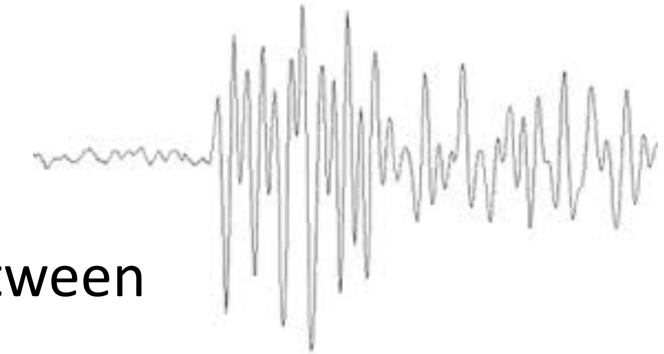
Statistical properties of earthquakes

T. Budavari & M. Graham

In addition to their recorded strengths, earthquakes have spatial and temporal measurements: where and when they occur. Let's build a statistical model to analyze the spatial and temporal correlations of these events.



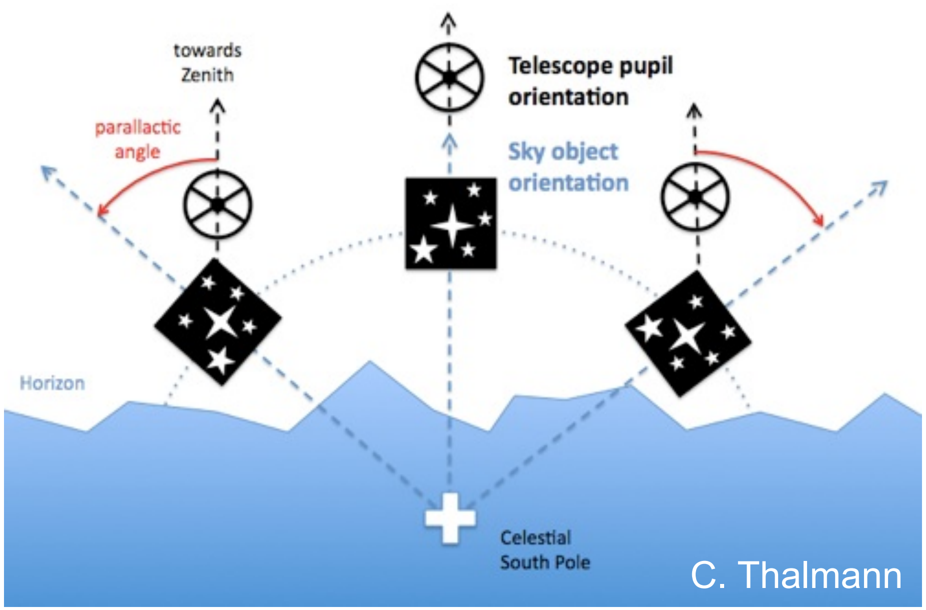
1) Download a relevant dataset with GIS information
2) Plot the events to visualize the fault line (or fault lines)
3) Write code to measure the distance along the fault line between two earthquakes
4) Map out the probability density of an new earthquake happening
    a) t time after an earthquake
    b) r miles away on the fault line from an earthquake within T time
5) How do the results change as a function of the strength of the earthquakes?

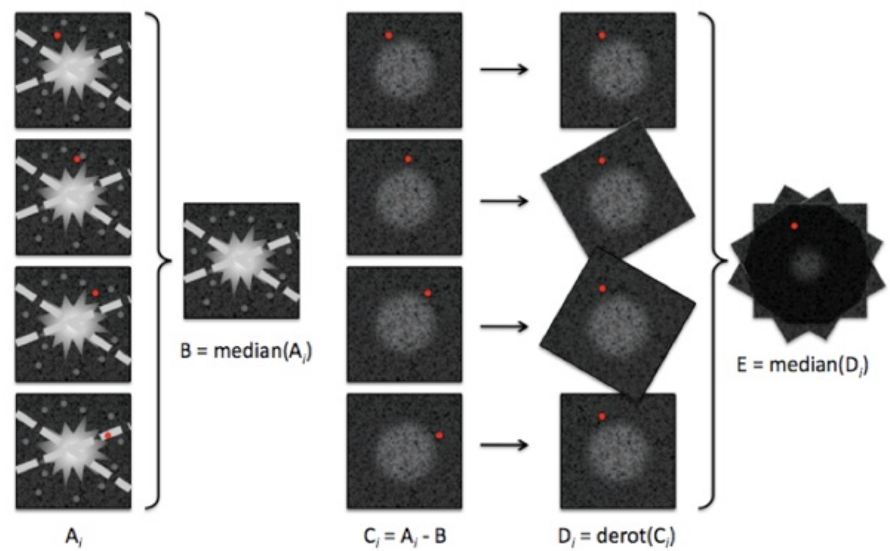
Looking for EM signatures of BH-BH mergers
M. Graham

The Zwicky Transient Facility (ZTF) observes the northern sky every 3 nights in g and r and produces a public alert stream of all transients detected in real time. LIGO has detected 18 BH-BH mergers in O3 but these not expected to have an EM signature. However, if the merger occurs in the accretion of disk of a SMBH then an EM signature may be expected (as predicted by McKernan et al. 2019). This project will involve looking for possible EM events in ZTF data associated with LIGO triggers.

Extended sources in ADI

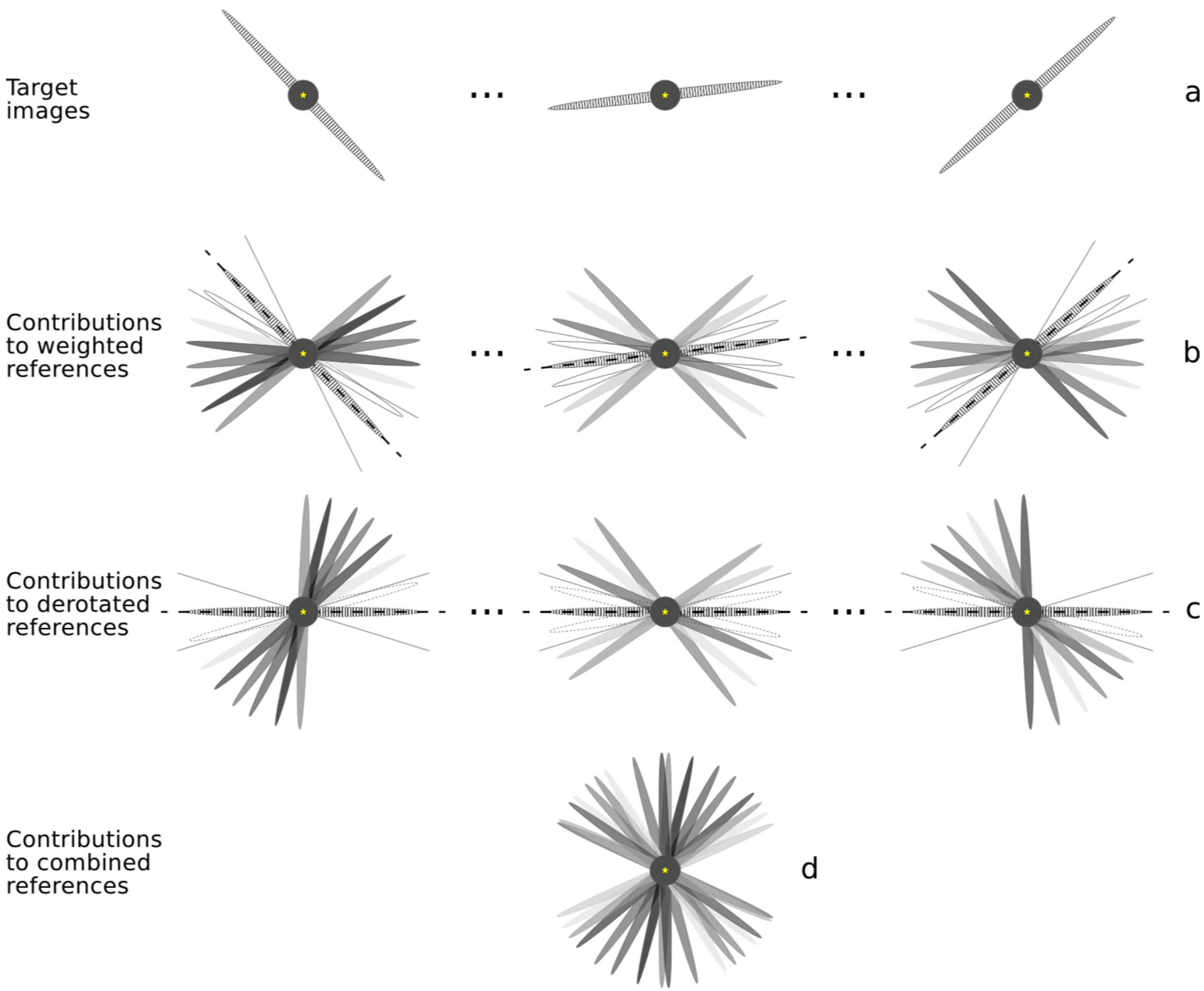
Directly detecting an exoplanet implies overcoming a contrast challenge of several orders of magnitude. Part of the solution to this problem passes by improving conoronographs, detectors, optics, etc (ExAO), but one can also use “natural improvements” like taking advance, for example of the rotation of the Earth to differentiate between speckles and “true” companions. Such technique is called “ADI” angular differential imaging, and the post-processing of images obtained with this technique work beautifully to increase our chances to detect point sources, but not so much extended ones. We will get familiar with ADI, classical median and PCA subtraction, but also try novel alternatives.





Extended sources in ADI

Esposito et al.

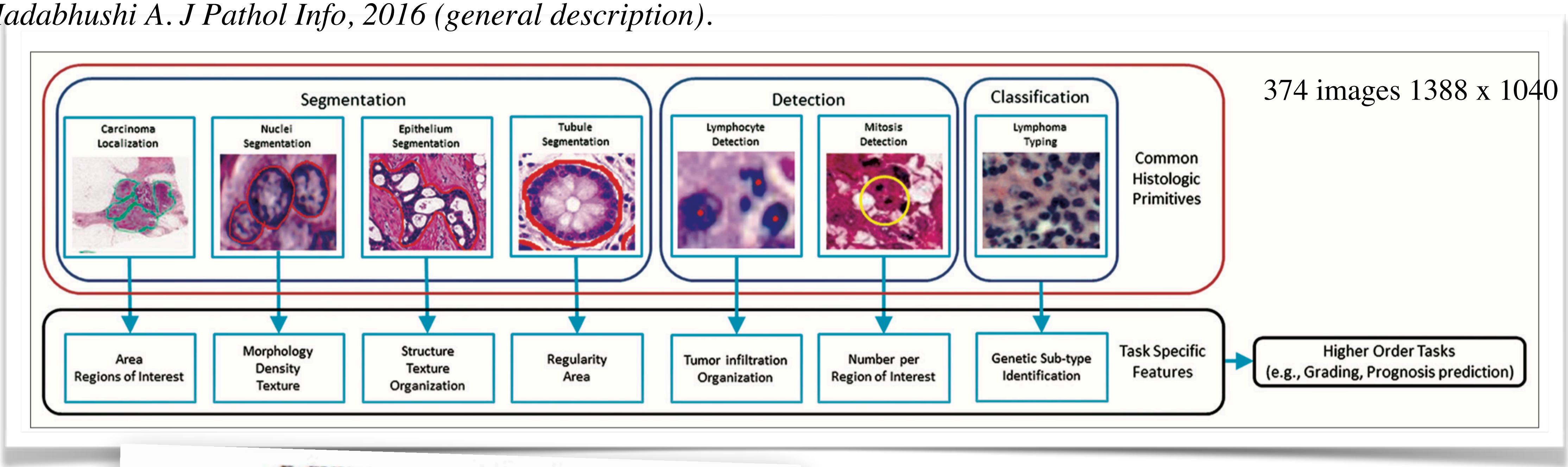


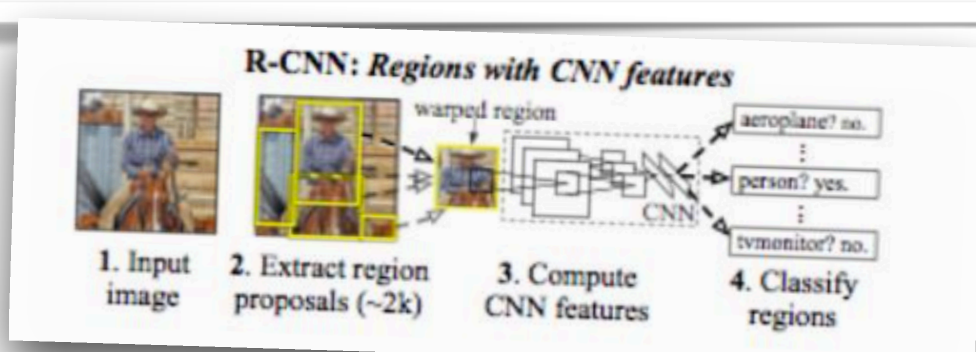
Deep learning for digital pathology image analysis.

FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

This project will investigate the use of deep learning techniques in digital pathology (DP) problems. The variety of image analysis tasks in the context of DP includes detection and counting (e.g., mitotic events), segmentation (e.g., nuclei), and tissue classification (e.g., cancerous vs. non-cancerous). After **selecting one of those tasks**, the aim of the project is to replicate/improve reference results, **exploring other network architectures** and/or optimize parameters. The used database is of public access. Technical aspects include to implement in a recent deep learning framework using a **GPU cluster available (CUDA).**

Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, Janowcyk A, Madabhushi A. J Pathol Info, 2016 (general description).





NVIDIA GPU RESEARCH CENTER

Analysis of United Nations' Food Sheets to understand changes in diets in the word

Preliminary unsupervised learning of worldwide diets show that, when the two principal components of food sales are plotted, some countries are close to neighboring countries, while others are close to distant countries (Chile for example in the figure below).

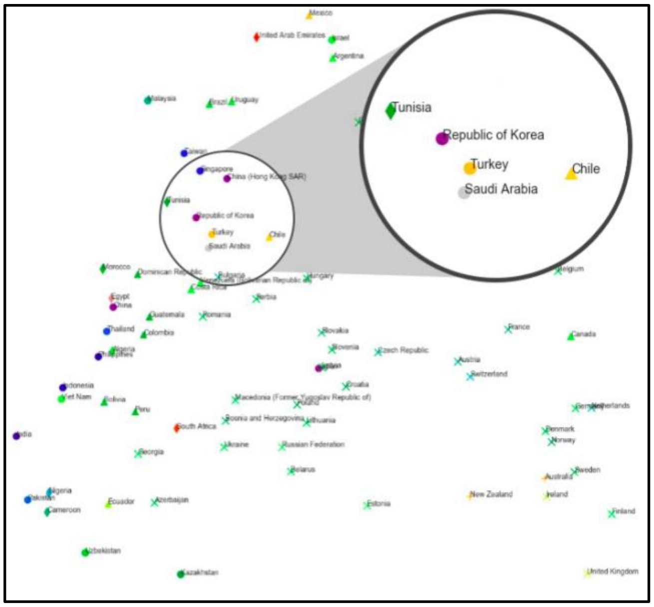


Figure 1: PCA analysis of food sales from 79 countries using Euromonitor data. The zoom shows the countries close to Chile. Taken from Dunstan et al, Health Informatics Journal (2019).

This project consists in exploring United Nations data for food production that contains diet information in 97 food categories for 175 countries between from 1961 to 2013. The idea is to investigate if we can observe migrations in 2D-projections of diets.

Was Chile always far from other Latin American countries? Is France always close to Belgium? Can we grasp why this is happening?

Astronomical time series classification

Francisco Förster

A new generation of survey telescopes studying the variable Universe is detecting millions of variable events every night. An important challenge is how to Identify the most interesting objects to be characterized by follow-up telescopes every night into a complex taxonomy. In order to solve this problem we are creating a new ecosystem of tools such as astronomical alert brokers and target & observation managers (TOMs). These tools will use a combination of machine learning methods, but their power will be greatly determined by the quality and volume of training sets.

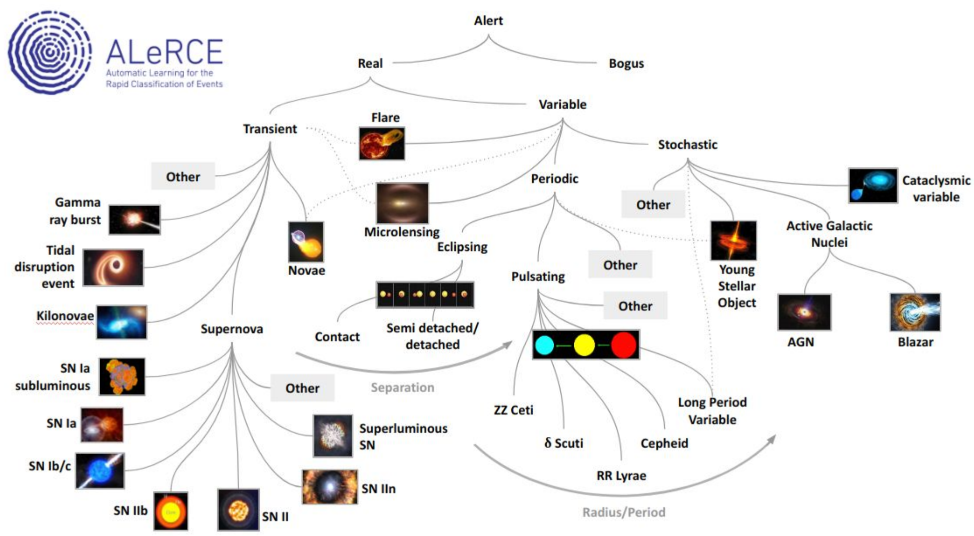


Fig. 1: A taxonomy of variable objects

In this project you will learn to classify astronomical light curves from the Zwicky Transient Factory (ZTF, https://www.ztf.caltech.edu/) using data curated by the ALeRCE broker (http://alerce.science/). The steps are the following:

1. **Visualization**: build a tool to load and visualize astronomical light curves from ZTF
2. **Feature extraction**: extract features from the light curves (e.g. using tsfresh)
3. **Dimensionality reduction**: use some dimensionality reduction algorithm (e.g. t-SNE, UMAP) to visualize your data in a lower number of dimensions.
4. **Clustering**: use some clustering algorithm to identify clusters. Overlay the known classes from labeled data in your clusters
5. **Supervised classification**: use some supervised classification method to classify your data, obtain different classification metrics.
6. **Semi-supervised classification**: apply some semi-supervised classification method to classify your light curves, combining labeled and unlabeled data. Compare the metric

Sumarizing ZTF alerts

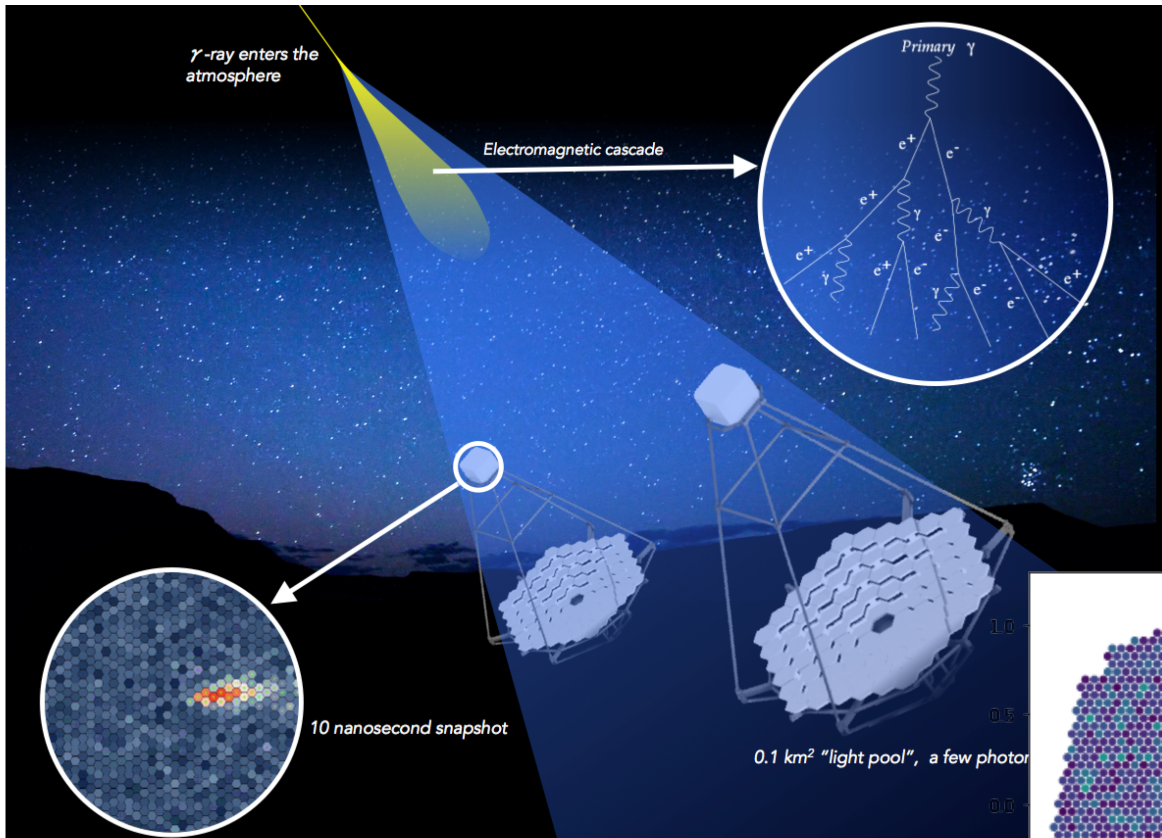
Prof. Juan-Carlos Maureira

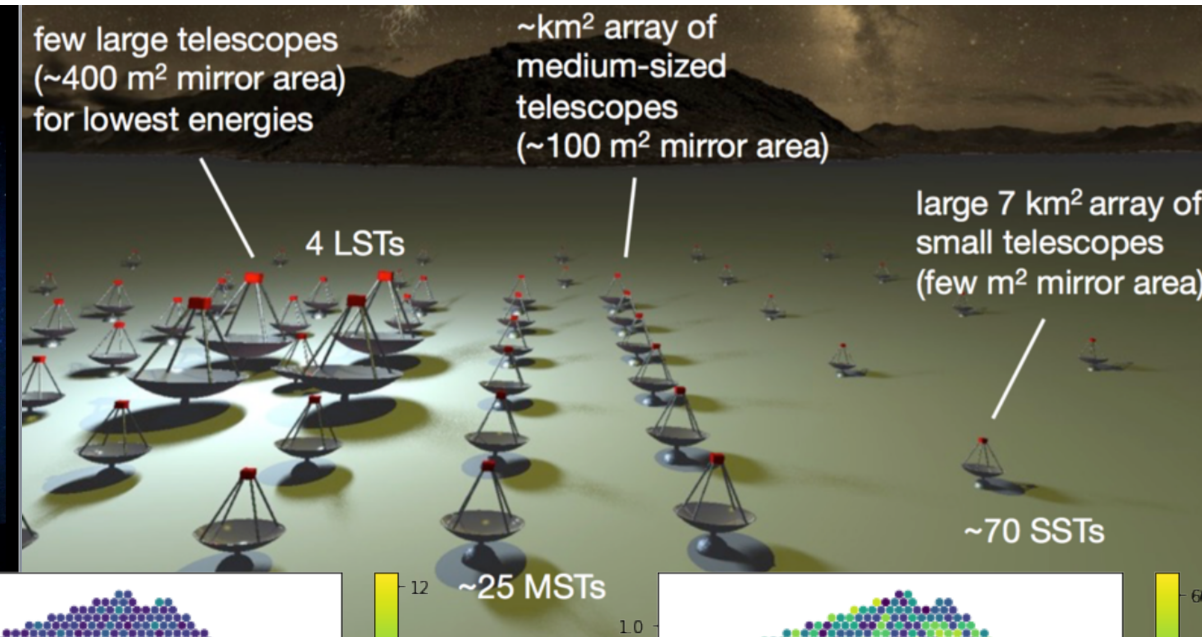
Working with trillions of files on a HPC system is a challenging task. This will be the case when dealing with a LSST alerts, and while you may work with alerts coming from a reduced window of time, the resulting number of files are large enough to stress any HPC filesystem. In this context and to prepare ourselves to make science with the LSST, this project proposes to apply some reductions functions to the ZTF data (2TB), starting with the acquisition of alerts (one tarball per day), filtering the easy to catch bogus detection, and reduce the data given certain criteria such as grouping per day or grouping per object ID. For this purpose, you will use the NLHPC system (leftraru) via a Jupyter notebook, building your reduction pipeline by using tasks and jobs orchestration together with asynchronous programming.

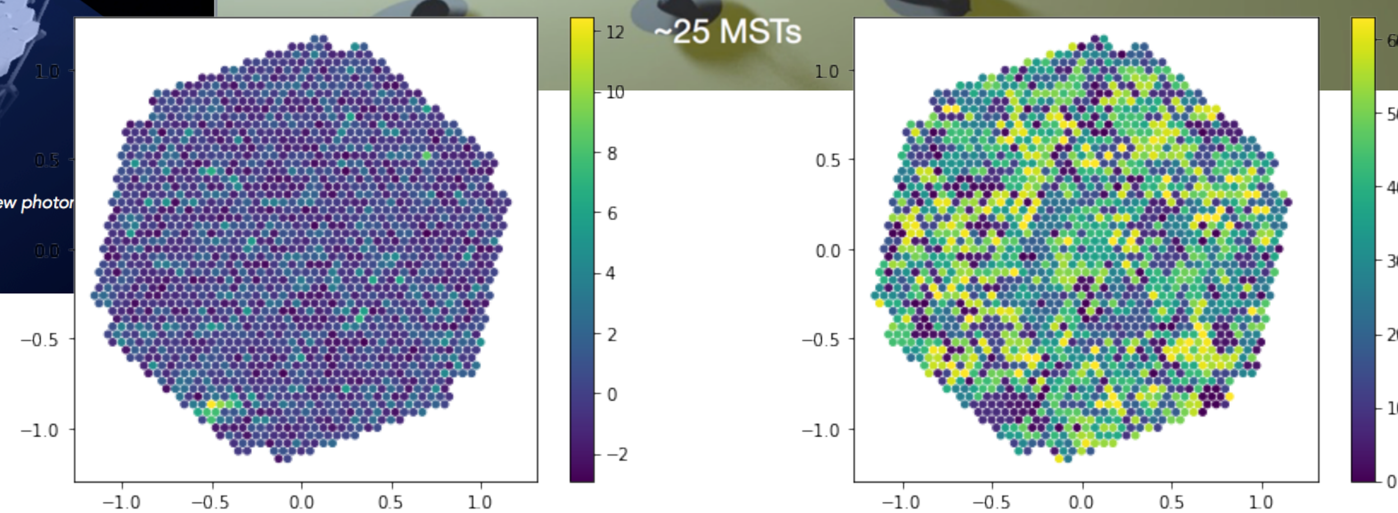
****

Gamma-Ray Reconstruction

Objective: reconstruct gamma-ray angular position and energy from simulated multiple observations of Cherenkov showers (i.e., regression from multiple images)







Challenges:

- **Data ~ 200 GB**
- **Hexagonal-grids**
- **Multiple observations**
- **Structured fileformat (HDF5)**
- **Imbalanced data**
- **Multi-criteria objective**