Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Optimization of Data Reduction for Future CMB Experiments

Avinash Anand[1] and Giuseppe Puglisi[2]

[1]University of Rome "Tor Vergata", [2]University of Catania

Spoke 3 Technical Workshop, Trieste October 9 / 11, 2023

Scientific Rationale : Map-making for CMB experiments

- Data model for CMB signal $d_{i,t} = [A_i]_{tp} s_p + n_{i,t}$
  - p → pixel label; i → detector label; t → time stamp
  - A_i → Pointing matrix for detector i
  - s_p → signal amplitude at pixel p
  - $n_{i,t}$ → noise amplitude at time t for detector i

- Maximum likelihood solution taking into account multiple detectors observing at same frequency

$$\hat{s} = \left(\sum_{ij} A_i^t \left[N^{-1}\right]_{ij} A_j\right)^{-1} \left(\sum_{ij} A_i^t \left[N^{-1}\right]_{ij} d_j\right)$$

  - $N = \langle nn^t \rangle$ is the noise correlation matrix in time and N_{ij} is the block of N corresponding to the noise correlation between detectors i and j

- For LiteBIRD-like mission, inversion direct inversion of N^{-1} is impractical due to matrix size

Scientific Rationale : Map-making for CMB experiments (contd.)

- For stationary and end-to-end continuous data segment, N_{ij} is **circulant**. So,

$$\left[N^{-1}\right]_{ij,tt'} = \mathcal{F}\left\{\left[P^{-1}(\omega)\right]_{ij}\right\}(t-t')$$

  - $P(\omega)$ is the power spectra of the noise amplitude vector and $\mathcal{F}$ is the Fourier transform operator
- **Inversion of matrix N is now reduced to computation of just the first rows of block matrices N_{ij}**

- The map-making method with circulant matrix approximation has already been implemented in **SANEPIC**[1] (**S**ignal **A**nd **N**oise **E**stimation **P**rocedure **I**ncluding **C**orrelations)
- **Goal**: **To optimize SANEPIC for LiteBIRD[2] mission data and wrap the code with LiteBIRD simulation framework**

[1]Patanchon, G., et al. "SANEPIC: A mapmaking method for time stream data from large arrays." The Astrophysical Journal, vol. 681, no. 1, 2008, pp. 708–725, https://doi.org/10.1086/588543.

[2]LiteBIRD Collaboration. Probing cosmic inflation with the LiteBIRD cosmic microwave background polarization survey, Progress of Theoretical and Experimental Physics, Volume 2023, Issue 4, April 2023, 042F01, https://doi.org/10.1093/ptep/ptac150

Technical Objectives, Methodologies and Solutions

- Complete data size: ~250 TB
- Estimated execution walltime on Marconi for complete dataset: >200,000 CPU hours
- Performance analysis for reduced dataset using `perf` tool:

Event	Counts	Summary
Cycles	$2.32x10^8$	1.851 GHz
Instructions	$2.58x10^8$	1.11 insn per cycle
Cache-references	$2.81x10^6$	22.424 M/sec
Cache-misses	$1.47x10^6$	52.446 % of all cache refs
Branches	$5.81x10^7$	463.160 M/sec
Branch-misses	$6.98x10^5$	1.20% of all branches

- **Large cache-misses: Possibility of improving memory access pattern**
- **Low branch-misses: Possibility of compile-time optimization**

Technical Objectives, Methodologies and Solutions (contd.)

- Profiling using `gprof` gives further insight:

% Time Elapsed	Self Time Elapsed	No. of Calls	Function Name
41.60	5087.74	17160	`do_PtNd(double*, double*, …)`
19.03	2327.77	16874	`deproject(double*, long*, …)`
13.30	1626.77		`__libm_cos_l9`
12.82	1567.37		`__libm_sin_l9`
2.62	320.47	286	`compute_diagPtNPCorr(double*, long*, …)`
2.42	295.86	16874	`write_tfAS(double*, double (*) [2], …)`
1.85	226.10		`main`
1.11	135.29		`compare_global_array_long(void const*, void const*)`

- **Optimization required only in four routines**

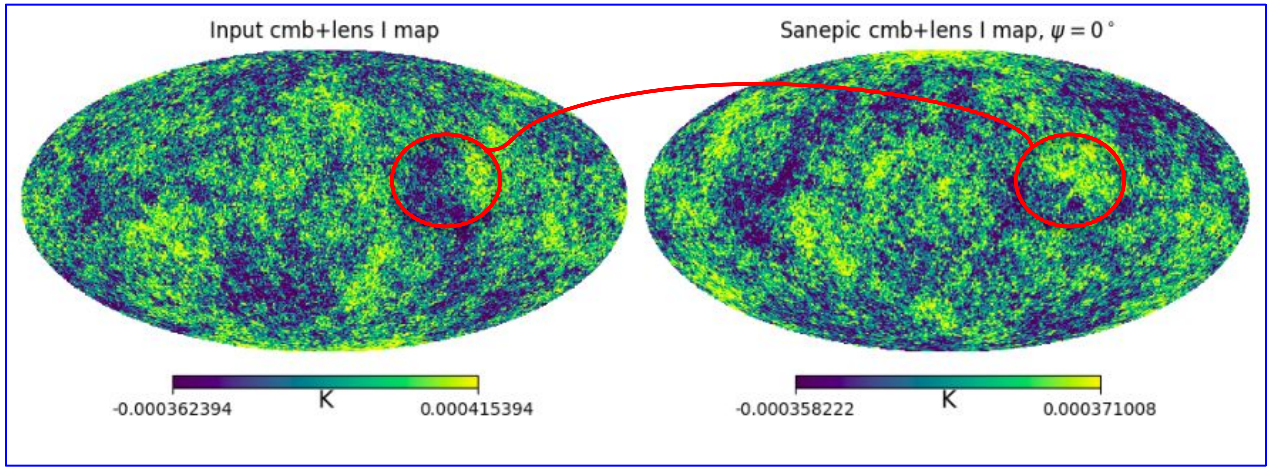
Technical Objectives, Methodologies and Solutions (contd.)

- Optimization target:
  - Code refactoring
  - Improving memory access pattern by redefining the data structures
  - To reduce the branching by utilizing compile-time optimization
  - To minimize process imbalance

- Further optimization:
  - Offloading Fourier transform and MatVec operations to GPU with suitable offloading technology

Timescale, Milestones and KPIs

- Started work on SANEPIC from July 2023
- Current status: Validation and debugging (~1 month)



- CPU optimization (~1 month)
- Validation, profiling and benchmark (~1 month)
- GPU optimization (~1 month)
- Validation, profiling and benchmark (~0.5 month)

Next Steps and Expected Results (by next checkpoint: April 2024)

- Optimization of SANEPIC for LiteBIRD
- Validation of the code with different noise components added to the sky signal
- Preparing a detailed documentation of the code
- Wrapping SANEPIC with LiteBIRD simulation framework
- Optimization of other modules of LiteBIRD simulation framework

Accomplished Work, Results

-

$$\hat{s} = \left(A^t N^{-1} A\right)^{-1} A^t N^{-1} d$$