Hopes and challenges in information science for cosmology

Seminar at Centre de Recherche Astrophysique de Lyon

Florent Leclercq

www.florent-leclercq.eu

Institut d'Astrophysique de Paris
CNRS & Sorbonne Université

In collaboration with the Aquila Consortium

ANR-23-CE46-0006: INFOCW

www.aquila-consortium.org

3 May 2024

Outline

01 Information science, but what for?

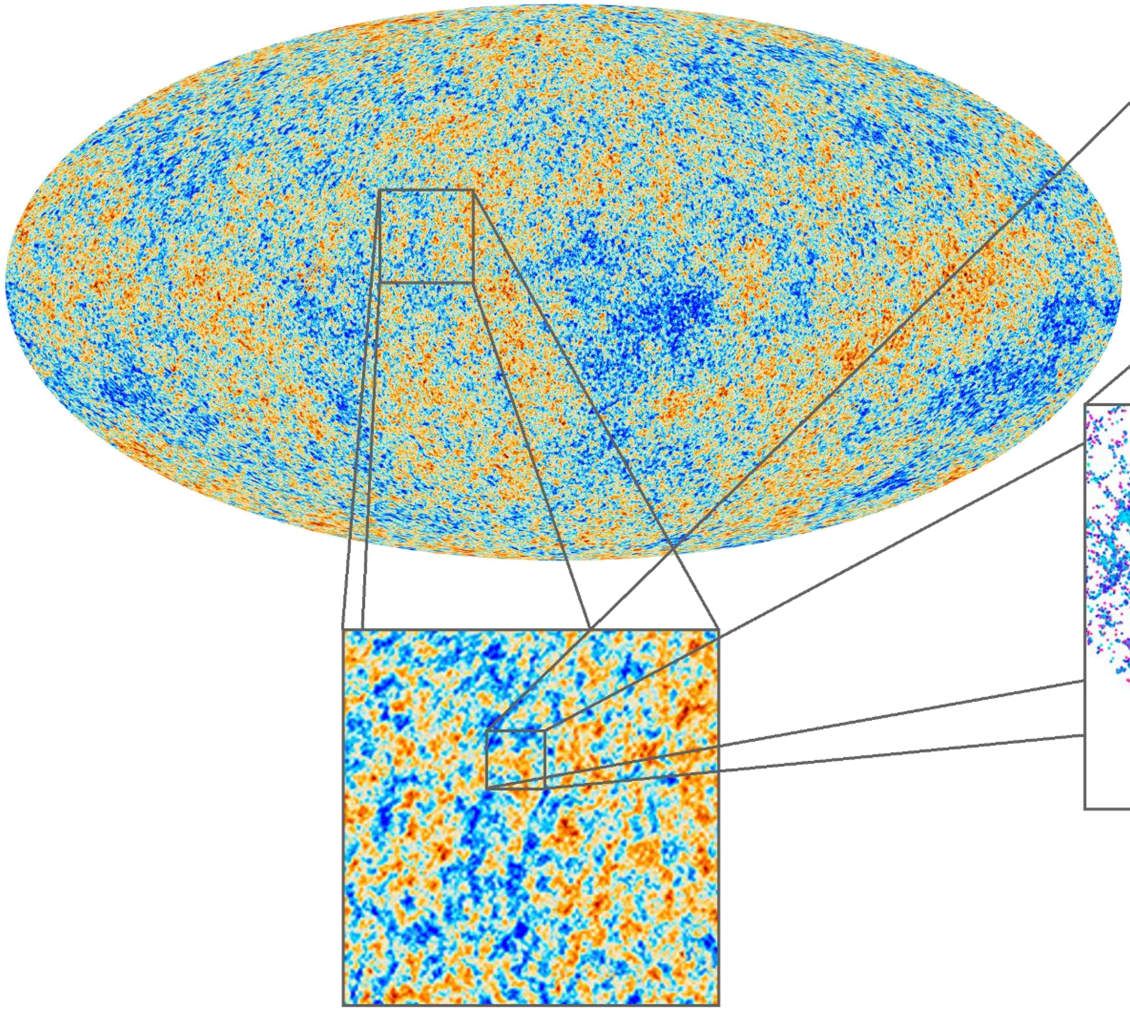
02 The forward problem: from theory to data

03 The inverse problem: from data to theory

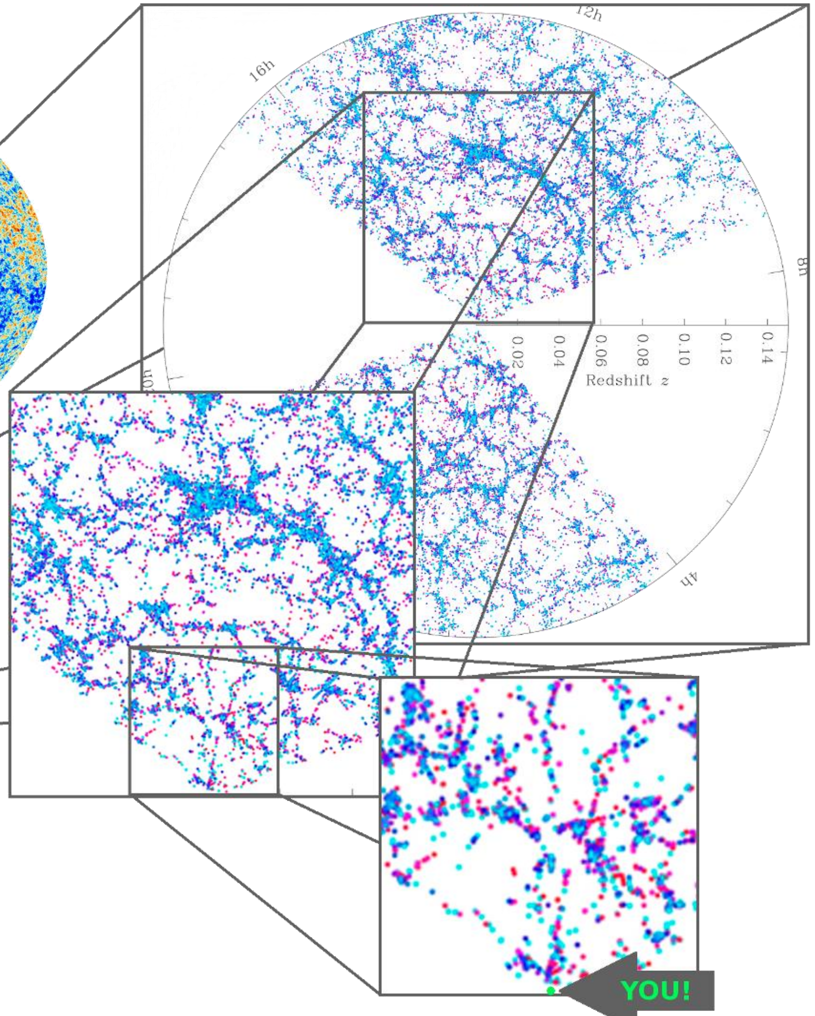
04 The imitation problem: algorithms beyond blind oracles?

INFORMATION SCIENCE, BUT WHAT FOR?

The big picture: the Universe is highly structured



Planck collaboration (2013-2015)



M. Blanton and the Sloan Digital Sky Survey (SDSS)

What we want to know from the large-scale structure

The large-scale structure is a vast source of knowledge:

- Cosmology:
  - ΛCDM: cosmological parameters and tests against alternatives,
  - Physical nature of the dark components,
  - Neutrinos: number and masses,
  - Geometry of the Universe,
  - Tests of General Relativity,
  - Initial conditions and link to high energy physics

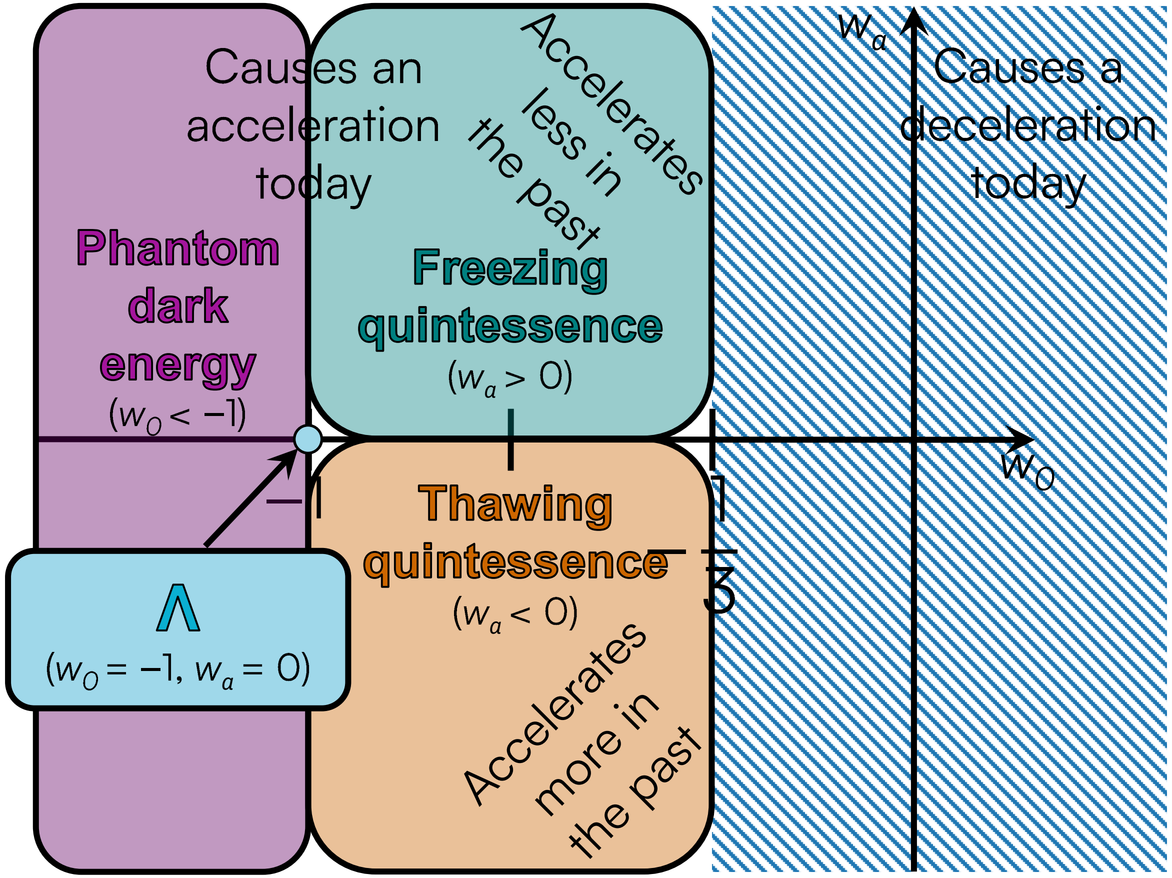
- Astrophysics: galaxy formation and evolution as a function of their environment
  - Galaxy properties (colours, chemical composition, shapes),
  - Intrinsic alignments, intrinsic size-magnitude correlations

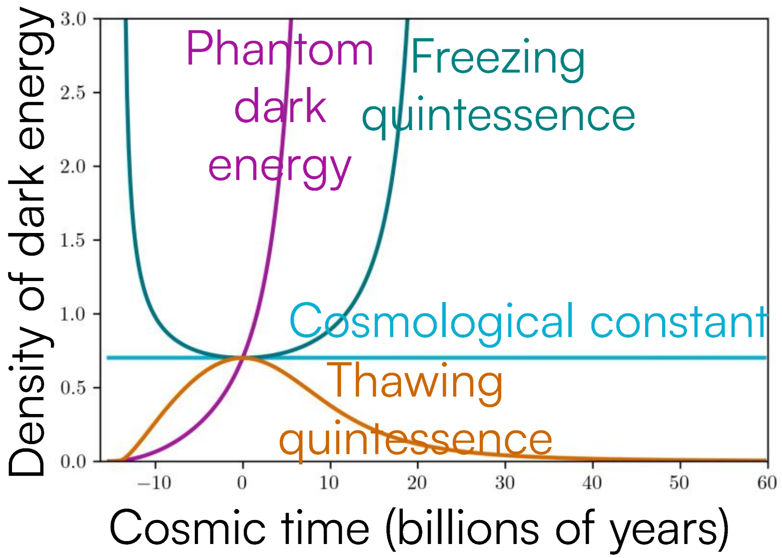
e.g. FL, Pisani & Wandelt, 1403.1260

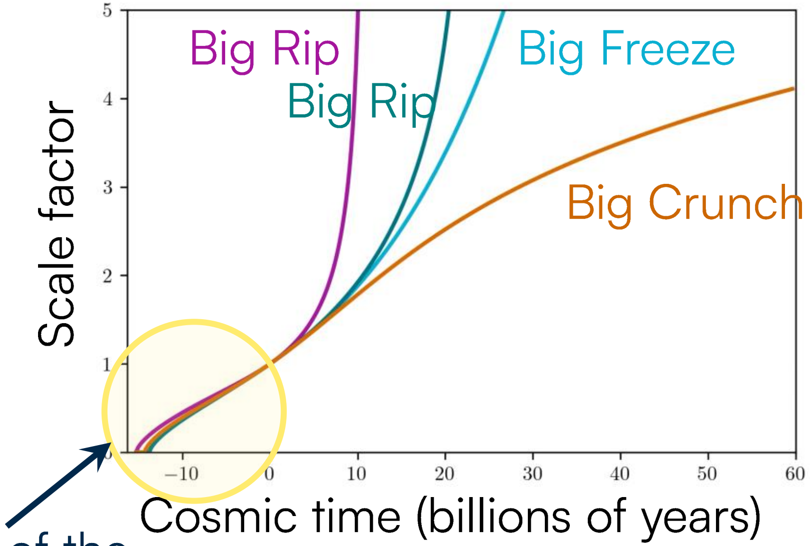
Interesting cosmological signals are faint

- Dynamic dark energy is usually parametrised as: $w(t) = w_0 + w_a \times (1 - a(t))$

Scale factor of the Universe

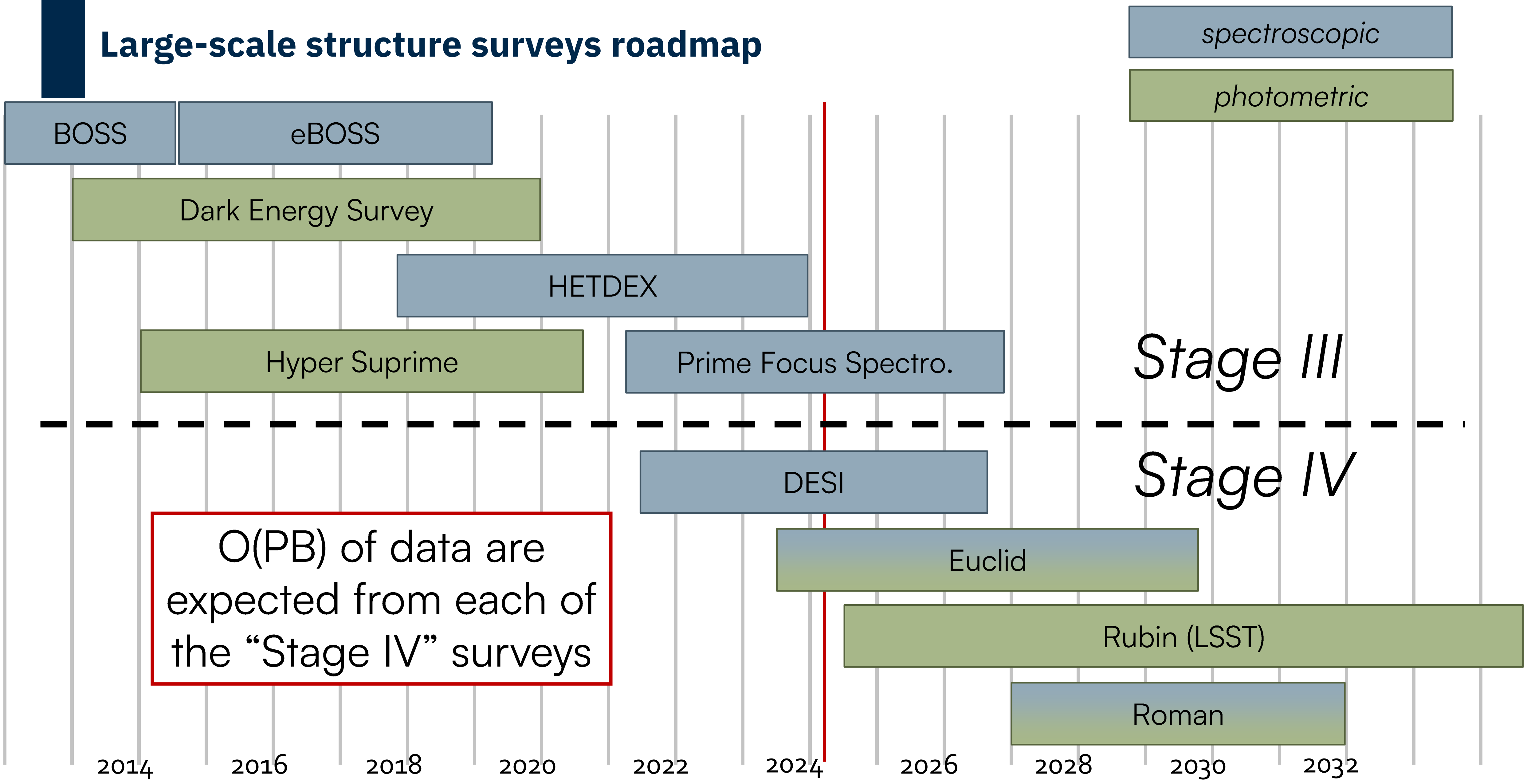






Predictions about the past of the Universe are almost identical!

Large-scale structure surveys roadmap



Euclid’s view of the Perseus cluster of galaxies, ESA, 07/11/2023

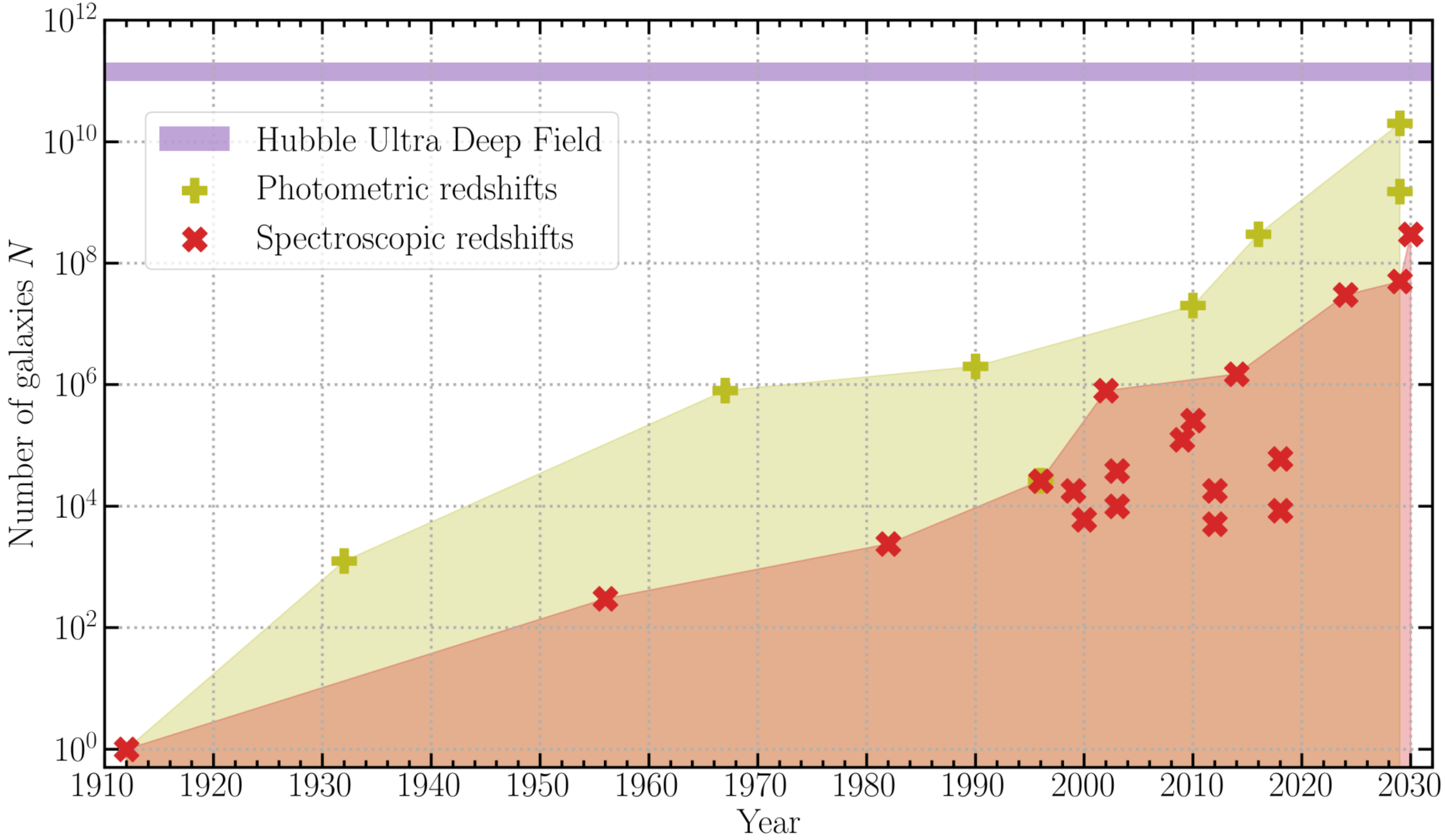
The growth of data, models, methods, and computers

- We live in an age where everything grows quickly. But what is growing the fastest?

Data?	Models?
Methods?	Computers?

The growth of data

- The number of **observed galaxies** has grown exponentially since 1910.

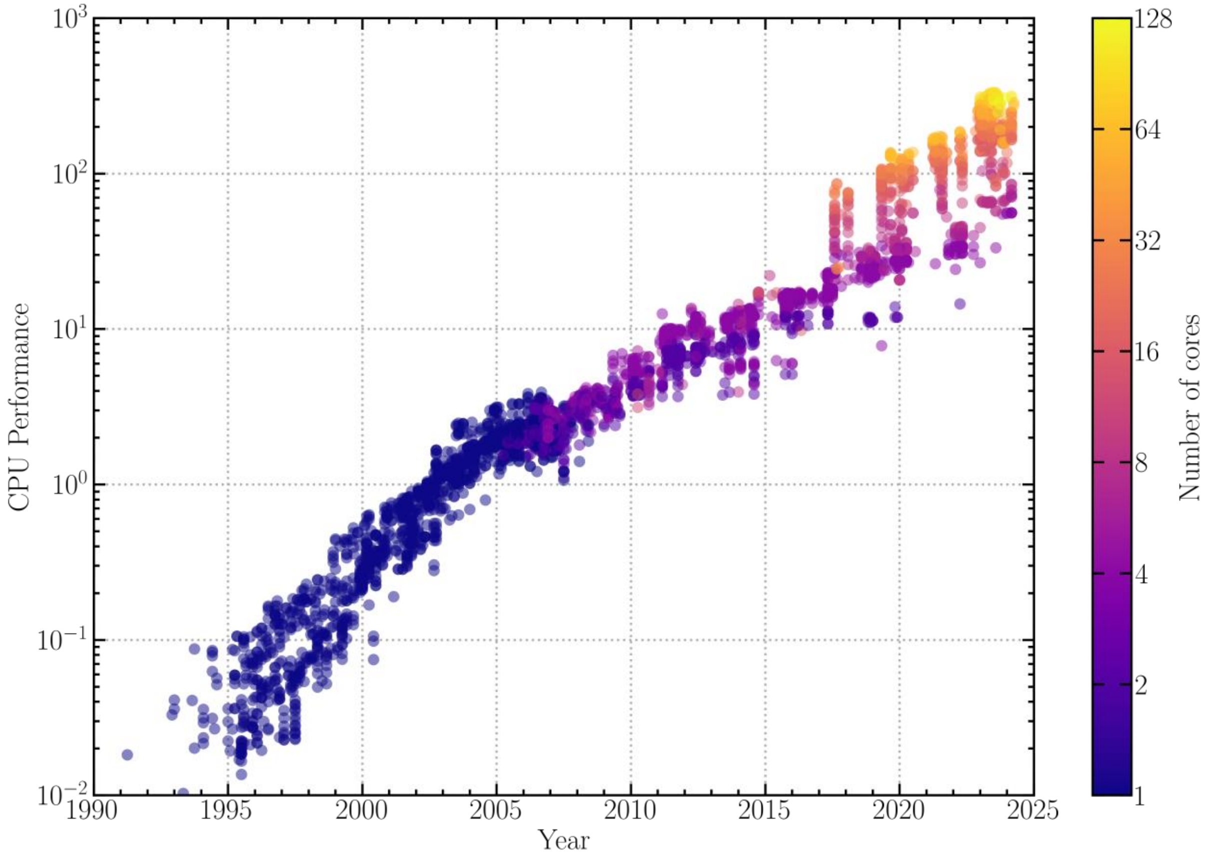


Galaxy surveys: figure inspired by J. Peacock, data collected by J. Jasche

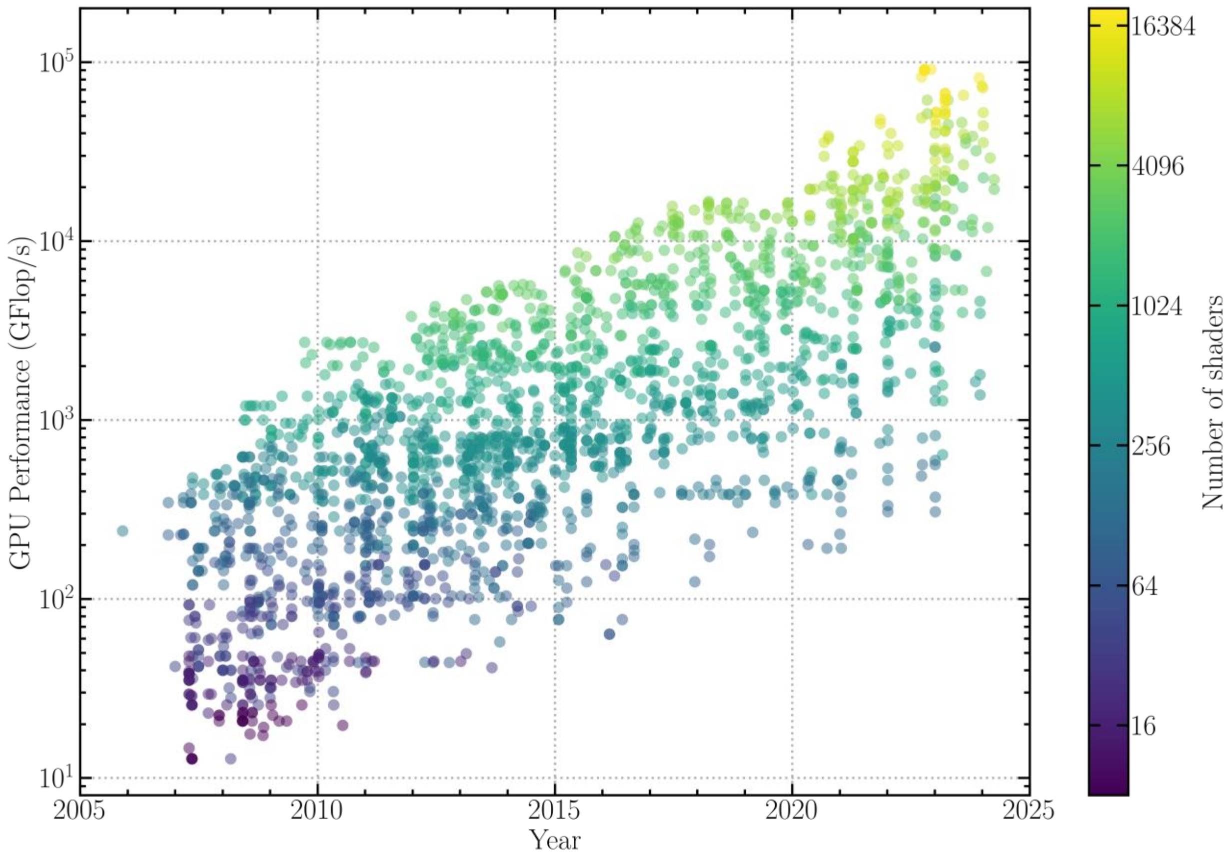
THE FORWARD PROBLEM: FROM THEORY TO DATA

The growth of computers

- Traditional hardware architectures are reaching their physical limit: per-core compute performance is slowing down.
- Modern architectures are hybrid: cores + hardware accelerators: GPUs, reconfigurable or dedicated chips (FPGAs/ASICs).



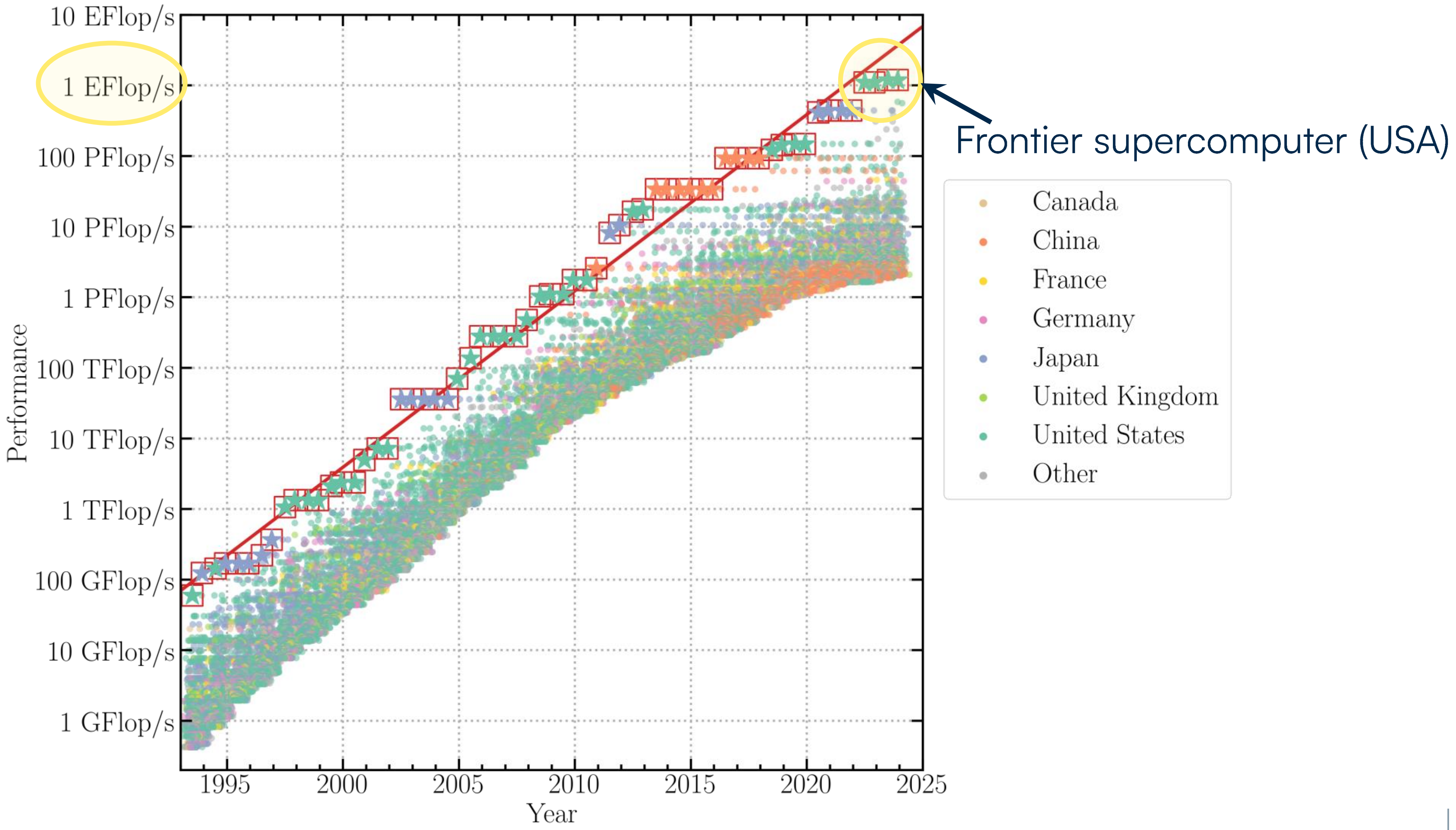
CPU Performance: based on adjusted SPECfp® results



GPU Performance: based on data from techpowerup.com

The growth of computers

- We have just entered the era of exascale computing.

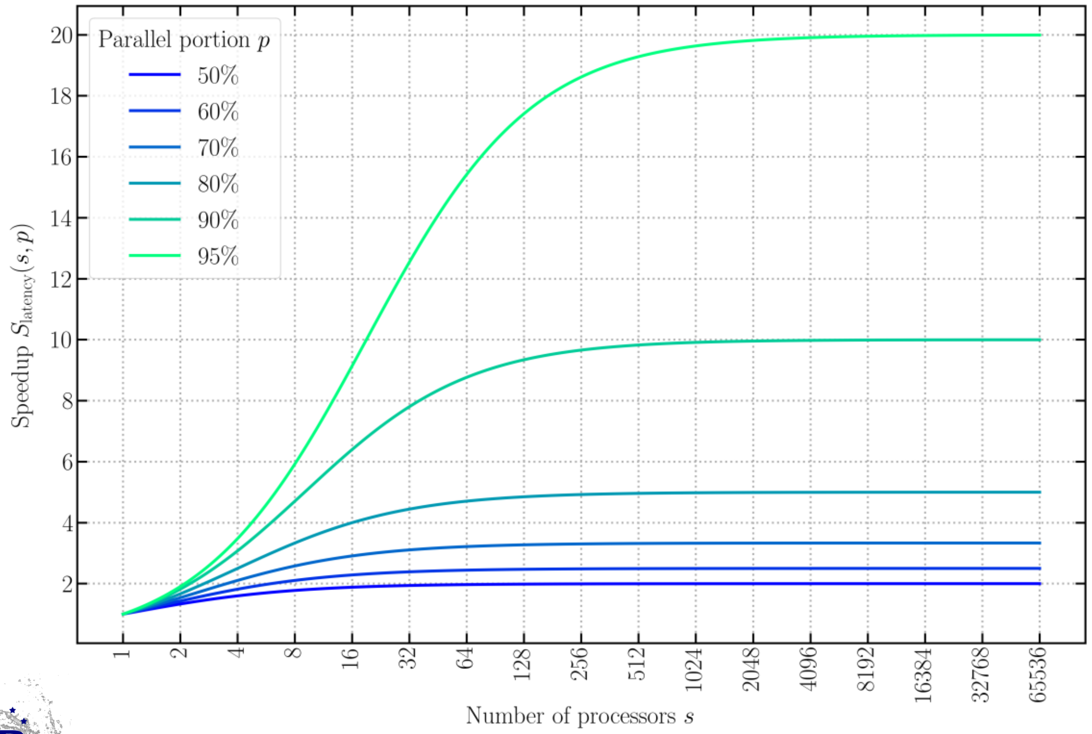


TOP 500 : data from top500.org

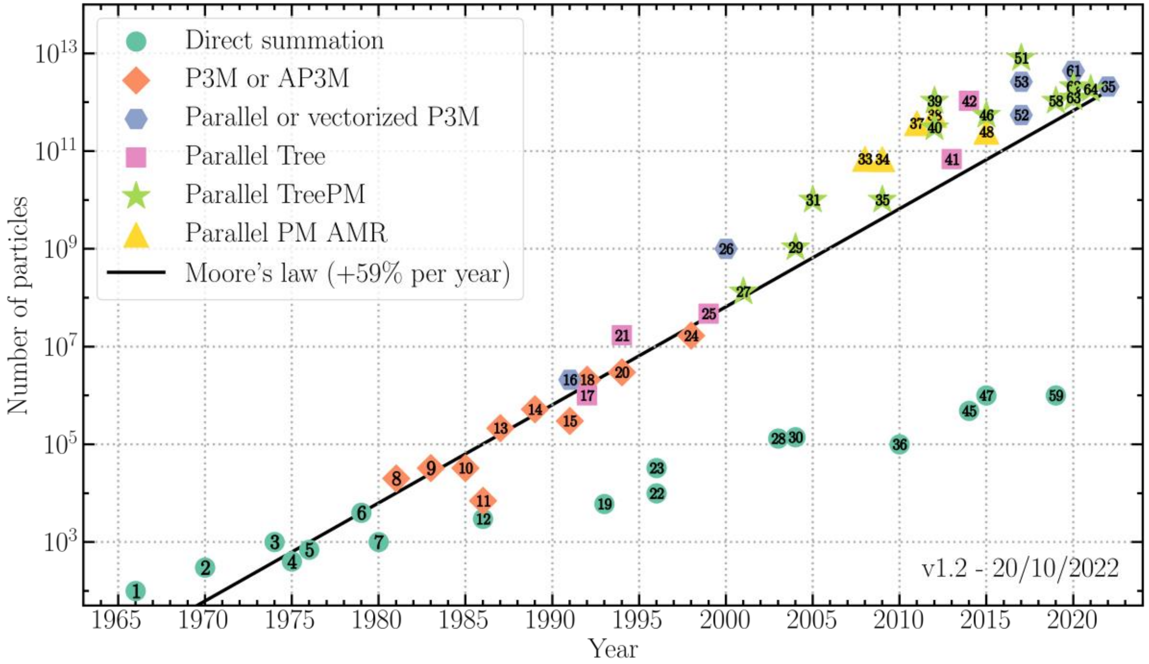
Parallelisation of *N*-body codes: the challenge

- Compute cycles are no longer the scarce resource. The cost is driven by interconnections.
- Amdahl's law: latency kills the gains of parallelisation.

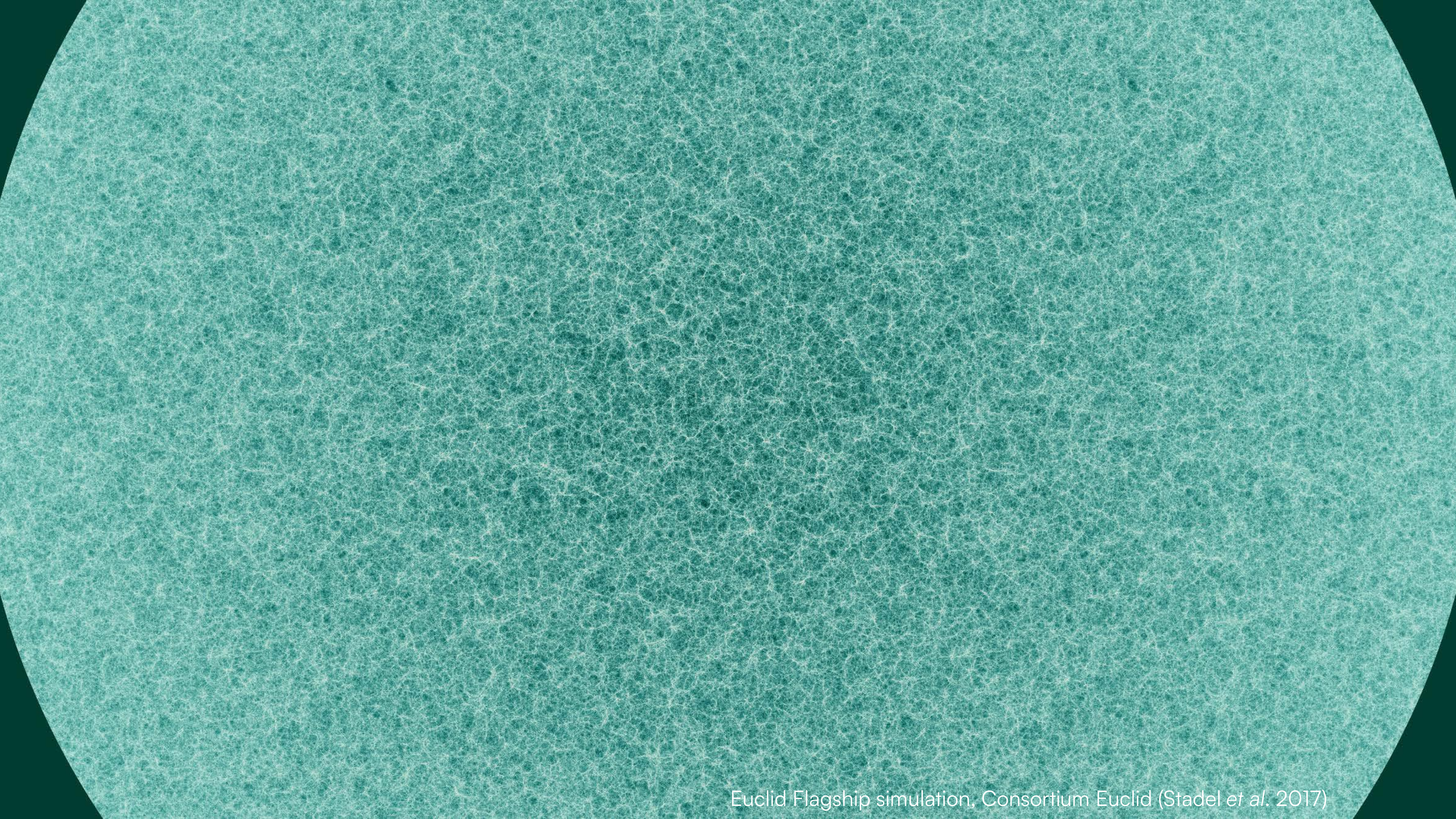
  Amdahl 1967, doi:10.1145/1465482.1465560



- Numerical data models cannot merely rely on computers becoming faster to reduce the computational time.
- Most of the work on numerical cosmology so far has focused on algorithms (such as tree, multipole, and mesh methods) that reduce the need for communications across the full volume



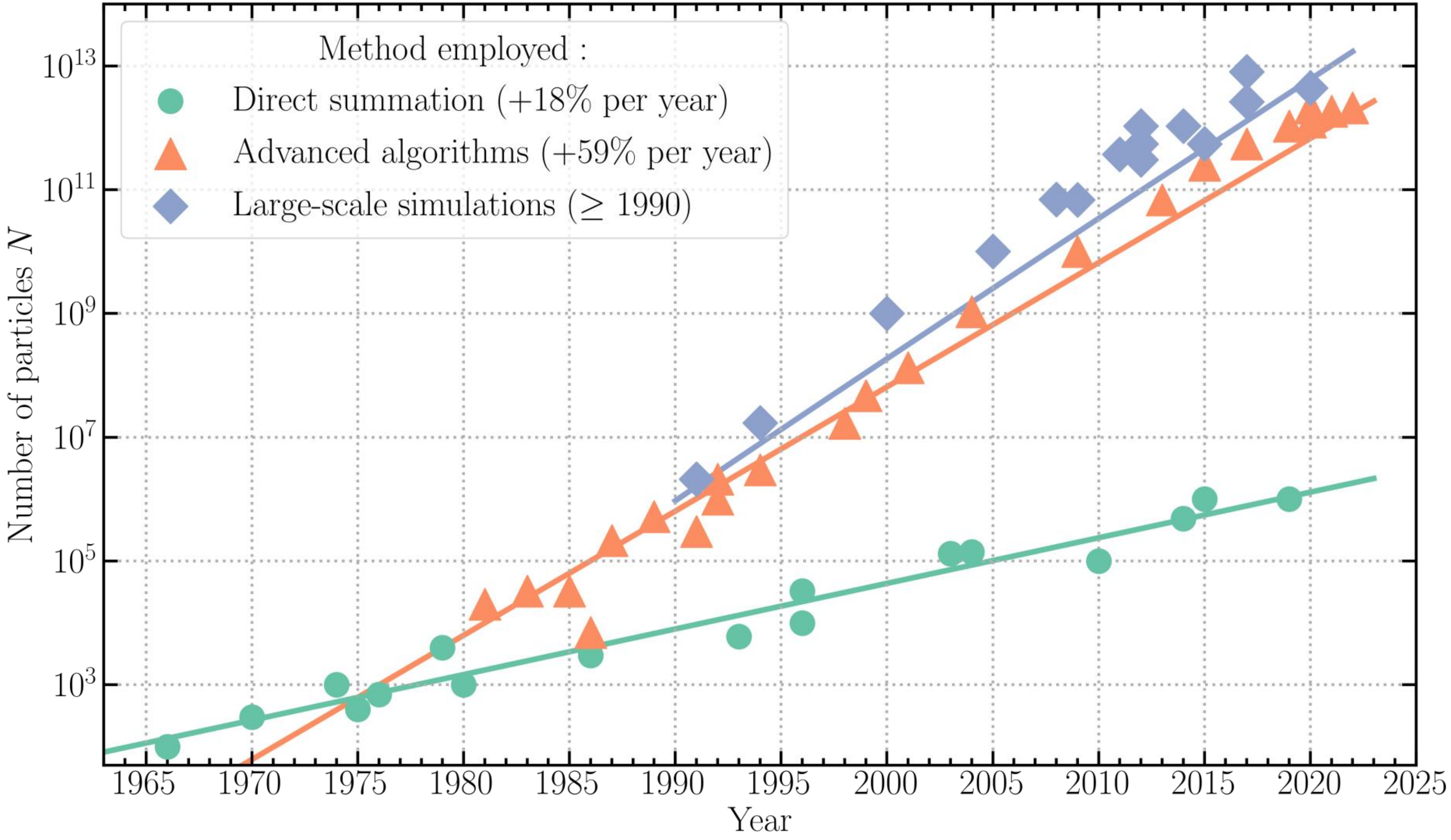
All references at Github:florent-leclercq/Moore_law_cosmosims



Euclid Flagship simulation, Consortium Euclid (Stadel *et al*. 2017)

The growth of models

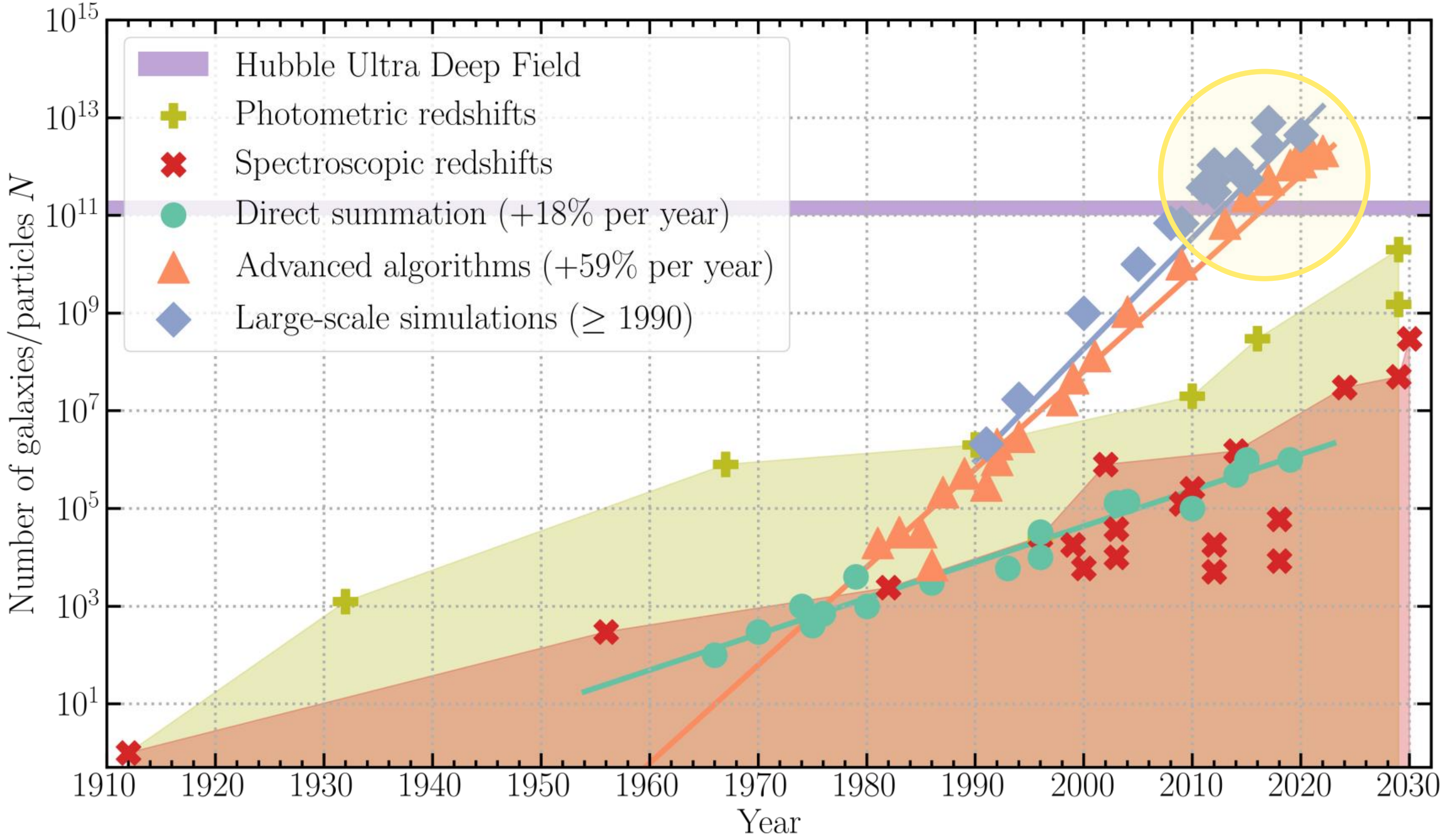
- Numerical simulations are the new way to express theoretical models.



Cosmological simulations: Github:florent-leclercq/Moore_law_cosmosims

Comparative growth of data and models

- We are already using more particles in simulations than there are galaxies in the observable Universe!

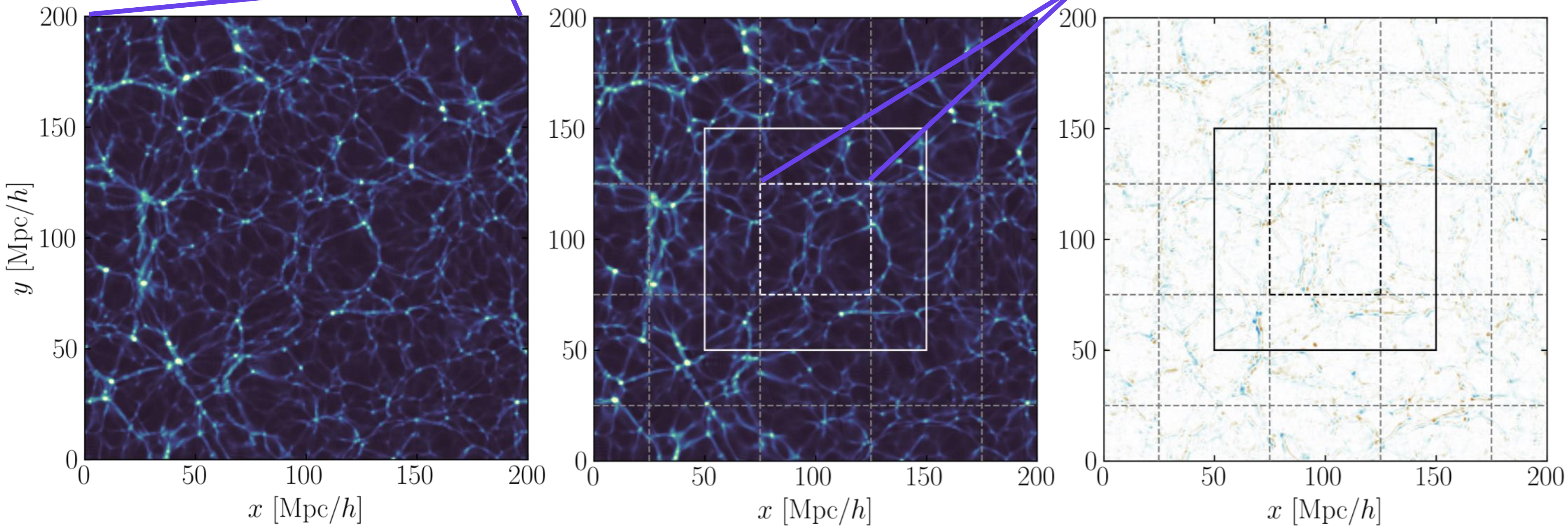


Galaxy surveys: figure inspired by J. Peacock, data collected by J. Jasche
Cosmological simulations: [Github:florent-leclercq/Moore_law_cosmosims]

Perfectly parallel cosmological simulations using spatial comoving Lagrangian acceleration (sCOLA)

- Can we decouple sub-volumes by using the large-scale analytical solution?

$$\frac{\partial^2 \mathbf{x}}{\partial t^2} = -\nabla \left[\underbrace{\Delta^{-1}\delta}_{\text{non-local}}\right] \Longleftrightarrow \frac{\partial^2}{\partial t^2}\left(\mathbf{x} - \mathbf{x}_{\mathrm{l.s.}}\right) = -\nabla \left[\underbrace{\Delta^{-1}\left(\delta - \delta_{\mathrm{l.s.}}\right)}_{\text{local}}\right]$$



tCOLA (reference) sCOLA Difference

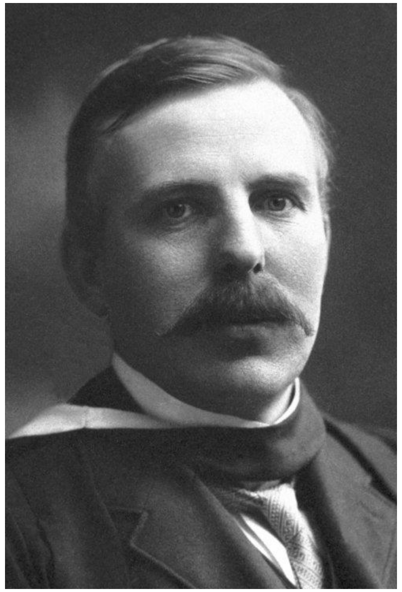
FL, Faure, Lavaux, Wandelt, Jaffe, Heavens, Percival & Noûs, 2003.04925

Publicly available implementation: Bitbucket:florent-leclercq/simbelmyne/

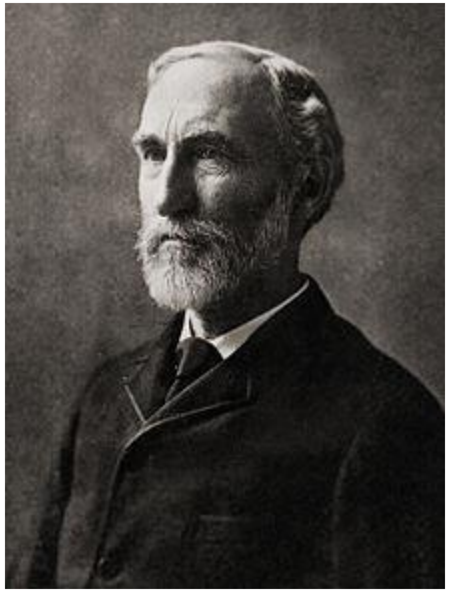
THE INVERSE PROBLEM: FROM DATA TO THEORY

Why proper statistics matter

> If your experiment needs statistics, you ought to have done a better experiment.
> Ernest Rutherford



Ernest Rutherford
(1871-1937)



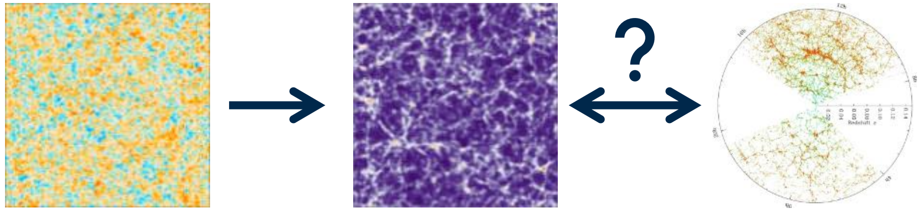
J. Willard Gibbs
(1839-1903)

- Gibbs's canonical and grand canonical ensembles, derived from the maximum entropy principle, fail to correctly predict thermodynamic properties of real physical systems (1884-1902).
- The predicted entropies are always larger than the observed ones… there must exist additional microphysical constraints:
  - Discreteness of energy levels: radiation: Planck (1900), solids: Einstein (1907), Debye (1912), Ising (1925), individual atoms: Bohr (1913)…
  - …Quantum mechanics: Heisenberg, Schrödinger (1927)

The first clues indicating the need for quantum physics were uncovered by seemingly "unsuccessful" application of statistics!

Why Bayesian inference in cosmology?

- Inference of signals: an ill-posed problem
  - Incomplete observations: finite resolution, survey geometry, selection effects
  - Noise, biases, systematic effects
  - Cosmic variance



➡ No unique recovery is possible!

- A natural progression in cosmology:
  - Observations of the homogeneous and isotropic expansion (supernovæ)
  - Anisotropies of linear perturbations (CMB)
  - Non-linear cosmic structure at small scales and late times (galaxy surveys)

- Additional challenges for next-generation data:
  - Difficult data analysis questions and/or hints for new physics will first show up as tensions between measurements
  - Non-linearity: 80% of the total signal will come from non-linear structures

    e.g. LSST Science Book, 0912.0201
  - Model misspecification: Next-generation surveys will be dominated by (unknown) systematics

A simple statement about building knowledge

This is (probably!) not the right person

- Bayes' theorem (1763): a statement about how we analyse evidence and change our minds at we get new information:

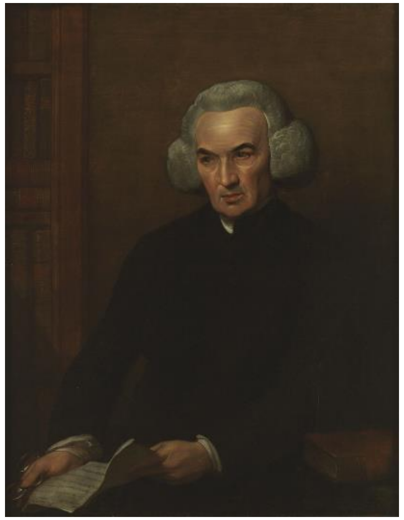
$$p(s|d) = \frac{p(d|s)p(s)}{p(d)}$$

Demonstration:

$$p(s,d) = p(s|d)p(d) = p(d,s) = p(d|s)p(s). \quad \square$$

- But why should we use it?
  - Bayes' theorem is trivial and outdated.
  - It measures belief. It says we can learn even from missing or incomplete data, from approximations, from ignorance. It runs counter to the conviction that science requires objectivity and precision.
  - After Laplace's death, it was pronounced dead and buried.

Thomas Bayes (1701-1761)



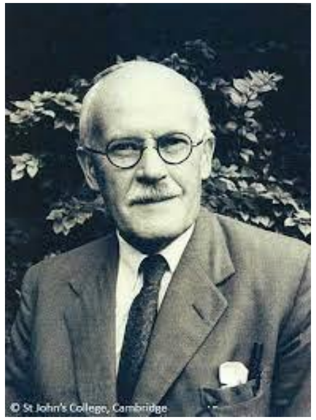
Richard Price (1723-1791)

Pierre-Simon de Laplace (1749-1827)

Picture taken at Bunhill Fields Burial Ground, City of London, 2021

Controversy: frequentism versus Bayesianism

- Two different conceptions of the nature of probability and of scientific questions:
  - Frequentism: “Objective” probabilities linked to the frequency of repetitive random phenomena. Questions related to specific and reproducible experiments.
  - Bayesianism: “Subjective” probabilities related to the degree of belief given to a measurement or a theory. Questions related to phenomena and choices not involving the idea of repetition.

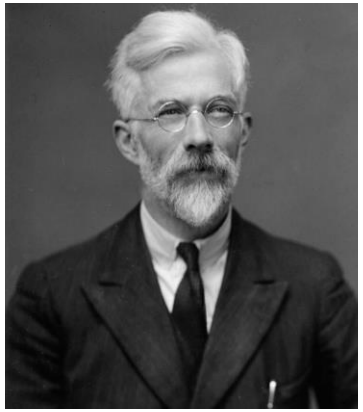


Harold Jeffreys
(1891-1989)

Leonard J. Savage
(1917-1971)

> [Fisher] sometimes published insults that only a saint could entirely forgive.
>
> Savage 1976, On Rereading R. A. Fisher

Karl Pearson
(1857-1936)



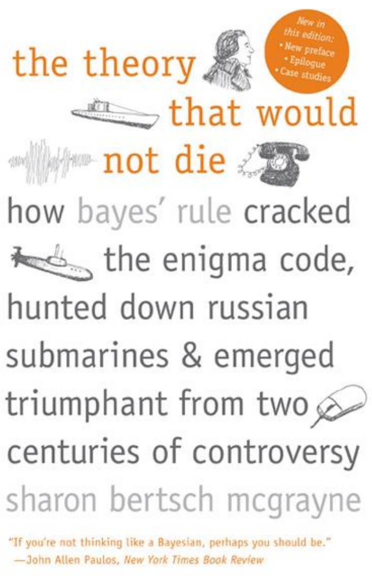
Ronald Aylmer Fisher
(1890-1962)

Jerzy Neyman
(1894-1981)

- Frequentist and Bayesian techniques give the same results when working on large samples. It is only on small numbers and low occurrences that frequentist estimation and Bayesian induction differ.

The theory that would not die

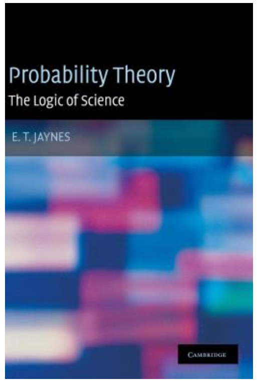
- And yet, after Laplace, Bayes' theorem helped in many practical situations:
  - Exonerate Alfred Dreyfus from miscarriage of justice (Henri Poincaré, 1899-1906),
  - Save the Bell Telephone system from financial panic (Edward C. Molina, 1907),
  - Predict earthquakes and tsunamis (Harold Jeffreys, 1930-1940),
  - Break the German navy's Enigma cipher (Alan Turing, 1940-1944),
  - Prove that smoking causes lung cancer (Jerome Cornfield, 1951),
  - Search for an H-bomb then a submarine lost at sea (John P. Craven, 1966-1968)



Sharon Bertsch McGrayne 2012

- The scientific battle lasted for 150 years, until computers arrived.

> The superiority of Bayesian methods is now a thoroughly demonstrated fact in a hundred different areas. One can argue with a philosophy; it is not so easy to argue with a computer printout, which says to us: "Independently of all your philosophy, here are the facts of actual performance."
>
> Jaynes 2002, Probability Theory — The logic of science



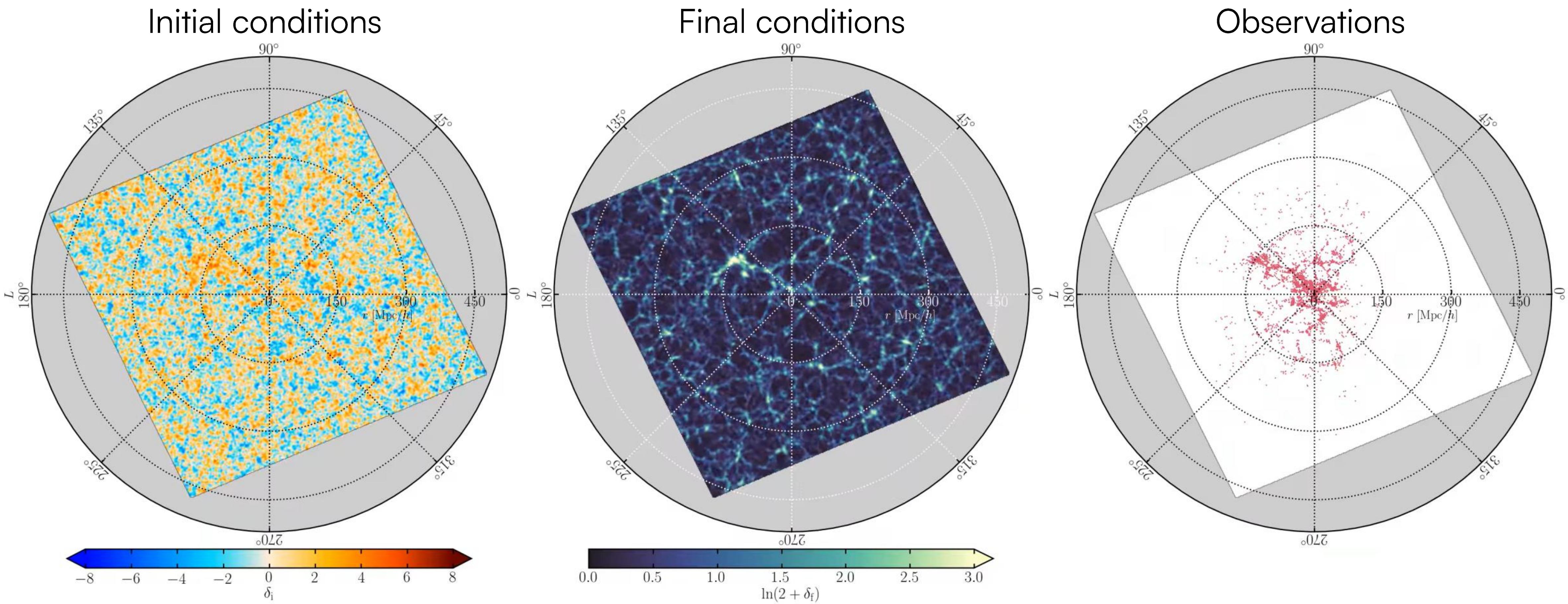
Jaynes 2002

Edwin Thompson Jaynes (1922-1998)

Richard Threlkeld Cox (1898-1991)

- Cox-Jaynes theorem (1946): Any system to manipulate "plausibilities", consistent with Cox's desiderata, is isomorphic to Bayesian probability theory.

Bayes at work in cosmology:
The BORG algorithm (*Bayesian Origin Reconstruction from Galaxies*)

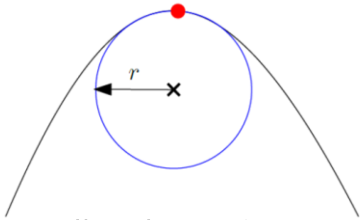


67,224 galaxies, ≈ 17 million parameters, 5 TB of primary data products, 10,000 samples, ≈ 500,000 forward and adjoint gradient data model evaluations, 1.5 million CPU-hours

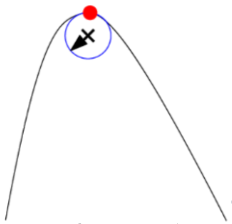
Jasche & Wandelt, 1203.3639; Jasche, FL & Wandelt, 1409.6308; Jasche & Lavaux, 1806.11117; Lavaux, Jasche & FL, 1909.06396

What there is to learn:
Notion of information in probability theory

- With Bayes' theorem we know *how* to learn. But how much *can* we learn? The Fisher information (1922) measures the amount of information that a random variable contains about an unknown parameter.

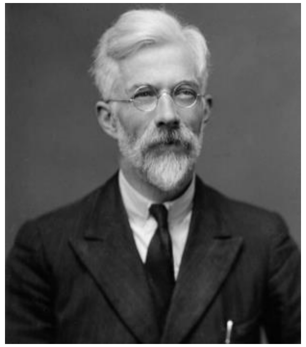


small Fisher information
flat likelihood peak
large variance, low accuracy



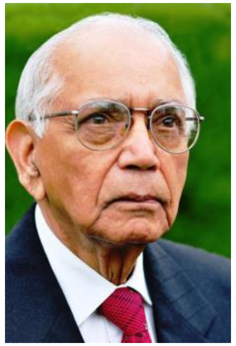
large Fisher information
sharp likelihood peak
small variance, high accuracy

Nielsen 2020, 1808.08271



Ronald Aylmer Fisher (1890-1962)

Prasanta Chandra Mahalanobis (1893-1972)



Calyampudi Radhakrishna Rao (1920-2023)

- Generalisations yield the field of information geometry:
  - The Mahalanobis distance (1927) measures the distance between a point and a distribution.
  - For a multi-dimensional problem, the Fisher information generalises to a matrix, and defines a metric: the Fisher-Rao metric (1945).

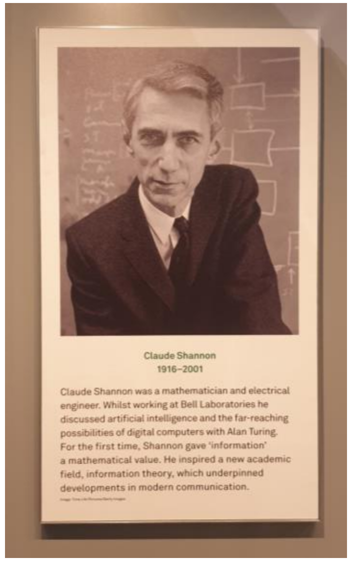
> (...) I suggested the ***differential geometric approach*** in my 1945 paper by considering the space of probability distributions. I used Fisher information matrix in defining the metric, so it was called Fisher-Rao metric. Differential geometry was not well known at that time, and in order to compute the geodesic distance from the metric, I had to learn the mathematics from papers on ***relativity describing Einstein metric***. It was only 30 years later, my work received attention (...).
>
> Calyampudi Radhakrishna Rao, Scholarpedia

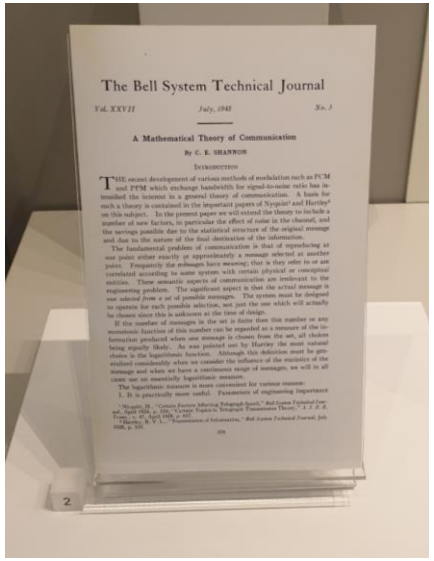
- But is information/learning really geometric?
  - Divergences: $D_{\mathrm{KL}}(P||Q) \neq D_{\mathrm{KL}}(Q||P)$

What information is missing: Notion of entropy in probability theory

- Information entropy: average amount of information conveyed by an event, when considering all possible outcomes.

Shannon 1948



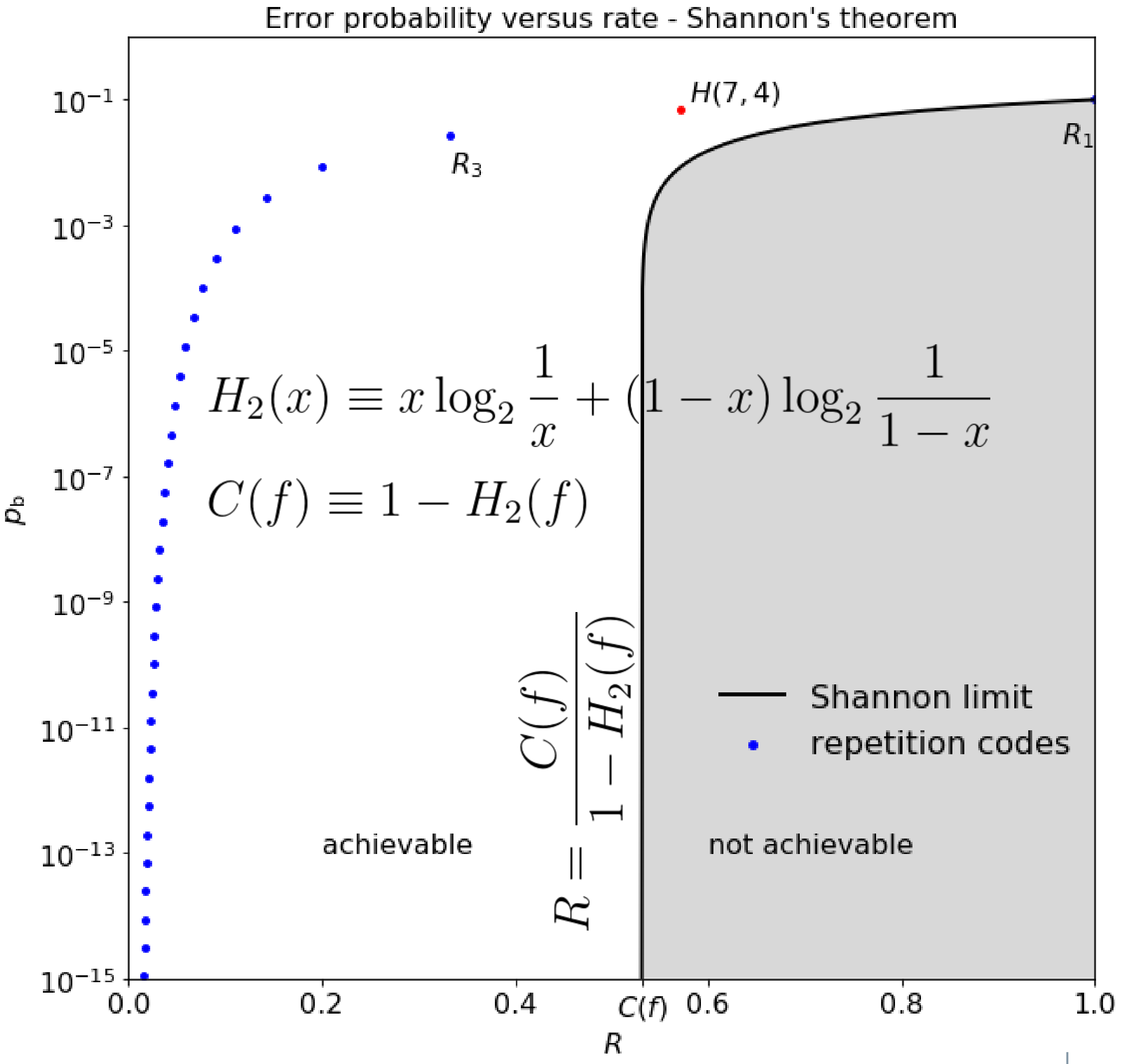


Pictures taken at the Science Museum, South Kensington, 2021

Why don't you call it entropy? In the first place, a mathematical development very much like yours already exists in Boltzmann's statistical mechanics, and in the second place, no one understands entropy very well, so in any discussion you will be in a position of advantage.

von Neumann to Shannon, about a name for "missing information"

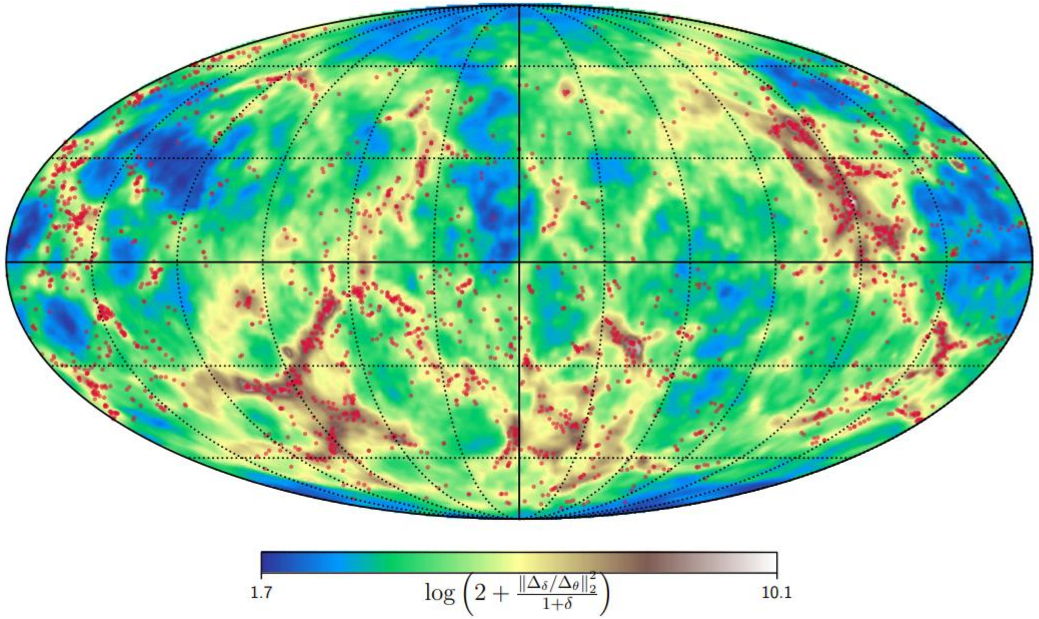
- The notion of entropy is useful to communication, but also data compression.



Shannon 1948, McKay 2003

Bayesian experimental design:
Information-optimal or entropy-maximal acquisition of future cosmological data

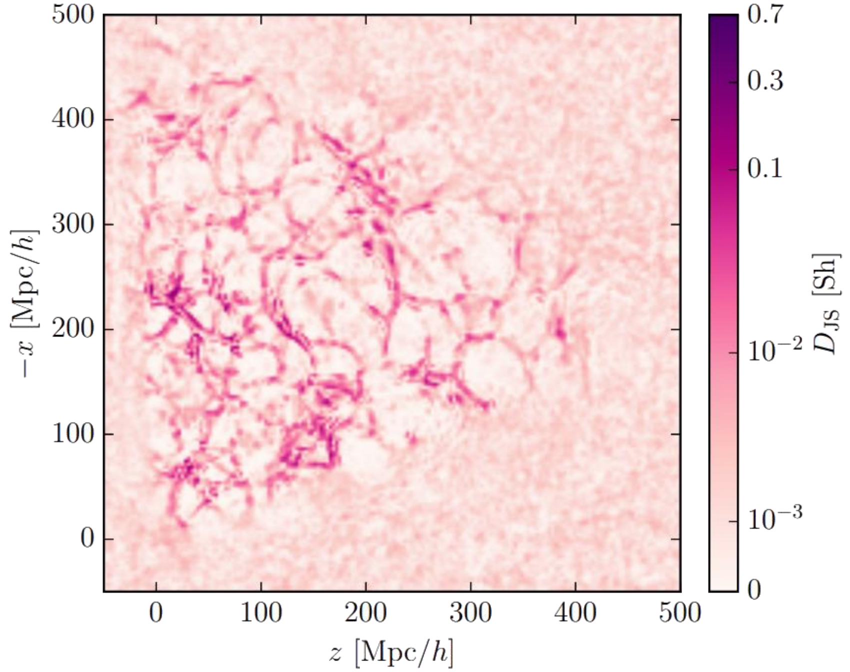
- This is where to look if we want to measure cosmological parameters of ΛCDM…



(Fisher information map for perturbative changes in the cosmological model)

Kostić, Jasche, Kodi Ramanah & Lavaux, 2107.00657

- And this is where to look if we want to learn about dark energy…



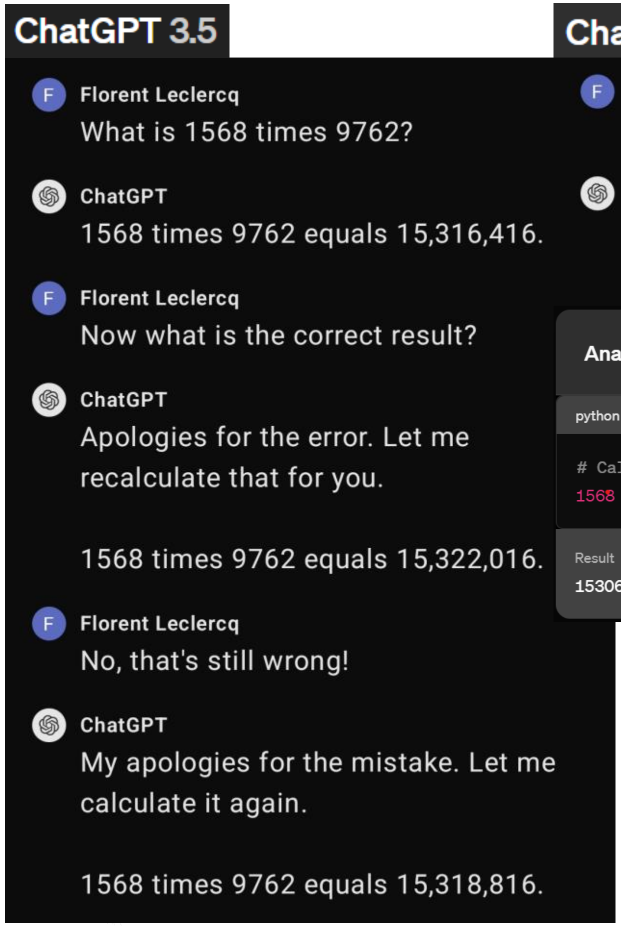
(Jensen-Shannon divergence between cosmic web-type posteriors for different values of the dark energy equation of state)

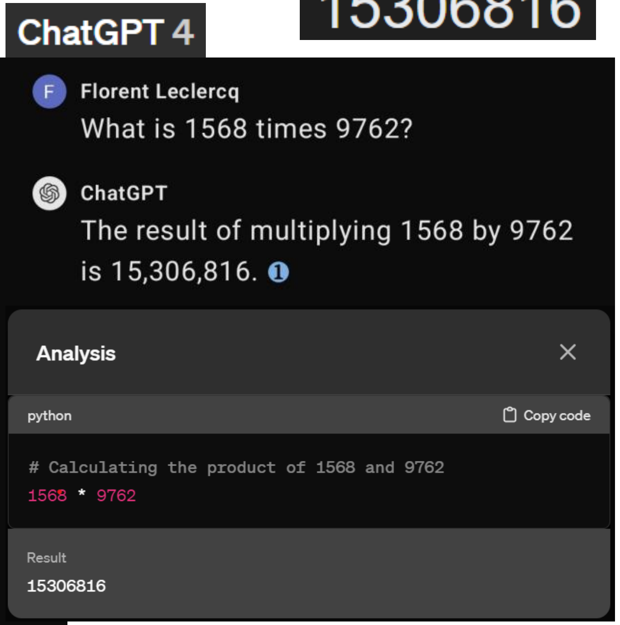
FL, Lavaux, Jasche & Wandelt, 1606.06758

THE IMITATION PROBLEM: ALGORITHMS BEYOND BLIND ORACLES?

Some examples of AI in 2024…

- Large language models (and their limitations…)





- Text-conditional image generation:

Exemples : Dall-E, Stable Diffusion, Midjourney

« Cute cats working on new AI research as digital art »

Generated with DALL-E 2 on 08/10/2022

Generated with DALL-E 3 on 08/03/2024

The birth of AI

- The Dartmouth Summer Research Project on Artificial Intelligence (1956).

Marvin Minsky, Claude Shannon, Ray Solomonoff and other scientists at the Dartmouth Summer Research Project on Artificial Intelligence (Photo: Margaret Minsky)

- The proposal (31 August 1955) states:

> We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that ***every aspect of learning or any other feature of intelligence*** can in principle be so precisely described that ***a machine can be made to simulate it***. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.
>
> J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon

AI algorithms: metaphors & methodology

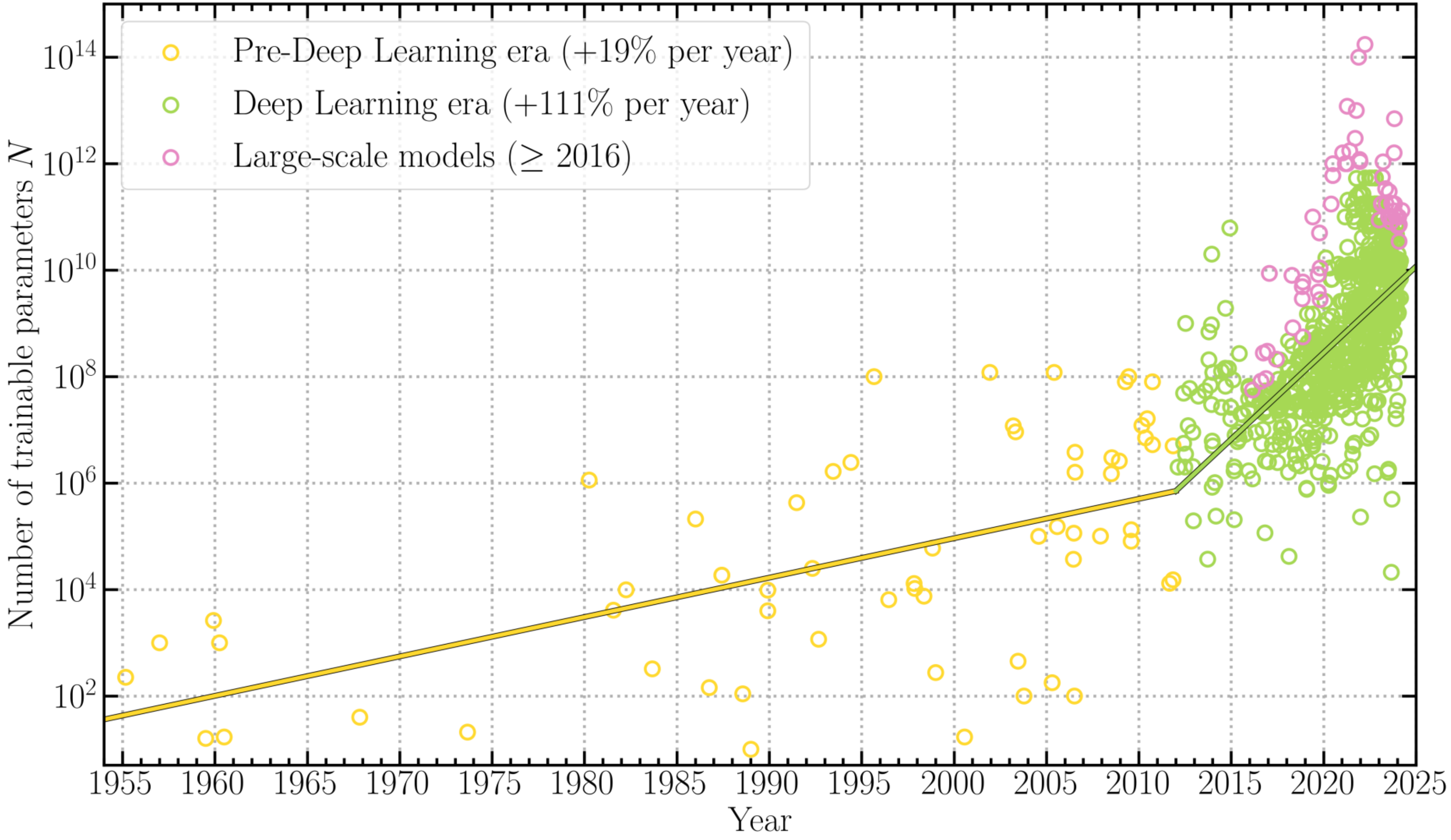
- Humanity: classical theories of learning
  - Rule-based models, case-based reasoning (Aamodt & Plaza 1994)
  - Learning by practice, “chunking” (Newell & Rosenbloom 1981)
  - Reinforcement learning (Samuel 1959)
  - Non-supervised learning (Feigenbaum 1963), e.g. auto-encoders (Kramer 1991)
- Physiology: the brain
  - Artificial neuron (McCullogh & Pitts 1943), perceptron (Rosenblatt 1958)
  - Multi-layer perceptrons (Rumelhart *et al.* 1986, Rumelhard & McClelland 1987), gradient back-propagation (Rumelhart *et al.* 1986)
  - Deep learning & convolutional neural networks (LeCun *et al.* 2015, Goodfellow *et al.* 2016)
- Nature: evolution
  - Genetic algorithms (Holland 1975)
- Culture: epistemology
  - Scientific discovery (Langley et al. 1987)
  - Ontologies (Powers & Turk 1989), semantic web
- Physics: statistical mechanics, thermodynamics, quantum physics
  - Decision trees (Quinlan 1975), Bayesian networks, graphs
  - Hamiltonian Monte Carlo (Duane et al. 1987)
  - Information theory, distributed AI (Demazeau & Müller 1989)
  - Hidden Markov Models (Baum 1966)

Symbolic AI: explainable but costly

Connectionist/numerical AI (machine learning): automatic but “black-box”

The growth of methods

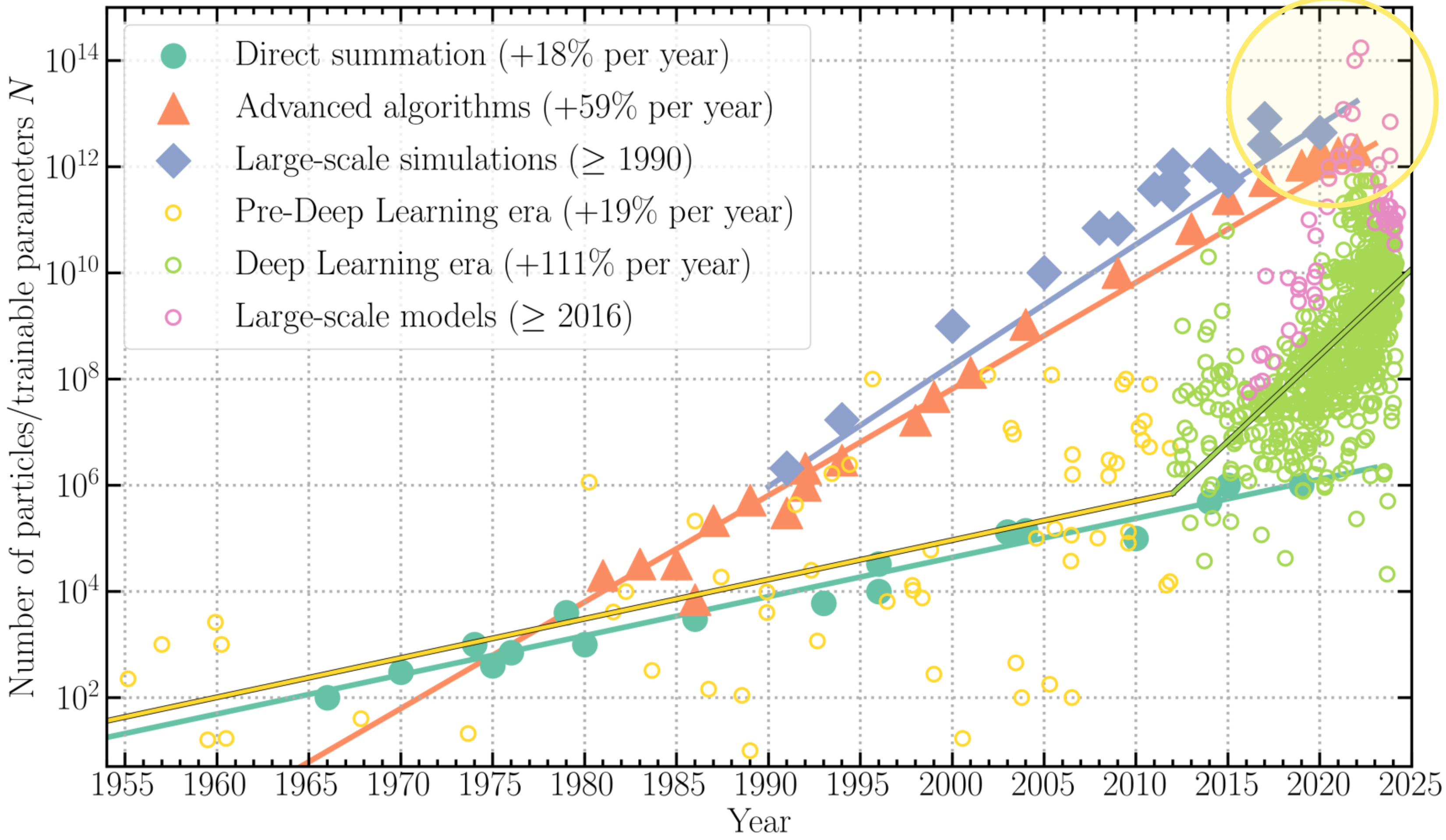
- ML methods are characterised by the number of trainable parameters.



IA models: data from epochai.org

Comparative growth of models and methods

- Machine learning methods have caught up with the largest cosmological simulations!



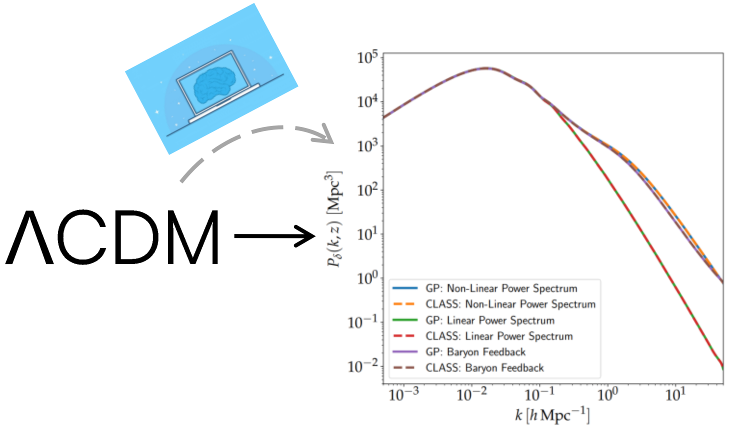
Cosmological simulations: Github:florent-leclercq/Moore_law_cosmosims
IA models: data from epochai.org

Why machine learning for cosmology?

Last conference at the IAP (November 2023)

Speed up & go beyond approximations

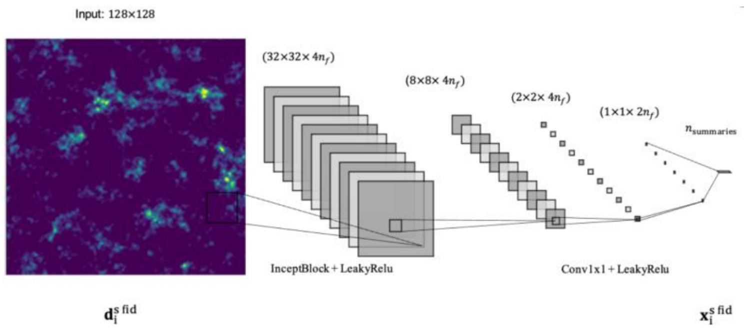
Emulators



emuPK: Mootoovaloo, Jaffe, Heavens & FL, 2105.02256

Find the information content

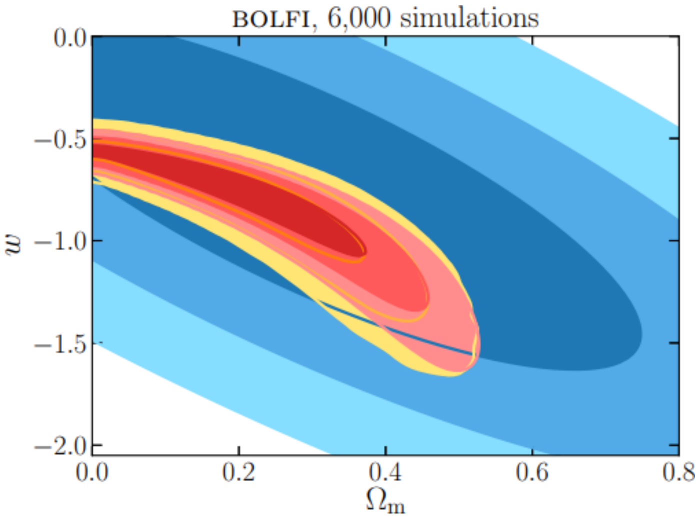
Automatic data compression



Information Maximising Neural Networks (IMNN): Charnock, Lavaux & Wandelt, 1802.03537; Makinen *et al.*, 2107.07405

Build a posterior/evidence approximator

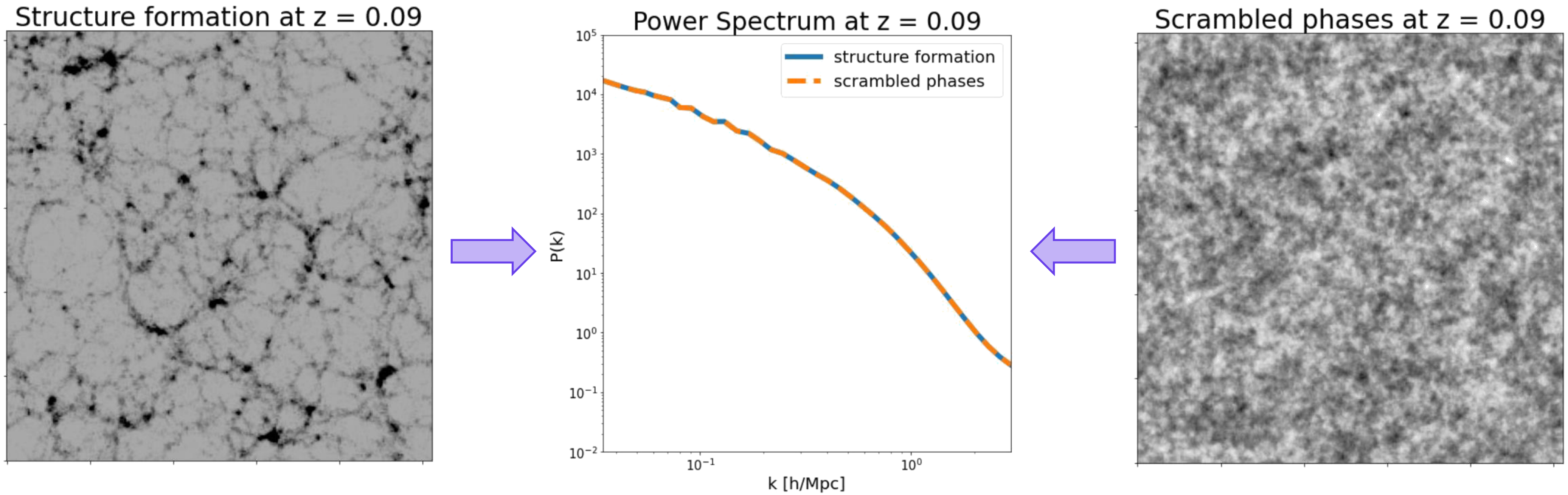
Implicit likelihood inference



Bayesian Optimisation for Likelihood-Free Inference (BOLFI): FL, 1805.07152

What makes a good AI/ML model?
Associative versus causal reasoning in scientific research

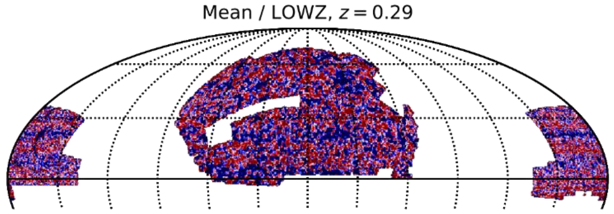
- With traditional machine learning, we obtain associative links between a latent space and data. But this doesn't mean we understand how nature works!



J. Jasche

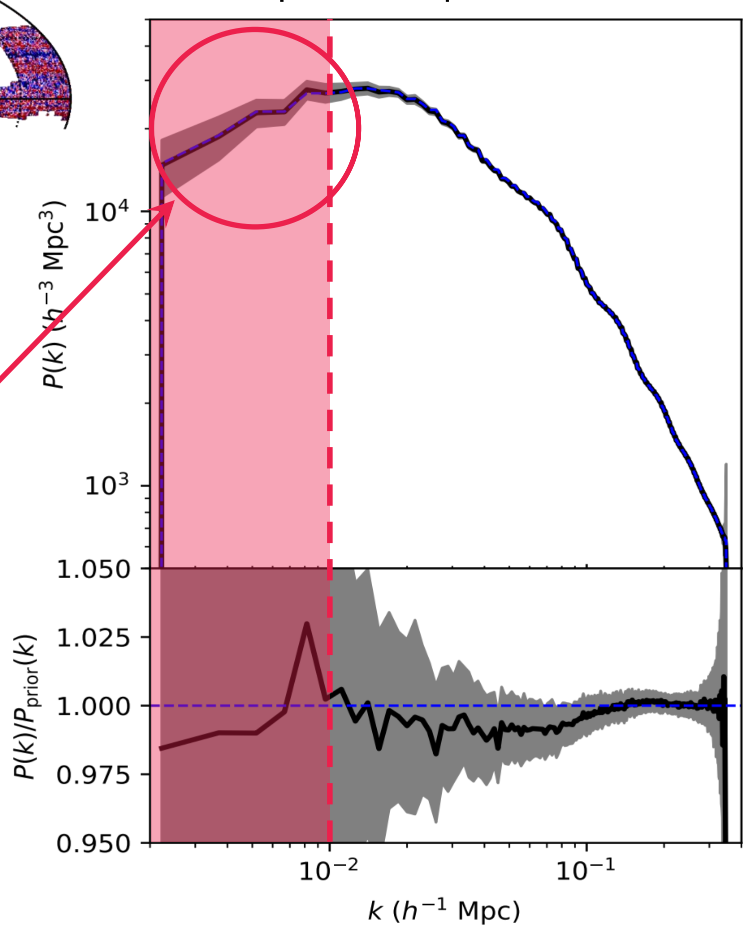
- Purely data-driven machine learning may not be sufficient for research!
  - Causally consistent generative AI/ML models produce explanations.
  - Predictive AI/ML models are assistance systems for hypothesis generation.

Machine-aided report of unknown data contaminations
Application to SDSS-III/BOSS (LOWZ+CMASS)



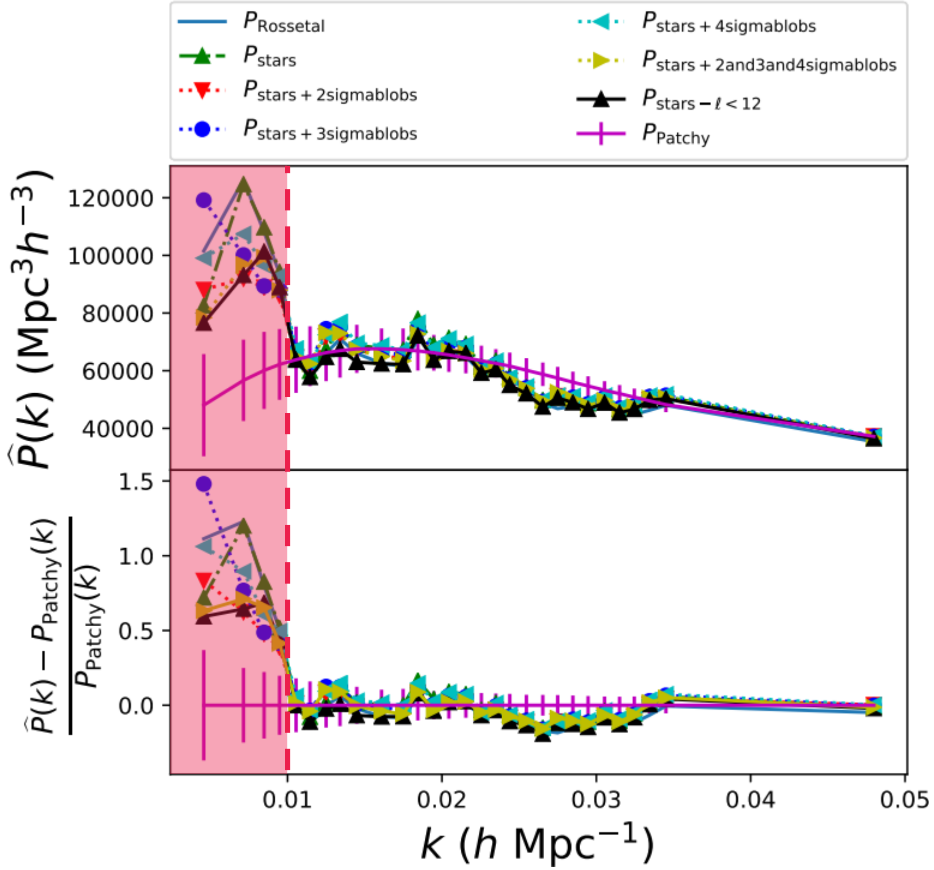
A map of an unknown foreground

BORG *a posteriori* power spectrum



No apparent contamination, even well beyond the turn-over

State-of-the-art with backward-modelling technique (mode subtraction)



Porqueres, Ramanah, Jasche & Lavaux, 1812.05113
Lavaux, Jasche & FL, 1909.06396

Kalus, Percival *et al.*, 1806.02789

Conclusion:
Hopes and challenges in information science for cosmology

The forward problem

- Hopes: Numerical models are the new way to formulate theory in data analysis.
- Challenges: Scalability (and energy cost!)

The inverse problem

- Hopes: Bayesian data analysis is established as a fundamental theory of learning.
- Challenges: Control of external components in modern Bayesian models (in addition to likelihood and prior) : training data, posterior approximator...

The imitation problem

- Hopes: Machine-driven scientific discovery becomes conceivable.
- Challenges: Interpretability & explainability, proof & certifiability...

Acknowledgements & credits

References:

- Leclercq, Pisani, Wandelt 2014, 1403.1260, *Cosmology: from theory to data, from data to theory*
- Leclercq *et al.* 2020, 2003.04925, *Perfectly parallel cosmological simulations using spatial comoving Lagrangian acceleration*
- Leclercq *et al.* 2016, 1606.06758, *Comparing cosmic web classifiers using information theory*
- Leclercq 2018, 1805.07152, *Bayesian optimisation for likelihood-free cosmological inference*
- Lavaux, Jasche, Leclercq 2019, 1909.06396, *Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data*

www.florent-leclercq.eu

www.aquila-consortium.org

The author acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-23-CE46-0006 (project INFOCW).