

DIGITISATION

L. BALKENHOL¹ AND C. L. REICHARDT¹

Draft version August 2, 2018

ABSTRACT

Observations of the Cosmic Microwave Background (CMB) are of immense value to modern cosmology. However, future CMB experiments must confront challenges in mission planning, hardware and analysis that arise from the sheer size of the time-ordered-data being recorded. These challenges are particularly significant for Antarctic and satellite experiments which depend on satellite links to transmit the data. We investigate using extreme digitisation to address these challenges. Unlike lossless compression, extreme digitisation introduces additional noise into the data. We present an optimal 1, 2 and 3 bit digitisation schemes and lay out how the added noise varies between the schemes and as a function of hits per pixel in the map for temperature and polarisation observations. We show that the noise penalty of 3 bit digitisation is at the percent-level for CMB power spectrum measurements. This is impressive considering that it would reduce the data volume by an order of magnitude. We argue that extreme digitisation is a promising strategy for upcoming experiments.

Subject headings: cosmic background radiation — polarization

1. INTRODUCTION

Observations of the Cosmic Microwave Background (CMB) have played a key role since 1964 (discovery paper). Current and future CMB experiments will continue to deliver new insights by studying the temperature and polarisation information contained in the CMB. These will constrain help putting tighter constraints on cosmological models. The most prominent science goal is the discovery of the imprint left by inflationary gravitational waves. Additionally studies of CMB lensing and the SZ effects will deliver new insight. CMB experiments however also provide a valuable counterpart to ground-based particle physics experiments, as they probe the relativistic number of species, the helium fraction and the neutrino mass sum.

The outstanding contribution of CMB science to modern physics has demanded a high standard of data analysis. The CMB community has developed a variety of compression and computational techniques to manage the increasing influx of data, while maximising the science output. These include the compression of time-ordered data into maps, bandpower estimation and the pseudo Cl method.

A growing hurdle for experiments at remote locations are the transmission limitations of satellite links. Space-based experiments have employed a combination of lossless and lossy compression techniques, including reduced bits in the time-ordered-data (TOD). Antarctica based experiments that transmit a portion of their data via a satellite link have downsampled their TOD in the past to meet their telemetry requirements. They have yet to exploit few bit digitisation of the TOD.

As we approach the next generation of ground-based experiments, Stage-4, and the launch of a new generation of space-based missions (liteBIRD, PIXIE, CORE+), we must treat the transmission bottleneck carefully. Without a review of the current compression techniques em-

ployed we are sure to lose information.

In this work we present the method of extreme, i.e. few-bit digitisation, and detail its effect when applied to the TOD. To highlight the consequences we investigate temperature and polarisation powerspectra. We find that an optimal 3-bit digitisation scheme adds as little as $\sim 2\%$ to the map noise level. While the digitisation schemes described here are primarily laid out for ground-based Stage-4 experiments, future space-based missions must employ a similar digitisation as a step of their compression algorithm.

This work is structured as follows. We present the digitisation schemes in §??, by formulating them mathematically, motivating their use and describing the framework employed to test their performance. In §?? we present the results obtained. We summarize our findings in §??.

2. DIGITISATION

2.1. Problem

The science goals of upcoming CMB experiments naturally lead to a large influx of data. To achieve them longer observations with more detectors are needed. However, the best observations sites for CMB observations are at remote locations: in space and in the Antarctica. Experiments at these locations depend on satellite transmission. The next generation of ground-based experiments, Stage-4, will aim to collect 2 million detector years worth of data. A Stage-4 style experiment will face a data influx of $\sim O(10)Tb/d$. However, the current transmission allocation for the South Pole Telescope (SPT) is at $150Gb/d$. While this will likely see an increase going into the next stage of experiments it will not be able to over the entire observation. In fact we expect that compression rates for transmission must increase by an order of magnitude. Similarly space-based missions aim to increase their detector count by at least an order of magnitude compared to PLANCK. It is questionable whether their telemetry specifications will allow for transmission of the data with PLANCK-style compression.

christian.reichardt@unimelb.edu.au

¹ School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

Telemetry drawbacks can be sought to be combated by increasing storage capacity. However, storage space becomes a considerable price point when considering the amount of data to be recorded.

Beyond these transmission hurdles detailed mission planning is becoming exceedingly expensive. As noted by (S4 science book) a full simulation of TOD over the entire parameter space of detection scenarios for numerous set-ups is the desired way to plan Stage-4 configuration. Similarly space-based missions must carry out a similar analysis to optimise their science out-put. However, given the sheer size of TOD this is not possible. We must turn to different planning strategies or aim to reduce the size of the TOD in order to maximise the productivity of planning and development stages.

Operations on the TOD, such as noise-removal or map-making are a vital part of CMB data analysis. Through the exponentially growing size of TOD CMB data analysis is becoming increasingly expensive.

Extreme Digitisation would tackle the challenges mentioned above by reducing the size of the TOD by an order of magnitude. Together with already exploited lossless compression techniques (such as FLAC) this will directly tackle transmission hurdles. While it needs to be investigated to what extent existing algorithms can be carried over, extreme digitisation has the possibility of solving planning and analysis problems.

2.2. Extreme Digitisation

Digitisation is a lossy compression technique. However the induced noise depends on the number of bits used as well as the the digitisation thresholds and output levels chosen. It is possible to calculate the an optimal set of digitisation thresholds and output levels for any number of bits given the nature of the input signal. This was laid out by Max in 1978.

For 1 bit digitisation we apply the sign function to the TOD

$$\hat{x}(t) = \begin{cases} 1, & \text{for } x(t) > 0 \\ -1, & \text{for } x(t) \leq 0 \end{cases}$$

For 2 bit digitisation we apply the optimal digitisation scheme suggested by Max as

$$\hat{x}(t) = \begin{cases} 1.51\sigma, & \text{for } x(t) \geq 0.9816\sigma \\ 0.4528\sigma, & \text{for } 0 \leq x(t) < 0.9816\sigma \\ -0.4528\sigma, & \text{for } 0.9816\sigma \leq x(t) < 0 \\ -1.51\sigma, & \text{for } 0.9816\sigma < x(t) \end{cases}$$

Finally the optimal 3 bit digitisation scheme is

$$\hat{x}(t) = \begin{cases} 2.152\sigma, & \text{for } x(t) \geq 1.748\sigma \\ 1.344\sigma, & \text{for } 1.05\sigma \leq x(t) < 1.748\sigma \\ 0.756\sigma, & \text{for } 0.501\sigma \leq x(t) < 1.05\sigma \\ 0.245\sigma, & \text{for } 0 \leq x(t) < 0.501\sigma \\ -0.245\sigma, & \text{for } 0.501\sigma \leq x(t) < 0 \\ -0.756\sigma, & \text{for } 1.05\sigma \leq x(t) < 0.501\sigma \\ -1.344\sigma, & \text{for } 1.748\sigma \leq x(t) < 1.05\sigma \\ -2.152\sigma, & \text{for } 1.748\sigma < x(t) \end{cases}$$

2.3. Methods

Template maps:

To investigate the performance of the aforementioned digitisation schemes we simulate many scans over CMB

template maps at the timestream level. We begin by creating said template maps by creating a realisation of one I, Q, U map, for which we use the power spectrum provided by the Planck collaboration XXX. The underlying cosmological parameters are summarised in table ??.

Scan strategy:

These template maps are formulated in the Healpix framework. We then proceed to simulate constant elevation scans (CES) over a square patch of sky using a single detector. The scan speed is tweaked to match a desired number of hits-per-pixel. We allow ourselves to repeat the entire scan 100 times slightly offset in RA and DEC each time, so that we may hit all pixels roughly uniformly. We assume that the noise is purely gaussian white noise. The simulation parameters are summarised in table ??.

Noise addition and digitisation:

While performing each CES the appropriate pixels that are being targeted are determined. The corresponding values from the template maps are then accessed and added to realisations of the detector noise of appropriate length. The constructed timestream for each channels is then compressed in 4 different ways: averaging of the timestream, respecting each datapoint as a 64bit floating point number and applying the introduced digitisation schemes before averaging the timestream, again withing a 64bit floating point framework. Through this procedure we slowly build up 4 different maps: one map that corresponds to a normal scan over the CMB and 3 maps that reconstruct the observed path from extremely reduced time-ordered data.

Compression:

To obtain maps from the digitised time-stream we simply use our knowledge of the scan strategy to recognise what pixel each recorded point hits. All hits of a pixel are then simply averaged to obtain the pixel value.

TABLE 1
INPUT COSMOLOGICAL PARAMETERS

| Ω | XX | XX | r |
|----------|----|----|-----|
|----------|----|----|-----|

NOTE. — comments

TABLE 2
MODEL PARAMETERS

| NSIDE | $f_{readout}$ | f_{sky} | $\sigma_{detector}^T$ | $\sigma_{detector}^{Pol}$ |
|-------|---------------|--------------|-----------------------|--------------------------------|
| 4096 | 200Hz | ~ 0.014 | $500K\sqrt{s}$ | $\sqrt{2} \times 500K\sqrt{s}$ |

NOTE. — comments

3. RESULTS

Get power spectra

Analysis procedure

Analysis results: Nbits to map noise

Analysis results: Nhits per pixel to map noise

Analysis result: Noise disappears in cross-spectra $-i$ actually cannot check this in current framework.

Any other results that are outstanding

? Analysis result: To uncertainty in cosmological parameters?

This is why it works well: low S/N many hits per pixel, jumps in CMB noise level are order of magnitude.

TABLE 3
NOISE LEVELS

| Model | XX | XX | r |
|-------|----|----|-----|
|-------|----|----|-----|

NOTE. — go go

4. CONCLUSIONS

In this work we have demonstrated the power of digitising CMB TOD from 64-bit down to few-bits. The additional noise induced through this step was at the percent-level for chosen optimal digitisation schemes for temperature and polarisation observations.

The motivation behind employing extreme digitisation is the reduction of TOD by an order of magnitude. This primarily addresses challenges in data transmission that

will face upcoming CMB experiments at remote locations. Benefits in planning and analysis seem feasible.

Future work on investigating this compression technique must aim to understand the nature of the induced noise better. For this the performance of cluster-finding algorithms is interesting.

While the X added noise is already impressive it should be laid out how this changes when moving to a more realistic noise profile.

We thank the **referee as well as** Srinivasan Raghunathan and Federico Bianchini for valuable feedback on the manuscript. We acknowledge support from an Australian Research Council Future Fellowship (FT150100074), and also from the University of Melbourne. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

TABLE 4
ASSUMED SURVEY PARAMETERS

| Experiment | Sky coverage | Polarized Noise level (μ K-arcmin) | $1/f$ knee | Beam FWHM (arcmin.) |
|---------------|--------------|--|------------|------------------------|
| CMB Stage III | | | | |
| SPT-3G | 6% | 3.0 | 200 | 1.2 |
| Simons Array | 36% | 9.5 | 200 | 3.5 |
| CMB Stage IV | 55% | 1.3 | 100 | 4.0 |

NOTE. — Key numbers about the planned stage III and IV experiments. The sky coverage percentages are after galactic cuts. Unless otherwise noted, the Fisher matrix forecasts in this work use these numbers. All forecasts also allow for beam and calibration uncertainties as noted in the text.