Handling of Neutrino Telescope Data

Kay Graf, ECAP, Friedrich-Alexander-Universität Erlangen-Nürnberg

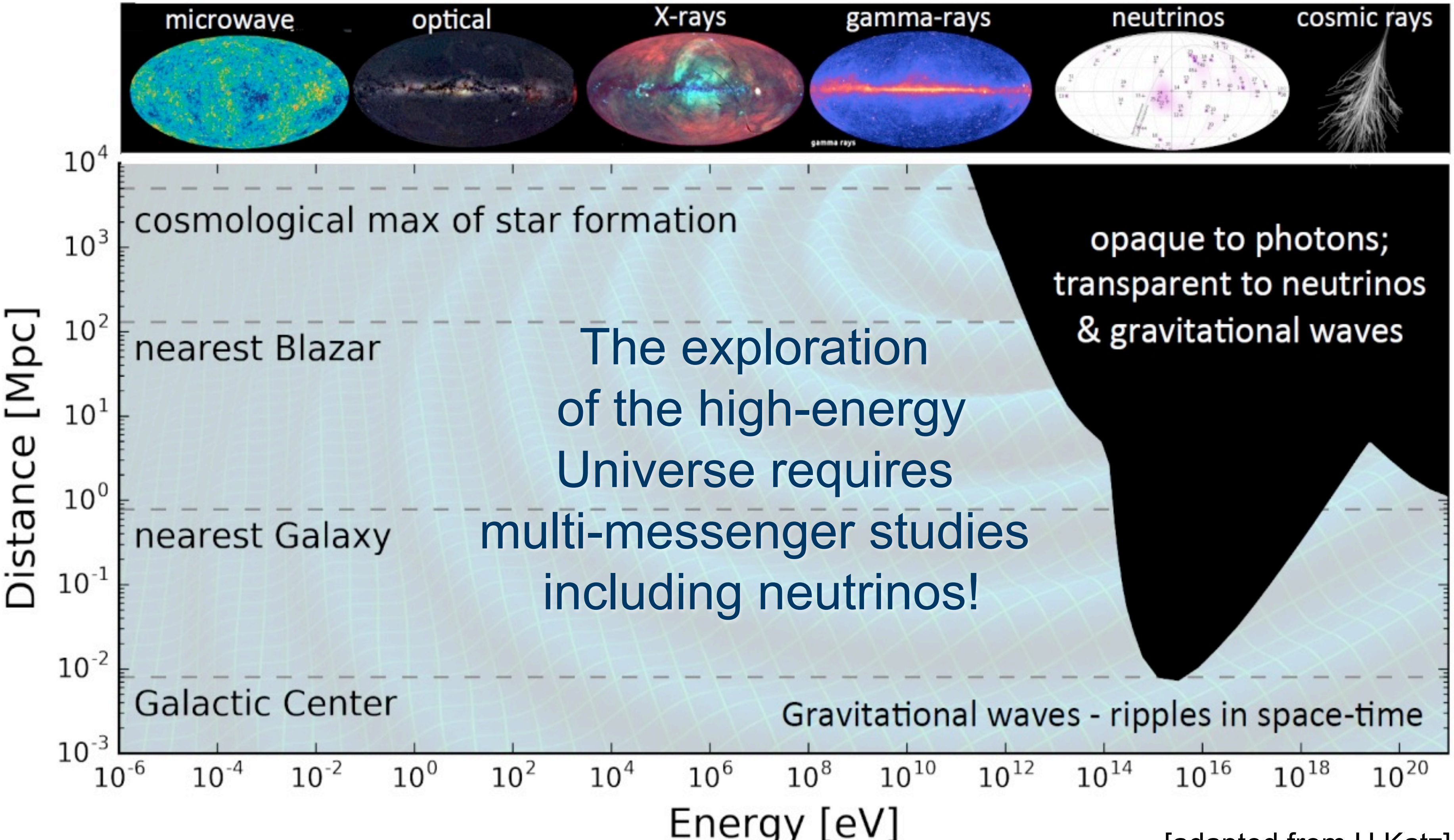
The Science Cloud
711. WE-Heraeus-Seminar
Physikzentrum Bad Honnef, 12th – 15th of January 2020

Overview

- Introduction – setting the scene
- Neutrino Telescopes
- Data Handling – from a cloud perspective
- Conclusion

Introduction –
The high-energy Universe

Different messengers → Multi-messenger



[adapted from U.Katz]

Neutrino astronomy: where are we?

✓ High-energy cosmic neutrinos discovered by IceCube

✓ Neutrino/X-ray/gamma-ray coincidence (IceCube): First hint of a neutrino source

– Neutrinos from Galactic accelerators

– “Real neutrino astronomy”

→ We need more statistics, increasingly precise data, and full sky coverage

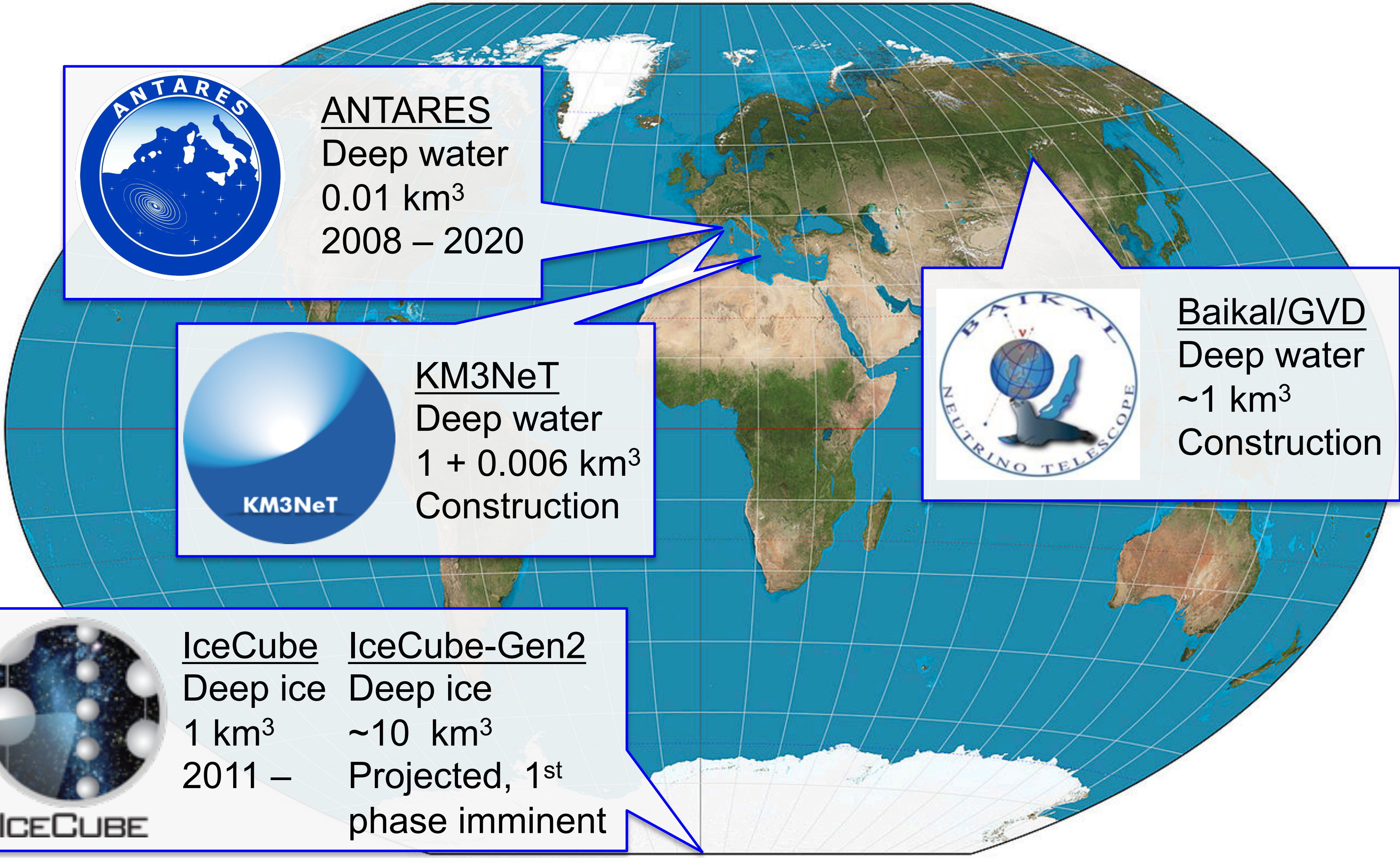
Rich science program:

→ Neutrino astronomy
→ Particle physics
→ Dark matter searches
→ Exotics
→ Earth and Sea sciences, glaciology, …

This talk:

Data handling for detecting GeV to PeV neutrinos using Cherenkov detectors in deep water or ice …

The HE neutrino telescope world map



ANTARES
Deep water
0.01 km^3
2008 – 2020

KM3NeT
Deep water
1 + 0.006 km^3
Construction

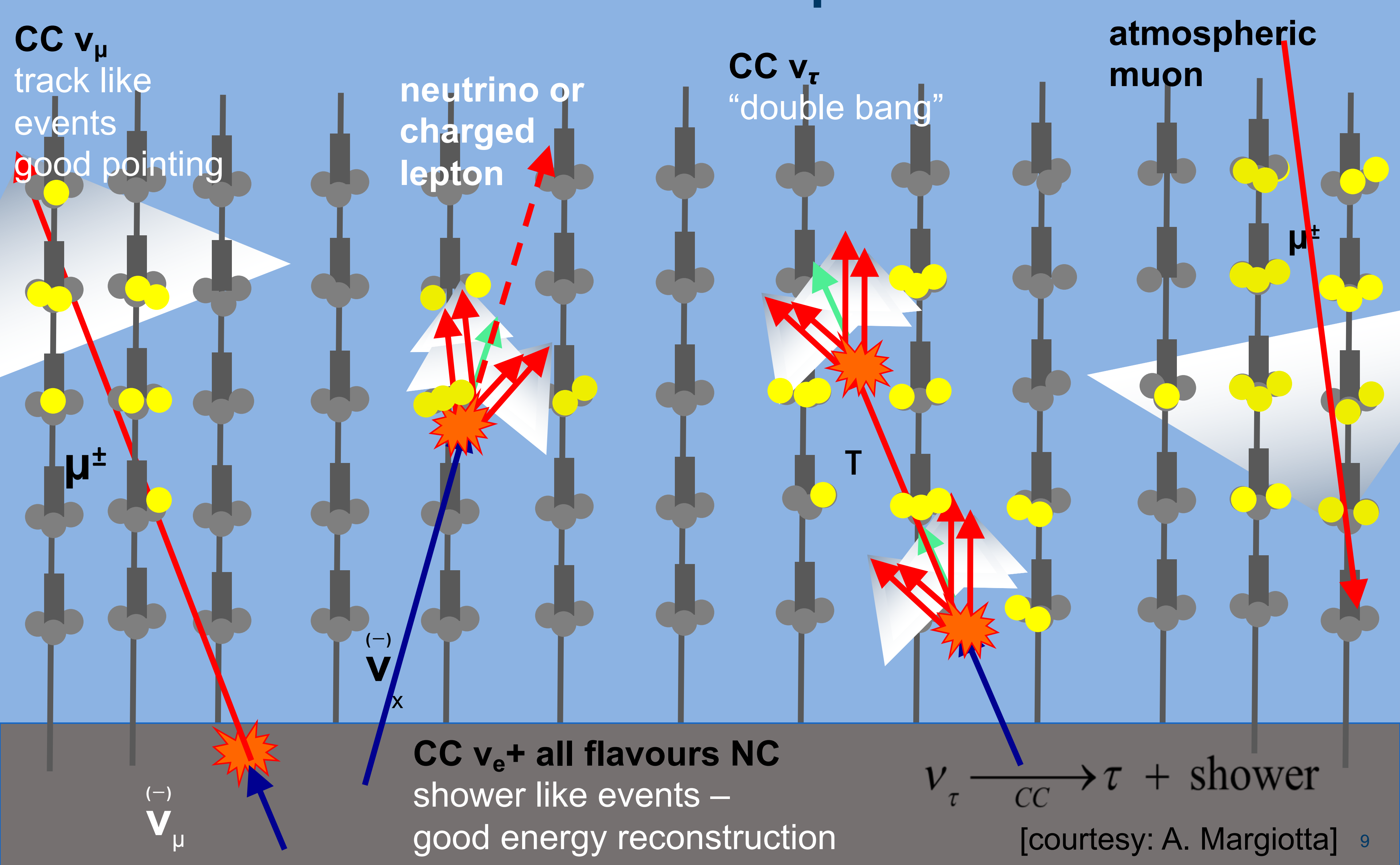
Baikal/GVD
Deep water
~1 km^3
Construction

IceCube
Deep ice
1 km^3
2011 –

IceCube-Gen2
Deep ice
~10 km^3
Projected, 1st phase imminent

Neutrino Telescopes – Working Principle and Experiments

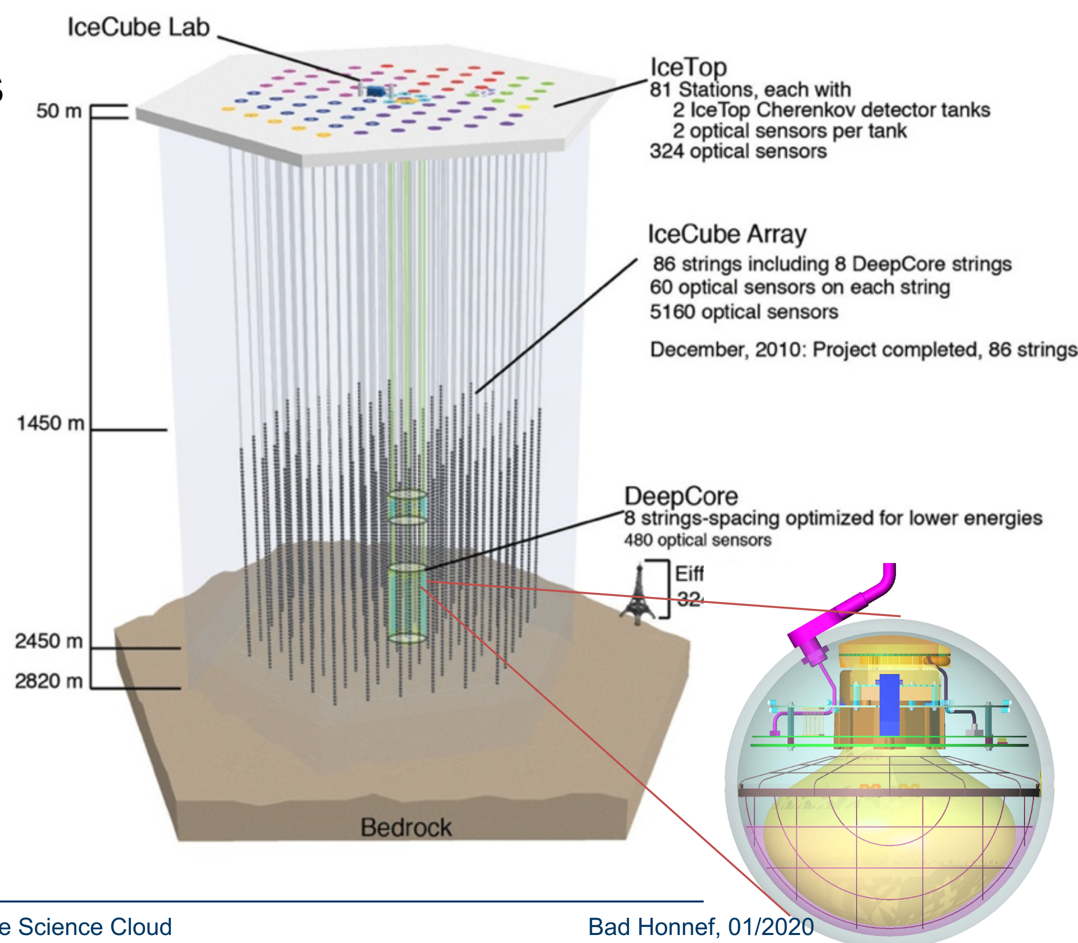
Events in a neutrino telescope



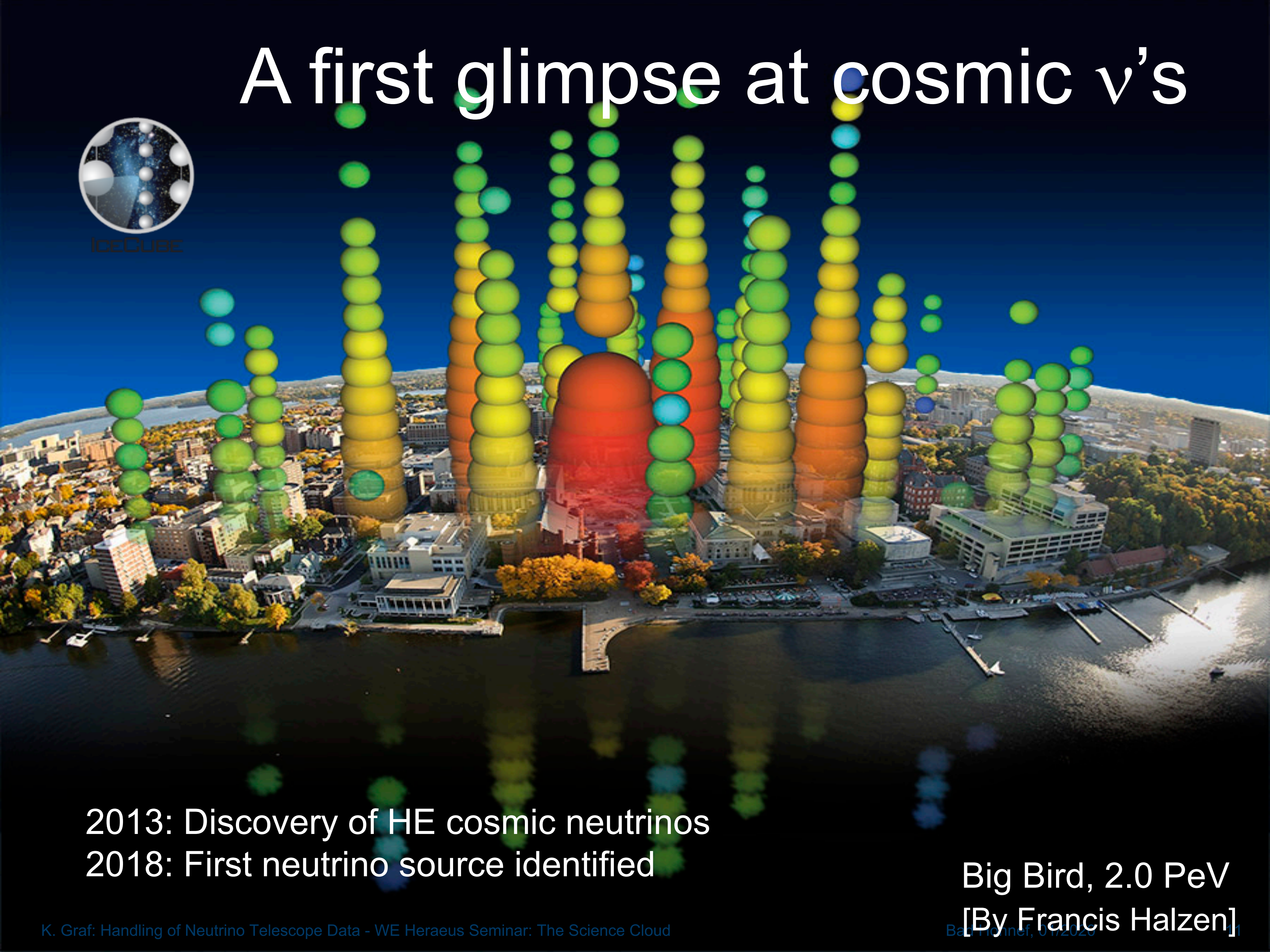
[courtesy: A. Margiotta]

IceCube: Most sensitive ν telescope to date

- 86 strings altogether
  - 125 m horizontal spacing
  - 17 m vertical distance between Optical Modules
  - **1 km³ instrumented volume**, depth 2450m
- Deep Core
  - densely instrumented region in clearest ice
- **Completed in Dec. 2010**
- **Extremely stable and efficient operation since then**



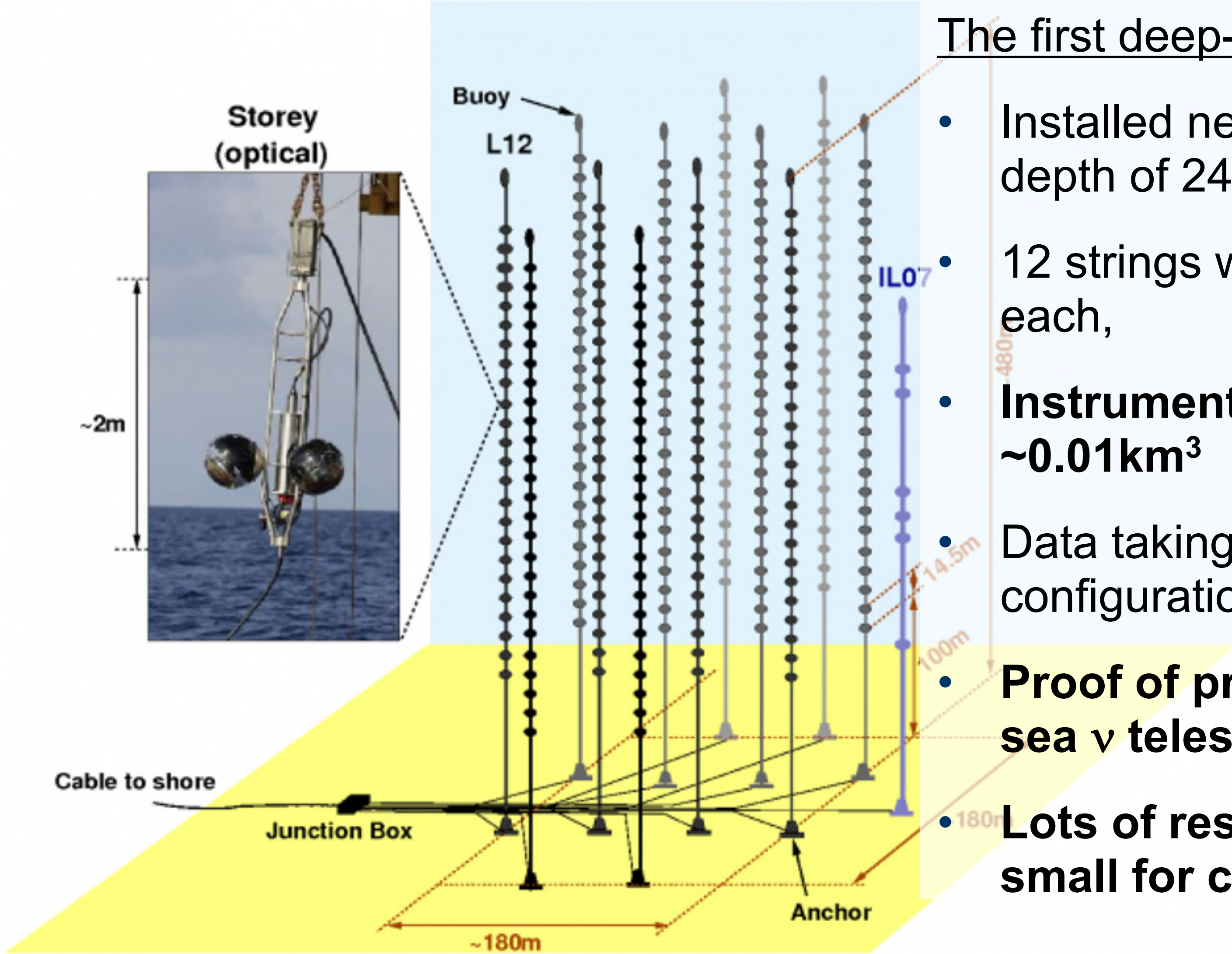
A first glimpse at cosmic ν's



2013: Discovery of HE cosmic neutrinos
2018: First neutrino source identified

Big Bird, 2.0 PeV
[By Francis Halzen]

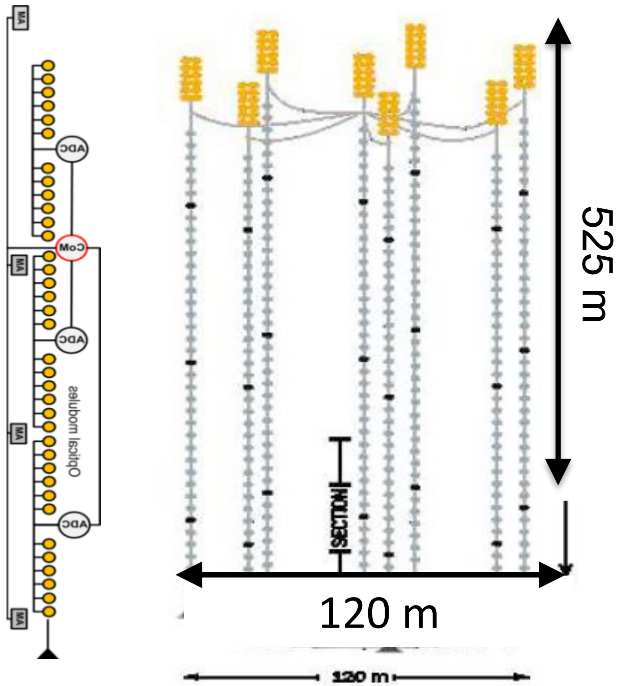
Currently taking data: ANTARES …



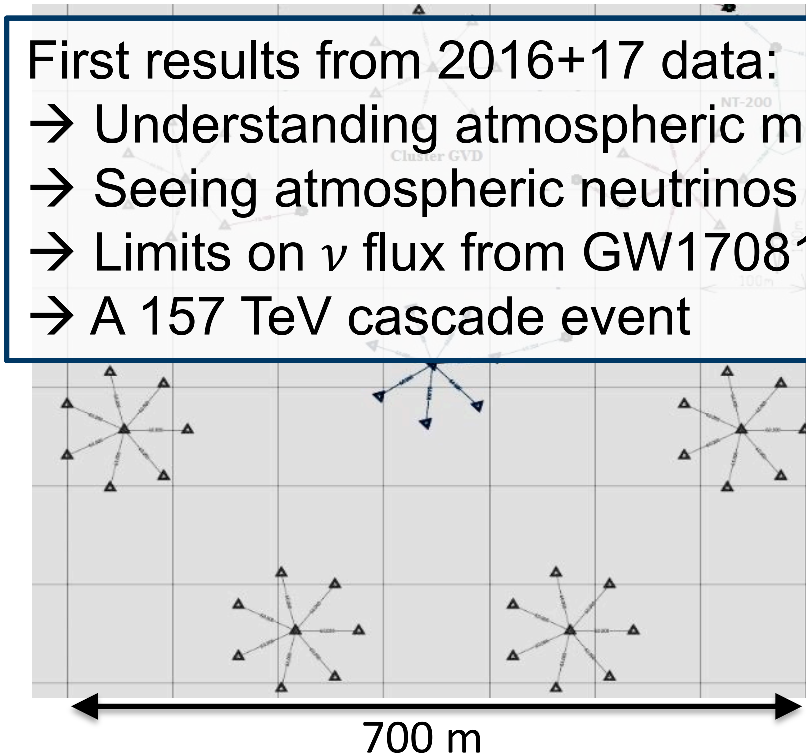
The first deep-sea ν telescope

- Installed near Toulon at a depth of 2475m
- 12 strings with 25 storeys each,
- **Instrumented volume ~0.01km^3**
- Data taking in full configuration since 2008
- **Proof of principle of deep-sea ν telescope**
- **Lots of results – but (too) small for cosmic neutrinos**

Baikal-GVD

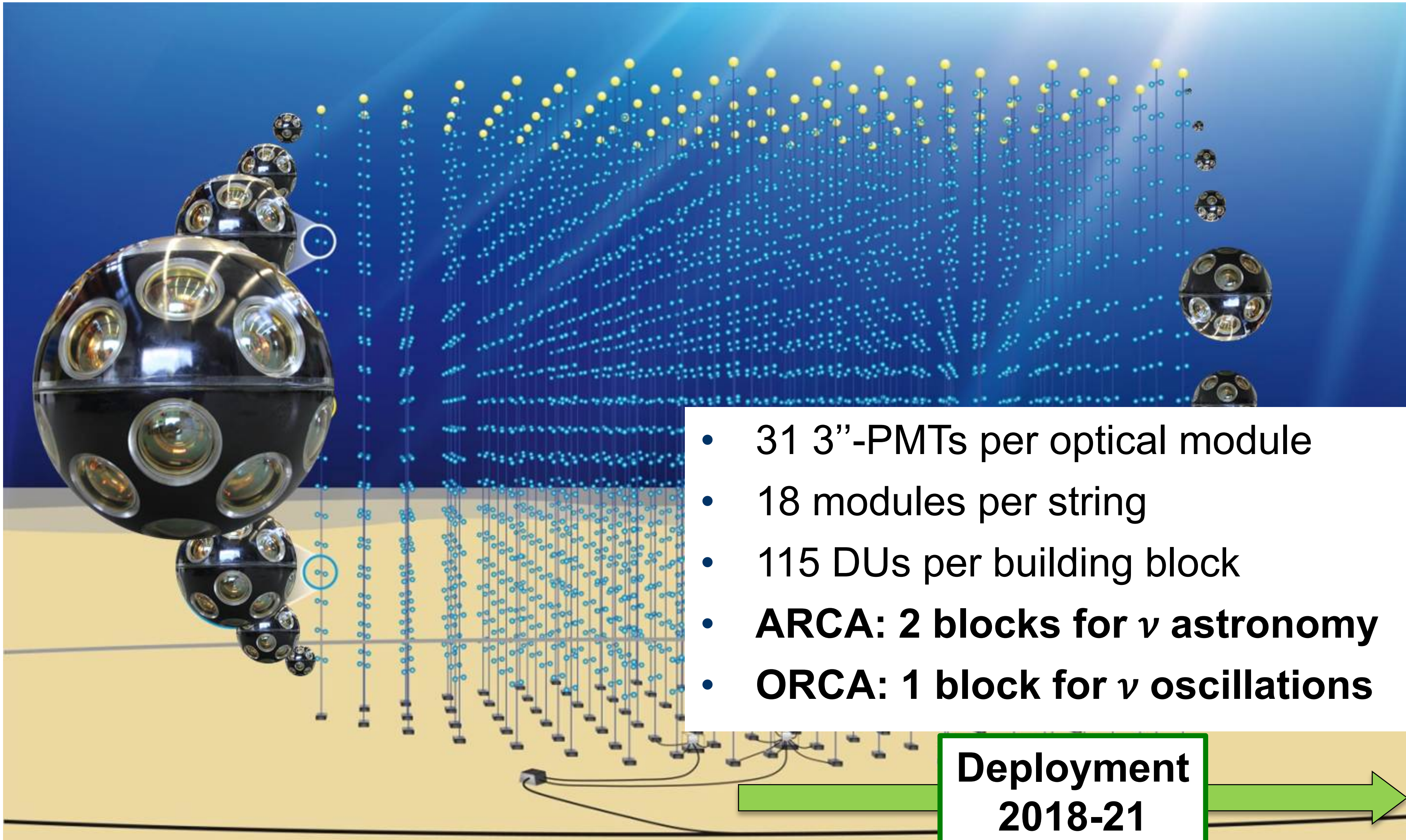


First results from 2016+17 data:
→ Understanding atmospheric muons
→ Seeing atmospheric neutrinos
→ Limits on ν flux from GW170817
→ A 157 TeV cascade event



- Project to construct a **Gigaton (=km^3) detector in Lake Baikal**
- Phase 1 (GVD-1): 8 clusters (see figure), 0.4 km^3
- 3 clusters operational, 1-2 more clusters to be deployed per season
- Commissioning, calibration, sensitivity studies in progress
- Final goal: 27 clusters, 1.5 km^3

The future in the Mediterranean: KM3NeT



- 31 3''-PMTs per optical module
- 18 modules per string
- 115 DUs per building block
- **ARCA: 2 blocks for ν astronomy**
- **ORCA: 1 block for ν oscillations**

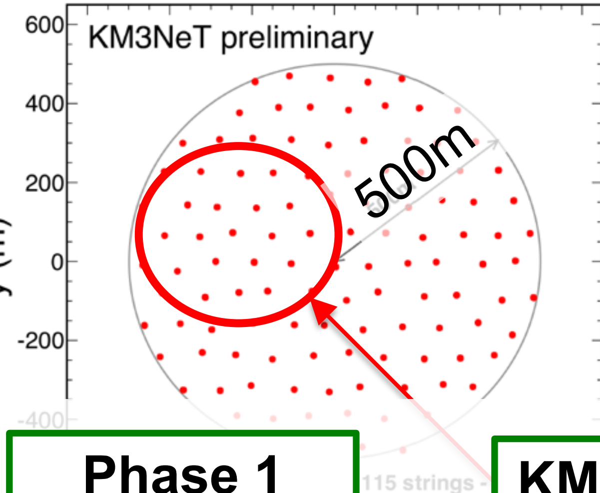
Deployment 2018-21

KM3NeT 2.0

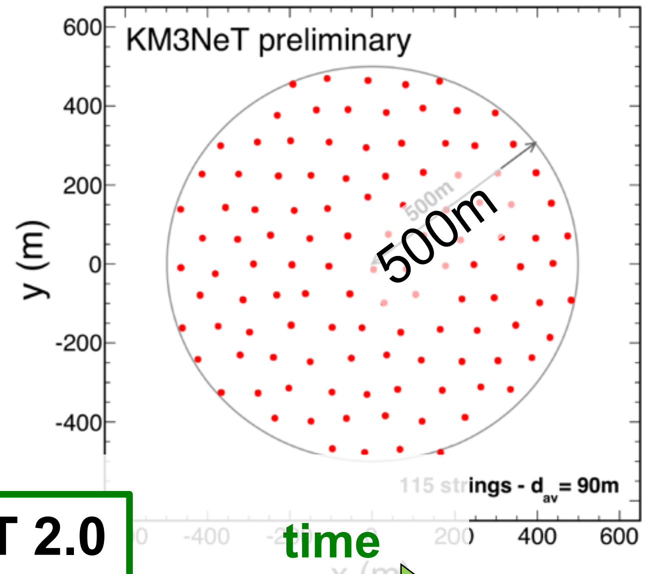
ARCA =

Astroparticle Research with Cosmics in the Abyss

Vertical DOM distance = 36 m



+



Phase 1 Deployment 2018-20

KM3NeT 2.0 Deployment 2020-25

time

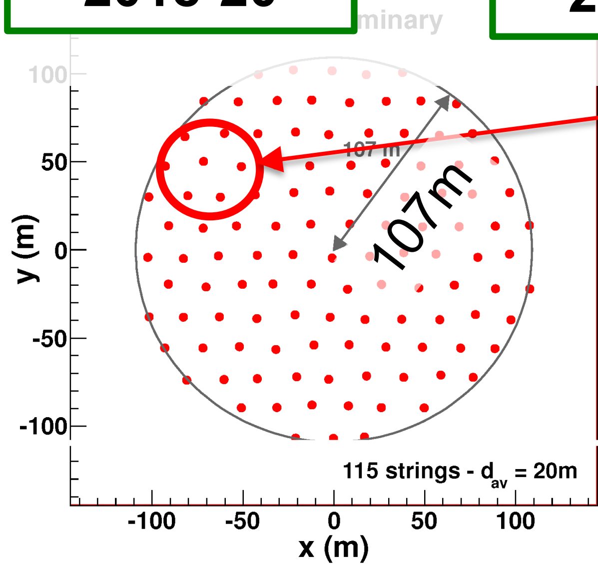
Phase 1 (fully funded)

Phase 2 partially funded

ORCA =

Oscillation Research with Cosmics in the Abyss

Vertical DOM distance = 9 m

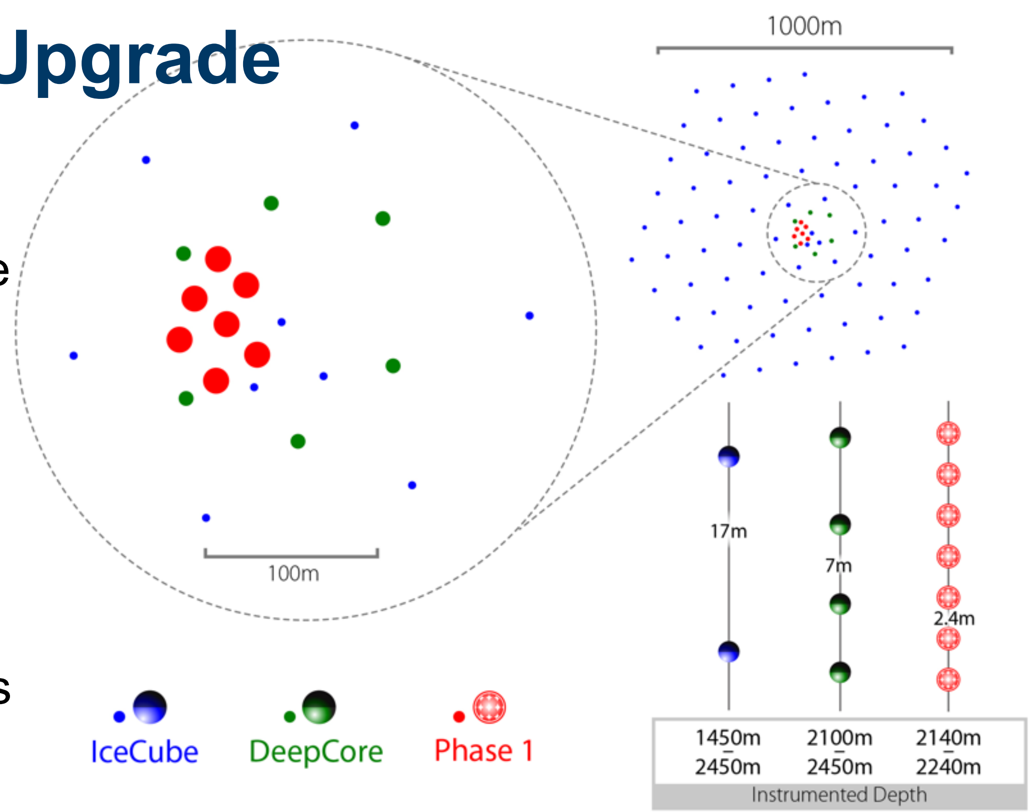


KM3NeT 2.0 Letter of Intent: arXiv:1601.07459 and J.Phys. G43 (2016) 084001

IceCube: next step = Upgrade

IceCube-Upgrade

- **7 additional strings** in Deep Core domain, **densely instrumented**
- Objectives:
  - GeV neutrinos: τ appearance, Dark Matter, …
  - Improved understanding of ice properties → better precision, reduced systematic uncertainties
  - Opportunity to test new hardware developments



Array	String Spacing	Module Spacing	Modules / String
IceCube	125 m	17 m	60
DeepCore	75 m	7 m	60
Upgrade	20 m	2 m	125

Deployment 2022/23

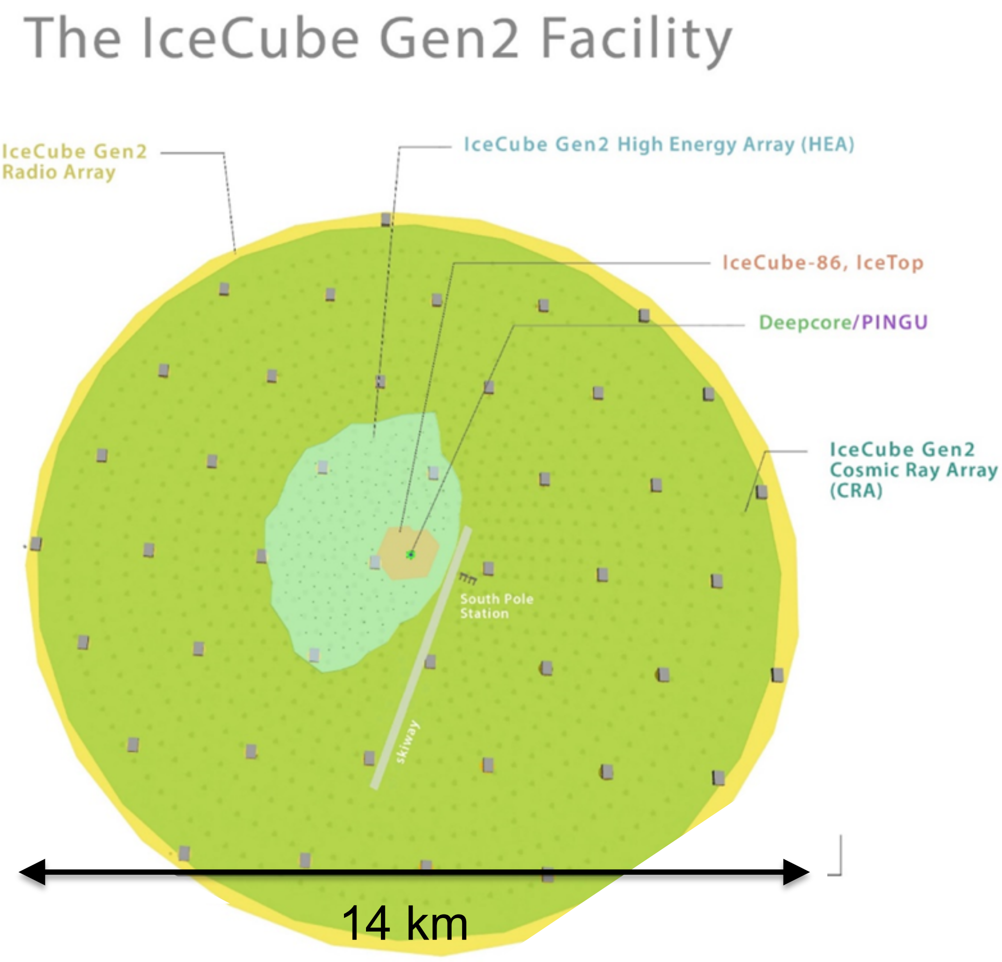
time →

IceCube Gen2

- **Next-generation neutrino observatory** at South Pole, with
  - High-energy deep-ice detector (High-energy array, HEA)
  - Cosmic-ray and veto surface array (CRA)
  - Radio array (RA)
  - High-density core for low-energy neutrinos (PINGU)
- Funding application expected in NSF MREFC scheme (~2020)

Deployment 2025-31

time



Data Handling of Neutrino Telescopes from a Science Cloud Perspective

Neutrino Telescopes: input to data handling

- Sparse, 3-dimensional arrays of photo sensors
  → sparse information
  - Lowest-level data type is single/multiple-photon **hits** as ntuple of:
    - Position and orientation of photo sensor – cm/degree precision
    - Time of arrival of photon(s) – ns precision
    - Photo sensor "charge" (integrated charge, time-over-threshold, waveform)
  - Aggregated type:
    - **Events** (position, direction, time, energy, quality information) + derived parameters (e.g. hit pattern)
    - Can include hits (dependent on data level)
- Different configurations and environments
  (but basically the same challenges)
  - Detector calibration
  - Background suppression

Data Handling: General Scheme

- (Classical) **tier-like structure**, mixed access: GRID, direct (batch), new: some tasks in the cloud; driven by HPC resources available to the experiments

Data Reduction

Tier	Computing Facility	Processing steps
Tier-0	at detector site	Triggering (in software or hardware), online-calibration, quasi-online reconstruction, alert handling
Tier-1	computing centres	calibration, reconstruction, simulation
Tier-2	local computing clusters	development, simulation, analysis

- Data distribution between the computing centers and data processing chains differ by experiment (based on resource providers, but also "home-made" code)

Some Rough Resource Estimates

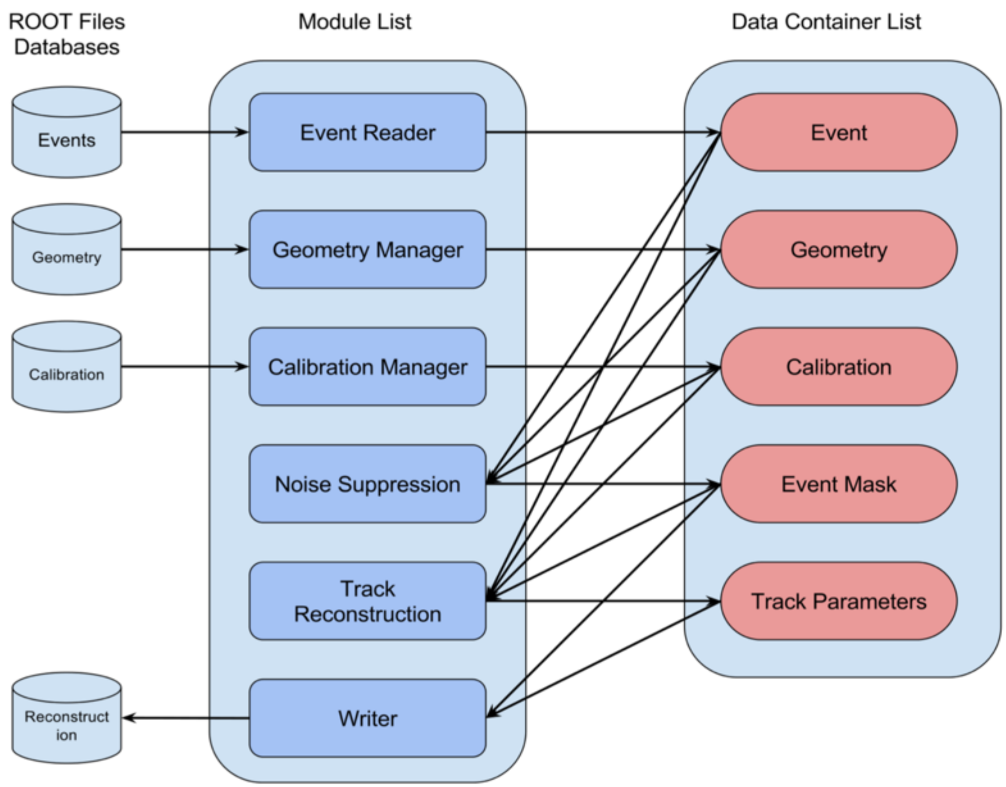
- Channels: 10k – 100k optical sensors
- Background: few 100 Hz – 10 kHz per channel
- Raw data rate: few 1 – few 10 GB/s
- Event rate: few 100 Hz ($\rightarrow$ few 10 MB/s)
- Necessary resources (for standard tasks including simulation, calibration, reconstruction):
  - CPUs: **few 1k to few 10k cores**
  - storage: **PBs/year** (disk and tape archive)
  - +burst and development activities
  - Analysis on aggregated data **$\mathcal{O}$(TB/year)**
- **Not so Big Data…**

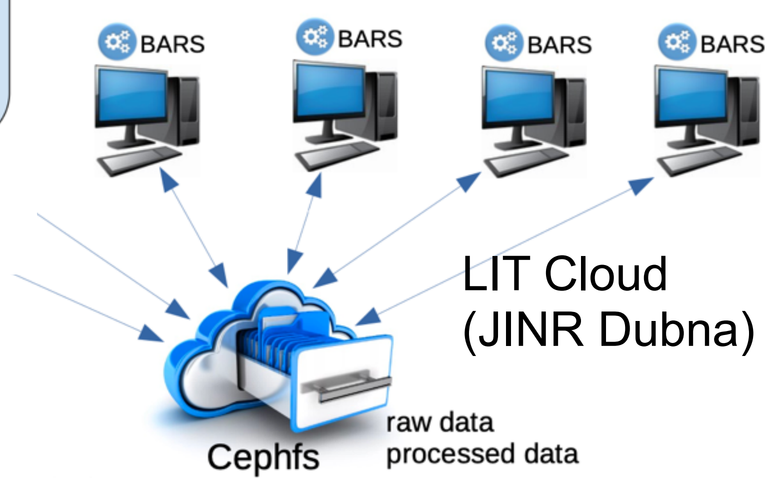
Data Management: General Scheme

- Making use of **existing, established** (mostly pledged) **HPC/computing resources**
- Following common practices:
  - For (meta-)data formats and storage (databases, HPSS, no large cloud storage as too expensive)
  - central services like software repository (CI/CD systems, e.g. containerized → ”cloud-ready”)
  - data processing by specialized service group within collaborations
  - data access: via WAN access tools
- One important goal:
  reproducibility (rather re-processibility) of all scientific results and data usability over full time of experiment
  (+ at least 10 years after end of operation)

Baikal-GVD Data Management

BARS (Baikal Analysis and Reconstruction Software)





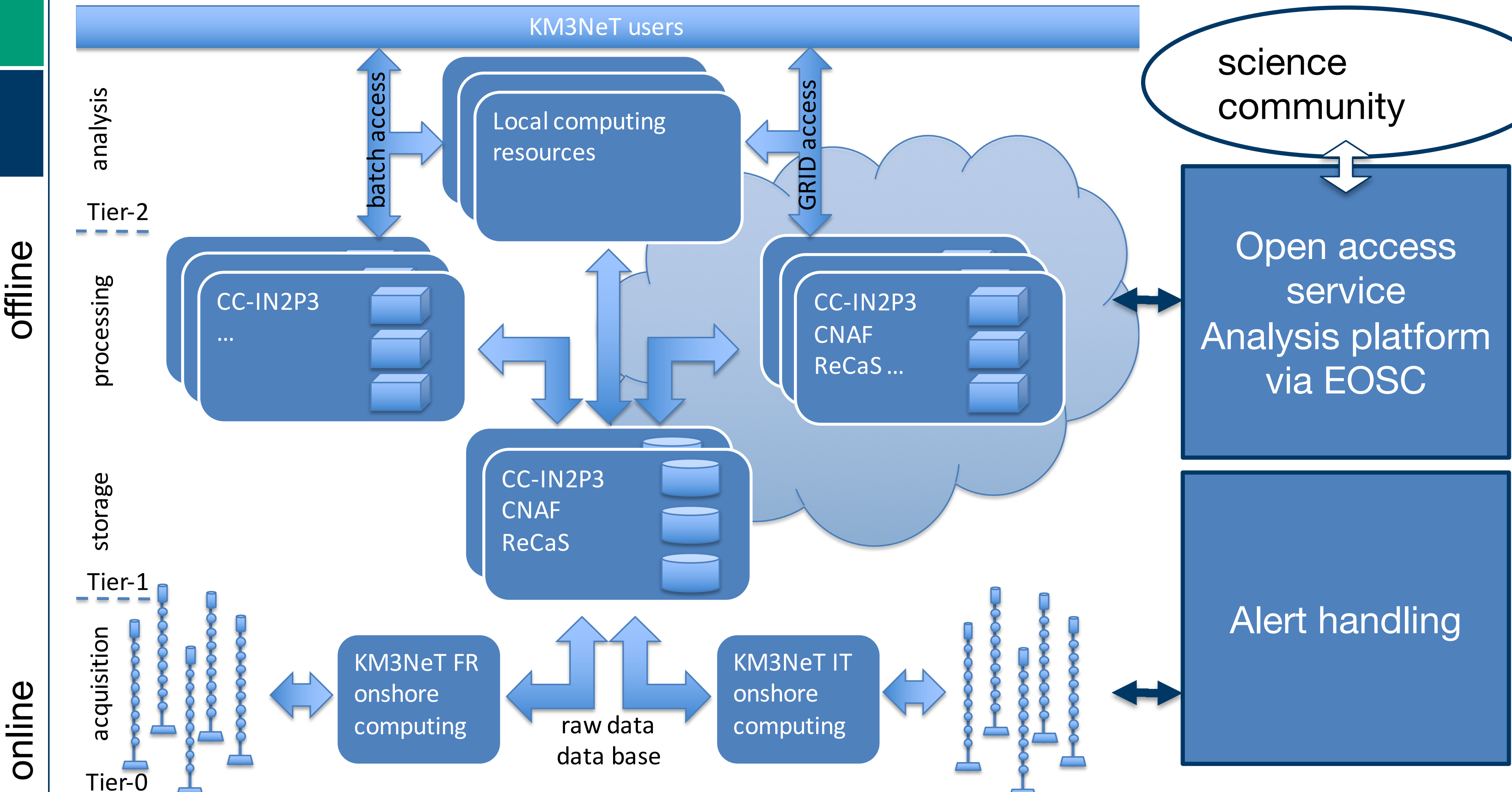
[PoS 301 (ICRC 2017) 1046]

[courtesy B. Shaybonov]

KM3NeT Data Management

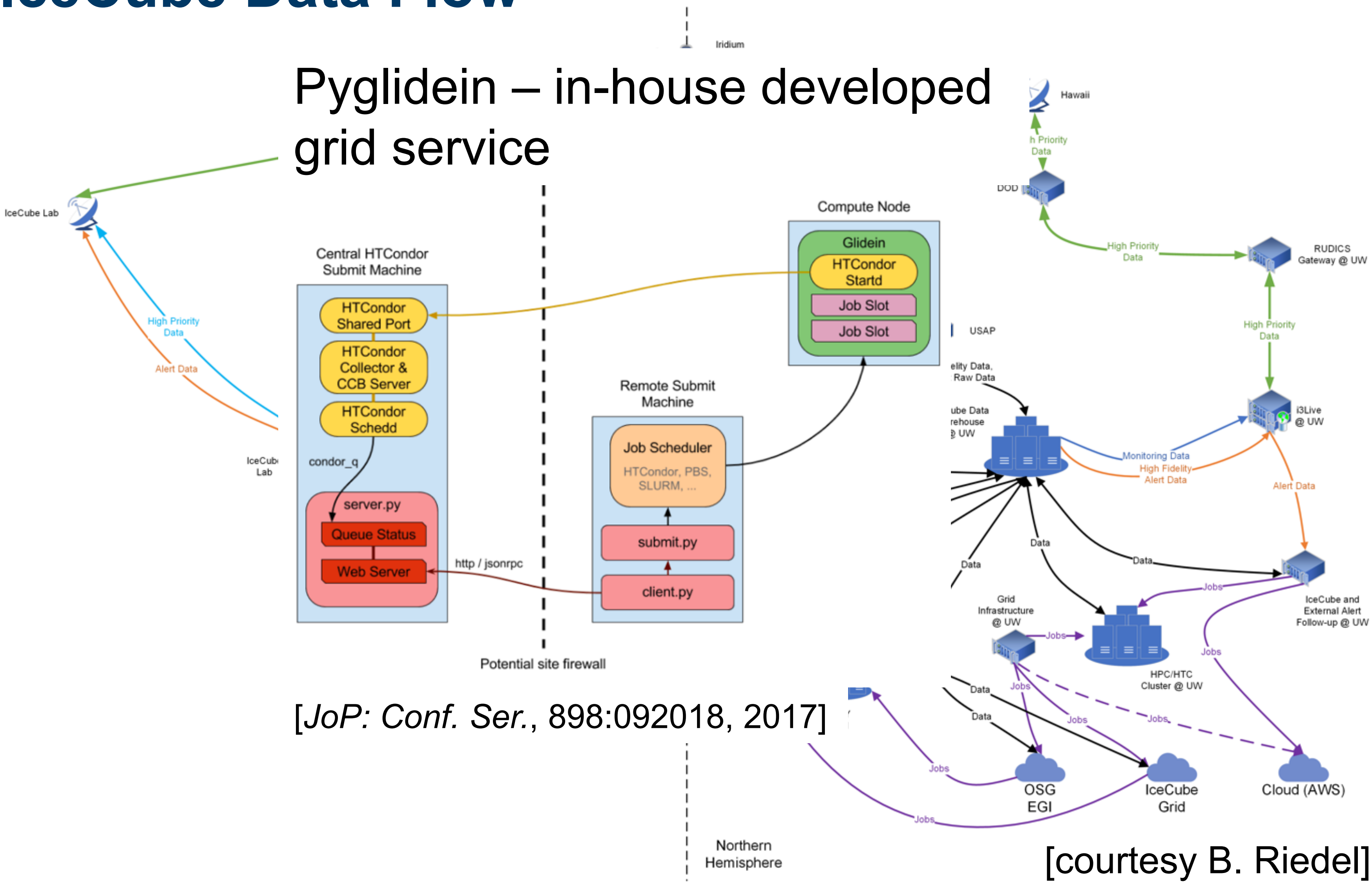
Internal:

External (under development):



IceCube Data Flow

Pyglidein – in-house developed grid service



[*JoP: Conf. Ser.*, 898:092018, 2017]

[courtesy B. Riedel]

Examples of Processing Steps

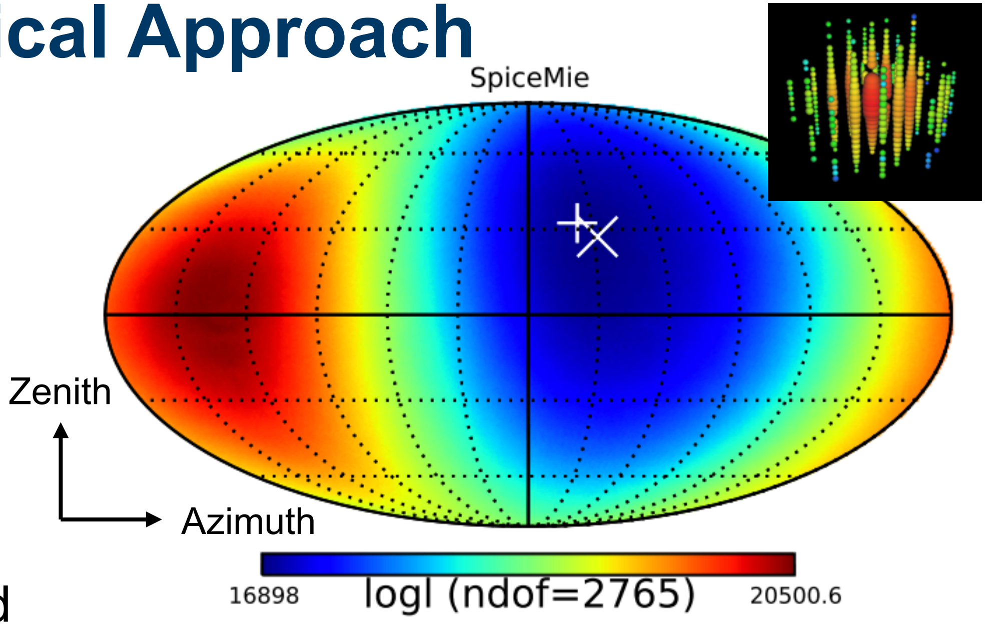
Simulations

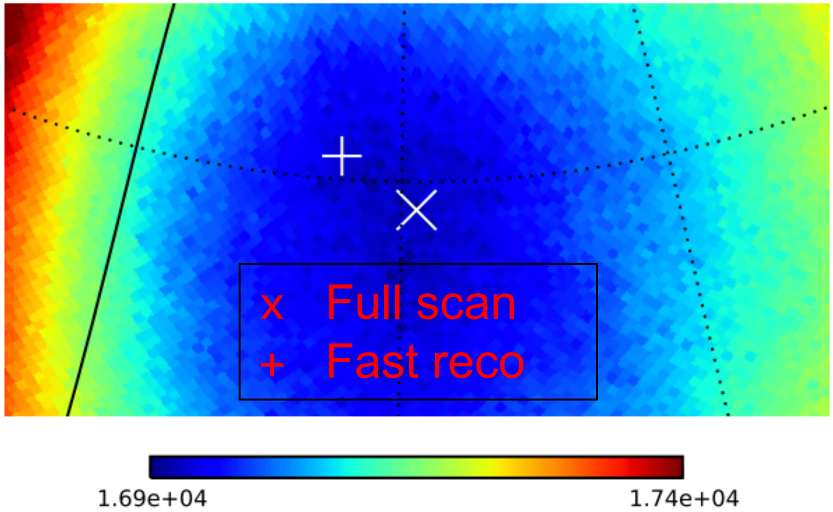
Extensive MC Simulations necessary (analysis based on data/MC comparisons)

- Background: lifetime ~ uptime of detector
  - Challenge (especially if environmental conditions change, i.e. MC needs to be adapted to conditions)
  - Can typically only be achieved for high energies
  - Here, extensive cloud resources could help
    - HPC cloud environment in use where available (Baikal-GVD)
    - Commercial clouds expensive
- Neutrinos: lifetime ≫ uptime of detector (feasible)
  - Classical resources typically sufficient

Reconstruction: Classical Approach

- **Minimisiation of a multi-dimensional PDF-based (log)likelihood** $\mathcal{L}(\boldsymbol{\Theta}|\mathrm{X}_{\mathrm{Data}})$ where $\boldsymbol{\Theta} = (x, y, z, \theta, \phi, E, t)$
- Computationally expensive (depending on strategy - e.g. if uncertainty estimation is included or brute-force ansatz is used)
- Standardly run on all events using the computing services
- Commercial cloud services used for speed up of interesting events by IceCube → alerts sent to community should have lowest possible uncertainty and delay





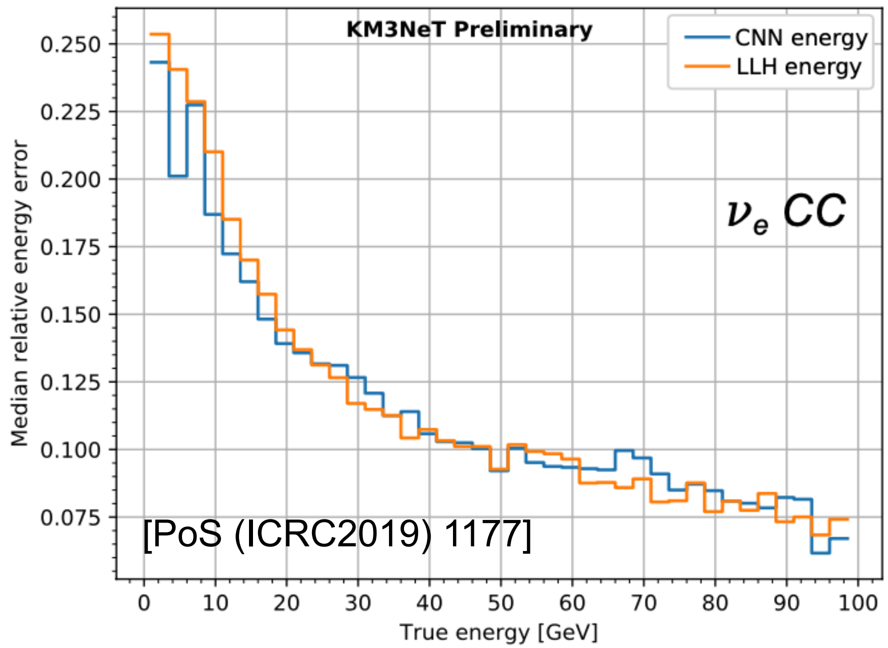
[https://events.icecube.wisc.edu/event/105/contributions/1077/attachments/846/917/190611_bootcamp.pptx]

Reconstruction: The Deep Learning Approach

- **Machine and Deep Learning techniques very popular** and widely used e.g. for event classification …

- Reconstruction via regression (energy and direction) competitive to the classical approach now
- Usually GPU resources in HPCs not sufficient or not flexible enough

- Special GPUs and TPUs in commercial cloud could be used, also full DL analysis platform envisioned (IceCube)



Median relative error, $|(E_{reco} - E_{true}) / E_{true}|$, of the reconstructed energy.

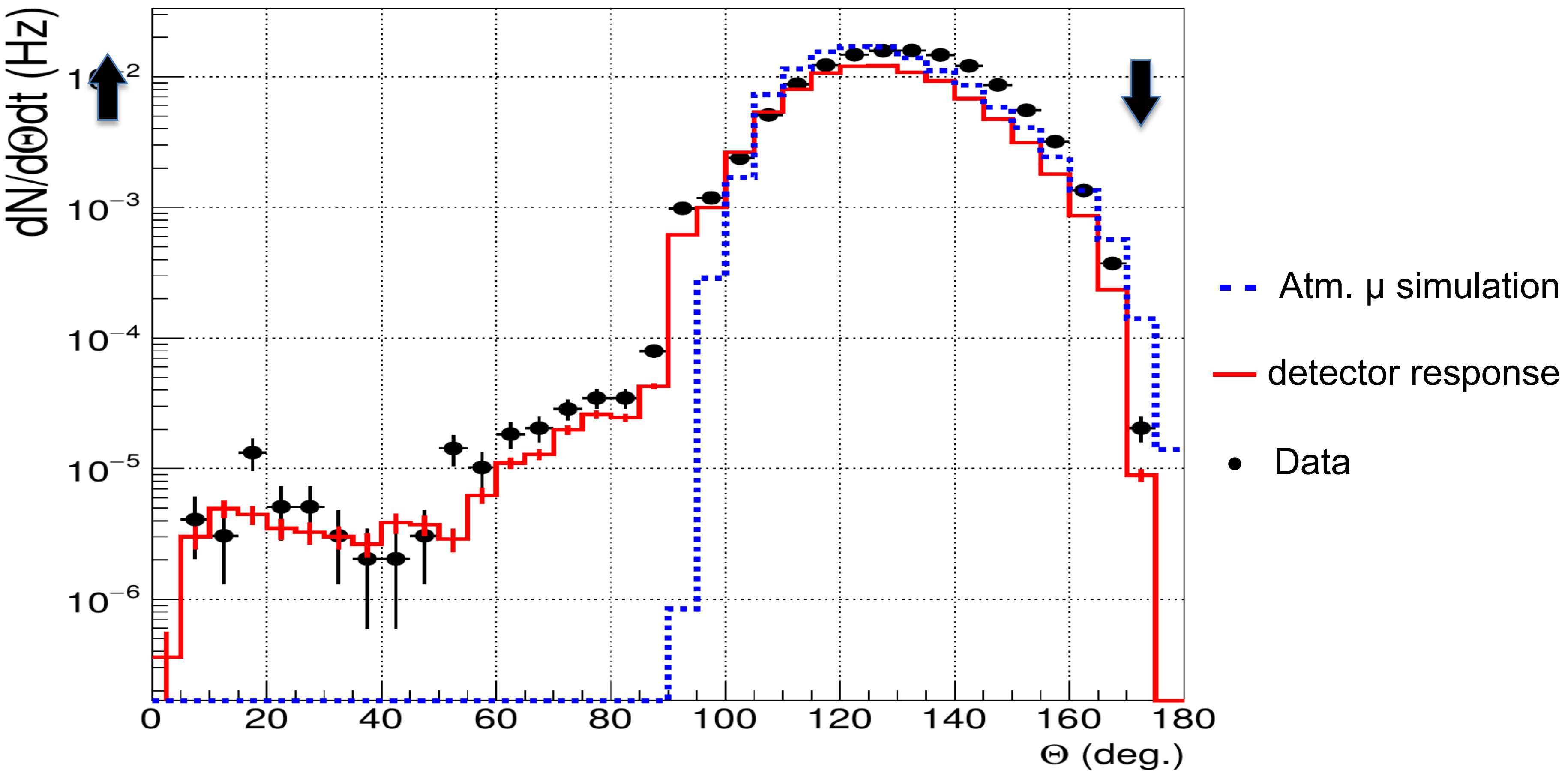
Analysis

Typical employed techniques:

- **Cut and count technique** (optimising MDP/MRF)
- **Binned and un-binned likelihood analysis** via test statistics (PDFs for signal and background)
- Standard procedures available and applied on standard computing resources (Tier-I or II centres)

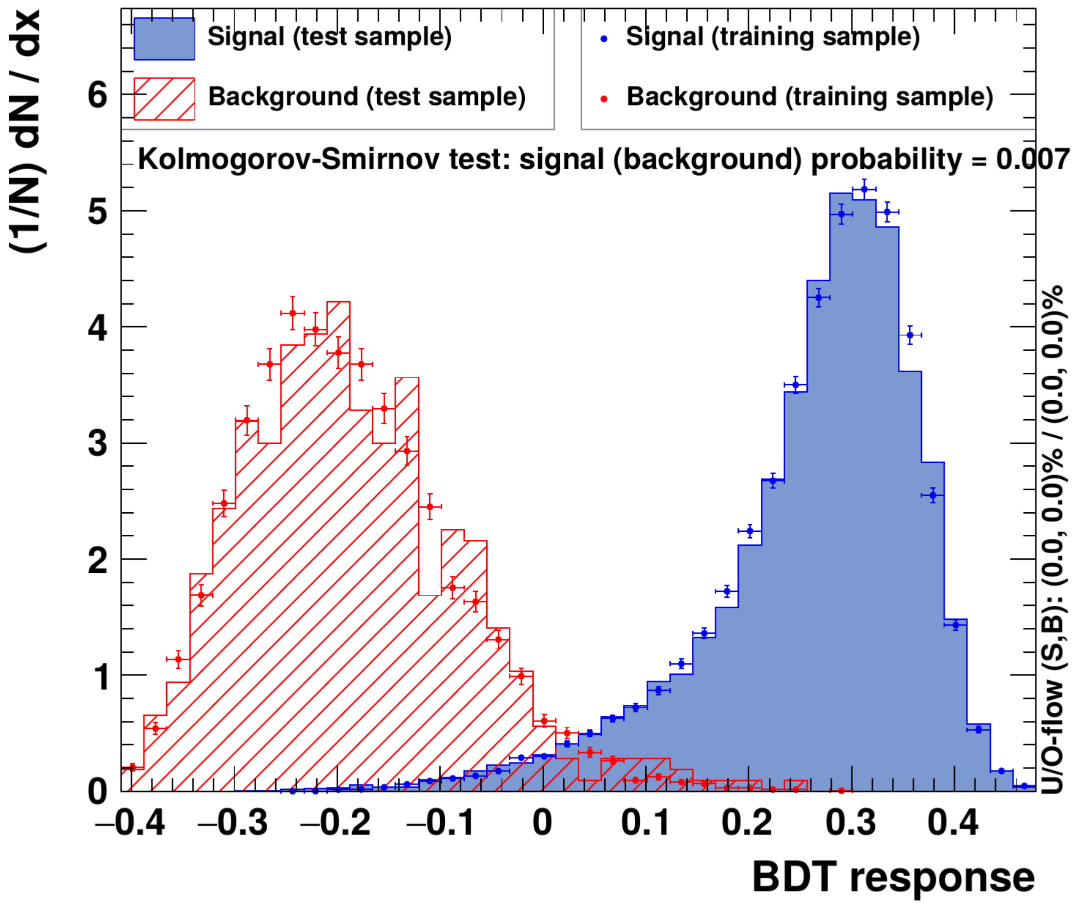
GVD: first data analysed, first ν's …

- 2016 data: select atmospheric muons (≥6 OMs at ≥3 strings)



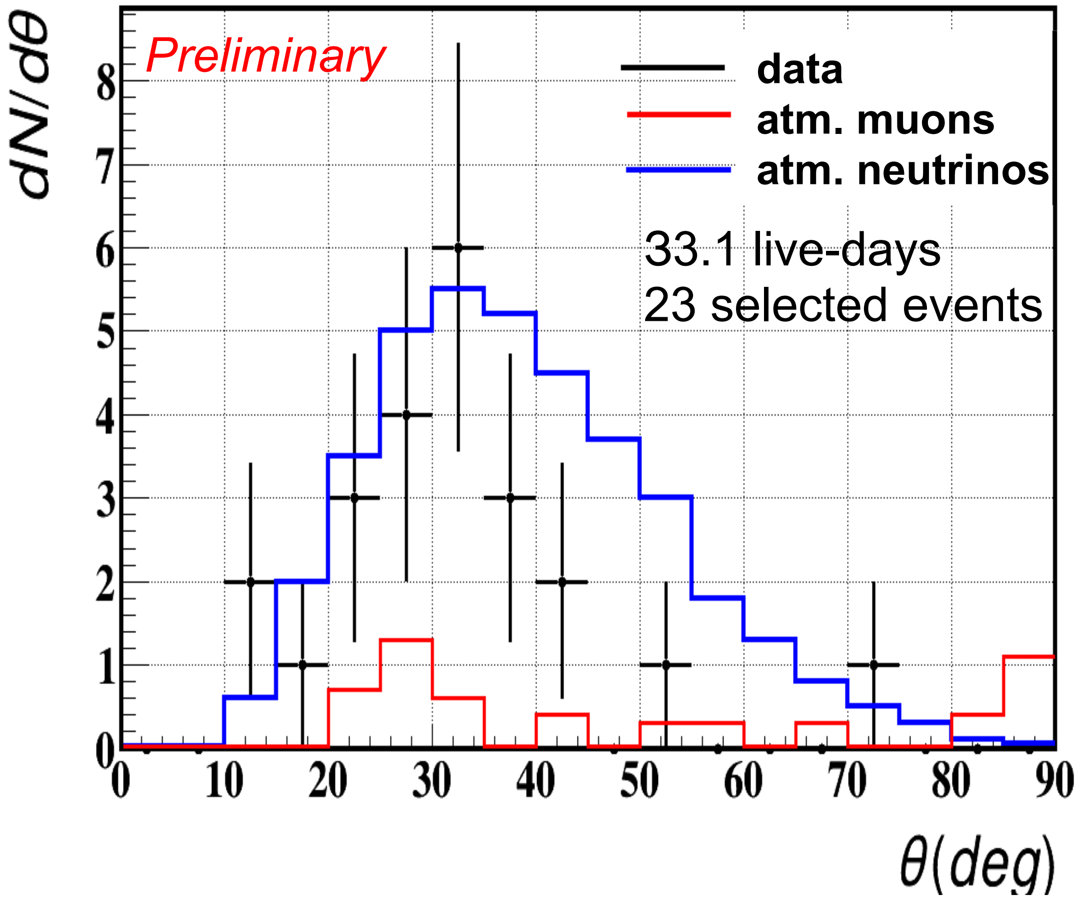
GVD: first data analysed, first ν's …

- 2016 data: select atmospheric muons (≥6 OMs at ≥3 strings)
- Apply quality cuts and boosted decision tree for ν/μ separation



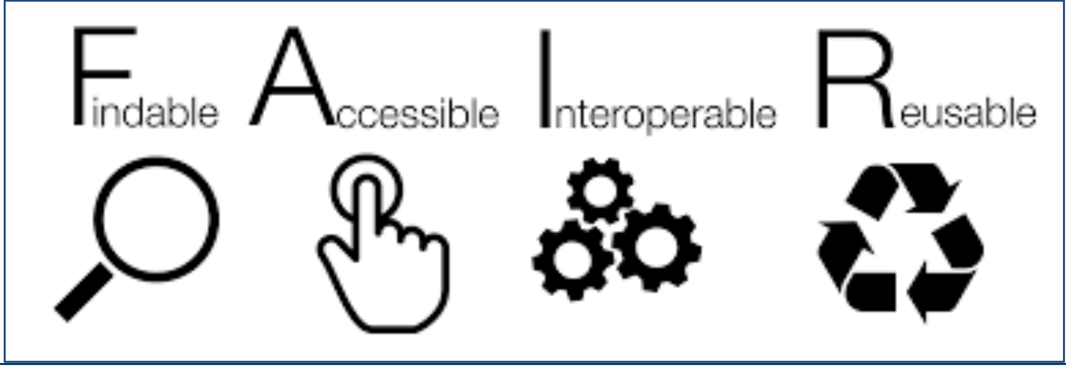
GVD: first data analysed, first ν's …

- 2016 data: select atmospheric muons (≥6 OMs at ≥3 strings)
- Apply quality cuts and boosted decision tree for ν/μ separation
- **… and observe first neutrinos!**



Open Access Data and Open Science

- Central goal of experiments: **prompt dissemination of scientific results**, new methods and implementations
- As observatories: making data accessible to the science (and public) community – e.g. for KM3NeT following the FAIR principle
- IceCube and KM3NeT following strategies for high-level data catalogue and MM data analysis platforms in the cloud
    - alert handling in the **multi-messenger ansatz**



Summary: (Science) Cloud Usage

- **Baikal-GVD:**
  - Standard processing and alert handling
- **IceCube:**
  - Speed up of reconstruction for special events
  - Machine learning applications (specialised hardware)
  - Open access data archive (foreseen)
- **KM3NeT:**
  - Open access and open science (analysis portal) → partner in the EOSC ESCAPE cluster project
  - Special purpose usage (foreseen)

EOSC and ESCAPE

EC started the EOSC (European Open Science Cloud) programme to overcome fragmentation of eInfrastructures

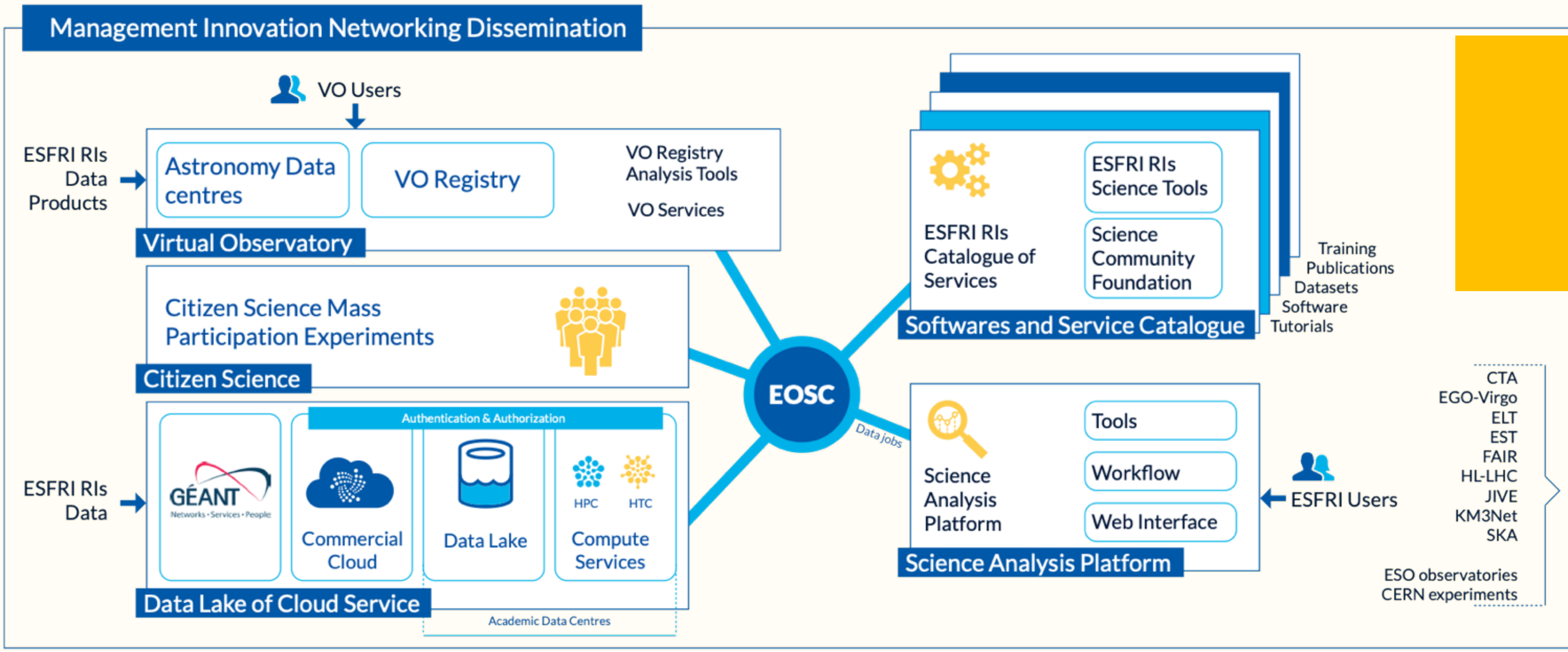
- Federated, open-access cloud environment
- EU funding of computing and (IT) services will be channelled through this project

ESCAPE project: ESFRI Cluster for AstroParticlePhysics

- Bring the communities' point of view into EOSC
- Solve open science issues based on RI needs and expectations
- Many of the examples above are in the ESCAPE work plan

⇒ EOSC will bring up many opportunities enriching open science output

ESCAPE Project Overview - https://projectescape.eu



ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

ESCAPE Partner RIs

Radio

JIVE-VLBI

SKA

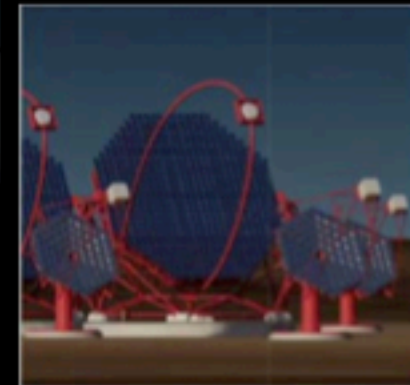
Visible light

ELT

ESO

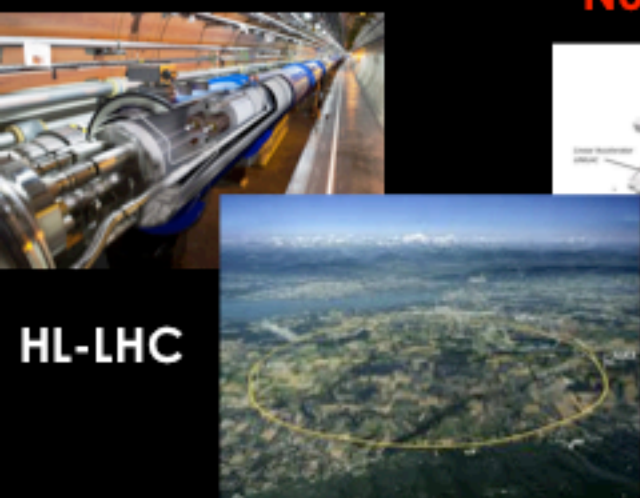
EST

Gamma rays



CTA

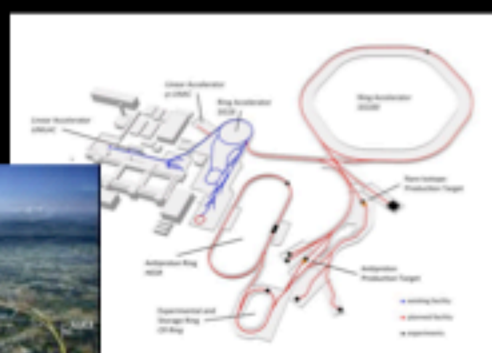
Accelerator-based Particle Physics



HL-LHC

CERN

Accelerator-based Nuclear Physics

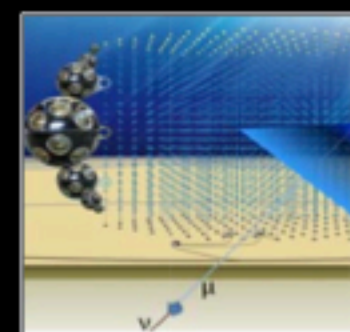


FAIR

Gravitational Waves

EGO-VIRGO

Cosmic-rays Neutrinos



KM3NeT

ESCAPE Partners

Conclusions

- Data handling in neutrino telescope follow mainly a (classical) **tiered structure**
- **Cloud ansatz followed in several parts** but not globally yet (mainly driven by available HPC resources)
  - Great opportunities seen both in commercial cloud solution and the science cloud initiatives
    (EOSC in the EU)
    - Speed up of resource-hungry processing steps
    - Servicing bursts of processing
    - Open access data/software & MM analysis platforms
  - However, also drawbacks (e.g. cost/benefit ratio)
    - Not all services centralisable in a/the cloud
    - Technical implementation not easy through fragmentation of services and specialised software

Backup Slides

The neutrino telescope timeline