Draft version April 21, 2019
Preprint typeset using LaTeX style emulateapj v. 08/13/06

CMB-S4 PERFORMANCE-BASED CONSTRAINTS ON PRIMORDIAL GRAVITATIONAL WAVES

Author List: TBD

Draft version April 21, 2019

ABSTRACT

ToDo: Write abstract.

Subject headings: cosmic background radiation - cosmological parameters

1. INTRODUCTION

ToDo: Add broad CMB-S4/Inflation intro.

In this paper we present the CMB-S4 strawperson concept, and our case that it achieves the desired science requirements. The concept evolved from the process we developed for the CMB-S4 Science Book, to flow down science requirements to measurement and instrument requirements. Here we have extended that work to improve our understanding of the impact of astrophysical foregrounds, instrumental systematics, delensing non-idealities, and analysis methodology.

We rely on semi-analytic tools for the setting of our measurement requirements, as such tools generally offer sufficient speed to allow for optimization. To ensure realism, our general forecast/simulation approach has been as follows:

1. Develop a (semi-) analytic spectral forecast that makes use of noise performance that is informed by scaling from actual analyses of real experiments from time-streams to power spectra.

2. Use this forecasting tool to optimize the allocation of detector effort across frequencies, determining certain baseline "checkpoints" in survey definition space.

3. Validate these checkpoint configurations with standardized, version-numbered map-based data challenges. If independent analyses show recovery of science parameters from these challenge maps that does not match analytic forecasts (either in terms of variance or bias), we revise the forecasts accordingly.

4. Iterate between 1 and 3, injecting increasing realism in the form of (a) sky model complexity informed by the latest data and modeling efforts, and (b) systematics whose form, parameterization, and likely amplitude is likewise guided by real-world experience.

2. PERFORMANCE-BASED FORECASTS

2.1. *Fisher Framework*

For the CMB-S4 Science Book work we developed a Fisher forecasting machinery specifically targeted towards optimizing tensor-to-scalar parameter constraints in the presence of Galactic foregrounds and gravitational lensing of the CMB. This machinery was based on scaling the bandpower covariance matrices and noise power spectra of published BICEP/Keck analyses – as such, it automatically builds in all real world inefficiencies including (but not limited to): imperfect detector yield, non-uniform detector performance, read-out noise, observing inefficiency, losses due to timestream filtering, beam smoothing, and non-uniform sky coverage.

At its core this code is a Fisherized version of the BICEP/Keck likelihood analysis (BICEP2, Keck Array and Planck Collaborations (2015), BICEP2 and Keck Array Collaborations (2016), BICEP2 and Keck Array Collaborations (2018)) and our belief in the projections is grounded in that connection to achieved performance and published results. In particular, we emphasize the importance of using map-level signal and noise simulations as a starting point. These simulations are a good description of the maps because they pass jackknife tests derived from them.

Figure 1 is a schematic representation of the framework, identifying the user inputs, code modules, and outputs of said modules. This section describes this framework in detail.

2.1.1. *Fisher Formalism*

Given a likelihood function of the form:

$$L(\theta; d) \sim \frac{\exp\left[-\frac{1}{2}(d-\mu(\theta))^T \Sigma(\theta)^{-1}(d-\mu(\theta))\right]}{\sqrt{det(\Sigma(\theta))}} \quad (1)$$

where d are the data bandpowers, θ are the theory parameters, $\mu(\theta)$ are the expectation values given the parameters, and $\Sigma(\theta)$ is the bandpower covariance matrix, which can also be a function of the parameters. We can introduce the Fisher Information Matrix:

$$F_{ij} = -<\frac{\partial^2 log(L(\theta; d))}{\partial\theta_i \partial\theta_j}> \quad (2)$$

which is the ensemble average of the log-likelihood curvature, evaluated at the position of the best fit model. This matrix measures how steeply the likelihood falls as we move away from the best fit model, and F^{-1} can be thought of as the best possible covariance matrix for the measurement errors on the parameters θ_i. It can be shown that $\sqrt{(F^{-1})_{ii}}$ is the minimum obtainable standard deviation on the desired parameters Tegmark et al. (1996).

After some algebraic manipulations we can arrive at:

$$F_{ij} = \frac{\partial\mu^T}{\partial\theta_i}\Sigma^{-1}\frac{\partial\mu}{\partial\theta_j} + \frac{1}{2}Tr(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}) \quad (3)$$

And we calculate our parameter constraints as

$$\sigma_{ii} = \sqrt{(F^{-1})_{ii}} \quad (4)$$

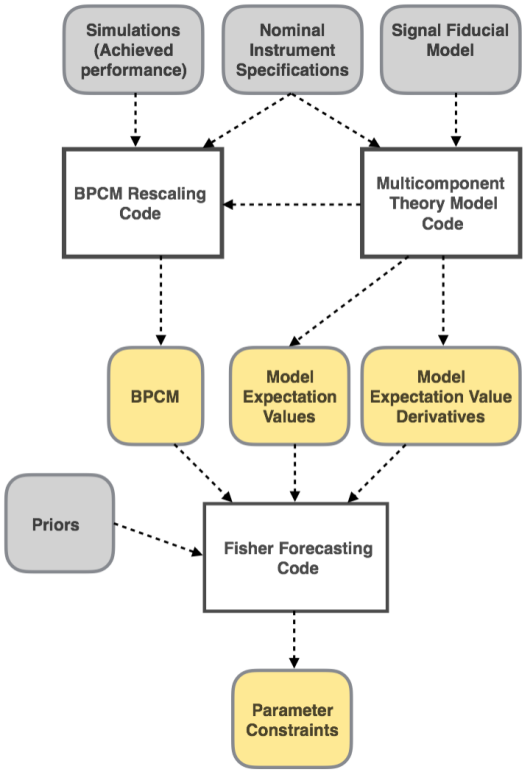

FIG. 1.— This is a schematic representation of our Fisher Machinery. Grey boxes represent *user inputs*, white boxes represent *code modules*, and yellow boxes represent *code outputs*.

Eq. 3 makes it clear that the construction of the covariance matrix Σ directly impacts the final constraints, and that a misestimation of it could lead to constraints that are far too optimistic. It is with this in mind that we have decided to base our calculations on scaled achieved performances.

2.1.2. *Inputs*

This section describes the inputs to the Fisher forecasting code.

Achieved Performance: the code takes information derived from signal and noise simulations of the BICEP/Keck dataset, but could be adapted to use similar information from another experiment. More specifically, we use signal-only, noise-only, and signal x noise bandpower covariance terms, as well as the ensemble-averaged signal and noise bandpowers. These inputs contain information about the actual map noise achieved from multiple receiver-years at {95, 150, 220} GHz, including the real penalties for detector yield, distribution of detector performance, weather, and observing efficiency. These inputs also fold in the incomplete mode coverage due to sky coverage, scan strategy, beam smoothing, and filtering in data analysis. For projections, we assume that we can scale down the noise based on increased detector count and integration time and that we can apply beamsize and NET rescaling to estimate performance at other frequencies. We operate under the assumption that detector noise is uncorrelated and we don't suffer heavily from common-mode atmospheric noise.

Instrument Specifications: number of frequencies and bandpass data specifying observing frequency response for each experiment included in the analysis, beam size, bandpower window functions specifying the response of each observed bandpower to each multipole on the sky, number of detectors per each frequency, and ideal per-detector noise-equivalent temperatures (NET's) per each frequency. The last two items can be used to perform an idealized calculation of the instrument sensitivity. These ideal performance numbers are only used for scaling purposes by comparing to the ideal sensitivity of BICEP/Keck and calculating the appropiate noise scaling factors. These factors are applied to the achieved sensitivities of BICEP/Keck to obtain performance-based sensitivities in our new bands.

Signal Fiducial Model: parameters used to calculate the multicomponent model. Our standard model has fourteen parameters, discussed in the section below.

Priors: if we have prior knowledge of a given parameter θ_i, we can easily introduce this information into the Fisher Matrix by simply adding $Pi = 1/\sigma_i^2$ to the diagonal of the Fisher Matrix, where σ_i is the width of the prior.

2.1.3. *Multicomponent Theory Model*

This code module calculates the parametrized model of one or more signal fields, including how the different signals scale to the various observing frequencies. Our model includes four signal components: a CMB component, parametrized by r and the lensing amplitude A_L, and components of dust, synchrotron, and correlated dust and synchrotron, which has a frequency scaling that depends on the relative strength of dust and synchrotron. We assume that the synchrotron scales as a simple power law in both frequency and l, while for dust we assume a power law scaling in l, and a greybody spectral energy distribution (SED) for the frequency dependence. In addition we also consider dust and synchrotron frequency decorrelation parameters which allow for dust/sync cross-spectral power supression. A detailed description of the full parametric model is presented in Appendix G of BICEP2 and Keck Array Collaborations (2018). While this model is easily extendable, other models could be substituted here.

The 14 model parameters are:
- r, tensor-to-scalar ratio, at pivot scale $k_t = 0.05$
- A_L, lensing amplitude
- A_{dust}, dust amplitude, in μK^2_{CMB}, at 353 GHz and $l = 80$
- β_{dust}, dust spectral index
- T_{dust}, dust greybody temperature
- α_{dust}, dust spatial spectral index
- Δ_{dust}, dust frequency decorrelation, at 217×353 GHz and $l = 80$
- EE/BB, ratio for dust
- A_{sync}, synchrotron amplitude, in μK^2_{CMB}, at 23 GHz and $l = 80$
- β_{sync}, synchrotron spectral index
- α_{sync}, synchrotron spatial spectral index
- Δ_{sync}, synchrotron frequency decorrelation, at 23×33 GHz and $l = 80$
- EE/BB, ratio for synchroton
- ϵ, synchrotron-dust spatial correlation

Since each signal component is independent, different signals have zero-cross power. Therefore, for a given spectra, the code just steps through the model components, combines the appropriate amplitude functions for the two experiments contributing to the spectrum, and applies the bandpower window functions to get the binned expectation values. Finally, it sums over model components to get the total expectation value for that spectrum. In addition, since a Fisher forecast requires knowledge about the response of the model expectation values with respect to the model parameters, the code also outputs the derivatives of the model expectation values.

The Fisher matrix we consider is usually 10-dimensional. The parameters we are constraining are: $\{r, A_{dust}, \beta_{dust}, \alpha_{dust}, \Delta_{dust}, A_{sync}, \beta_{sync}, \alpha_{sync}, \Delta_{sync}, \epsilon\}$. We fix $T_{dust} = 19.6K$ and A_L is varied separately to account for various levels of lensing power, as specified in Sec. 2.2.1. The EE/BB ratios are not relevant for the calculations in this manuscript because we are focusing on constraints from B-mode measurements only. The fiducial model for the Fisher forecasting is centered at r of 0 or 0.003, with $A_{dust} = 4.25\mu K^2_{CMB}$ (best-fit value from BK14) and $A_{sync} = 3.8\mu K^2_{CMB}$ (95% upper limit from BK14). The spatial and frequency spectral indices are centered at the preferred Planck values $\beta_{dust} = 1.59$ (with Gaussian prior of width 0.11), $\beta_{sync} = -3.10$ (with a Gaussian prior of width 0.30), $\alpha_{dust} = -0.42$, $\alpha_{sync} = -0.6$, and the dust/sync correlation is centered at $\epsilon = 0$. The central dust decorrelation value is taken to be $\Delta_{dust} = 0.97$ and the synchrotron decorrelation value is assumed to be $\Delta_{sync} = 1$. Unless otherwise stated, the parameters have flat unbounded priors.

2.1.4. *BPCM (Bandpower Covariance Matrix) Rescaling*

Signal Scaling: the output model expectation values can also be useful in the construction of our bandpower covariance matrix. To construct the bandpower covariance matrix components we use lensed-ΛCDM + dust + noise BICEP/Keck simulations (described in Appendix H of BICEP2 and Keck Array Collaborations (2018)). However, because we have the individual signal-only, noise-only, and signal x noise terms, we can record all the BPCM components:

sig = signal-only terms $Cov(S_i \times S_j, S_k \times S_l)$
noi = noise-only terms $Cov(N_i \times N_j, N_k \times N_l)$
sn1 = signal×noise terms $Cov(S_i \times N_j, S_k \times N_l)$
sn2 = signal×noise terms $Cov(S_i \times N_j, N_k \times S_l)$
sn3 = signal×noise terms $Cov(N_i \times Sj, S_k \times N_l)$
sn4 = signal×noise terms $Cov(N_i \times S_j, N_k \times S_l)$

and then rescale and combine them to create a bandpower covariance matrix for a new desired multicomponent model. Here, S are signal simulations, N are noise simulations, and the indices i, j, k, l run over fields in the analysis, i.e. all combinations of a map type (T, E, B) and an experiment ($S4_{20}$, $S4_{30}$, etc.). For many combinations of indices, we set certain covariance terms to be identically zero - $Cov(N_{S4_{20}} \times N_{S4_{20}}, N_{S4_{30}} \times N_{S4_{30}})$, for example. It is worth noting that having all of these terms allows us to have different numbers of degrees of freedom per bandpower for noise than for signal, a complication that is often ignored in other analyses by setting the noise and signal degrees of freedom to be identical.

While calculating the covariances from the signal and noise simulations, we record the average signal bandpowers from the simulations. For a new signal model, we can calculate the new bandpower expectation values, and rescale the signal components in the bandpower covariance matrix by the appropriate power of the ratio of the recorded average signal bandpowers, and the newly calculated expectation values.

The ability to get a BPCM for any model means only one set of simulations is necessary and one does not have to run simulations for any and all conceivable scenarios. In all the projections below we choose to fix $\Sigma(\theta) = \Sigma$ and hence we only apply the step above once, i.e. we do not rescale our BPCM at every step of the way.

Noise Scaling: In addition to scaling from one signal model to the other, recording all of the covariance terms allows us to rescale the noise parts as well. Given a dataset for which we have simulations, the noise scaling can go one of two ways: the first allows one to take a frequency present in the dataset and scale down the noise in the BPCM by the desired amount; the second one allows one to add an additional frequency by taking an existing one, scaling down the noise by the desired amount, and then expanding the BPCM and filling it in with the appropriate variance and covariance terms between this new band and all the existing ones. These two tools allow us to set up a new data structure to explore any combination of frequency bands, with any sensitivity in each band.

We would like to base our noise scaling factors on achieved sensitivities rather than ideal performances. To that end, we use the achieved survey weights from BICEP/Keck at $\{95, 150, 220\}$ GHz to obtain projected weights for any of the S4 channels. The average achieved per detector-year survey weights, calculated directly from final multi-year BICEP/Keck maps, and therefore include all the non-idealities described in the sections above, are $\{83.3, 58.6, 3.9\}\mu K^{-2}$ (BICEP2 and Keck Array Collaborations (2018)).

Once have these numbers, we can write the achieved per-detector-year survey weight as follows:

$$w^{per-det-yr}_{BK,achieved} = \frac{t_{obs}}{\alpha_{ideal/BK,achieved} NET^2_{BK,ideal}} \quad (5)$$

where $\frac{t_{obs}}{\alpha_{ideal/BK,achieved}}$ is the reality factor that encapsulates the less-than-ideal observing time, receiver performance, cuts, etc. It is the factor that takes us from the ideal scenario to reality. We never have to actively calculate this factor because we operate with ratios of survey weights. To obtain the achieved survey weight at a new frequency, one rescales the achieved survey weight as follows:

$$w_{S4,achieved} = w^{per-det-yr}_{BK,achieved} \mathcal{N}^{det-yr}_{S4} \frac{NET^2_{BK,ideal}}{NET^2_{S4,ideal}} \quad (6)$$

Where $\mathcal{N}^{det-yr}_{S4}$ is the number of detector-years assumed for CMB-S4 at any particular frequency. The implicit

assumption that is made here is that the reality factor for this new frequency is exactly the same as the one from which we are scaling. To not abuse this assumption, the survey weight scaling is always done from the closest frequency that we have an achieved survey weight for, as that is the performance that should guide us.

Therefore, to obtain N_l's for one channel by scaling the achieved N_l's of another channel we have to first scale them by the ratio of the survey weight and then scale by the ratio of the beams squared. Achieved performance inputs fold in the actual BICEP/Keck B_l's, but we rescale based on Gaussian approximations of the beams, which closely follow the BICEP/Keck beams. Under the assumption of Gaussian beams, we can write the full relation as:

$$N_{l,S4} = N_{l,BK} \frac{w_{BK,achieved}}{w_{S4,achieved}} \frac{B^2_{l,S4}}{B^2_{l,BK}} \quad (7)$$

where $B^2_{l,\nu} = \exp \frac{-l(l+1)\theta^2_\nu}{8\log(2)}$, and θ_ν is the FWHM (in radians) of the Gaussian beam. With the N_l's scalings on hand, we can perform the above mentioned BPCM operations to arrive at a scaled CMB-S4 BPCM that encompasses all the intricacies of reality that BICEP/Keck does.

2.2. *Experiment Specification*

We start our CMB-S4 optimization by assuming eight low-resolution channels at {30, 40, 85, 95, 145, 155, 215, 270} GHz and one high-resolution channel at 20 GHz, in four atmospheric windows, see Figure 2. The procedure used to come up with the split in each window was to separate the overlapping bands as far as possible while still keeping the calculated per-detector NET within 10-15% of the NET for a detector that spans the full window.

The ideal NET's per detector were calculated with `NETlib.py` [1] and are assumed to be {214, 177, 224, 270, 238, 309, 331, 747, 1281} $\mu K_{CMB}\sqrt{s}$. These NET's are calculated for a 100mK thermal bath, as opposed to 250mK for the Science Book, and are therefore lower. We want to emphasize that these NET numbers are only used to determine the scalings between different channels, and not to calculate sensitivities.

For the low-resolution instruments we pick $0.52m$ apertures. For the 20 GHz channel, a $0.52m$ aperture would result in a very broad beam which would dominate the noise at the relevant scales; to circumvent this we place this low frequency channel on a large aperture instrument, and while the scaling of the noise is still done from achieved performance, we choose a more appropriate $l_{knee} = 200$ and $\theta_\nu = 11'$ FWHM, keeping the slope γ the same as for the small-aperture noise (see Eq.7). In addition to the low-resolution effort, we assume a separate high-resolution instrument for delensing purposes, described in Section 3.4. We also fold in information from two WMAP channels: {23, 33} GHz and seven Planck channels: {30, 44, 70,100, 143, 217, 353} GHz, though this extra information is only relevant in the early stages of the optimization.

The optimization unit of effort is equivalent to 500 det-yrs at 150 GHz. For other channels, the number of detectors is calculated as $n_{det,150} \times \left(\frac{\nu}{150}\right)^2$, i.e. assuming comparable focal plane area. The projections run out to a total of 3×10^6 det-yrs, which, if all at 150 GHz, would be equivalent to 500,000 detectors operating for 6 yrs – this seems like a comfortable upper bound for what might be conceivable for S4. S4 scale surveys seem likely to be in the range of 10^6 to 3.0×10^6 det-yrs.

The multipole range is assumed to be $l = [30, 330]$, and using BICEP/Keck bandpower window functions and binning, we obtain 9 bins with nominal centers at l of {37.5, 72.5, 107.5, 142.5, 177.5, 212.5, 247.5, 282.5, 317.5}.

2.2.1. *Delensing Treatment*

As mentioned in Section 2.2, we assume a separate, high-resolution instrument dedicated to measuring the medium- and small-scale information necessary to construct a template of lensing B modes, so that their effect can be removed. In the semi-analytic optimization process, this instrument is assumed to have 1-arcminute resolution and detector weight at a single frequency. The translation between detector effort and map noise in the delensing instrument is based on the method used for the low-resolution instrument (as described in Sec. 2.1.4 and Eq. 8), but with certain non-idealities specific to low-resolution instruments and low-ℓ analysis (such as mode removal and non-uniform coverage) removed. Following the formalism in Smith et al. (2012), we convert the map noise in the delensing survey to a delensing efficiency, or equivalently a fractional residual in lensed B-mode power, specified by setting A_L to the corresponding residual.

The detector effort dedicated to the delensing instrument comes out of the total detector effort budget for the small-area r survey, and the distribution of effort between the low-resolution and delensing instruments is part of the optimization process.

2.3. *Optimization and Parameter Constraints*

In this section we try to answer the following question: given a fixed amount of effort, and the instrument specifications offered in the previous sections, what is the optimal effort distribution for foreground cleaning and delensing such that a minimal constraint on r is achieved. To do so we set-up an optimization that calculates the steepest descent through the 10 dimensional space (9 foreground channels + 1 channel for delensing). At each step of the algorithm, the code tries to allocate a fixed amount of effort in each dimension iteratively, rescaling the BPCM accordingly, computing a new Fisher matrix, inverting it to marginalize over all the parameters, and computing the resulting σ_r. Finding the dimension that offers the minimal constraint, the code finally permanently assigns the step of effort to the particular dimension. Though it is generally prohibitive to calculate the full 10D hypercube of σ_r's, we have validated our approach with the detailed calculation at various points in the optimization. An example of the final path is presented in Figure 3.

An additional dimension that is of different nature is the selection of sky fraction. Since we are using BICEP/Keck products, which are defined for a particular mask with $\sim f^{BK}_{sky} = 1\%$, we must scale them appropriately for all different fractions. We propagate the effects

[1] cmb-s4.org/wiki/index.php/New_NET_Calculator_and_Validation

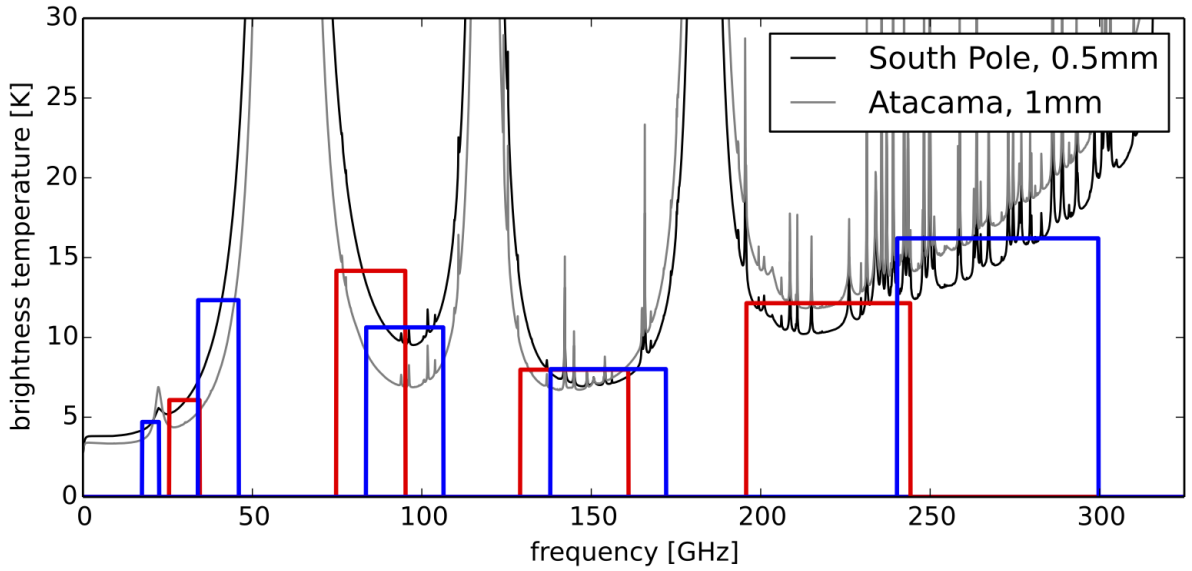

FIG. 2.— This figure shows calculated atmospheric brightness spectra (at zenith) for South Pole at 0.5 mm PWV and Atacama at 1.0 mm PWV (both are near median values). Atmospheric spectra are generated using Paine (2017). The tophat bands are plotted on top of these spectra, with the height of each rectangle equal to the band-averaged brightness temperature using the South Pole spectrum.

of sky fraction in the noise spectra and BPCM in two ways: first, we inflate the N_l's by a factor $\beta = f_{sky}^{S4}/f_{sky}^{BK}$ which boosts the (signal × noise) and (noise × noise) terms of the covariance matrix by β and β^2, to take into account the redistribution of the achieved sensitivity onto a larger patch. Note that the (signal × signal) component remains unchanged in this step. Second, we scale down the entire covariance matrix by a factor of β to increase the number of degrees on freedom in the BPCM, accounting for the fact that we are now observing more modes. This procedure scales the signal and noise degrees of freedom independently, preserving the relative effects that filtering, non-uniform coverage, etc have on the covariance structure. In the examples below we pick $f_{sky}^{S4} = 3\%$ as we have done for the Science Book.

Upon obtaining the optimized detector count distribution we can calculate the final S4 noise spectra according to Eq.7. In order to be able to use these spectra for the design of noise simulations (see 3.1) it is useful to distill them to a few numbers, to that end we fit them to the following formula:

$$N_{l,fit} = \frac{l(l+1)}{2\pi}\frac{\Omega_{pix}}{B_l^2}\left(1+\left(\frac{l}{l_{knee}}\right)^{\gamma}\right)\sigma_{map}^2 \qquad (8)$$

and obtain the map depth σ_{map}, slope γ, and l_{knee}. For the small-aperture data, we find $\ell_{\rm knee} = 50$–60 with γ of -2 to -3, depending on the frequency. The optimized map depths are presented in Figure 3.

Figure 4 shows the optimized constrains on r as a function of total effort as well as the fraction of effort spent on removing the lensing sample variance and the resulting map rms lensing residual. To reach the desired science goal of $\sigma(r) = 5\times10^{-4}$ we note that 1.2×10^6 150 GHz equivalent detector-years are necessary. The optimal distribution of this effort is summarized in Table 1.

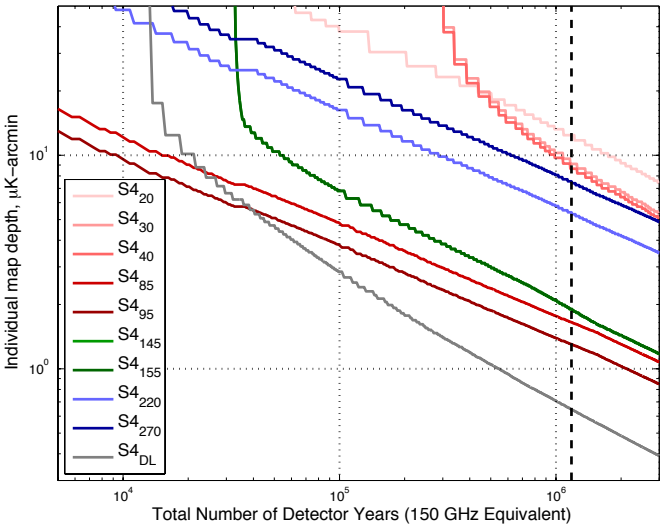

FIG. 3.— Optimized map-depth in each of the small-aperture channels as well as in the delensing channel, for an $f_{sky} = 3\%$.

3. SIMULATIONS

Using simulations to optimize the design of a CMB experiment inevitably involves a trade-off between the degree of detail that the simulations are able to capture and the computational (and human) cost of generating and analyzing them. This trade-off includes the choice of domain in which the simulation is generated, ranging from the most detailed but most expensive time domain through the map domain to the most simplified but most flexible spectral domain. Inclusion of additional detail can help to validate general results, to explore their sensitivity to assumptions about foreground models, sky coverage, and instrumental noise and systematics, and in more mature stages of design can inform specific instrument and survey strategy choices.

Here we review the methods used to explore parameter space for the small-area survey, including map level

TABLE 1
INSTRUMENT CONFIGURATION SATISFYING THE MEASUREMENT REQUIREMENTS.

Science	Item	Frequency [GHz] 20	30	40	85	95	145	155	220	270	Total
r	**14 x 0.5-m cameras**										
	# detectors	...	260	470	17 k	21 k	18 k	21 k	34 k	54 k	168 k
	Angular resolution[FWHM]		77′	58′	27′	24′	16′	15′	11′	8.5′	
	1 x 6-m telescope										
	# detectors	130	250	500	...	25 k	25 k	...	8.7 k	8.7 k	68 k
	Angular resolution [FWHM]	11′	7.0′	5.2′	...	2.2′	1.4′	...	1.0′	0.8′	

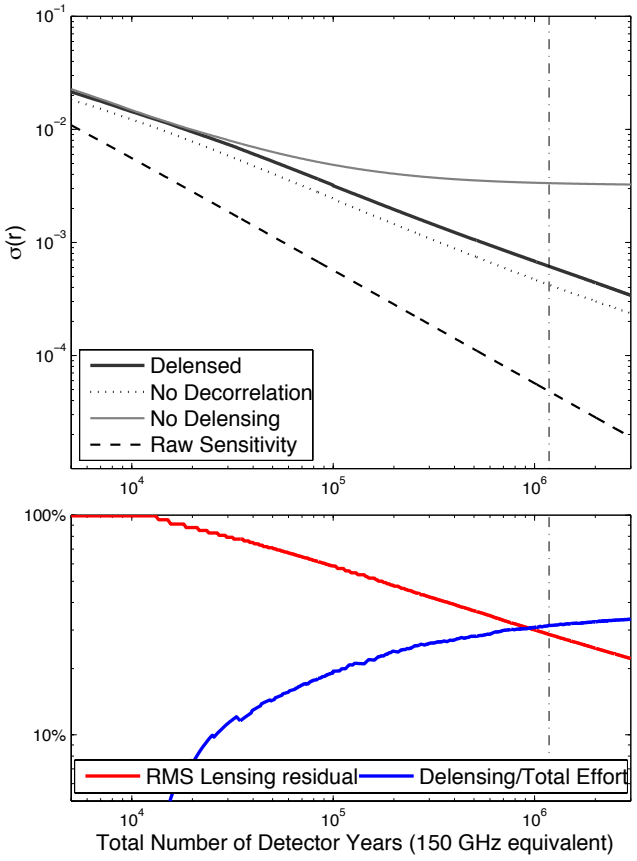

FIG. 4.— Optimized constrains on r as a function of total effort as well as the fraction of effort spent on removing the lensing sample variance and the resulting rms lensing residual.

noise simulations, sky models, and observation strategy. We also describe our approach to modeling instrumental systematics, the delensing survey, and the analysis methods.

We also use these simulations to validate the spectral domain forecasts for configurations where the approaches are directly comparable.

3.1. *Map Noise Simulations*

As in the Science Book, we have mostly used a sky area of 3%, with some simulations also done for 1% and 10%. We use 8 to obtain the desired noise presctiption and then generate Gaussian noise realizations at each band. Small-aperture cameras have a very wide instantaneous field of view and hence the observed sky region necessarily has a large edge taper. For the nominal 3% sky coverage simulations, we assumed a circular sky patch with full coverage at $r < 12\,\mathrm{deg}$ and "relative hits" tapering to zero with a cosine-squared shape for $12\,\mathrm{deg} < r < 27\,\mathrm{deg}$. The noise realizations are divided by the square-root of this coverage pattern such that the noise "blows up around the edge" as it does in real maps.

Future timestream simulations will include an explicit scan strategy on the sky and will produce more realistic sky coverage patterns, but for the moment we regard the above as a reasonable compromise between idealism and realty.

3.2. *Foreground Models*

To make simulated sky maps we add realizations of lensed CMB both without and with an r component to models of the Galactic foregrounds. So far we have run simulations with seven foreground models:

0. Simple Gaussian realizations of synchrotron and dust with power-law angular power spectra at amplitudes set to match the observations in the BICEP/Keck field, and simple uniform SEDs (power law for synchrotron, greybody for dust).

1. The PySM[2] model `a1d1f1s1`, where the letters refer to anomalous microwave emission, dust, free-free and synchrotron respectively, and the numbers are the base models described in Thorne et al. (2016).

2. The PySM model `a2d4f1s3`, where the models have been updated to variants that are also described in Thorne et al. (2016). Note that these include 2% polarized AME, a curvature of the synchrotron SED, and a two-temperature model for dust.

3. The PySM model `a2d7f1s3`, where the dust model has been updated to a sophisticated physical model of dust grains as described in Hensley (2015). This model is interesting in that it does not necessarily conform to the greybody SED.

4. The dust model in 3 is replaced by a model of polarized dust emission that incorporates HI column density maps as tracers of the dust intensity structures, and a phenomenological description of the Galactic magnetic field as described in Ghosh et al. (2017). The model is expanded beyond that described in the paper to produce a modest amount

[2]https://github.com/bthorne93/PySM_public

of decorrelation of the dust emission pattern as a function of frequency motivated by the analysis of Planck data in Planck Collaboration (2017).

5. A toy model where the dust decorrelation suggested in Fig. 3 of Planck Collaboration (2017) is taken at face value ($\Delta_{80}^{217\times353} = 0.85$, at $\ell = 80$) and scaled to other frequencies using the functional form given in appendix B of Vansyngel et al. (2017), with a linear scaling in ℓ. While such a model is not ruled out by current data it appears to be very hard to produce such strong decorrelation in physics-based models. We also note that Sheehy & Slosar (2017) have re-analyzed the same Planck data and, while they find that the high level of decorrelation in this model is consistent with the data, their best fit to that same data has no decorrelation.

6. A model based on MHD simulations (Kritsuk et al. 2017) of the Galactic magnetic field, which naturally produces non-Gaussian correlated dust and synchrotron emission.

Models 1 to 4 use the actual large-scale modes of the real sky as measured above the noise in the Planck data. This means that these models are intrinsically "single-realization," and this must be borne in mind when interpreting the results. The PySM models fill in the small-scale structure with power-law Gaussian extrapolations, while models 4 and 6 naturally produce non-Gaussian small-scale structure. However, all of these models are consistent with current data, and we should be careful not to necessarily associate nominal sophistication with greater probability to more closely reflect reality.

3.3. *Instrumental Systematics*

Control of instrumental systematics is a critical design consideration. However, predicting and modeling these effects realistically is a difficult task that is dependent on actual instrument and survey design details, and in any case their impact on an actual result comes not through the modeled effects but through unmodeled residuals. For this study we have taken the first steps in simulating various generic classes of additive systematic by injecting additional noise-like components into the maps, and then re-analyzing them without knowledge of what was put in. We have experimented with components that are both correlated and uncorrelated across frequency bands, and which have white, $1/\ell$, and white + $1/\ell$ spectra, at varying levels compared to single-frequency map noise or, for correlated cases, combined map noise. Examples of mechanisms that might produce map residuals within this class, after modeling them and either correcting or filtering their leading-order effects, include bandpass mismatches, beam and pointing variations, calibration variations, cross-talk effects, half-wave-plate leakages, ground pickup, and readout irregularities.

Other classes of systematics can be simulated by manipulating the reanalysis procedure only. Examples of such effects include bandpass, polarization angle, calibration, and beam shape uncertainties.

For assessing the impact of instrumental systematics on measurement requirements, in determining both the required survey depths and the maximum allowable levels of systematic effects in the final single-frequency survey maps, our general procedure is to feed parameterizations of various systematic effects into semi-analytic forecasts to judge at what levels classes of systematics introduce parameter biases or additional uncertainties that are significant compared to science targets for those parameters.

3.4. *Delensing*

We have generated high-resolution simulated maps on which we can run explicit lensing reconstruction and then include that information in the analysis. However, that process is not yet converged, and so for the present we approximate delensing by scaling down the ΛCDM lensing signal by the appropriate factor, as described in section 2.2.1.

4. ANALYSIS METHODS

To make simulated maps the noise realizations described in Section 3.1 are added to the sky models described in Section 3.2. For each realization one then has a stack of multi-frequency $I/Q/U$ maps containing non-uniform noise, foregrounds and signal, and the challenge is to re-analyze them to recover the parameter of interest (in this case r). This can be done by different teams using different methods, and could be done in a blind manner, although we have not done this yet.

So far we have experimented with two methods. The first is a map-based ILC cleaning method (e.g., Eriksen et al. 2004), which seeks the linear combination of maps that minimizes the remaining CMB signal, followed by a marginalization over residual foregrounds. This method has the advantage that it does not need to know the bandpasses of the frequency channels.

The second method is an evolution of the parametric multi-component fit to the ensemble of auto- and cross-spectra as used for the BICEP/Keck analysis to date (BICEP2, Keck Array and Planck Collaborations (2015), BICEP2 and Keck Array Collaborations (2016), BICEP2 and Keck Array Collaborations (2018)). This method fits the observed bandpowers to a model composed of the lensing expectation plus dust and synchrotron contributions and a possible r component. Dust and synchrotron each have an amplitude (A_d and A_s), a spatial spectral parameter (α_d and α_s), and a frequency spectral parameter (β_d and β_s). We also allow dust/synchrotron correlation (ϵ), and decorrelation of the foreground patterns over frequency (Δ_d and Δ_s).

Both of these analysis methods are only close to optimal when the foreground behavior is close to uniform across the observing field. For analysis of larger fields, algorithms that fit, for example, the frequency spectral indices individually in (large) pixels, will be required.

5. RESULTS

Table 2 summarizes the results of the analysis for simulations of the optimized configuration obtained in 2.3 (1.2×10^6 150-GHz-equivalent detector-years) and residual lensing power $A_L = 0.1$. The results from the parametric analysis naturally depend on whether a marginalization over decorrelation is performed, while the ILC does not attempt to capture the effects of decorrelation on the recovery of r and $\sigma(r)$. This is evidenced by the

TABLE 2
RESULTS OF TWO ANALYSIS METHODS APPLIED TO MAP-BASED SIMULATIONS ASSUMING THE STRAWPERSON CONFIGURATION AND OUR SUITE OF SKY MODELS. ALL SIMULATIONS ASSUME AN INSTRUMENT CONFIGURATION INCLUDING A (HIGH-RESOLUTION) 20 GHZ CHANNEL, A SURVEY OF 3% OF THE SKY WITH $\sim 1.2 \times 10^6$ 150-GHZ-EQUIVALENT DETECTOR-YEARS, AND $A_L = 0.1$.

		ILC		Parametric (no decorrelation)		Parametric (incl. decorrelation)	
r value	Sky model	$\sigma(r) \times 10^{-4}$	r bias $\times 10^{-4}$	$\sigma(r) \times 10^{-4}$	r bias $\times 10^{-4}$	$\sigma(r) \times 10^{-4}$	r bias $\times 10^{-4}$
0	0	5.4	0.3	4.4	0.2	5.6	0.3
	1	8.6	1.1	4.7	6.9	6.4	5.3
	2	8.0	−2.3	4.8	4.0	6.6	2.0
	3	6.8	0.5	4.7	6.1	6.7	0.7
	4	7.5	5.0	7.8	42.0	8.1	−5.8
	5[a]	16	18	33.6	341.8	12.8	−0.3
	6	5.7	−0.4	4.8	0.6	6.5	1.8
0.003	0	7.2	−4.0	6.2	0.3	8.1	0.4
	1	10	0.2	6.5	7.0	8.5	5.5
	2	10	−2.8	6.5	4.1	8.0	2.1
	3	7.4	−1.4	6.6	6.8	8.7	1.1
	4	10	5.8	9.9	49.7	10.3	−4.0
	5[a]	20	20	36.1	351.7	14.1	−0.5
	6	8.0	−0.4	7.1	1.6	8.6	2.6

[a] An extreme decorrelation model—see Section 3.2. The parametric analysis includes a decorrelation parameter. No attempt is made in the ILC analysis to model decorrelation. The middle column shows the parametric analysis when we don't include the deccorelation parameter.

large bias for the ILC method for Model 5 when compared to the parametric analysis that directly accounts for a possible decorrelation (last column). In general, we see that for $r = 0$ the simple Gaussian foreground Model 0 gives $\sigma(r) \sim 5 \times 10^{-4}$, exactly as expected from the semi-analytic code. As we progress to the more complex foreground models, $\sigma(r)$ is generally in the range 5–8$\times 10^{-4}$.

The level of biases is generally at $\lesssim 0.3\sigma$ for all the models. These simulations are sets of 500 realizations, so the statistical uncertainty on the bias is $\approx 0.04\sigma$. However, the strong decorrelation in Model 5 as well as the high-significance detection of decorrelation in Model 4 does significantly increase $\sigma(r)$. While the parametric method is able to account for the decorrelation, by construction information is lost, and in fact if one believed in such a scenario, re-optimization to concentrate the sensitivity at closer-in frequencies would be called for.

Table 3 shows results on detection significance for the strawperson configuration for sky Model 6. For $r = 0$, the 95% upper limit is about $2.1\sigma(r)$. The value of the tensor-to-scalar ratio for which we expect a 5σ detection is expected after 4 years of operation is $r = 0.004$. For a tensor-to-scalar ratio of $r = 0.003$, the median detection significance after 4 years is expected to be 4σ. If a detection were to be emerging at this point, extending the run time to 8 years would be justified to reach a 5σ detection.

While for given assumptions $\sigma(r)$ can be precisely forecast, the achieved detection level for r depends on the realization of the B-mode field in the observed patch of sky and the instrument noise. Therefore we can only forecast a distribution of detection levels. For a tensor-to-scalar ratio of $r = 0.003$ and the strawperson concept with 8 years of observing we expect to achieve more than 3σ detection with a probability of 0.99, more than 4σ with a probability of 0.93, more than 5σ with a probability of 0.53, and more than 6σ with a probability of 0.14. For simplicity we focus on $\sigma(r)$, and on median detection levels as well as median 95% confidence upper limits to state the typical outcome.

The numbers in Table 2 clearly show dependence on the foreground model used in the simulation. If the actual foregrounds are substantially different from any of these cases, then the biases could be larger. To get some understanding of how large the biases could be, and what instrument modifications might help to reduce them, we have also looked at ILC biases in the extreme case that the foreground residuals are not modeled, but simply absorbed into the estimated B-mode power spectrum. Doing so with simulations based on Sky Model 6 increases the magnitude of the bias on r from 0.4×10^{-4} to 1.3×10^{-4}. The dominant contribution to the bias comes from synchrotron residuals, which motivated placing one lower-frequency channel on the large-aperture.

Table 4 summarizes the results of analysis of simulations including additive systematic effects, in different combinations of uncorrelated and correlated contamination with varying spectra, added on top of foreground model 3. The levels of systematic contamination for these simulations were chosen to predict biases on r of 1×10^{-4} in semi-analytic forecasts. We can see that the different combinations explored increase biases on r by amounts that typically vary from 0.5–1.5$\times 10^{-4}$ for the two different analyses, over the different cases. We find that to restrict bias on r to this level, the sum of additive contamination effects needs to be controlled to 3–7% of the single-frequency survey noise, or (in the case of correlated systematics) 6–11% of the total combined noise levels. Such percentages are consistent with the upper limits currently achieved for residual additive systematic contamination compared to survey noise by small-aperture experiments (e.g., BICEP2 and Keck Array Collaborations 2016). Assuming CMB-S4 will include a sustained effort to continue to control, understand, and model systematic effects down to levels limited by survey noise, these percentages provide reasonable benchmark requirements.

TABLE 3
RESULTS ON DETECTION SIGNIFICANCE FOR THE STRAWPERSON CONCEPT SELECTED FOR CMB-S4, USING THE TWO ANALYSIS METHODS.

			ILC		Parametric (incl. decorrelation)	
r value	Duration	Sky model	95% CL UL	Detection Significance	95% CL UL	Detection Significance
0	4 years	6	1.0×10^{-3}	. . .	1.0×10^{-3}	. . .
0.003	4 years	6	. . .	4.0	. . .	4.2
	8 years	6	. . .	5.1	. . .	5.6

TABLE 4
MAP-BASED SIMULATION RESULTS FOR SIMULATIONS CONTAINING SYSTEMATICS. SIMULATIONS ARE AS IN TABLE 2 FOR SKY MODEL 3 AND $r = 0$, WITH ADDITIVE SYSTEMATIC EFFECTS IN VARYING COMBINATIONS, THE AMPLITUDES OF WHICH ARE SPECIFIED AS PERCENTAGES OF SURVEY NOISE.

	Uncorrelated		Correlated		ILC		Parametric	
Systematic	A [%]	B [%]	A [%]	B [%]	$\sigma(r) \times 10^{-4}$	r bias $\times 10^{-4}$	$\sigma(r) \times 10^{-4}$	r bias $\times 10^{-4}$
None	0	0	0	0	5.3	0.0	7.2	0.0
Uncorrelated white	3.3	0	0	0	6.0	0.84	8.0	0.63
Uncorrelated $1/\ell$	0	6.8	0	0	5.0	0.99	7.0	0.85
Correlated white	0	0	5.8	0	6.3	1.2	7.3	1.41
Correlated $1/\ell$	0	0	0	10.5	5.2	1.0	6.7	0.97
Uncorrelated white + $1/\ell$ . . .	1.6	3.5	0	0	5.6	0.89	7.5	0.76
Correlated white + $1/\ell$	0	0	2.9	5.3	5.5	0.98	6.9	1.04
Both, white + $1/\ell$	0.8	1.7	1.5	2.6	5.6	1.1	7.9	0.98

Results of simulating systematic errors in the determination of bandpasses vary by analysis method. The construction of the ILC method makes it largely insensitive to such errors. The parametric analysis, which includes parametrized models of the frequency spectra of different foregrounds, shows biases on r at the 1×10^{-4} level for uncorrelated random deviations in bandcenter determination of 0.8%, or for correlated deviations of 2%, which we adopt as reasonable benchmark requirements to accommodate a variety of both blind and astrophysical foreground modeling approaches.

In the current simulations, the treatment of delensing is still somewhat crude. We have explored alternate methods of translating detector effort to map noise, including scaling directly from the noise in a fielded, high-resolution experiment (SPTpol), but with some assumed modifications in per-detector sensitivity. The scaling using this alternate method is slightly more pessimistic than the default scaling, and we adopt this more pessimistic scaling for the numbers in Table 1 and consequently the Science and Measurement Requirements section of the CDT report[3]. We have also included multiple frequency bands in the delensing instrument because we cannot conclusively rule out the possibility that non-Gaussianity in small-scale Galactic foregrounds will cause a bias in a single-frequency delensing survey. The level of delensing assumed in the optimization code is reproduced with the configuration in Table 1 if the information from the two "CMB channels" (95 and 145 GHz) can be combined optimally for CMB sensitivity. If instead we assume we have to combine frequency bands to explicitly project out a dust component and a synchrotron component, the delensing efficiency degrades by 5–10%. Effort is currently underway to include delensing with fully non-Gaussian small-scale foregrounds in the Data Challenges. Separate simulations with one model of fully non-Gaussian dust (Vansyngel et al. 2017) indicate that biases to delensing are negligible (also see Challinor et al. 2017), but a larger parameter space of foreground models will need to be explored.

6. CONCLUSION

To Do: Write Conclusion

ACKNOWLEDGEMENTS

To Do: Define and write the Acknowledgements

REFERENCES

BICEP2 and Keck Array Collaborations. 2018, Phys. Rev. Lett. 121, 221301 2.1, 2.1.3, 2.1.4, 4

BICEP2 and Keck Array Collaborations. 2016, Phys. Rev. Lett., 116, 031302 2.1, 4, 5

BICEP2, Keck Array and Planck Collaborations. 2015, Phys. Rev. Lett., 114, 101301 2.1, 4

Tegmark, M., et al. 1996, Astrophys.J., 480, 22 2.1.1

Challinor, A., et al. 2017, arXiv e-prints 5

Eriksen, H. K., Banday, A. J., Gorski, K. M., & Lilje, P. B. 2004, Astrophys. J., 612, 633 4

Ghosh, T., et al. 2017, A&A, 601, A71 4

Hensley, B. 2015, PhD thesis, Princeton University 3

Kritsuk, A. G., Ustyugov, S. D., & Norman, M. L. 2017, New Journal of Physics, 19, 065003 6

Paine, S. 2017, The am atmospheric model (v. 9.2) 2

Planck Collaboration. 2017, A&A, 599, A51 4, 5

Sheehy, C., & Slosar, A. 2017, ArXiv e-prints, 1709.09729 5

Smith, K. M., Hanson, D., LoVerde, M., Hirata, C. M., & Zahn, O. 2012, JCAP, 1206, 014 2.2.1

Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2016, arXiv e-prints 1, 2

Vansyngel, F., et al. 2017, Astron. Astrophys., 603, A62 5, 5