Kavli Institute
for Cosmological Physics
at The University of Chicago

CosmoSIS: Cosmological Survey Inference System

Alessandro Manzotti*, Douglas Rudd, Sarah Bridle, Scott Dodelson, Elise Jennings, Jim Kowalkowski, Marc Paterno, Saba Sehrish, Joe Zuntz*

THE UNIVERSITY OF CHICAGO | Research Computing Center

https://bitbucket.org/joezuntz/cosmosis

21st Century Cosmology: Big data and experiments

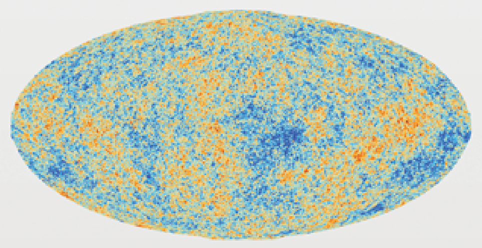
In the past several decades, cosmology has made great strides in measuring the dynamical history of the universe revolutionizing the way we see it. Using observational data, we discovered that the universe is flat and expanding at an increasing speed due to a mysterious component called dark energy.

Understanding the nature and properties of this dark energy is a major goal of cosmology in the coming decades. Numerous large scale experiments, both planned and currently ongoing, are designed to study the dark energy at different epochs and through different observational probes.

A major challenge for the field is to enable researchers to maximize the information extracted by properly analyzing and combining together datasets from all the experiments.

[*Right: example surveys*]

DARK ENERGY SURVEY
Collaboration of 25 institutions world wide with over 200 scientists. Optical survey of 5000 deg^2 (1/8 of the sky) with the goal of using multiple probes to identify the physics driving the acceleration of the Universe



PLANCK European Space Agency space telescope designed to image the temperature and polarization anisotropies of the Cosmic Background Radiation Field over the entire sky.

New era ⟶ new generation of computational tools

CosmoSIS is a flexible and extensible framework for the joint analysis of cosmological datasets and the exploration of cosmological models. It was designed with modularity at its heart to enable researchers to tackle the computational challenges of working with modern cosmological data.

Using CosmoSIS we help scientists:

- **Easily test different theories with different datasets**: its modular design allows easy modification of models, datasets, and sampling algorithms.
- **Get answers fast without worrying about the implementation**: it has a simple, "one-click" installation on a variety of platforms, from laptops to supercomputers and fast parallelized pipeline.
- **Easily contribute with new and existing code**: it supports a variety of languages (C/C++, Fortran, and Python) so developers can contribute in the most appropriate language for their task and integrate existing legacy code.
- **Get credit and be reproducible:** it implements modern code provenance and result tracking to emphasize reproducibility and ensure developers and experiments receive credit for their contributions.

Cosmologist and software engineers working together

CosmoSIS is developed by software engineers and researchers at the Kavli Institute for Cosmological Physics (KICP) and the Research Computing Center (RCC) at the University of Chicago, Fermilab, and Manchester University. The RCC has provided developer support and hosts a development installation on the Midway cluster.

In May 2014 the KICP and RCC hosted a breakout workshop designed to introduce CosmoSIS to the community. Over 25 researchers and CosmoSIS developers attended the two day workshop during which they were introduce to the software and its goals, and were able to install, run, and start contributing to CosmoSIS. RCC provided all participants access to the Midway cluster during the workshop to test CosmoSIS.

The first version of the software is now available to the entire community at **https://bitbucket.org/joezuntz/cosmosis** and a detailed description can be found in the submitted code paper **http://arxiv.org/abs/1409.3409**.

CosmoSIS : a modular and powerful framework

Structure:
Typically we want to compute, given a cosmological model, different quantities needed to obtain physical observables and finally compare those with real data to extract how likely a particular model and parameters are given the measurements that we have.

In CosmoSIS a parameter estimation problem is represented by various components:

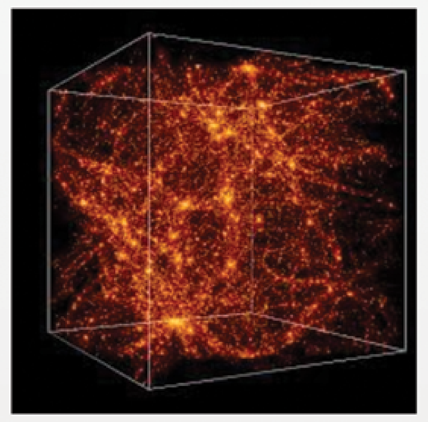
Pipeline a sequence of calculations that computes a joint likelihood from a series of parameters and a given theoretical model.

Modules the individual "pipes" in the pipeline, each of which performs a separate step in the calculation. Some do physics calculations, others interpolation, and at the end some generate likelihoods. Many modules are shipped with CosmoSIS as part of a standard library and users can easily write and distribute more.

Datablock the object passed down the pipeline. For a given set of parameters all module inputs are read from the datablock and all module outputs are written to it.

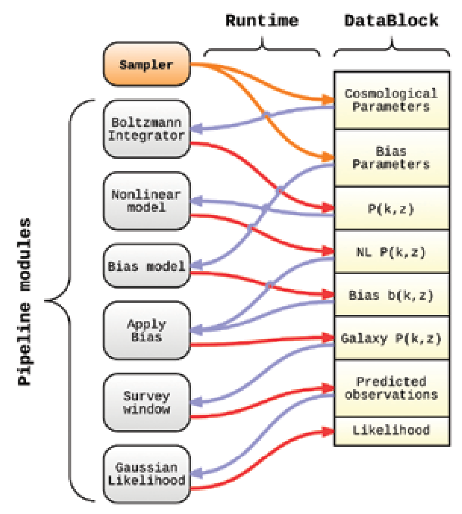
Sampler (generically) anything that generates sets of cosmological and other parameters to analyze. It puts the initial values for each parameter into the datablock.

Runtime the code layer that connects the above components together, coordinates execution, and provides an output system that saves relevant results and configuration.



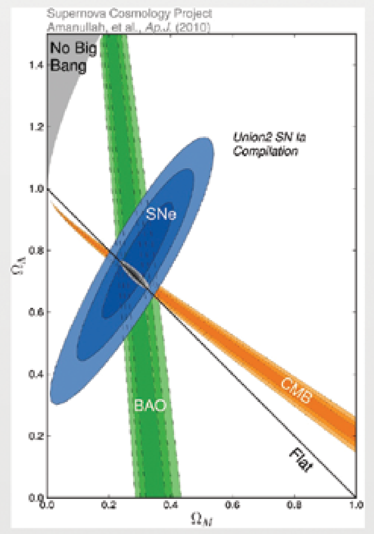
[galaxies simulation]

Given the model and parameters it generates the physical observable



[*Real galaxies from Hubble Deep Field*]

The output for different parameters is then compared with real data to compute a likelihood



[*Results:the typical CosmoSIS results consist of likelihood of different parameters: in this example the amount of dark energy and gravitational matter in the universe*]

Example: understanding dark energy from the distribution of galaxies.

The distribution of galaxies in the universe can help us to better understand dark energy. Given the initial distribution of matter measured by the cosmic microwave background we can, assuming a model for the evolution of the universe, follow the gravitational collapse that led to the formation of galaxies, and predict their distribution. This tells us how good our model is and the value of its parameters when compared to real data.
[*Left: example of a CosmoSIS pipeline described below*]

Sampler: it generates the parameters of our model for the universe, for example, how much dark energy there is in the universe. We may need to sample a lot of parameters, for this reason CosmoSIS uses sophisticated statistical samplers and a fast parallel implementation.

Boltzmann Integrator a physical calculation needed to evolve initial conditions. We may want to just modified this step of the pipeline. CosmoSIS modularity allow us to do so and share it easily with collaborators.

Non linearities and bias It accounts for the last complex step of evolution, the formation of galaxies in dense region. What we observe, the light from galaxies can be a biased tracers of what we have obtained in the previous step. Different models are necessary for different observational samples.

Survey peculiarities characteristics of the survey like the diameter of the telescope or the optics used have to be taken into account to go from the light emitted by galaxies to what we observed in our camera. How complete is the sample?

Likelihood How well does the galaxies distribution estimated from the model fit the real observed distribution of galaxies? this information may also be useful for the sampler to choose the new set of parameters.

MANCHESTER 1824
The University of Manchester

THE UNIVERSITY OF CHICAGO

Fermilab