CONSTRUCTING COSMIC MICROWAVE BACKGROUND POWER SPECTRA FROM FEW BIT TIMESTREAMS

L. Balkenhol[1] and C. L. Reichardt[1]
Draft version August 3, 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 Sunyaev-Zeldovich (SZ) effects will deliver new insight. CMB experiments however also provide a valuable counterpart to ground-based particle physics experiments, as they probe the relativisitc 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. With-

christian.reichardt@unimelb.edu.au
[1] School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

out a review of the current compression techniques employed we are sure to lose information.

In this work we present the method of extreme digitisation, which compresses a rich digital input signal and compresses it into a few bits. We apply extreme digitisation to the TOD and detail its effect on temperature and polarisation observations. We find that an optimal 3-bit digitisation scheme adds as little as $\sim 2\%$ to the map noise level. While the digtisation schemes described here are primarily laid out for ground-based Stage-4 experiments, future space-based missions must inevitably study lossy compression techniques.

This work is structured as follows. We detail the arising challenges in handling large TOD in §2.1. We subsequently formulate extreme digitisation formerly in §2.2 and lay out the framework used to test the performance of this compression technique in §2.3. In §3 we present the results obtained. We summarize our findings in §4.

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. Experimetns 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 SPT3G is at $150Gb/d$. SPT3G currently deals with its transmission bottleneck by downsampling the TOD, i.e. discarding high frequency information. While this will likely see an increase going into the next stage of experimetns it will not be able to over the entire observation. Going into Stage-4 we anticipate that compression rates for transmission must increase by an order of magnitude. Continued use

of heavy downsampling will narrow the informatino window decisively - prohibiting high l science to be carried out on the transmitted dataset. This also means that any potential faults or errors in the experiment that only become visible in high frequencies will go unnoticed for longer.

Similarly, future space-based missions (PIXIE, COrE+, liteBIRD) 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. Methods of storing large amounts of data on upcoming satellites will likely be prohibited by financial decisions: the amount of storage space required becomes financially relevant at the scales targeted.

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 expontentially growing size of TOD CMB data analysis is becoming increasingly expensive.

Extreme Digitisation would tackle the challenges mentioned above by recuding 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 possiblity 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, the digitisation thresholds and the output levels chosen. To minimise the noise induced through this process one must know the nature of input signal. A theoretical framework to obtain these levels was laid out by Max in 1978. We review the key aspects of his work relevant to our findings below.

We divide the space $-\infty \to \infty$ into N ranges. One then has endpoints x_k and output levels y_k, such that an input between x_i and x_{i+1} produces an output at y_i. Conventionally one chooses $x_1 = -\infty$ and $x_{N+1} = \infty$. In order to determine an optimal set of x_k and y_k we would like to minimise the distortion

$$D = \left\langle (s_{in} - s_{out})^2 \right\rangle$$

For a given input amplitude probability density $p(x)$ we may rewrite the above as

$$D = \sum_{i=1}^{N} \int_{x_i}^{x_{i+1}} (x - y_i)^2 p(x) dx$$

Seeing as we wish to minimise the distortion we differentiate the above with respect to x_i and y_i and set the derivatives to zero.

$$\frac{\partial D}{\partial x_i} = (x_i - y_{i-1})^2 p(x_i) - (x_i - y_i)^2 p(x_i) = 0$$

$$\frac{\partial D}{\partial x_j} = -2 \int_{x_i}^{x_{i+1}} (x - y_i) p(x) dx = 0$$

From the first equation above we deduce that

$$x_i = \frac{y_i + y_{i+1}}{2} \tag{1}$$

The second one

$$\int_{x_i}^{x_{i+1}} (x - y_i) p(x) dx = 0 \tag{2}$$

implies that we should choose y_i, such that it halves the area underneath $p(x)$ in the interval $x_i \to x_{i+1}$.

We are interested in the case where our input signal can fairly be approximated by gaussian white noise, i.e. $p(x) = 1/sqrt2\pi e^{-x^2/2}$. Even with these assumptions we cannot solve the resulting equations analytically. Instead one can begin by picking y_1 and calculating the remaining x_i and y_i using equation 1. One then observes weather this choice of values satisfy the conditions given by equation 2. If that is the case the x_k and y_k were chosen appropriately. This is unlikely to succeed on the first try. We rather solve the above system iteratively.

This was carried out by Max. We incorporate his results by formulating the multi-level functions we use for our 1, 2 and 3 bit digitisation process. For 1 bit digitisation we apply the sign function

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

to the TOD. For 2 bit digitisation we apply the four-level function

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

Finally the optimal 3 bit digitisation is described by the eight-level function

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

2.3. *Methods*

Summary:
To investigate the performance of the aforementioned digitisation schemes we simulate many scans over CMB template maps at the timestream level, where we add detector noise, for which we use gaussian white noise .

We obtain maps that use 64bit TOD and maps that have undergone 1, 2 and 3 bit digitisatino at the timestream level. We obtain the TT, EE, BB, TE powerspectra of each map. Through comparison we then determine the additional noise induced through the extreme digitisation process.

Template maps:

The I, Q, U template maps used were created through healpy's synfast. We rely on the power spectrum provieded by the Planck collaboration XXX. The underlying cosmological parameters are summarised in table 1.

Scan strategy:

We simulate observing a ~ 600 deg^2 path of the sky. To do so we perform constant elevation scans over the observation region. We repeat the observation strategy 100 times with a slight offset in RA and DEC each time, such that all pixels within the path are hit roughly uniformly. The simulation parameters are summarised in table 2.

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 timestreams for I, Q and U are then compressed in 4 different ways: averaging of the timestream, representing each datapoint as a 64bit floating point number and applying the three introduced digitisation schemes before averaging the timestream.

Compression:

We average all datapoints that lie within one pixel for each observation scheme and each channel. Through this procedure we slowly build up 12 different maps: three maps that corresponds to a normal scan over the polarised CMB and 9 maps that reconstruct the observed patch from an extremely compressed time-ordered data. Additionally we keep hold of the hitmap.

TABLE 1
Input Cosmological Parameters

Ω	XX	XX	r

NOTE. — Cosmological parameters used to create the template maps.

TABLE 2
Assumed Survey Parameters

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

NOTE. — Parameters used in the simulated sky strategy. The RA speed is tweaked to match the desired hits per pixel.

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 -¿ 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.

TABLE 4
Assumed survey parameters

Experiment	Sky coverage	Polarized Noise level (μK-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.