Simulations in LiteBIRD

E. Hivon (IAP)

Simulation pipeline

- Inputs:
  - Cosmological model (CMB)
  - Sky model
    - extragalactic (consistent with CMB)
    - Galactic
    - Solar system
  - Instrument model (IMo)
    - Scanning, focal plane, beams, noises, … + known unknowns, …
- Code parameters:
  - Complexity knobs
  - Random sequence
- Outputs:
  - Fake observations (data streams and/or maps) with controllable level of realism for arbitrary set of detectors

Why do we need simulations ?

- Compare alternatives and guide choices in instrumental design
- Test, compare, validate, improve algorithms and implementations of the analysis pipeline in real life conditions (mapmaking, C(l) computation, component separation, …)
- Propagate via Monte-Carlo up to final products (C(l) and cosmological parameters)
  - uncertainties on instrumental performance (gain, beam, …),
  - non idealities (mismatch beams, rHWP imperfections, complex foregrounds …), and *interactions* between systematics (eg, beams + foregrounds),
  - filtering, data alteration, … -> effective transfer functions or transfer matrices,
  - complex noise features, (eg non-white, non-uniform, lines, …), replace $N_{pix}xN_{pix}$ matrices
  - sample + noise variance -> C(l) covariance matrices
- Calibrate null-test results (residual and error bars) of consistency checks of (many) data splits, especially for blind analysis
- Validate possible numerical / ML-based / analytical short-cuts
- Staples of likelihood-free / simulations-based approaches

What it takes to run simulations (and analyze data) ?

- Planck HFI experience
  - Data volume: 52 det * 180 Hz * 3 yr = 0.8 x 10^{12} detector samples ~ 3TB of raw data,
    (6 IQU maps @ Nside=2048 ~ 3GB of final maps)
    Yet, **300TB** of ancillary + intermediate + simulated data required (***100** input data)

  - CPU time: (mostly at NERSC and CSC)
    - FFP8 simulations: 10^6 maps (300TB) ~ 25 M cpu.h (Planck 2015 XII)
    - FFP10 simulations: ??

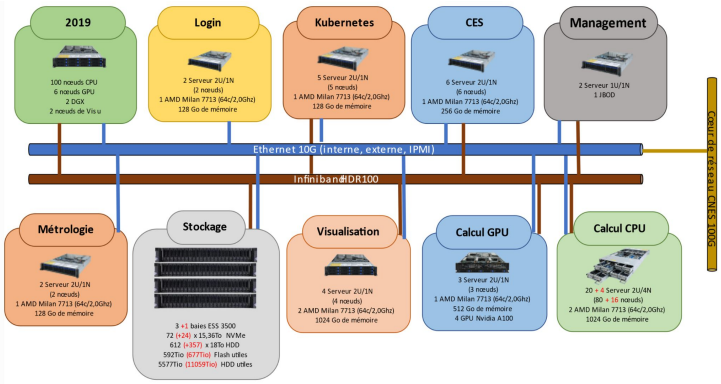
- LiteBIRD
  - Data volume: 4508 det * 19.1 Hz * 3 yr = 8 x 10^{12} detector samples ~ 30TB of raw data,
    (22 IQU maps @ Nside=512 ~ 1GB of final maps)
    If *100 holds: 3PB probably required

  - Will need to optimize time *and* energy
    - See (post)PTEP simulations

- Good integration of simulation and analysis pipelines
  - Eg, to bypass TOD writing and reading

Available computing facilities

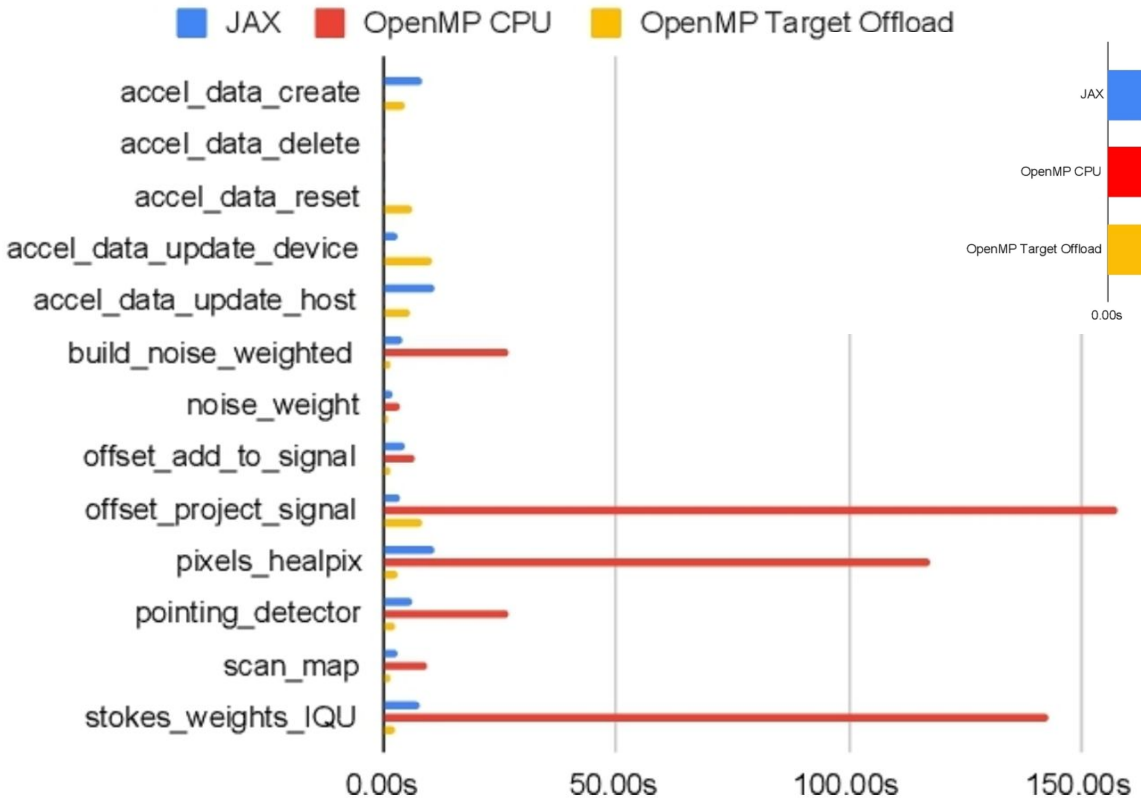
- NERSC supercomputers (US): no more free lunch in mp107, priority to CMB-S4
- Cineca supercomputers (IT): used for LBsims
- TREX @ CNES (France):
  - 16,000+ cores: 96 nodes * 128 cores (960 GB/node) + 100 nodes * 40 cores (180 GB/node)
  - 44 GPUs (Nvidia Tesla A100 and V100) (reserved to “level2” or higher users)
  - 14PB of disc space, 5 available
  - Used at <50%
  - Tempted (and already part of LiteBIRD) ?: contact L. Montier and/or M. Tristram http://litebird.in2p3.fr/pmwiki/pmwiki.php/DA/Hands-onCNESHAL
- Infinity @ IAP (France):
  - 4,700+ cores, 7 GPUs, overbooked, for prototyping
- CC @ IN2P3 (France)
- RURI @ JAXA (Japan) JSS3 -> JSS4
- CC @ KEK (Japan)
- + Oslo, Garching, …

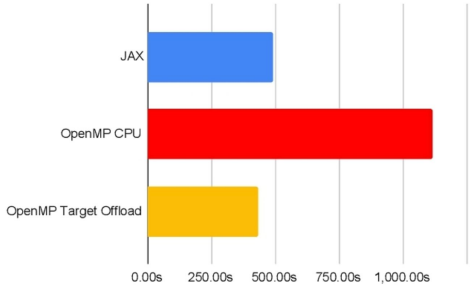


Available pipelines

- Stand alone pipelines
  - TOAST https://github.com/hpc4cmb/toast (public)
    - US based (LBL developers, among NERSC flagships)
    - now mostly focused on ground experiments (eg, CMB-S4)
    - Larger data volumes
    - Python front-end
      - OpenMP-C++ back-end so far,
      - now switching to full python GPU-optimized JAX (TOAST3)
  - LBsims https://github.com/litebird/litebird_sim (public)
    - Large italian involvement + external contributions
    - Designed from scratch for LiteBIRD, different data model with TOAST
    - On-going interfacing with TOAST map-maker
    - Python only
  - Both interfaced with LiteBIRD IMo https://github.com/litebird/IMo_LiteBIRD (private)
- Niche pipelines, modules, …
  - Falcon (Okayama U., Japan)
    - In Julia
  - Realistic HWP simulations
    - G. Patanchon

TOAST and JAX

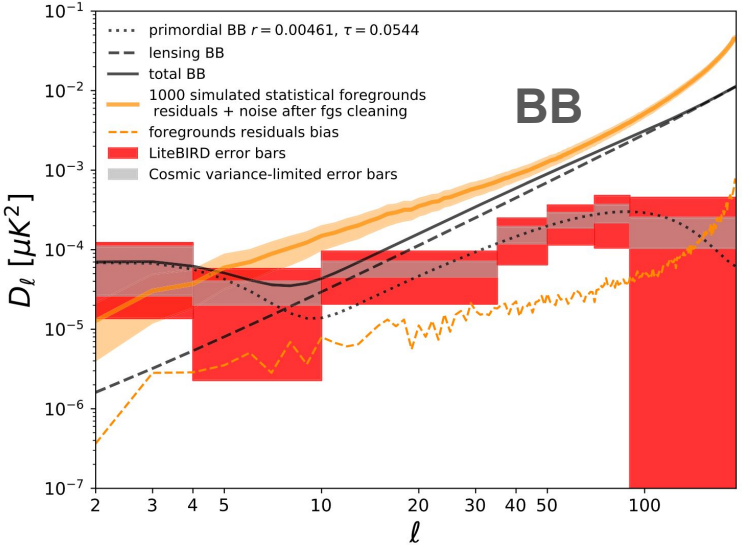




https://dl.acm.org/doi/fullHtml/10.1145/3624062.3624186

Existing/on going simulations

- PTEP (TOAST) PTEP 2023 042F01
  - CMB + PySM + radio sources
  - Gaussian beams, tophat bandpass
  - ~⅓ of detectors, 1yr of data, ideal HWP



- post-PTEP (LBsims @ CINECA) on going
  - Lensed CMB + PySM3 + orbital dipole
  - Gaussian beams, tophat bandpass
  - ~⅓ of detectors, 1yr of data, ideal HWP
  - Scaled 1/f noise or white noise
  - Nsims = 500 realisations, w/ different noise models
  - 1 TOD + Nsims * 22 IQU (`madam` produced) frequency maps,
  - ~1.5M CPUh, 50 TB (slowed down by I/O in mapmaking)

- LBxCMB-S4 joint simulations for study of synergies (LBsims) on going

What can be done today ?

Data simulation

- CMB (with *r* and τ)
- Foregrounds:
  - From PySM
- Systematics
  - Ephemerides,
  - Orbital dipole,
  - 4Pi Beams convolution (from GRASP simulations) with `ducc0`
  - 1/f Noise, gain drift, band pass,
  - rHWP, with imperfections (ongoing),
  - Cosmic rays,
  - See talks by
    - Sophie
    - Guillaume (summarised by Ludo)
  - …

Data reduction

- Map making:
  - Simple binning codes,
  - TOAST3 destriper,
  - `madam`,
  - `mappraiser`,
  - `SANEPIC` (Patanchon et al, https://arxiv.org/abs/0711.3462)
  - `SRoll3` (python interface to C++ code used in Planck-HFI)

Data analysis

- Component separation
  - See Ariana's talk
- Delensing
  - Joint simulations with CMB-S4
- C(l) calculation and parameter estimation
  - Birefringence angle β (see Josquin's talk)

Help wanted !

- Science Ground Segment (SGS, see Ludo's talk)
  - Definition
  - Involvement

- Many open questions on achieving LiteBIRD main and secondary science goals
  - CNES Phase B0 starting early 2025
  - Simulations telecons and MaPLes (Map making, Power spectra & Likelihood algorithms) telecons on alterning Thursdays at 9AM

- Now hiring: *Ingénieur d'étude en CDD pour simulations LiteBIRD* / Temporary engineering position for LiteBIRD simulations https://emploi.cnrs.fr/Offres/CDD/UMR7095-ERIHIV-001/Default.aspx


LiteBIRD

Extra slides

What it takes to run simulations (and analyze data) ?

- Data volume
  - Planck-HFI: 52 det * 180 Hz * 3 yr = 0.8 x 10^{12} detector samples ~ 3TB of raw data,
    (6 IQU maps @ Nside=2048 ~ 3GB of final maps)
    Yet, **300TB** of ancillary + intermediate + simulated data required (***100** input data)
  - SPT: 10,000 det * 152 Hz * 5 yr = 240 x 10^{12} detector samples ~ 1PB of raw data,
    1000 PB of disc space available
  - LiteBIRD: 4508 det * 19.1 Hz * 3 yr = 8 x 10^{12} detector samples ~ 30TB of raw data,
    (15 IQU maps @ Nside=512 ~ 0.7GB of final maps)
    If *100 holds: 3PB probably required
- CPU time
  - Planck-HFI: (mostly at NERSC and CSC)
    - FFP8 simulations: 10^6 maps (300TB) ~ 25 M cpu.h
    - FFP10 simulations: ??
  - LiteBIRD:
    - Will need to optimize time *and* energy
    - See (post)PTEP simulations
- Good integration of simulation and analysis pipelines
  - Eg, to bypass TOD writing and reading

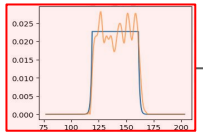
Toast 3

Simulation Workflow

Satellite scanning strategy

Pointing info

Solar Dipole



Per-detector sky alm (pysm3- low complexity)

Per-detector GRASP blm

Beam Convolution

Detector Time constant convolution

Calibration systematics (Gain errors, FP thermal fluctuations, slow drifts)

Noise (1/f+ white+ cosmic rays)

Cross-Talk

Scan Template for Calibration (T-only foregrounds, top-hat bpass)

Reduction Workflow

Toast3 destriper

Mitigate Cal. systematics

Mitigate Cross-talk

Deconvolve detector time constant

Map-Maker

Analysis Workflow

Comp-Sep (HILC)

Delensing

Delta-r from systematics residuals