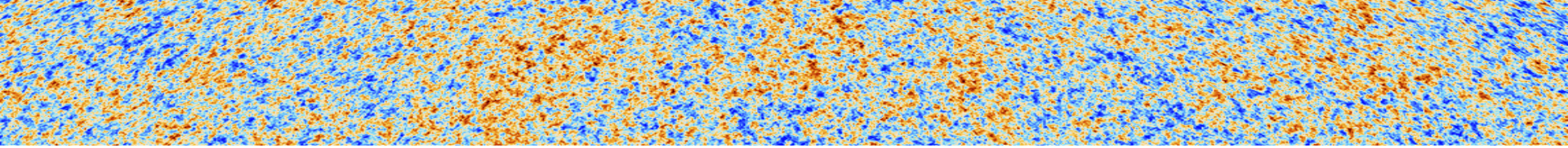


Science exploitation of CMB data

Paolo Natoli

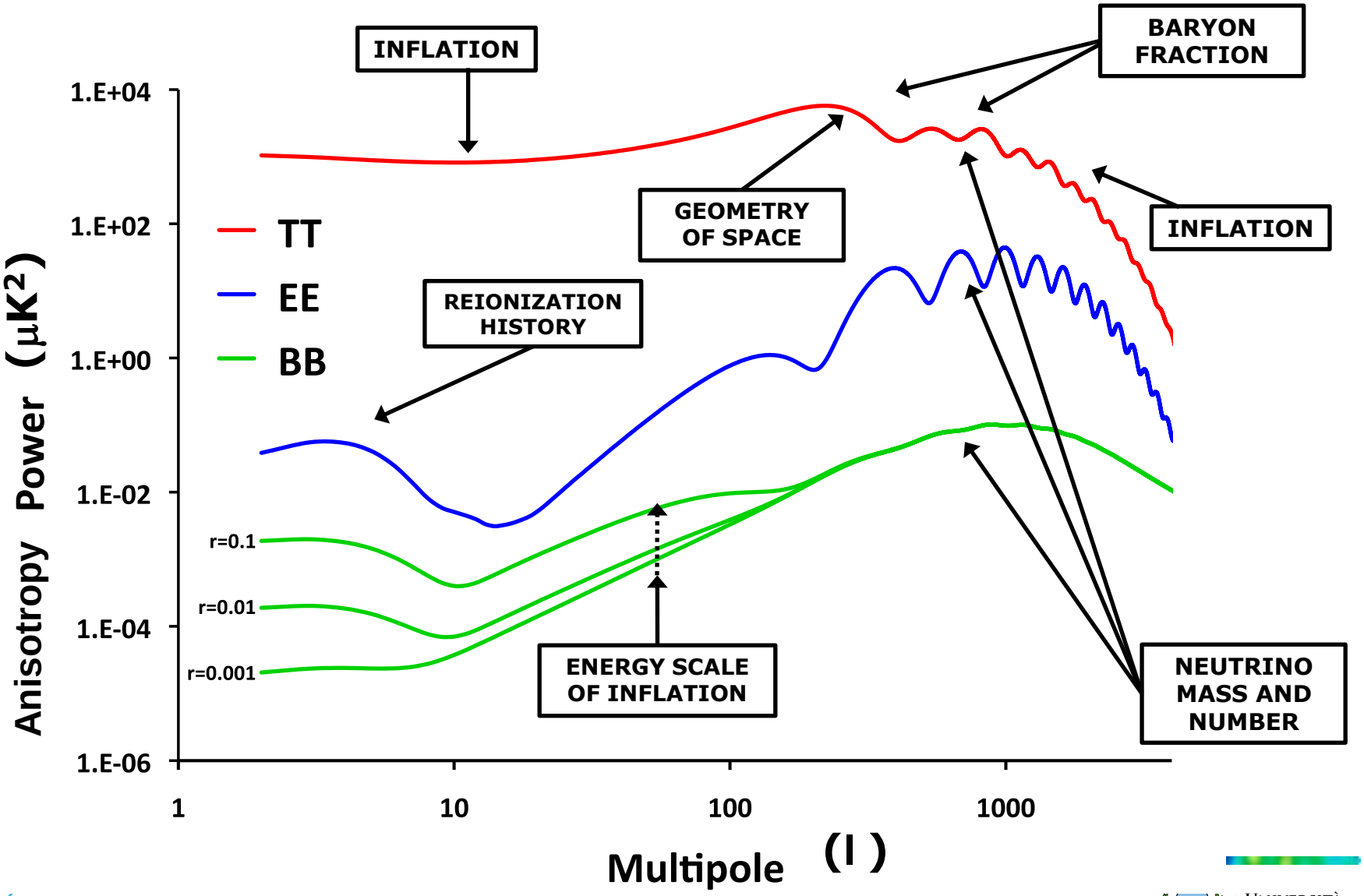
Università di Ferrara and INFN

on behalf of the Italian CMB community

Overview

- Rich, complex datasets for many science targets
- State of the art
- Role and expertise of the Italian community
- Open issues and criticalities

Rich science, faint signals, wide angular range

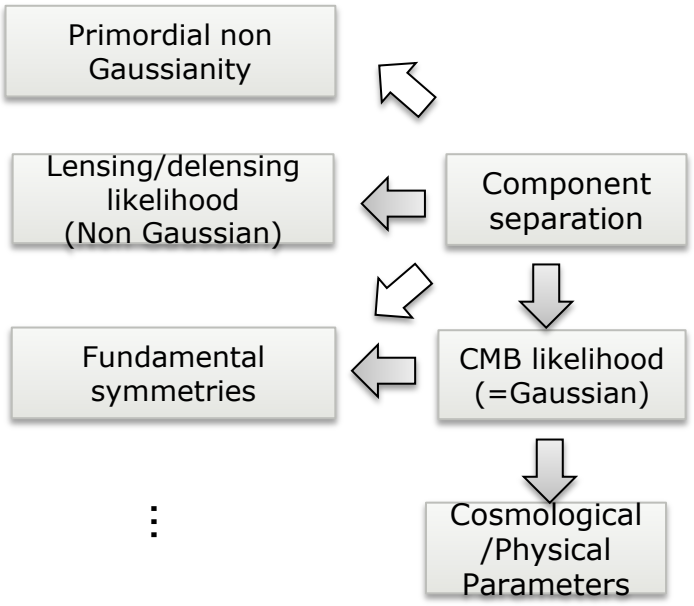


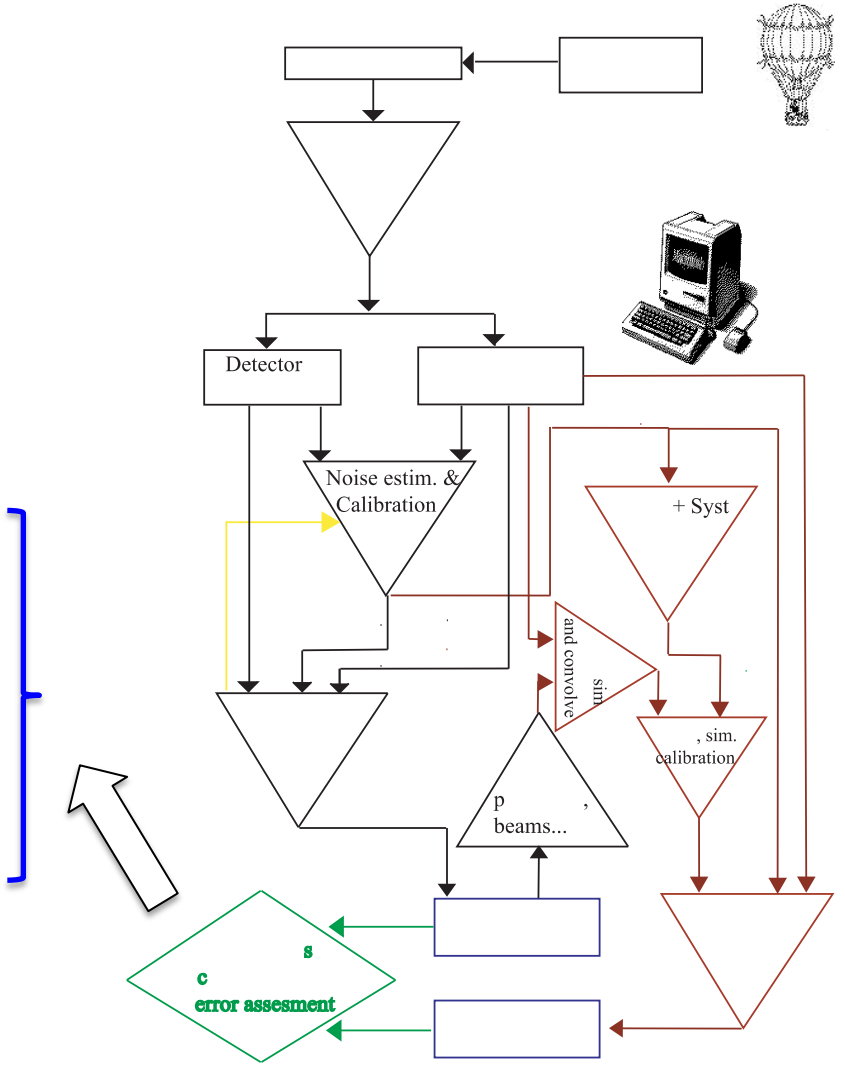
Exploitation issues in brief

1. Long list of science targets in cosmology and fundamental physics
2. Wide range of angular scales:
    - Large datasets: full sky maps (Mpix)
3. Signals ranking from faint to extremely faint
    - Large datasets: many detectors, long observations (Tb to Pb)
4. Large not huge. But analysis is *extremely* challenging:
    - Statistically optimal techniques needed: dense problem
    - Error budget dominated by systematics

CMB data analysis pipeline

- Complex pipeline
- Relies on simulated data



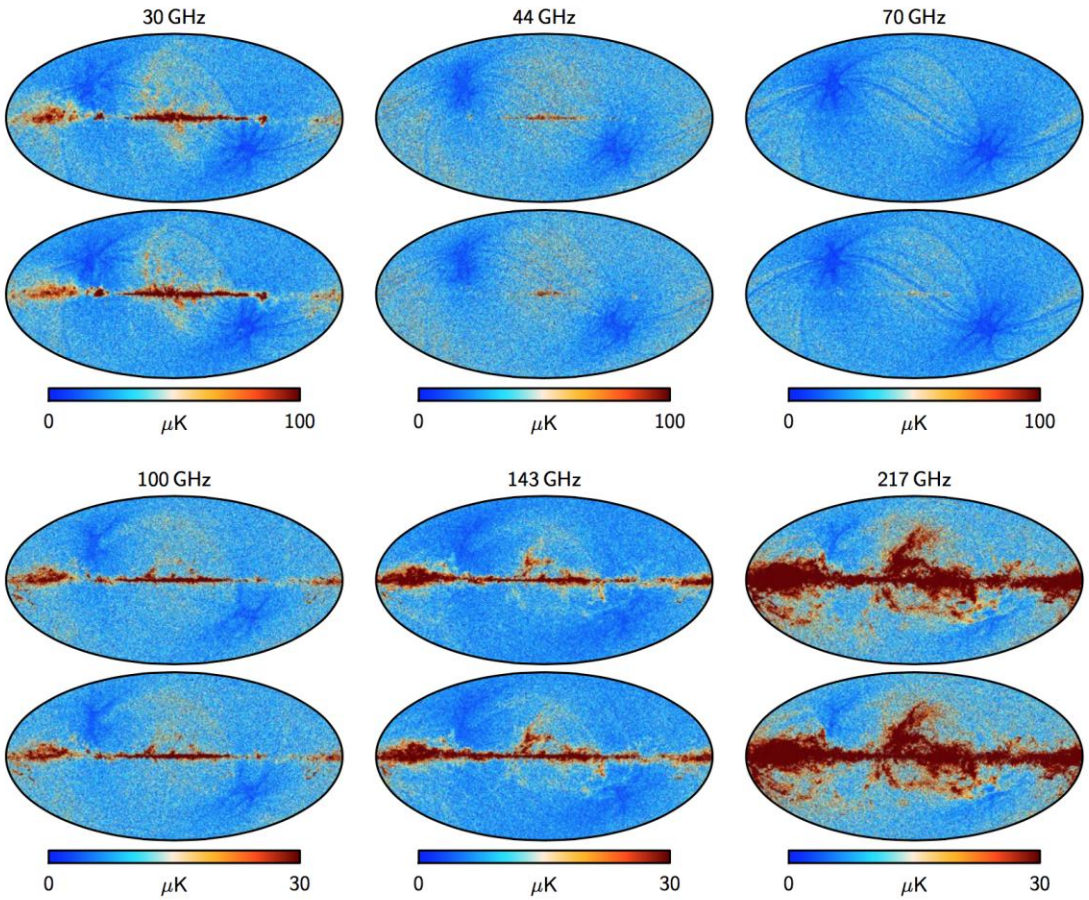


Approach vs computational cost

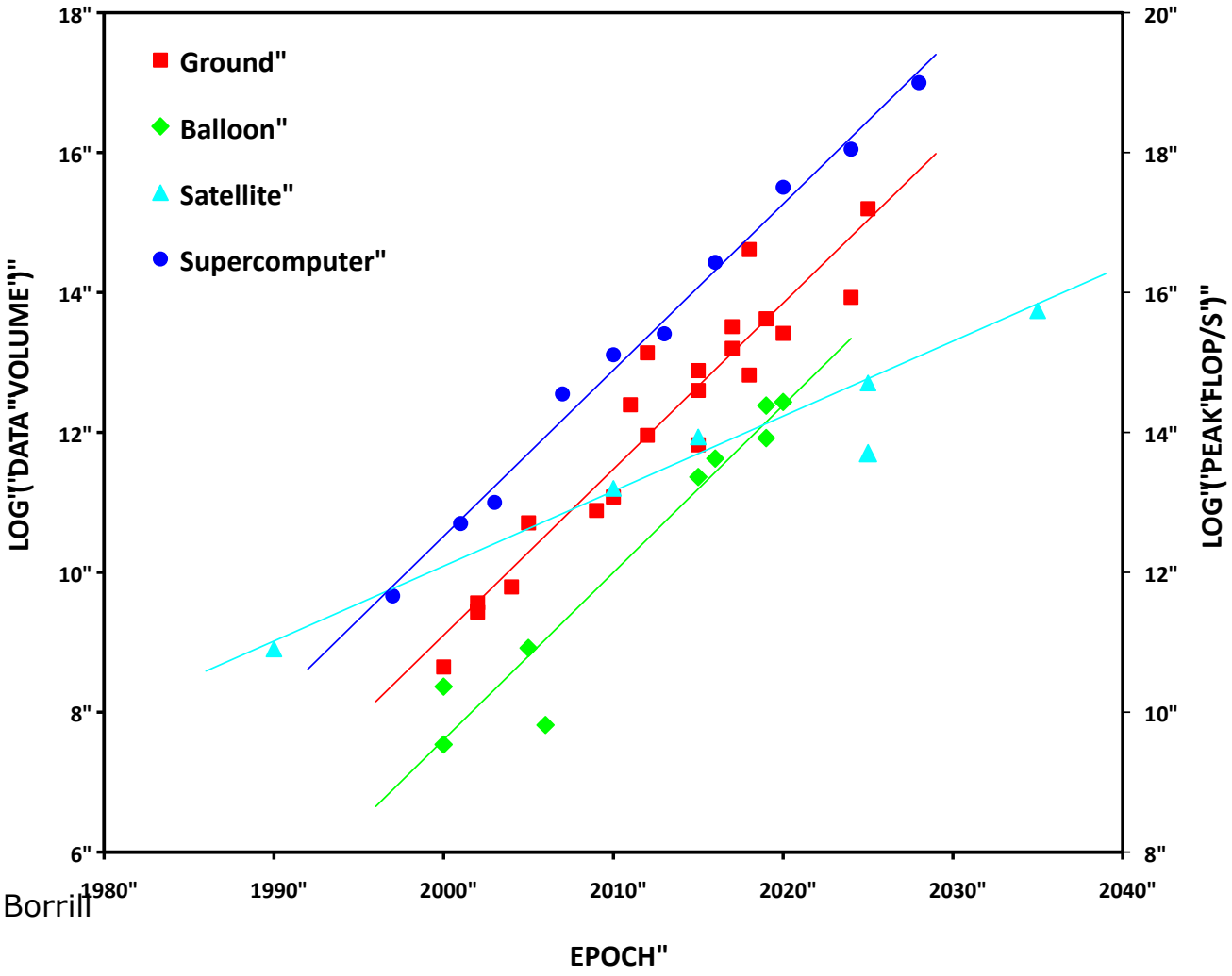
- "Exact" treatment totally unfeasible
  - Too costly (N_{pix}^3 or worse, for megapixel maps)
  - Error budget dominated by systematics, no analytic model cheap and easy!
- Have to rely on Monte Carlo methods
  - Computational cost dominated by simulation/map making level.
  - Scales as timeline length times number of simulations
  - Propagating systematics through MC is very costly and not always straightforward (c.f. Planck)
- Heavy dependence on supercomputers. Needs High Performance Computing:
  - Low latency, high bandwidth communication
  - Significant storage, fast I/O
  - No grid or share-at-home!

Planck Full Focal Plane simulations

1. End to end effort for all Planck Channels [arXiv:1509.06348]
2. Major computational burden was set of 10^4 Monte Carlo maps: 1 million CPU-days on world class super computer (NERSC and CSC)
3. Supported Planck cosmological analysis



Projected computional needs



Plot/data by J. Borrill

Projected computional needs

1. Computational cost driven up by sheer number of detectors packed in focal plane.
2. Moore's law provides some margin *but*
3. Untold part of the story: sheer size of data limits human direct intervention. Automatization is a must.

Expertise of the Italian Community

1. Very Good tradition/expertise in CMB data analysis.
    a. Heritage from earlier and current experiments (BOOMERanG, Maxima, Beast, Planck)

Expertise of the Italian community

What	Where	How well (vs. international top of class)	Remarks
Systematics sims	Milano, Roma I	Very good	
Signal, noise, sims	Distributed (Planck Level-S)	Good but…	Polarized foregrounds?
Map Making and calibration	Roma II, ASDC, Trieste	Good	Projected comput. cost?
Component separation	Ferrara/Bologna, Trieste	Good	Polarization still an issue (worldwide)
Harmonic analysis/Power Spectrum	ASDC Ferrara/Bologna, ASDC	Very good	
CMB likelihood and parameters	Ferrara/Bologna, ASDC, Roma I	Good but…	Small scale expertise
Non likelhood (primordial NG, birefringence etc)	Padova, Ferrara/Bologna, Roma II	Good	
Delensing	?	To be built	Theoretical expertise present

Expertise of the Italian Community

1. Very Good tradition/expertise in CMB data analysis.
    - Heritage from earlier and current experiments (BOOMERanG, Maxima, Beast, Planck)
2. Computational needs mostly outsourced, NERSC in California and CSC in Finland (to lesser extent)
    - Very good "internal" support to produce timeline and maps (e.g. Planck/DPC SGS)
    - Little support for Monte Carlo analysis (weakness to be overcome).
3. A new generation has to gain expertise on post Planck forefront issues (education)
    - Polarized foreground and related component separation
    - B mode de-lensing

Conclusions

- Science can be exploited provided one is able to analyze the data. Size and accuracy requirements for CMB targets make this challenging
- We need to exploit future HPC architectural developments to our needs. A national/shared infrastructure support is critical. NERSC/DOE may well not support us forever for free.
- Critical expertise has to be gathered through specific education of PhD students/young postdocs. This can realistically be gained only by working on future and ongoing experiments.