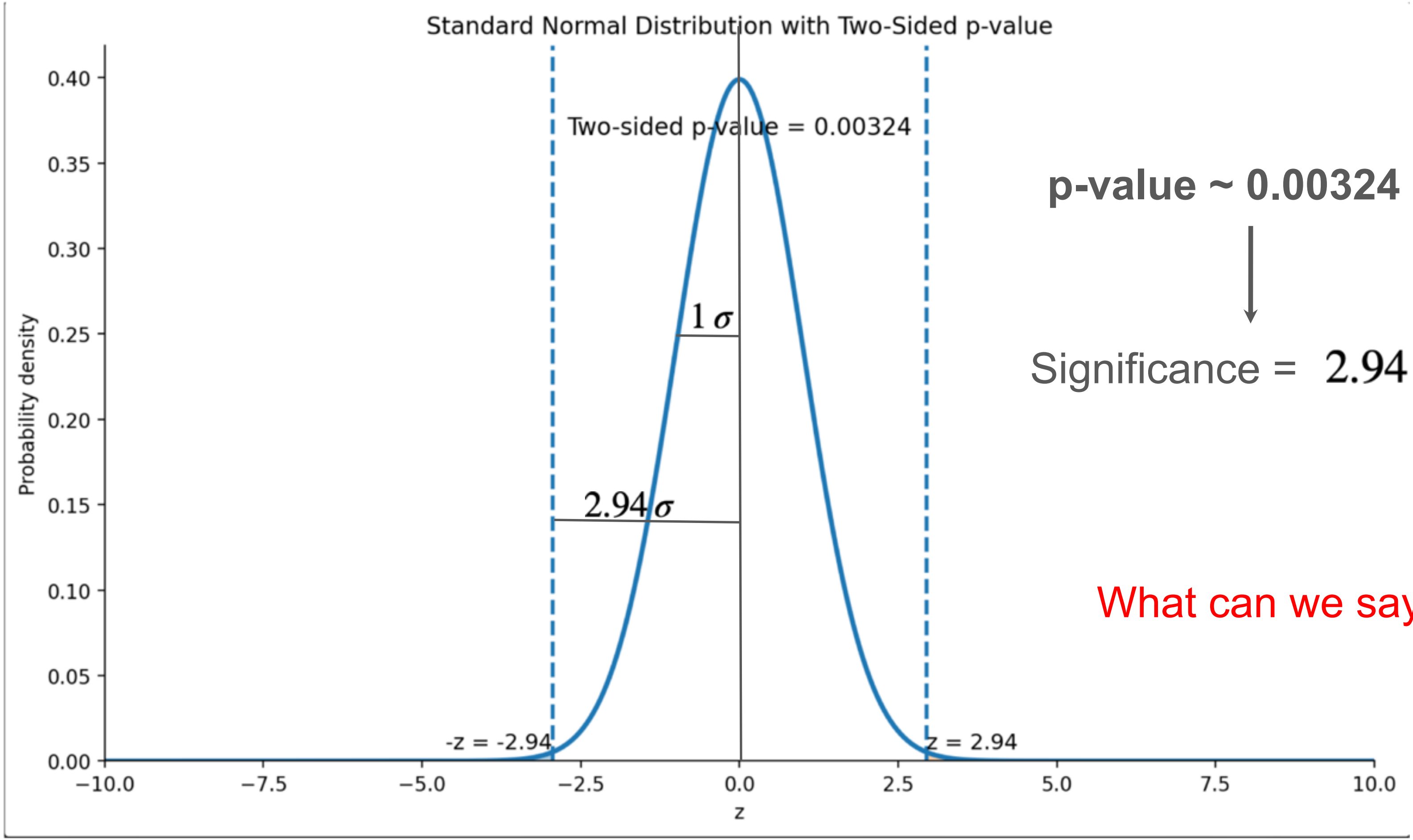


Now…: **one student gets *9.5***.
How likely is it, given this **hypothesis** for the probability distribution of students?
The question is: how is it probable to get a Grade ***equal or better than 9.5?***



$$P(X \geq \bar{X}) = \int_{\bar{X}}^{X_{max}} f(x)dx$$

So, can we claim that such a Grade=9.5 **is actually significant**?



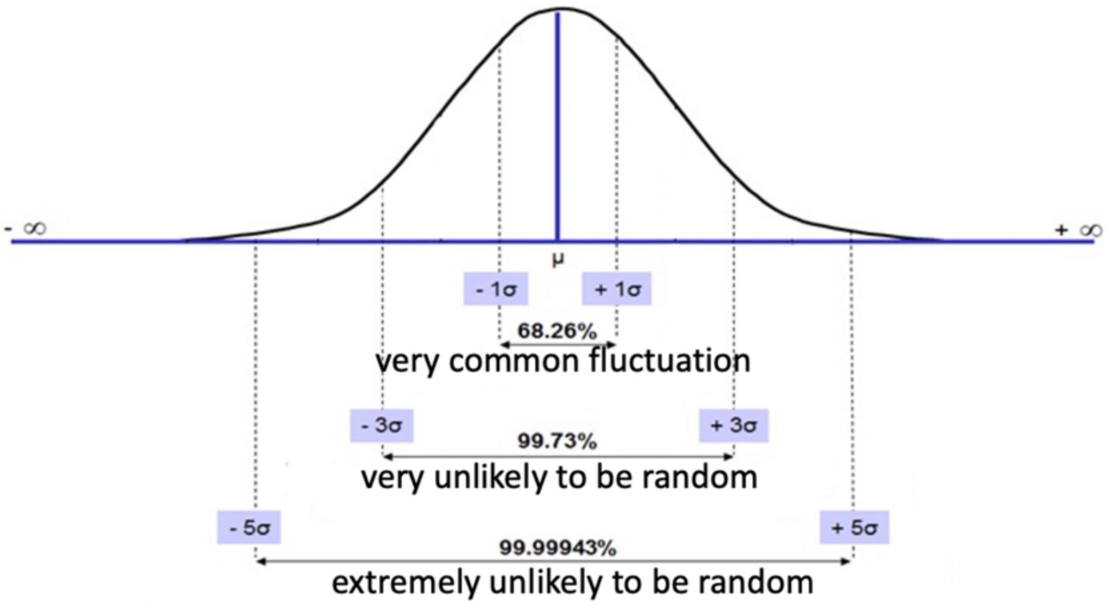
p-value ~ 0.00324

Significance = $2.94\,\sigma$

What can we say of that?

What is “statistical significance”?

- When we measure something, the result is never exactly the same
- Small differences come from:
  - Random fluctuations
  - Measurement errors
- Statistics helps us decide:
  - Is this result real, or just random chance?



3σ → evidence

5σ → discovery

Introduction to the skymap exercise

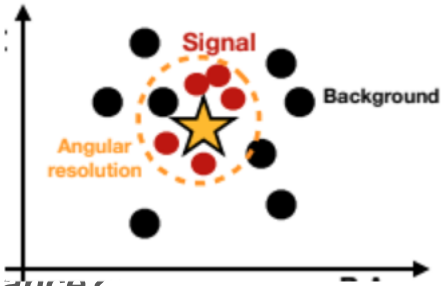
Significance estimation with ON/OFF zones

- Neutrino detectors see many events
- Most of them are background (not from a source)
- We need a way to tell if a source is really producing neutrinos

- ***ON zone***:
  - a 0.3° circle around the astrophysical source
  - If the source emits neutrinos, extra events should appear here
  - In the ON Zone: mix of source + background

- ***OFF zone***:
  - the rest of the sky outside the ON zone (at the same altitude)
  - It is assumed to contain:
    - Only background neutrinos
    - No contribution from the source
- It tells us how many background events we expect

Significance estimation with ON/OFF zones



- Compare ON and OFF
  - Count events in the ON zone
  - Use the OFF zone to estimate the background
  - **Ask: "Are there more events in the ON zone than those expected *by chance*?**
- Statistical significance
  - If ON zone events ≈ expected background → no detection
  - If ON zone events > background by 3σ → evidence of a signal
  - If ON zone events > background by 5σ → discovery of a source

$$p\text{-value} = P(n \geq n_{\text{on}} \mid N, H_0) = \sum_{k=n_{\text{on}}}^{N} \binom{N}{k} \left(\frac{\alpha}{1+\alpha}\right)^k \left(\frac{1}{1+\alpha}\right)^{N-k}$$

$$\alpha = \frac{A_{\text{on}}}{A_{\text{off}}} \cdot \frac{T_{\text{on}}}{T_{\text{off}}} \cdot \frac{\epsilon_{\text{on}}}{\epsilon_{\text{off}}}$$

$A_{\text{on}}, A_{\text{off}}$: effective area (or solid angle)

$T_{\text{on}}, T_{\text{off}}$: exposure time

$\epsilon_{\text{on}}, \epsilon_{\text{off}}$: detection efficiency (or acceptance)

Significance estimation with likelihood

Why go beyond ON/OFF?

- ON/OFF counts how many events we see
- But it ignores where exactly the events are inside the region
- Likelihood uses all the information we have

- Signal hypothesis
  - Cosmic neutrinos come from the source direction
  - Signal centered on the source
  - Signal width → KM3NeT angular resolution
  - Events closer to the source are more likely signal

- Background hypothesis
  - Background events come from everywhere
  - Their directions are uniformly spread
  - Same probability everywhere in the region

Significance estimation with **likelihood**

- For each detected event, we ask:

"Is this event more likely signal or background?"

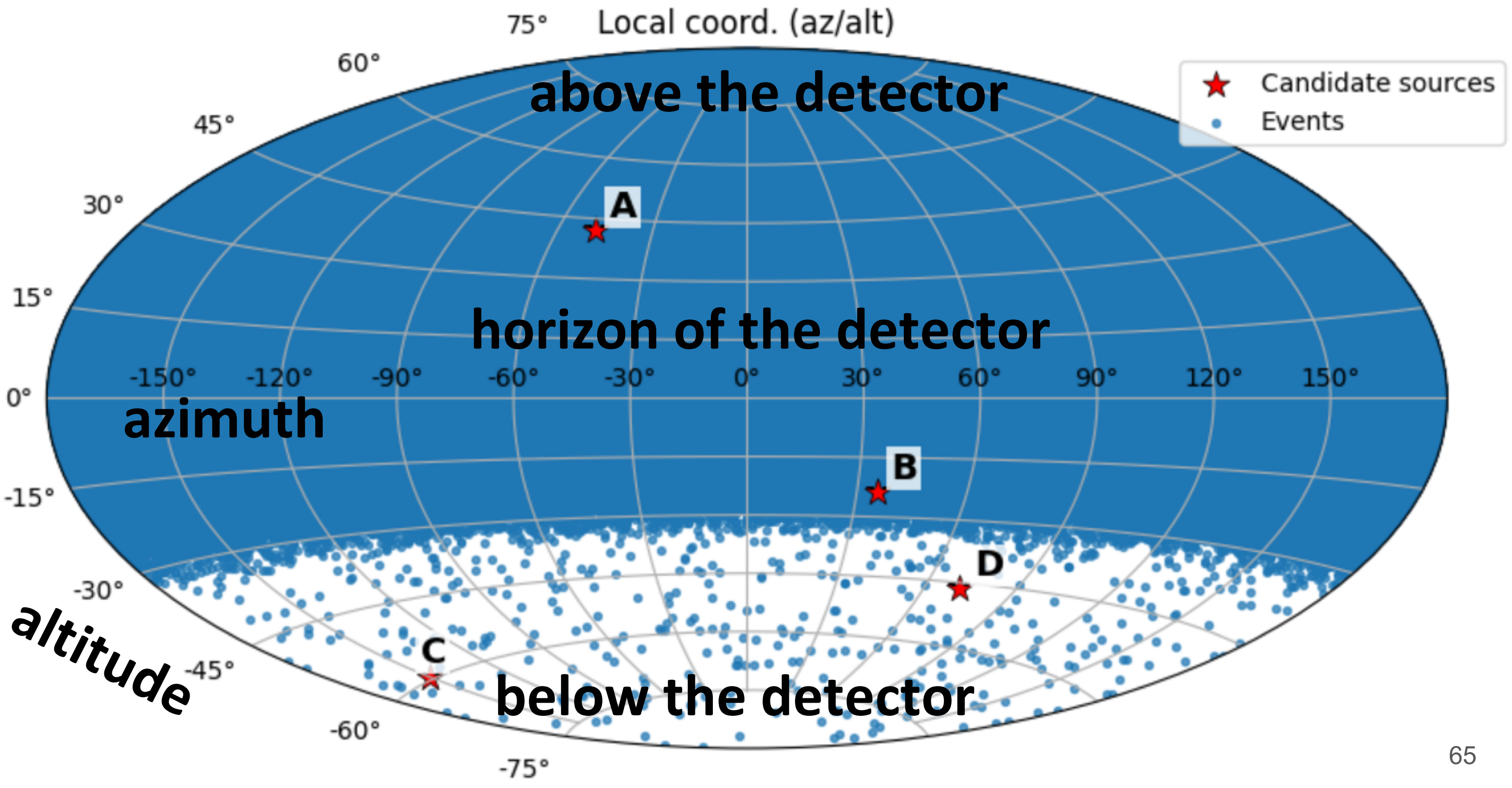
- We combine all events together

- We estimate the amount of signal that best explains the data

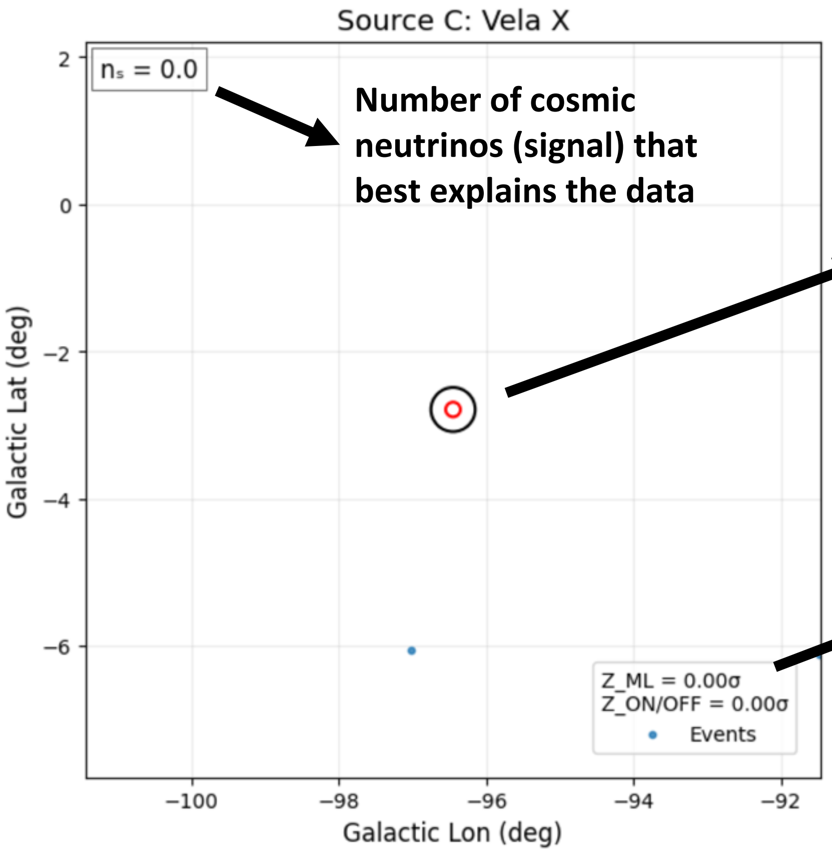
- We compute the significance of our observation

$$\ln \mathscr{L}(\mu) = \sum_{i=1}^{N_{\text{bin}}} \left[n_i \ln(B_i + \mu S_i) - (B_i + \mu S_i) - \ln(n_i!) \right]$$

How to interpret the plots



How to interpret the plots



Number of cosmic neutrinos (signal) that best explains the data

Red circle: 0.1° distance from the source
Black circle: 0.3° distance from the source

Significance computed with the likelihood method
Significance computed with the ON/OFF method

3σ → evidence
5σ → discovery!

GRAZIE !