University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Abazajian, K;Addison, GE;Adshead, P;Ahmed, Z;Akerib, D;Ali, A;Allen, SW;Alonso, D;Alvarez, M;Amin, MA;Anderson, A;Arnold, KS;Ashton, P;Baccigalupi, C;Bard, D;Barkats, D;Barron, D;Barry, PS;Bartlett, JG;Thakur, RB;Battaglia, N;Bean, R;Bebek, C;Bender, AN;Benson, BA;Bianchini, F;Bischoff, CA;Bleem, L;Bock, JJ;Bocquet, S;Boddy, KK;Bond, JR;Borrill, J;Bouchet, FR;Brinckmann, T;Brown, ML;Bryan, S;Buza, V;Byrum, K;Caimapo, CH;Calabrese, E;Calafut, V;Caldwell, R;Carlstrom, JE;Carron, J;Cecil, T;Challinor, A;Chang, CL;Chinone, Y;Cho, H-MS;Cooray, A;Coulton, W;Crawford, TM;Crites, A;Cukierman, A;Cyr-Racine, F-Y;de Haan, T;Delabrouille, J;Devlin, M;Di Valentino, E;Dierickx, M;Dobbs, M;Duff, S;Dvorkin, C;Eimer, J;Elleflot, T;Errard, J;Essinger-Hileman, T;Fabbian, G;Feng, C;Ferraro, S;Filippini, JP;Flauger, R;Flaugher, B;Fraisse, AA;Frolov, A;Galitzki, N;Gallardo, PA;Galli, S;Ganga, K;Gerbino, M;Gluscevic, V;Goeckner-Wald, N;Green, D;Grin, D;Grohs, E;Gualtieri, R;Gudmundsson, JE;Gullett, I;Gupta, N;Habib, S;Halpern, M;Halverson, NW;Hanany, S;Harrington, K;Hasegawa, M;Hasselfield, M;Hazumi, M;Heitmann, K;Henderson, S;Hensley, B;Hill, C;Hill, JC;Hlozek, R;Ho, S-PP;Hoang, T;Holder, G;Holzapfel, W;Hood, J;Hubmayr, J;Huffenberger, KM;Hui, H;Irwin, K;Jeong, O;Johnson, BR;Jones, WC;Kang, JH;Karkare, KS;Katayama, N;Keskitalo, R;Kisner, T;Knox, L;Koopman, BJ;Kosowsky, A;Kovac, J;Kovetz, ED;Kuhlmann, S;Kuo, C-L;Kusaka, A;Lahteenmaki, A;Lawrence, CR;Lee, AT;Lewis, A;Li, D;Linder, E;Loverde, M;Lowitz, A;Lubin, P;Madhavacheril, MS;Mantz, A;Marques, G;Matsuda, F;Mauskopf, P;McCarrick, H;McMahon, J;Meerburg, PD;Melin, J-B;Menanteau, F;Meyers, J;Millea, M;Mohr, J;Moncelsi, L;Monzani, M;Mroczkowski, T;Mukherjee, S;Nagy, J;Namikawa, T;Nati, F;Natoli, T;Newburgh, L;Niemack, MD;Nishino, H;Nord, B;Novosad, V;O'Brient, R;Padin, S;Palladino, S;Partridge, B;Petravick, D;Pierpaoli, E;Pogosian, L;Prabhu, K;Pryke, C;Puglisi, G;Racine, B;Rahlin, A;Rao, MS;Raveri, M;Reichardt, CL;Remazeilles, M;Rocha, G;Roe, NA;Roy, A;Ruhl, JE;Salatino, M;Saliwanchik, B;Schaan, E;Schillaci, A;Schmitt, B;Schmittfull, MM;Scott, D;Sehgal, N;Shandera, S;Sherwin, BD;Shirokoff, E;Simon, SM;Slosar, A;Spergel, D;St Germaine, T;Staggs, ST;Stark, A;Starkman, GD;Stompor, R;Stoughton, C;Suzuki, A;Tajima, O;Teply, GP;Thompson, K;Thorne, B;Timbie, P;Tomasi, M;Tristram, M;Tucker, G;Umilta, C;van Engelen, A;Vavagiakis, EM;Vieira, JD;Vieregg, AG;Wagoner, K;Wallisch, B;Wang, G;Watson, S;Westbrook, B;Whitehorn, N;Wollack, EJ;Wu, WLK;Xu, Z;Yang, HYE;Yasini, S;Yefremenko, VG;Yoon, KW;Young, E;Yu, C;Zonca, A

Title:
CMB-S4: Forecasting Constraints on Primordial Gravitational Waves

Date:

2022-02-01

Citation:
Abazajian, K., Addison, G. E., Adshead, P., Ahmed, Z., Akerib, D., Ali, A., Allen, S. W., Alonso, D., Alvarez, M., Amin, M. A., Anderson, A., Arnold, K. S., Ashton, P., Baccigalupi, C., Bard, D., Barkats, D., Barron, D., Barry, P. S., Bartlett, J. G. ,... Zonca, A. (2022). CMB-S4: Forecasting Constraints on Primordial Gravitational Waves. ASTROPHYSICAL JOURNAL, 926 (1), https://doi.org/10.3847/1538-4357/ac1596.

Persistent Link:
https://hdl.handle.net/11343/303141

License:
CC BY

The Astrophysical Journal, 926:54 (23pp), 2022 February 10 https://doi.org/10.3847/1538-4357/ac1596
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

CMB-S4: Forecasting Constraints on Primordial Gravitational Waves

Kevork Abazajian[1], Graeme E. Addison[2], Peter Adshead[3], Zeeshan Ahmed[4], Daniel Akerib[4], Aamir Ali[5], Steven W. Allen[6], David Alonso[7], Marcelo Alvarez[5,8], Mustafa A. Amin[9], Adam Anderson[10], Kam S. Arnold[11], Peter Ashton[5], Carlo Baccigalupi[12,13,14], Debbie Bard[8], Denis Barkats[15], Darcy Barron[16], Peter S. Barry[17], James G. Bartlett[18], Ritoban Basu Thakur[19], Nicholas Battaglia[20], Rachel Bean[20], Chris Bebek[8], Amy N. Bender[17,21], Bradford A. Benson[10,21,22], Federico Bianchini[23], Colin A. Bischoff[24], Lindsey Bleem[17,21], James J. Bock[19,25], Sebastian Bocquet[26], Kimberly K. Boddy[2,27], J. Richard Bond[28], Julian Borrill[5,8], François R. Bouchet[29], Thejs Brinckmann[30], Michael L. Brown[31], Sean Bryan[32], Victor Buza[15,21,93], Karen Byrum[17], Carlos Hervias Caimapo[33], Erminia Calabrese[34], Victoria Calafut[20], Robert Caldwell[35], John E. Carlstrom[17,21,22], Julien Carron[36,37], Thomas Cecil[17], Anthony Challinor[38,39,40], Clarence L. Chang[17,21,22], Yuji Chinone[41], Hsiao-Mei Sherry Cho[4], Asantha Cooray[1], Will Coulton[42], Thomas M. Crawford[21,22], Abigail Crites[19,43,44], Ari Cukierman[4,6], Francis-Yan Cyr-Racine[16], Tijmen de Haan[45], Jacques Delabrouille[18,46,47], Mark Devlin[48], Eleonora Di Valentino[31], Marion Dierickx[15], Matt Dobbs[49], Shannon Duff[50], Cora Dvorkin[51], Joseph Eimer[2], Tucker Elleflot[8], Josquin Errard[18], Thomas Essinger-Hileman[52], Giulio Fabbian[37], Chang Feng[3], Simone Ferraro[8], Jeffrey P. Filippini[3], Raphael Flauger[11], Brenna Flaugher[10], Aurelien A. Fraisse[53], Andrei Frolov[54], Nicholas Galitzki[11], Patricio A. Gallardo[20], Silvia Galli[29], Ken Ganga[18], Martina Gerbino[14], Vera Gluscevic[53,55], Neil Goeckner-Wald[6], Daniel Green[11], Daniel Grin[56], Evan Grohs[5], Riccardo Gualtieri[17], Jon E. Gudmundsson[57], Ian Gullett[58], Nikhel Gupta[23], Salman Habib[17], Mark Halpern[59], Nils W. Halverson[60], Shaul Hanany[61], Kathleen Harrington[62], Masaya Hasegawa[45], Matthew Hasselfield[63], Masashi Hazumi[45], Katrin Heitmann[17], Shawn Henderson[4], Brandon Hensley[53], Charles Hill[5], J. Colin Hill[64], Renée Hložek[43,44], Shuay-Pwu Patty Ho[53], Thuong Hoang[20], Gil Holder[3], William Holzapfel[5], John Hood[21,22], Johannes Hubmayr[50], Kevin M. Huffenberger[33], Howard Hui[19], Kent Irwin[6], Oliver Jeong[5], Bradley R. Johnson[65], William C. Jones[53], Jae Hwan Kang[6], Kirit S. Karkare[21,22], Nobuhiko Katayama[66], Reijo Keskitalo[5,8], Theodore Kisner[5,8], Lloyd Knox[67], Brian J. Koopman[68], Arthur Kosowsky[69], John Kovac[15,51], Ely D. Kovetz[70], Steve Kuhlmann[17], Chao-lin Kuo[6], Akito Kusaka[8], Anne Lähteenmäki[71], Charles R. Lawrence[25], Adrian T. Lee[5,8], Antony Lewis[37], Dale Li[4], Eric Linder[8], Marilena Loverde[30], Amy Lowitz[21,22], Phil Lubin[72], Mathew S. Madhavacheril[73], Adam Mantz[6], Gabriela Marques[33], Frederick Matsuda[66], Philip Mauskopf[32], Heather McCarrick[53], Jeffrey McMahon[21,22], P. Daniel Meerburg[42], Jean-Baptiste Melin[46], Felipe Menanteau[3], Joel Meyers[74], Marius Millea[75], Joseph Mohr[26], Lorenzo Moncelsi[19], Maria Monzani[4], Tony Mroczkowski[76], Suvodip Mukherjee[29,77], Johanna Nagy[78,79], Toshiya Namikawa[38], Federico Nati[80], Tyler Natoli[21,22], Laura Newburgh[68], Michael D. Niemack[20], Haruki Nishino[45], Brian Nord[10], Valentine Novosad[17], Roger O'Brient[19,25], Stephen Padin[19], Steven Palladino[24], Bruce Partridge[56], Don Petravick[3], Elena Pierpaoli[55], Levon Pogosian[54], Karthik Prabhu[67], Clement Pryke[61], Giuseppe Puglisi[5,8], Benjamin Racine[15], Alexandra Rahlin[10], Mayuri Sathyanarayana Rao[5], Marco Raveri[48], Christian L. Reichardt[23], Mathieu Remazeilles[31], Graca Rocha[25], Natalie A. Roe[8], Anirban Roy[20], John E. Ruhl[58], Maria Salatino[6,81], Benjamin Saliwanchik[68], Emmanuel Schaan[8], Alessandro Schillaci[19], Benjamin Schmitt[15], Marcel M. Schmittfull[82], Douglas Scott[59], Neelima Sehgal[30], Sarah Shandera[83], Blake D. Sherwin[38], Erik Shirokoff[21,22], Sara M. Simon[10], Anže Slosar[84], David Spergel[53], Tyler St. Germaine[51], Suzanne T. Staggs[53], Antony Stark[15], Glenn D. Starkman[58], Radek Stompor[18], Chris Stoughton[10], Aritoki Suzuki[8], Osamu Tajima[85], Grant P. Teply[11], Keith Thompson[6], Ben Thorne[67], Peter Timbie[86], Maurizio Tomasi[87], Matthieu Tristram[88], Gregory Tucker[89], Caterina Umiltà[3], Alexander van Engelen[32], Eve M. Vavagiakis[20], Joaquin D. Vieira[3], Abigail G. Vieregg[21,22], Kasey Wagoner[53], Benjamin Wallisch[11,82], Gensheng Wang[17], Scott Watson[90], Ben Westbrook[5], Nathan Whitehorn[91,92], Edward J. Wollack[52], W. L. Kimmy Wu[21], Zhilei Xu[48], H. Y. Eric Yang[6], Siavash Yasini[55], Volodymyr G. Yefremenko[17], Ki Won Yoon[6], Edward Young[4], Cyndia Yu[6], and Andrea Zonca[11]

The CMB-S4 Collaboration

[1] University of California, Irvine, CA 92697, USA
[2] Johns Hopkins University, Baltimore, MD 21218, USA
[3] Illinois Center for Advanced Studies of the Universe & Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[4] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[5] Department of Physics, University of California, Berkeley, CA 94720, USA
[6] Stanford University, Stanford, CA 94305, USA
[7] University of Oxford, Oxford OX1 3RH, UK
[8] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[9] Department of Physics & Astronomy, Rice University, Houston, TX 77005, USA
[10] Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
[11] University of California, San Diego, La Jolla, CA 92093, USA
[12] SISSA—International School for Advanced Studies, Via Bonomea, Italy
[13] IFPU—Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy
[14] INFN—National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy
[15] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA; vbuza@kicp.uchicago.edu
[16] Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
[17] Argonne National Laboratory, Lemont, IL 60439, USA
[18] Laboratoire Astroparticule et Cosmologie (APC), CNRS/IN2P3, Université Paris Diderot, F-75205 Paris Cedex 13, France

[19] California Institute of Technology, Pasadena, CA 91125, USA
[20] Cornell University, Ithaca, NY 14853, USA
[21] Kavli Institute for Cosmological Physics, Chicago, IL 60637, USA
[22] University of Chicago, Chicago, IL 60637, USA
[23] School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia
[24] University of Cincinnati, Cincinnati, OH 45221, USA
[25] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91011, USA
[26] Ludwig-Maximilians-Universität, D-81679 Munich, Germany
[27] Theory Group, Department of Physics, University of Texas at Austin, Austin, TX 78712, USA
[28] Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON M5S 3H8, Canada
[29] Institut d'Astrophysique de Paris (IAP), CNRS & Sorbonne University, Paris, France
[30] Stony Brook University, Stony Brook, NY 11794, USA
[31] Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
[32] Arizona State University, Tempe, AZ 85287, USA
[33] Florida State University, Tallahassee, FL 32306, USA
[34] School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, UK
[35] Department of Physics & Astronomy, Dartmouth College, Hanover, NH 03755, USA
[36] Département de Physique Théorique et CAP, Université de Genève, CH-1211 Genève 4, Switzerland
[37] Astronomy Centre, School of Mathematical and Physical Sciences, University of Sussex, Brighton BN1 9QH, UK
[38] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK
[39] Kavli Institute for Cosmology, University of Cambridge, Cambridge CB3 0HA, UK
[40] Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK
[41] University of Tokyo, 277-8583 Kashiwa, Japan
[42] Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, 9747 AG Groningen, The Netherlands
[43] David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, ON, M5S 3H4, Canada
[44] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, ON M5S 3H4, Canada
[45] High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan
[46] Département d'Astrophysique, CEA Saclay DSM/Irfu, F-91191 Gif-sur-Yvette, France
[47] Department of Astronomy, School of Physical Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China
[48] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[49] McGill University, Montreal, QC H3A 2T8, Canada
[50] National Institute of Standards and Technology, Boulder, CO 80305, USA
[51] Department of Physics, Harvard University, Cambridge, MA 02138, USA
[52] Goddard Space Flight Center, Greenbelt, MD 20771, USA
[53] Princeton University, Princeton, NJ 08544, USA
[54] Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[55] University of Southern California, CA 90089, USA
[56] Haverford College, Haverford, PA 19041, USA
[57] Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, AlbaNova, Stockholm SE-106 91, Sweden
[58] Case Western Reserve University, Cleveland, OH 44106, USA
[59] University of British Columbia, Vancouver, BC V6T 1Z1, Canada
[60] Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Science, University of Colorado, Boulder, CO 80309, USA
[61] University of Minnesota, Minneapolis, MN 55455, USA
[62] University of Michigan, Ann Arbor, MI 48109, USA
[63] Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
[64] Columbia University, New York, NY 10027, USA
[65] University of Virginia, Charlottesville, VA 22903, USA
[66] Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, Kashiwa, Japan
[67] University of California, Davis, CA 95616, USA
[68] Department of Physics, Yale University, New Haven, CT 06520, USA
[69] University of Pittsburgh and PITT PACC, Pittsburgh, PA 15260, USA
[70] Department of Physics, Ben-Gurion University, Be'er Sheva 84105, Israel
[71] Aalto University, FI-00076, Aalto, Finland
[72] University of California, Santa Barbara, CA 93106, USA
[73] Perimeter Institute, Waterloo, ON N2L 2Y5, Canada
[74] Southern Methodist University, Dallas, TX 75275, USA
[75] Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA
[76] European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching b. Munchen, Germany
[77] GRAPPA Institute, University of Amsterdam, 1098 XH Amsterdam, The Netherlands
[78] Washington University in St Louis, St. Louis, MO 63130, USA
[79] McDonnell Center for the Space Sciences, Washington University, St. Louis, MO 63130, USA
[80] Department of Physics, University of Milano—Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy
[81] Kavli Institute for Particle Astrophysics and Cosmology, Stanford, CA 94305, USA
[82] Institute for Advanced Study, Princeton, NJ 08540, USA
[83] The Pennsylvania State University, University Park, PA 16802, USA
[84] Brookhaven National Laboratory, Upton, NY 11973, USA
[85] Department of Physics, Kyoto University, Kyoto 606-8502, Japan
[86] Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA
[87] Dipartimento di Fisica "Aldo Pontremoli," Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy
[88] Université Paris-Saclay, CNRS/IN2P3, IJCLab, F-91405 Orsay, France
[89] Brown University, Providence, RI 02912, USA
[90] Syracuse University, Syracuse, NY 13244, USA

[91] University of California, Los Angeles, Los Angeles, CA 90095, USA
[92] Michigan State University, East Lansing, MI 48824, USA
Received 2020 August 26; revised 2021 June 15; accepted 2021 July 8; published 2022 February 11

Abstract

CMB-S4—the next-generation ground-based cosmic microwave background (CMB) experiment—is set to significantly advance the sensitivity of CMB measurements and enhance our understanding of the origin and evolution of the universe. Among the science cases pursued with CMB-S4, the quest for detecting primordial gravitational waves is a central driver of the experimental design. This work details the development of a forecasting framework that includes a power-spectrum-based semianalytic projection tool, targeted explicitly toward optimizing constraints on the tensor-to-scalar ratio, r, in the presence of Galactic foregrounds and gravitational lensing of the CMB. This framework is unique in its direct use of information from the achieved performance of current Stage 2–3 CMB experiments to robustly forecast the science reach of upcoming CMB-polarization endeavors. The methodology allows for rapid iteration over experimental configurations and offers a flexible way to optimize the design of future experiments, given a desired scientific goal. To form a closed-loop process, we couple this semianalytic tool with map-based validation studies, which allow for the injection of additional complexity and verification of our forecasts with several independent analysis methods. We document multiple rounds of forecasts for CMB-S4 using this process and the resulting establishment of the current reference design of the primordial gravitational-wave component of the Stage-4 experiment, optimized to achieve our science goals of detecting primordial gravitational waves for $r > 0.003$ at greater than 5σ, or in the absence of a detection, of reaching an upper limit of $r < 0.001$ at 95% CL.

Unified Astronomy Thesaurus concepts: Cosmic background radiation (317); Cosmological parameters (339); Gravitational waves (678); Cosmic inflation (319)

1. Introduction

Determining the origin of structure in the universe is one of the most important open problems in cosmology. CMB anisotropies sourced by early-universe density perturbations are currently the most powerful observational probe of the earliest mechanisms of structure formation. It is possible that the same processes that produced the density perturbations also sourced tensor perturbations, or primordial gravitational waves (PGWs). If this is the case, detecting a PGW signal would yield insight into physics far earlier than the epoch of recombination, and allow us to build an unprecedented understanding of the earliest moments of time.

PGWs leave imprints on the polarization of the CMB (Hu & White 1997; Kamionkowski et al. 1997; Seljak & Zaldarriaga 1997). In particular, the sensitivity of CMB measurements to gravitational waves arises from the generation of polarization at the surface of last scattering: to first order, scalar perturbations produce only even-parity E-mode polarization, while tensor perturbations produce odd-parity B-mode polarization as well. Thus, a measurement of primordial B-mode polarization in the CMB, parameterized by the tensor-to-scalar ratio r, is a direct measurement of the amplitude of tensor perturbations. Detecting r has profound implications for high-energy physics and the quantum nature of gravity (Krauss & Wilczek 2014), and the potential to shed light on the mechanism that produced these primordial perturbations.

Cosmic inflation is our current leading paradigm for what occurred in the very early universe. It was first put forward to explain the lack of observed magnetic monopoles and to solve the flatness and horizon problems (Kazanas 1980; Starobinsky 1980; Sato 1981; Guth 1981; Linde 1982, 1983; Albrecht & Steinhardt 1982) and has since been an active field of research. The theory describes a period of exponential expansion in which quantum fluctuations are magnified to cosmic size and become the seeds for all structure in the universe (Mukhanov & Chibisov 1981, 1982; Guth & Pi 1982; Hawking 1982; Starobinsky 1982; Mukhanov 1985; Bardeen et al. 1986). In addition to the production of PGWs (for a recent review, see Kamionkowski & Kovetz 2016), inflation makes several predictions, most of which—superhorizon fluctuations, Gaussian perturbations, adiabatic fluctuations, spatial flatness, and a nearly scale-invariant scalar spectral tilt—have been confirmed, most recently by the Planck collaboration (Planck Collaboration et al. 2020b).

There are currently a number of ground-based experiments measuring the CMB polarization to high precision on a range of scales, and attempting to constrain the tensor-to-scalar ratio, including ACT (Aiola et al. 2020), BICEP/Keck (BICEP2/Keck Array Collaborations X 2018), CLASS (Harrington et al. 2016), POLARBEAR/Simons Array (Suzuki et al. 2016; Hasegawa et al. 2018), and SPT (Bender et al. 2018; Sayre et al. 2020), with Simons Observatory to follow soon (Ade et al. 2019). Additionally, there are current and future balloon and satellite missions such as SPIDER (Gualtieri et al. 2018) and LiteBIRD (Hazumi et al. 2019), which we expect to complement ground-based measurements. The current best constraints are $r(k_s = 0.05\,\mathrm{Mpc}^{-1}, n_t = 0) < 0.06$ at 95% confidence (BICEP2/Keck Array Collaborations X 2018), where k_s is the scalar pivot scale and n_t is the spectral index of the tensor modes. Ongoing efforts in both the Atacama Desert and at the South Pole, between now and the start of CMB-S4, will steadily improve these constraints while continuing to prove the methodologies on which CMB-S4 will rely.

CMB-S4, anticipated to start observations in 2027, is intended to be the definitive ground-based CMB polarization experiment. It is designed to cross critical thresholds in constraining the B-mode polarization signature of primordial

[93] Author to whom any correspondence should be addressed.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

gravitational waves and in sensitivity to new light relics, while also improving our understanding of the nature of dark energy and General Relativity on large scales (Abazajian et al. 2016). To achieve these goals requires a significant increase in sensitivity, from 2 to 4×10^4 detectors in Stage 3 experiments to roughly 5×10^5 detectors, and an unparalleled control over other sources of signal (e.g., Galactic foregrounds, gravitational lensing, etc.) and of systematics. Therefore, CMB-S4 will require telescopes at multiple frequencies, each with a maximally outfitted focal plane of pixels utilizing superconducting, photon-noise-limited detectors, and likely novel analysis techniques. To understand the optimal design for achieving the desired science goals, in particular focusing on primordial gravitational waves, we present the development of the *CMB-S4 r forecasting framework* and its application toward determining the CMB-S4 baseline *r* survey.

The sensitivity achieved by a CMB experiment, which observes for a given number of years with a given number of detectors, is subject to a number of efficiency factors. These include the fraction of detectors that are actually functional, the achieved sensitivity per detector versus model prediction, the fraction of days per year spent observing, the fraction of observing time spent "on field," and the fraction of data passing weather and other cuts. As an illustrative example, if each of the above efficiency factors were 0.65 then the product is $0.65^5 \sim 0.1$. Forecasts for the sensitivity to *r*, where excess low-frequency noise or systematic contamination can lead to additional (potentially large) sensitivity loss, are particularly challenging. Most previous forecasts of the sensitivity of CMB experiments to *r* have been *ab initio*—requiring assumptions as to the value of all efficiency factors. Possibly due to the human tendency toward optimism, many of these forecasts have not in practice been achieved once the data have been taken and analyzed: ABS (Essinger-Hileman 2011; Kusaka et al. 2018); BICEP/Keck (Hivon & Kamionkowski 2002; Yoon et al. 2006; Nguyen et al. 2008; BICEP2/Keck Array Collaborations VI 2016); EBEX (Reichborn-Kjennerud et al. 2010; Abitbol et al. 2018); Planck (Planck Collaboration 2006; Planck Collaboration et al. 2020b); POLARBEAR (Lee et al. 2008; Kermish et al. 2012; Polarbear Collaboration et al. 2020); QUIET (Lawrence et al. 2004; Araujo et al. 2012); SPIDER (Montroy et al. 2006; Fraisse et al. 2013; Gualtieri et al. 2018); SPTpol (McMahon et al. 2009; Austermann et al. 2012; Sayre et al. 2020). Therefore, for CMB-S4, we take an alternate approach, scaling from the overall performance achieved in the best available existing experimental results. We scale directly from published *B*-mode noise spectra and bandpower covariance matrices derived from multiyear maps that have passed systematics null tests. This bypasses the need for an unbiased, individual accounting of the various efficiency factors, and naturally incorporates all effects that impact real-world experiments. This also differentiates our current approach from forecasts other groups have made for CMB-S4 prior to the formal existence of our collaboration (Wu et al. 2014; Errard et al. 2016; Barron et al. 2018).

1.1. The CMB-S4 r Forecasting Loop

The CMB-S4 baseline *r* survey, described below, has been continuously evolving by translating science requirements to measurement and instrument requirements, based on our understanding of the impact of astrophysical foregrounds, instrumental systematics, delensing nonidealities, and analysis methodology. To maintain forecasting realism as complexity increases, our general approach has been an iterative one. We rely on a closed forecasting loop, presented in Figure 1, to tie the semianalytic tools (which allow for fast optimizations) to map-based studies (which can include multiple layers of additional complexity). To establish our measurement requirements and the baseline experiment configurations that can achieve them, we perform multiple passes through this loop. In the figure, the section number accompanying each stage of the loop indicates the paper section in which that stage is described in detail.

The main steps describing this process are as follows.

1. Develop a semianalytic power-spectrum-level forecast, assuming noise performance that is scaled from analyses of real experiments.
2. Use this forecasting tool to optimize the allocation of detector effort across observing frequencies, determining certain "checkpoints" in survey definition space.
3. Use the checkpoint configurations to create standardized, version-numbered, map-based data challenges (DCs) for validation.
4. Estimate science parameters from the DC maps with independent component-separation analysis methods.
5. Check that independent analyses show recovery of science parameters from these challenge maps that match analytic forecasts, either in terms of variance or bias. If they do not, we revise the forecasts accordingly.
6. Iterate steps 1–5, injecting increasing realism in the form of: (a) sky model complexity informed by the latest data and modeling efforts; (b) survey coverage based on proven observing strategies; (c) systematics whose form, parameterization, and likely amplitude are likewise guided by real-world experience; and (d) treatment of lensing.

1.2. CMB-S4 r Forecasting Workflow and Evolution

Developing the forecasting machinery for CMB-S4, and increasing the robustness and realism of its results, has been an ongoing, ever-evolving, group effort, performed under the auspices of the CMB-S4 *r* Forecasting Working Group. This work has undergone three major iterations, with results presented in the CMB-S4 Science Book (Abazajian et al. 2016), the CMB-S4 Concept Definition Task Force (CDT) Report (Lawrence et al. 2017), and most recently in the CMB-S4 Reference Design Report (Abazajian et al. 2019).

Each stage in the evolution of this framework has been collectively considered, and carefully documented in the CMB-S4 Simulation and Forecasting Logbook.[94] The resulting baseline survey definitions have been translated to publicly available, version-numbered, map-based Data Challenges (DCs).[95] Thus far, two independent groups have participated in testing the strengths and shortcomings of different analysis methods on these simulations, as well as improving the simulations themselves, as described in Sections 5 and 6.

Given the difficulty in describing a continuously evolving process, for this paper we present several snapshots of our forecasting work and describe the relevant details. Sections 2–6 represent a full pass through the forecasting loop, as presented in Figure 1. In Sections 2 and 3, we present the full semianalytic forecasting framework and optimization process, as used in the CMB-S4 CDT Report. In Section 4, we discuss

[94] https://cmb-s4.org/wiki/index.php/Simulation_and_Forecasting_Logbook
[95] https://cmb-s4.org/wiki/index.php/Data_Challenges

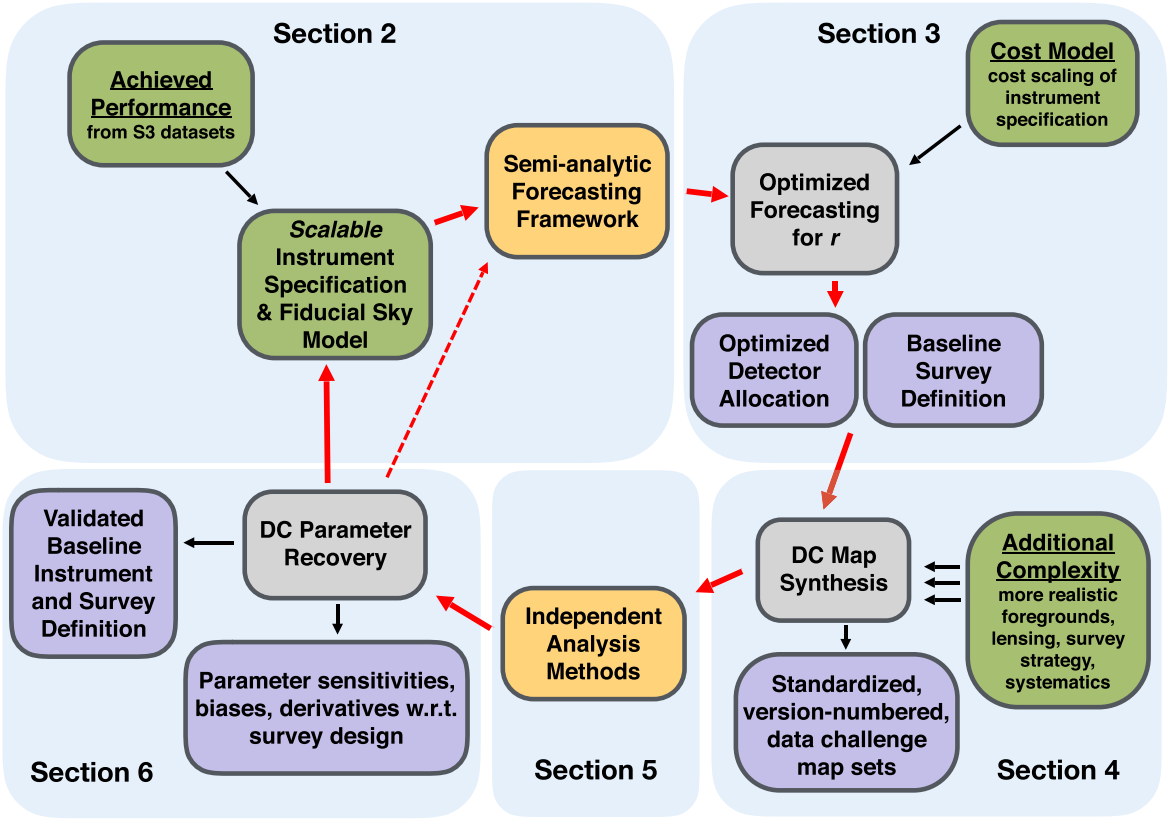


Figure 1. Schematic representation of the CMB-S4 r forecasting loop. Green boxes represent inputs, purple boxes represent outputs, yellow boxes represent large code frameworks, and gray boxes represent iterable code modules. For each stage of the loop, we identify the sections of this paper in which more detail is available. We start with the achieved performance of Stage 3 data sets, in the form of full covariance matrices and noise spectra, and a set of scalable instrument specifications, as well as a fiducial sky model. These are fed as inputs to the semianalytic optimization framework, yielding an optimized detector allocation and a baseline survey definition. Based on this definition, we develop standard data challenge (DC) noise maps, as well as a suite of signal maps with various degrees of complexity. We proceed by analyzing these maps with multiple independent component-separation analysis methods, and check for parameter recovery and the presence of biases. If the results suggest a necessary change in survey definition, or additional studies are required, the process is iterated as needed.

the creation of the DC4 simulation suite, corresponding to the baseline presented in that report. Sections 5 and 6 describe two independent analysis methods and the results obtained from applying these methods to DC4. In Section 7, we describe several recent updates to the framework and the resulting findings presented in the CMB-S4 Reference Design Report. We conclude this paper with a discussion of the path forward.

2. Semianalytic Forecasting Framework

For the CMB-S4 Science Book (Abazajian et al. 2016), we developed a semianalytic forecasting framework specifically targeted toward optimizing sensitivity to the tensor-to-scalar ratio, r, in the presence of Galactic foregrounds and gravitational lensing of the CMB. Currently, only small-aperture telescopes (SATs) have reached the level of systematics control and noise performance necessary to pursue a ground-based, high-precision measurement of B-mode polarization down to low multipoles ($\ell \simeq 30$), targeting the $\ell \simeq 80$ peak from the polarization signature generated by PGWs at the epoch of recombination. Therefore, to forecast the performance of next-generation SATs, this machinery is based on scaling the bandpower covariance matrices (BPCMs) and noise power spectra (N_ℓ) of current published SAT analyses, such as those from BICEP/Keck (BICEP2/Keck Array Collaborations VI 2016; BICEP2/Keck Array Collaborations X 2018). This automatically builds into the forecast all real-world inefficiencies including (but not limited to): imperfect detector yield; nonuniform detector performance; read-out noise; observing inefficiency; losses due to timestream filtering; beam smoothing; and nonuniform sky coverage.

At its core, this code is based on the BICEP/Keck parametric power-spectrum-based likelihood analysis. Such types of parametric analyses have also been extensively used by the Planck collaboration for $\ell \geqslant 50$ (Planck Collaboration XV 2014; Planck Collaboration XI 2016; Planck Collaboration et al. 2020a). We validate this approach using map-based simulations as part of the forecasting loop, and present detailed results in Section 6. Our confidence in the projections is grounded in the connection to achieved performance and published results.

Figure 2 presents a schematic representation of the semianalytic forecasting framework, identifying the user inputs, code modules, and outputs. The subsections that follow describe this framework in detail.

2.1. Fisher Formalism

Given a likelihood function of the form

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

where d are the data bandpowers, θ are the theory parameters, $\mu(\theta)$ and $\Sigma(\theta)$ are the bandpower expectation values and the bandpower covariance matrix given the parameters, and T

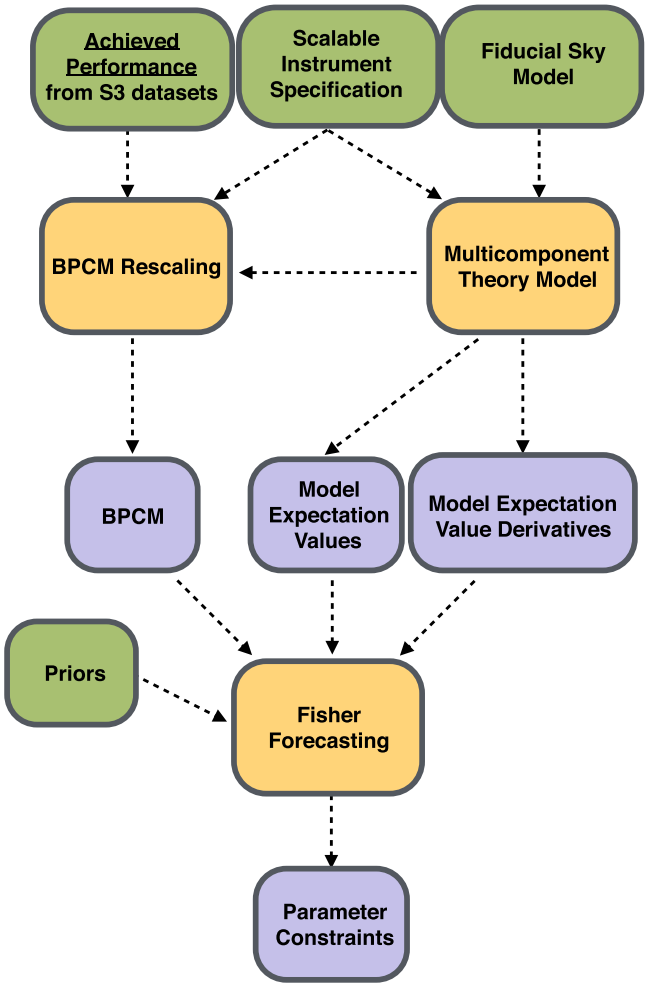


Figure 2. Schematic representation of our semianalytic forecasting framework. Green boxes represent user inputs, yellow boxes represent code modules, and purple boxes represent outputs. BPCM = bandpower covariance matrix.

denotes the transpose, we can calculate the expectation value of the log-likelihood curvature, evaluated at the position of the best-fit model:

$$F_{ij} = -\left\langle \frac{\partial^2 \log L(\theta; d)}{\partial\theta_i \partial\theta_j} \right\rangle. \quad (2)$$

This quantity, called the Fisher information 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 (e.g., Cramér 1946; Kendall & Stuart 1979; Tegmark et al. 1997).

Inserting Equation (1) into Equation (2) yields

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

We then calculate our parameter constraints as

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

In all the projections below, for each fiducial model considered, we choose to fix the covariance matrix with respect to the theory parameters, i.e., $\Sigma(\theta) = \Sigma$, making the second term of Equation (3) identically zero.[96] Equation (3) provides a clear picture of how the construction of the covariance matrix Σ impacts the final constraints, and how its misestimation could lead to constraints that are far too optimistic. It is with this in mind that we have decided to compute our BPCMs by directly scaling the values in achieved covariance matrices (see Section 2.5).

2.2. Forecasting Inputs

In this section, we briefly describe the set of inputs to the forecasting code, shown as green boxes in Figure 1. We discuss the key inputs in more detail in Sections 2.3 and 2.4.

Achieved Performance: The code takes signal and noise simulations of the BICEP/Keck data sets (BICEP2/Keck Array Collaborations VI 2016; BICEP2/Keck Array Collaborations X 2018) and derives the bandpower covariance matrix and the ensemble-averaged signal and noise bandpowers. These inputs contain information about the actual on-sky map noise achieved from multiple receivers, over multiple years, at 95, 150, and 220 GHz, including all real-world penalties. A more detailed description of these simulations is available in Appendix H of BICEP2/Keck Array Collaborations X (2018). Similar information from another experiment could easily be substituted. To project the performance of CMB-S4 channels, we assume that we can scale down the achieved noise based on increased detector count and integration time, and that we can apply beam size and noise-equivalent temperature (NET) rescalings to account for the differences in experimental design.

Scalable Instrument Specification: To specify the forecast instrument, we start by selecting a set of observing frequency bands, with bandpass functions describing the response of each band. Then, for each observing band, we must provide the beam size, number of detectors, and ideal per-detector NET. The last two items can be used to make an idealized calculation of the instrument sensitivity in each band. We emphasize that we use these ideal performance numbers *only* for scaling between frequency bands, by comparing to similarly calculated ideal sensitivities of BICEP/Keck. The scaling factors are ultimately applied to the achieved sensitivities, as described in Section 2.5, to obtain performance-based sensitivities in our desired bands. We note that, for the frequency bands in which we do not currently have existing data, we extrapolate from the closest frequency for which we do. A test of such extrapolations has been presented in Buza (2019), where the BICEP2/Keck Array Collaborations VI (2016) data set has been used to forecast the BICEP2/Keck Array Collaborations X (2018) results.

Fiducial Sky Model: A fiducial parametric model describing the foreground and CMB signal. Our standard model, which has 14 parameters, is discussed in Section 2.4 and in Appendix A.

Priors: If we have external prior knowledge for a given parameter θ_i, we can introduce this information by adding $P_i = 1/\sigma_i^2$ to the diagonal of the Fisher matrix, where σ_i is the width of the prior.

[96] This recipe produces constraints that are equivalent to the standard deviation of maximum-likelihood parameter estimates derived from an ensemble of simulations with bandpower covariance Σ (Buza 2019).

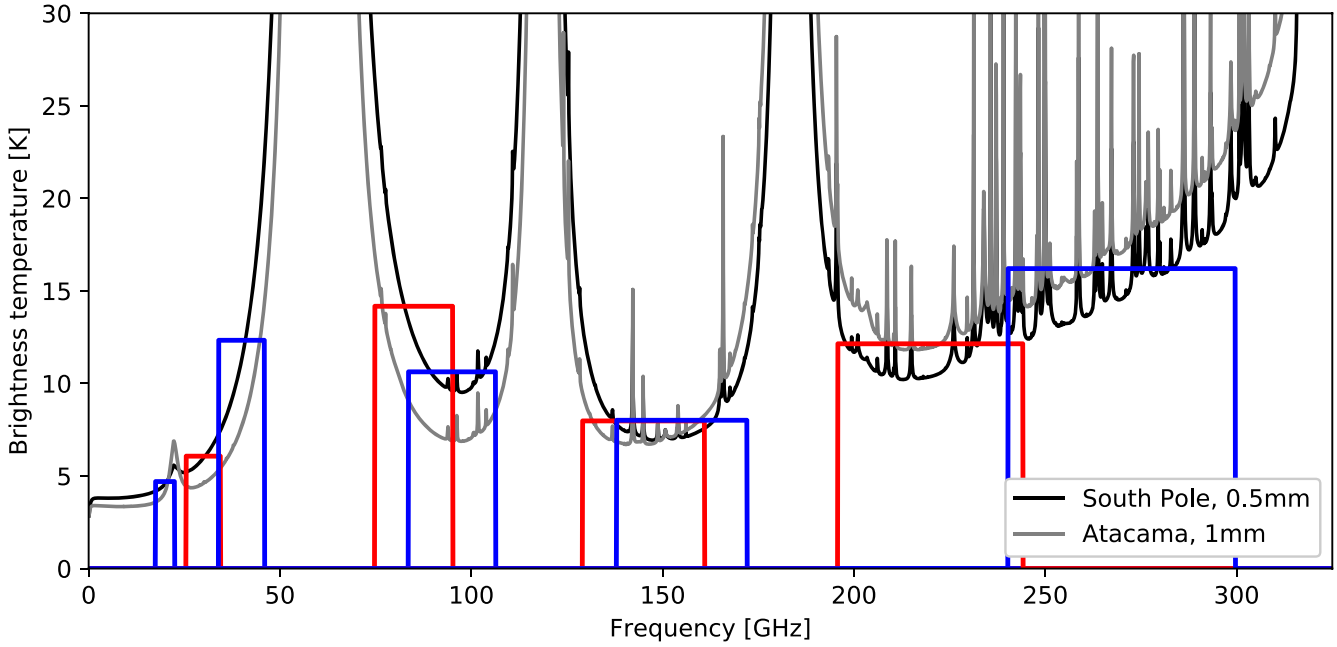


Figure 3. Calculated atmospheric brightness temperature spectra (at zenith) for the South Pole at 0.5 mm precipitable water vapor (PWV) and Atacama at 1.0 mm PWV (both are near the median values). Atmospheric spectra are generated using the `am` Atmospheric Model (Paine 2017). The top-hat bands, in red and blue, 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. Details about the bands, such as fractional bandwidths, are presented in Table 3-1 of Abazajian et al. (2019).

Table 1
Optimized Instrument Configuration for the PGW Survey, as Presented in the CMB-S4 CDT Report

		Frequency (GHz)										
Science Goal	Item	20	30	40	85	95	145	155	220	270	DL	Total
r	No. of detectors	130	260	470	17 k	21 k	18 k	21 k	34 k	54 k	84 k	250 k
	Angular resolution [FWHM]	11′	77′	58′	27′	24′	16′	15′	11′	8.′5	1.′0	

2.3. CMB-S4 Scalable Instrument Specification

To span the four atmospheric windows available to ground-based experiments (Figure 3) and have enough channels to mitigate against complex foregrounds, we assume eight channels at 30, 40, 85, 95, 145, 155, 220, and 270 GHz, which are placed on small-aperture telescopes. For these low-resolution instruments, we pick 0.52 m apertures, motivated by proven SAT Stage-3 experiments (Kang et al. 2018), yielding the beams presented in Table 1. In addition, we also include a 20 GHz channel on a large-aperture telescope (LAT).[97]

We derive the split in each window by separating 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 per-detector NETs were calculated with `NETlib.py`[98,99] at the South Pole and Chile, using the 10 yr MERRA2 median atmospheric profiles (Gelaro et al. 2017). We use the average of NET calculated for detectors at the two sites, which are 214, 177, 224, 270, 238, 309, 331, 747, and 1281 $\mu K_{CMB}\sqrt{s}$ for our nine channels at 20–270 GHz, respectively.[100] These NETs are lower than similar BICEP/Keck channels (which are 287.6, 313.1, and 837.7 $\mu K_{CMB}\sqrt{s}$ for 95, 150, and 220 GHz, predominantly because they are calculated for a 100 mK thermal bath, as opposed to 250 mK (which was also used in the CMB-S4 Science Book forecasts).

We also fold in information from two WMAP channels, 23 and 33 GHz, and seven Planck channels, 30, 44, 70, 100, 143, 217, and 353 GHz, though this extra information is only relevant in the early stages of CMB-S4 observation.

[97] The inclusion of a 20 GHz channel is the result of insight gained from an earlier iteration through the forecasting loop (performed for the CMB-S4 Science Book), which demonstrated that, for specific foreground models, sizable biases were present due to synchrotron residuals. To mitigate against such biases, the reference design was updated to include this additional channel. Placing this low-frequency band on an SAT would result in a very broad beam, which would dominate the noise at the relevant scales; to circumvent this, as mentioned above, we place it on an LAT. This means that, while the scaling of the noise for this channel is still calculated from achieved performance, we use noise parameters (see Equation (7)) that are more in tune with what has been achieved by LATs at the time of this forecasting (Louis et al. 2017; Henning et al. 2018), i.e., the $1/f$ noise component is characterized by an $\ell_{knee} = 200$, keeping γ—the slope of this component—the same as for the small-aperture noise, and a beam of $\Theta_\nu = 11'$ full width at half maximum (FWHM) (assuming a 6 m aperture).

[98] https://cmb-s4.org/wiki/index.php/New_NET_Calculator_and_Validation

[99] https://github.com/dbarkats/NET_forecast_python

[100] Going beyond the number of bands for the current reference design increases manufacturing complexity and costs, and decreases the per-channel NET, thereby reducing the overall sensitivity in each channel; while exploring alternative options with five bands (20, 30, 95, 155, 270 GHz) and seven bands (20, 30, 95, 155, 220, 270 GHz) has shown that choosing fewer bands leads to statistically significant biases on r after marginalization over foreground residuals (an increase in bias of up to $\Delta r = 1.5 \times 10^{-4}$ when compared to the current count). Balancing these considerations, we have chosen the proposed configuration with nine frequency bands for this work. Going forward, as instrumentation choices are finalized, we anticipate a possible revision of this design.

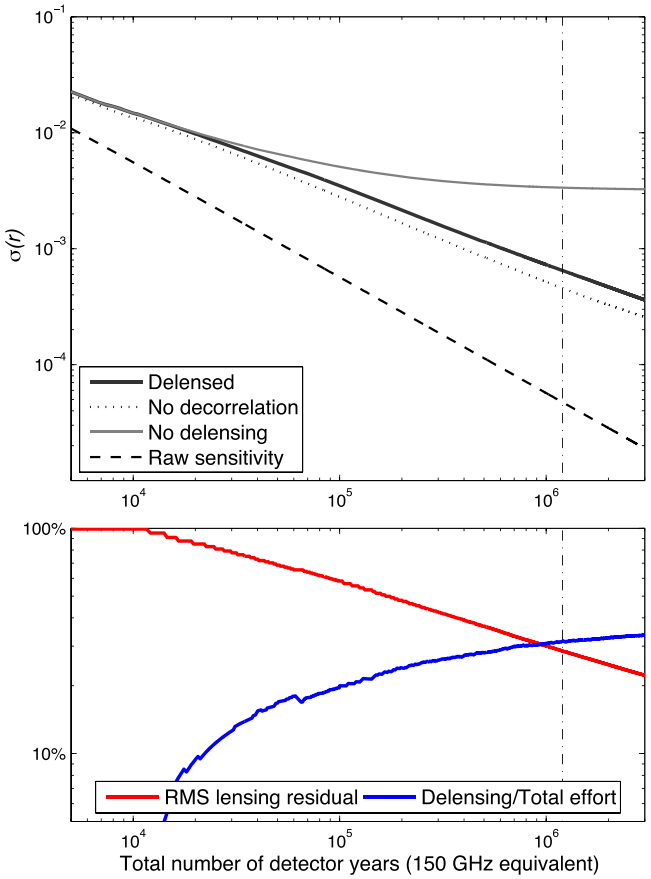


Figure 4. Top: optimized constraints on r for 3% sky fraction as a function of total effort. We include in solid black the case with delensing, allowing for decorrelation of the foregrounds, in solid gray the case without delensing, in dotted gray the case where no decorrelation is allowed in the model (with delensing), and in dashed black the raw sensitivity in the absence of foregrounds and lensing. Bottom: for the delensed case (with decorrelation), we show the fraction of effort spent on removing the lensing sample variance and the resulting rms lensing residual.

Since the forecasts use scaled BICEP/Keck bandpower statistics, they also use the same bandpower window functions and uniform binning: nine multipole bins with $\Delta\ell = 35$, spanning a multipole range of $21 \leqslant \ell \leqslant 335$.

2.3.1. Delensing Treatment

One of the main challenges for detecting primordial B modes is the lensing B-mode contribution from the weak lensing of E modes as the CMB photons travel to us. For instrument noise levels below 5 μK-arcmin, this lensing signal becomes an important source of contamination and its sample variance significantly worsens our constraining power on PGWs (Smith et al. 2012). Unlike Galactic foregrounds, the lensing signal is achromatic and cannot be mitigated with multifrequency observations. However, its contribution can be reduced by knowing the cumulative gravitational lensing potential ϕ along the line of sight and having a high-fidelity E-mode map. Together, the two can be combined to form a lensing B-mode template by lensing the E-mode map with the ϕ field and subtracting this template from the measured B-mode map. This technique is known as *delensing* (Knox & Song 2002; Seljak & Hirata 2004; Carron & Lewis 2017; Carron et al. 2017).

Reconstructing ϕ with high S/N requires high-sensitivity, high-angular-resolution CMB polarization maps (Lewis & Challinor 2006). Therefore, in addition to the low-resolution effort, we assume a separate high-resolution large-aperture instrument dedicated to measuring the intermediate- and small-scale information necessary to delens the B-mode map. This instrument is assumed to have 1′ resolution and noise performance equivalent to the 145 GHz channel from the small-aperture telescopes. These experiment specifications are revised for the CMB-S4 Reference Design Report and updated in Section 7. 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 Section 4.1 and Equation (7)), but using SPTpol achieved performance (Sayre et al. 2020), i.e., without incurring penalties from nonidealities specific to low-resolution instruments and low-ℓ analysis, such as low-ℓ mode filtering and nonuniform coverage. Following the iterative formalism in Smith et al. (2012),[101] using $\ell_{\rm min} = 300$ and $\ell_{\rm max} = 4000$ for the ϕ reconstruction and $\ell_{\rm min} = 30$ for the E-mode map, we convert the map noise in the delensing survey to a delensing efficiency, or equivalently a fractional residual in the lensed B-mode power, specified by setting the residual lensing amplitude $A_{\rm L}$ to the corresponding level.

The detector effort dedicated to the delensing instrument comes out of the total detector effort budget for the r survey, and the distribution of effort between the low-resolution and delensing instruments is part of the optimization process, as shown in Figures 4 and 5.

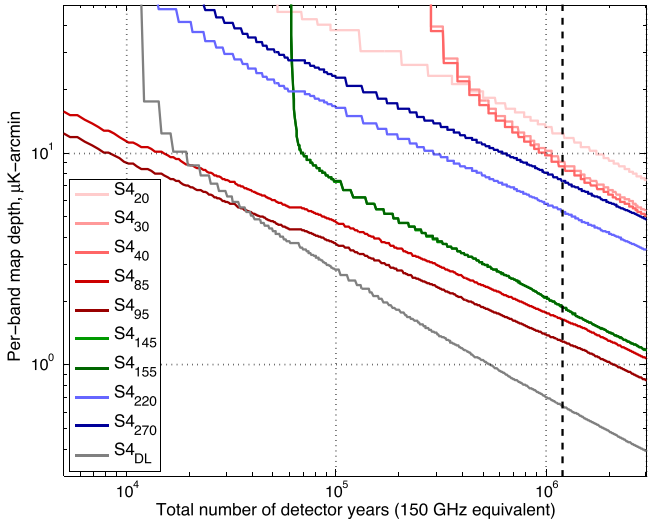


Figure 5. Optimized map depth in each of the small-aperture channels as well as in the delensing channel, for $f_{\rm sky} = 3\%$, corresponding to the delensed (with decorrelation) case in Figure 4.

2.4. Multicomponent Theory Model

Our model includes a CMB component (Planck Collaboration et al. 2014) parameterized by r and the residual lensing amplitude, $A_{\rm L}$, and components of polarized dust and synchrotron emission. We assume that the synchrotron scales as a simple power law in both frequency and ℓ. For the dust, we assume a power-law scaling in ℓ and a modified blackbody spectral energy distribution (SED). We allow for spatially correlated synchrotron and dust, parameterized by a single

[101] Delensing estimators that are technically more optimal have been introduced in Carron (2019) and Millea et al. (2020), and we are currently developing tools to test their feasibility in further iterations of our forecasting.

correlation parameter ϵ (Choi & Page 2015; Krachmalnicoff et al. 2018); the effective frequency scaling of this correlation depends on the relative strength of the two components. In addition, we also consider dust and synchrotron frequency decorrelation parameters, which allow their spatial pattern to change with frequency, suppressing the correlation of foreground signals between observing bands. A detailed description of the full parametric model is presented in Appendix A. The current model is easily extendable to accommodate additional complexities that have not yet been captured; alternatively, other models could be substituted here as well.

The model parameters are:

1. r, tensor-to-scalar ratio, at pivot scale $k_s = 0.05\,\mathrm{Mpc}^{-1}$;
2. and spectral index of the tensor modes $n_t = 0$;
3. A_L, residual lensing amplitude;
4. $A_{\rm dust}$, dust amplitude, in $\mu\mathrm{K}^2_{\rm CMB}$, at 353 GHz and $\ell = 80$;
5. β_d, dust spectral index;
6. T_d, dust temperature;
7. α_d, dust spatial spectral index;
8. Δ_d, dust frequency correlation, between 217 and 353 GHz, at $\ell = 80$;
9. $EE/BB_{\rm dust}$, power ratio for dust;
10. $A_{\rm sync}$, synchrotron amplitude, in $\mu\mathrm{K}^2_{\rm CMB}$, at 23 GHz and $\ell = 80$;
11. β_s, synchrotron spectral index;
12. α_s, synchrotron spatial spectral index;
13. Δ_s, synchrotron frequency correlation, between 23 and 33 GHz, at $\ell = 80$;
14. $EE/BB_{\rm sync}$, power ratio for synchrotron;
15. ϵ, dust/synchrotron spatial correlation.

For a given auto- or cross-spectrum, we step through the model components, combine the appropriate amplitude functions for the two bands contributing to the spectrum, and apply the bandpower window functions to obtain the binned expectation values. Finally, we sum over model components to find the total expectation value for that spectrum. In addition, because a Fisher forecast requires knowledge of the response of the model expectation values with respect to the model parameters, we also output the derivatives of the model expectation values.

The Fisher matrix that we consider is usually 10-dimensional. The parameters we constrain are r, $A_{\rm dust}$, β_d, α_d, Δ_d, $A_{\rm sync}$, β_s, α_s, Δ_s, and ϵ. We fix $T_d = 19.6\,\mathrm{K}$ because this parameter is mostly degenerate with $A_{\rm dust}$ for observations below 300 GHz, where the SED is in the Rayleigh–Jeans limit. The parameter A_L is assumed to be known, but its value is adjusted to represent varying levels of delensing, as discussed in Section 2.3.1. The EE/BB ratios are not relevant for calculations presented here, because we are focusing on constraints from the BB spectrum only; however, these parameters are left in for possible future forecasting. The fiducial model used for forecasting is centered at either $r = 0$ or 0.003, with $A_{\rm dust} = 4.25\ \mu\mathrm{K}^2_{\rm CMB}$ (best-fit value from BICEP2/Keck Array Collaborations VI 2016) and $A_{\rm sync} = 3.8\ \mu\mathrm{K}^2_{\rm CMB}$ (95% upper limit from BICEP2/Keck Array Collaborations VI 2016). In Sections 2–6, we assume no variation of these foreground amplitudes over the sky fractions relevant to this study, i.e., they are always pinned to the values listed above. We revisit this assumption in Section 7. The spatial and frequency spectral indices are centered at the preferred Planck and WMAP values (Fuskeland et al. 2014; Planck Collaboration XXII 2015): $\beta_d = 1.59$ (with Gaussian prior of width 0.11); $\beta_s = -3.10$ (with a Gaussian prior of width 0.30); $\alpha_d = -0.42$; $\alpha_s = -0.6$; and the dust/synchrotron correlation is centered at $\epsilon = 0$. Following Planck Collaboration (2017), the central dust correlation value is taken to be $\Delta_d = 0.97$ (3% decorrelation) and the synchrotron correlation value is assumed to be $\Delta_s = 1$ (no fiducial synchrotron decorrelation). Unless otherwise stated, the parameters have flat unbounded priors.

2.5. Bandpower Covariance Matrix Rescaling

2.5.1. Signal Scaling

The output model expectation values are also useful in the construction of our bandpower covariance matrix. To construct the BPCM components, we use lensed-ΛCDM + BICEP/Keck noise simulations. However, because we have the individual signal-only, noise-only, and signal × noise terms, we can record all the individual BPCM components:

sig = signal-only terms $\mathrm{Cov}(S_i \times S_j,\ S_k \times S_l)$;
noi = noise-only terms $\mathrm{Cov}(\mathcal{N}_i \times \mathcal{N}_j,\ \mathcal{N}_k \times \mathcal{N}_l)$;
sn1 = signal × noise terms $\mathrm{Cov}(S_i \times \mathcal{N}_j,\ S_k \times \mathcal{N}_l)$;
sn2 = signal × noise terms $\mathrm{Cov}(S_i \times \mathcal{N}_j,\ \mathcal{N}_k \times S_l)$;
sn3 = signal × noise terms $\mathrm{Cov}(\mathcal{N}_i \times S_j,\ S_k \times \mathcal{N}_l)$;
sn4 = signal × noise terms $\mathrm{Cov}(\mathcal{N}_i \times S_j,\ \mathcal{N}_k \times S_l)$.

Here, S are signal simulations, $\mathcal{N}$ are noise simulations, and the indices i, j, k, and l run over the experimental frequency channels.

While calculating the covariances from the signal and noise simulations, we also record the average signal bandpowers from the simulations. For a new signal sky 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 full BPCM construction and rescaling procedure is presented in Appendix B.

When we do this, we set to zero any term that has an expectation value of zero (under the assumption that signal and noise are uncorrelated, and different signals are uncorrelated) to reduce the Monte Carlo error in the resulting covariance matrix, given the relatively modest number of 499 realizations used. We also set to zero the covariance between bandpowers that are separated by more than one bin in ℓ, but importantly, preserve the covariance between the auto- and cross-spectra of the different frequency bands.

It is worth noting that this procedure allows us to have different numbers of degrees of freedom per bandpower for noise than for signal, which is a byproduct of signal and noise entering differently in a real analysis. This complication is often ignored in other forecasts by setting the noise and signal degrees of freedom to be identical.

The ability to estimate a BPCM for any model means that only a single set of simulations is necessary, and one does not have to run simulations for any and all conceivable scenarios. As already mentioned, in all the projections below we choose to fix $\Sigma(\theta) = \Sigma$, and hence we only apply the rescaling step once per fiducial model considered, i.e., we do not rescale our BPCM at every step along the way.

2.5.2. Noise Scaling

In addition to scaling from one signal model to another, recording all the covariance terms allows us to rescale the noise components as well. Given a data set for which we have simulations, the noise scaling can be performed in one of two ways. The first is to take a frequency present in the data set and scale down the noise in the BPCM by the desired amount. In particular, each BPCM component is scaled independently by $\sqrt{N_{\ell,\mathrm{S4}}/N_{\ell,\mathrm{BK}}}$ for each factor of $\mathcal{N}$ present. The second way is to add an additional frequency, for which we do not yet have data, by taking the covariance structure of an existing frequency, scaling down the noise by the amounts described above, and then expanding the BPCM by filling it in with the appropriate variance and covariance terms between the new band and all the existing ones. These tools facilitate the construction of a new data structure to explore any combination of frequency bands, with any sensitivity in each band.

To obtain N_ℓ values for a CMB-S4 channel by scaling the achieved N_ℓ values of a Stage-3 channel, we have to first scale by the ratio of their respective survey weights and then scale by the ratio of beam window functions, B_ℓ^2. The survey weight is defined as $w = 2f_{\mathrm{sky}}/\sigma^2_{\mathrm{map}}$, where f_{sky} is the effective sky area and σ_{map} is the Q/U map noise level. The input simulations use measured non-Gaussian B_ℓ shapes, but we rescale based on Gaussian approximations that are close to the true functions. We can write the noise spectrum of a CMB-S4 channel as

$$N_{\ell,\mathrm{S4}} = N_{\ell,\mathrm{BK}} \frac{w_{\mathrm{BK,achieved}}}{w_{\mathrm{S4,projected}}} \frac{B^2_{\ell,\mathrm{S4}}}{B^2_{\ell,\mathrm{BK}}}, \tag{5}$$

where $B^2_{\ell,\nu} = \exp \frac{-\ell(\ell+1)\Theta^2_\nu}{8\log(2)}$, Θ_ν is the FWHM, in radians, of the Gaussian beam, and $w_{\mathrm{i,achieved}}$ is the achieved integrated survey weight of a particular instrument.

To obtain the projected achieved survey weight for any of the CMB-S4 channels, we rescale the achieved survey weights as

$$\frac{w_{\mathrm{S4,projected}}}{w_{\mathrm{BK,achieved}}} = \frac{n_{\mathrm{S4}}^{\mathrm{det-yr}}}{n_{\mathrm{BK}}^{\mathrm{det-yr}}} \frac{NET^2_{\mathrm{BK,ideal}}}{NET^2_{\mathrm{S4,ideal}}}, \tag{6}$$

where $n_{\mathrm{S4}}^{\mathrm{det-yr}}$ is the number of detector-years assumed for CMB-S4 at any particular frequency, and $n_{\mathrm{BK}}^{\mathrm{det-yr}}$ is the number of detector-years in the BK15 data set, with the instruments in their final state (BICEP2/Keck Array Collaborations X 2018).

The implicit assumption in this step is that the performance of this new CMB-S4 frequency channel falls short of idealized performance by the same factor as the real map from which we are scaling. The survey weight scaling is always performed from the closest frequency for which we have available simulation inputs: 20–95 GHz are scaled from BICEP/Keck 95 GHz data; 145 and 155 GHz are scaled from BICEP/Keck 150 GHz data; and 220 and 270 GHz are scaled from BICEP/Keck 220 GHz data.

Since we are using BICEP/Keck products, which are calculated with a particular sky mask m (with $f_{\mathrm{sky}}^{\mathrm{BK}} = \frac{\Omega_{\mathrm{pix}}}{4\pi}\sum_i m_i \simeq 1\%$, where the sum is over the pixel i), we must also scale these products appropriately to evaluate the effect of different sky fractions. We propagate the effects of f_{sky} in the noise spectra and BPCM in two ways: first, we inflate the N_l values by a factor $\beta = f_{\mathrm{sky}}^{\mathrm{S4}}/f_{\mathrm{sky}}^{\mathrm{BK}}$, 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 of 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 and nonuniform coverage have on the covariance structure. We revisit the way this scaling is performed in Section 7, where we separately take into account the impact of realistic observing strategies on the various components.

With the N_ℓ scalings in hand, we can perform the aforementioned BPCM operations to arrive at a scaled CMB-S4 BPCM that encompasses the intricacies of realistic observing conditions.

3. Optimized Forecasting for r

In this section, we answer the following question: given a fixed amount of effort and the instrument specifications offered in the previous sections, what is the optimal distribution of effort for foreground cleaning and delensing such that the tightest constraint on r is achieved? To do this, we set up an optimization process that calculates the steepest descent through the ten-dimensional space (effort in the nine single-frequency low-resolution channels plus one high-resolution channel for delensing).

We operate in discrete units of effort, with a single unit defined to be equivalent to 500 detector-years at 150 GHz (similar to a BICEP/Keck 150 GHz receiver observing for 1 yr). For other channels, the number of detector-years per unit of effort is calculated as $n_{\mathrm{det},150} \times (\nu/150\,\mathrm{GHz})^2$. We define "effort" in these units because it is proportional to focal plane area, which is one of the strongest drivers of the overall project cost.

At each step of the algorithm, we allocate a unit of effort in each dimension. For each separate allocation, we rescale the BPCM, compute a new Fisher matrix, and calculate the resulting $\sigma(r)$. We then compare the constraints and permanently assign one half of the unit of effort to the channel that produces the largest improvement in $\sigma(r)$ and the other half to its atmospheric window counterpart (the groupings are 30/40, 85/95, 145/155, and 220/270 GHz). This last step, motivated by earlier iterations through the forecasting loop, is enforcing a split in order to reduce the foreground biases obtained under various foreground models. Projections run to a total of 6000 units of effort, which would be equivalent to 500,000 150 GHz detectors operating for six years. Stage-4 scale surveys seem likely to be in the range of (1–3) × 10^6 detector-years, an order of magnitude increase from Stage-3 experiments. Though it is generally prohibitive to calculate the entire 10-dimensional hypercube of $\sigma(r)$, we have validated our approach with a full-grid calculation at various points in the optimization.

Figure 4 shows the optimized constraints 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, for the no-detection scenario (i.e., $r = 0$).[102] To reach

[102] The correlation $\rho_{r\theta_j}$ between r and the foreground parameters, given by $\rho_{r\theta_j} = \mathrm{Cov}(r, \theta_j)/(\sigma(r)\sigma(\theta_j))$, is on the order of a few percent for the case with no decorrelation, and increases to a few tens of percent for the case which includes decorrelation. Given the chosen decorrelation parameterization and its strength, the increase in correlation values for the latter case is driven by the arising degeneracies between the dust decorrelation parameter and the rest of the dust parameters and r when $\Delta_{\mathrm{d}} < 1$. While the parameters are more correlated, we fully marginalize over them, resulting in a <40% increase in $\sigma(r)$.

the desired science goal of $\sigma(r) = 5 \times 10^{-4}$, for $f_{\rm sky} = 3\%$ (see discussion on sky fraction below), requires 1.2×10^6 150 GHz equivalent detector-years (or 1.8×10^6 when including marginalization over the decorrelation parameters).

Upon obtaining the optimized detector count distribution, we obtain the input noise spectra according to Equation (5). To use these spectra to create noise simulations (discussed in Section 4.1), it is useful to distill them to a few input parameters. To that end, we fit them to the formula

$$N_{\ell,\rm fit} = \frac{\ell(\ell+1)}{2\pi}\frac{\Omega_{\rm pix}}{B_\ell^2}\left(1 + \left(\frac{\ell}{\ell_{\rm knee}}\right)^{\gamma}\right)\sigma_{\rm map}^2 \quad (7)$$

and obtain the map depth $\sigma_{\rm map}$, slope γ, and $\ell_{\rm knee}$ values. For the small-aperture data, we find $\ell_{\rm knee} = 50$–60 with γ of -2 to -3, depending on the frequency. The optimal distribution of effort is presented in Figure 5, and the configuration that achieves the science goal in 4 yr (or 6 yr when marginalizing over decorrelation) is summarized in Table 1.

As mentioned, it is also necessary to optimize the sky fraction. The trade-off between raw sensitivity, ability to remove foregrounds, and ability to delens results is a complicated optimization problem with respect to sky coverage. Figure 6 shows the r sensitivity forecast for CMB-S4 as a function of the observed sky fraction for the case that we only have an upper limit ($r = 0$). In this case, the optimization prefers a deep survey that targets as small an area as possible. This conclusion, of course, depends on the forecasting assumptions; to that end, we would like to draw attention to several key factors.

First, holding the desired constraint on r fixed, the level to which we rely on delensing to decrease the sample variance increases at smaller sky fractions, as expected. For example, as shown in Figure 6, achieving the forecasted sensitivity on r for a survey targeting 1% of the sky will require a >80% reduction in the map rms level of the CMB lensing B-modes. While from a sensitivity standpoint it is possible to achieve these levels, the extent to which systematic effects and small-scale foregrounds will need to be constrained may become too stringent (Carron et al. 2017; Polarbear Collaboration et al. 2020).

Second, the current optimization assumes uniform foreground behavior across the sky (with amplitude equal to that in the BICEP2/Keck region), while in reality the average amplitude, and possibly the complexity of foregrounds, increase as larger sky areas are targeted. This effect would steepen the optimization curve at larger sky fractions and increase our preference for small amounts of sky.

Third, in the case of a detection, a practical consideration for the robustness of the final r result is its reproducibility across the sky. It is therefore useful to observe multiple 1% patches from which we can derive and compare separate cosmological constraints.

Finally, technical aspects of E/B separation of CMB maps may heavily disfavor patches smaller than about 1% of the sky due to cut-sky effects (Bunn et al. 2003; Smith & Zaldarriaga 2007; BICEP2/Keck Array Collaborations VII 2016).

Balancing the forecasting results with these concerns, we have chosen 3% as the default sky fraction for CMB-S4 r constraints (assuming a true value of $r = 0$). This choice was made for both the CMB-S4 Science Book (Abazajian et al. 2016) and CMB-S4 CDT Report (Lawrence et al. 2017) versions of the forecasts. As mentioned in Section 2.5.2, we revisit the issue of sky coverage in Section 7 with updated assumptions about our survey strategy and how we perform the $f_{\rm sky}$ scaling from the achieved products.

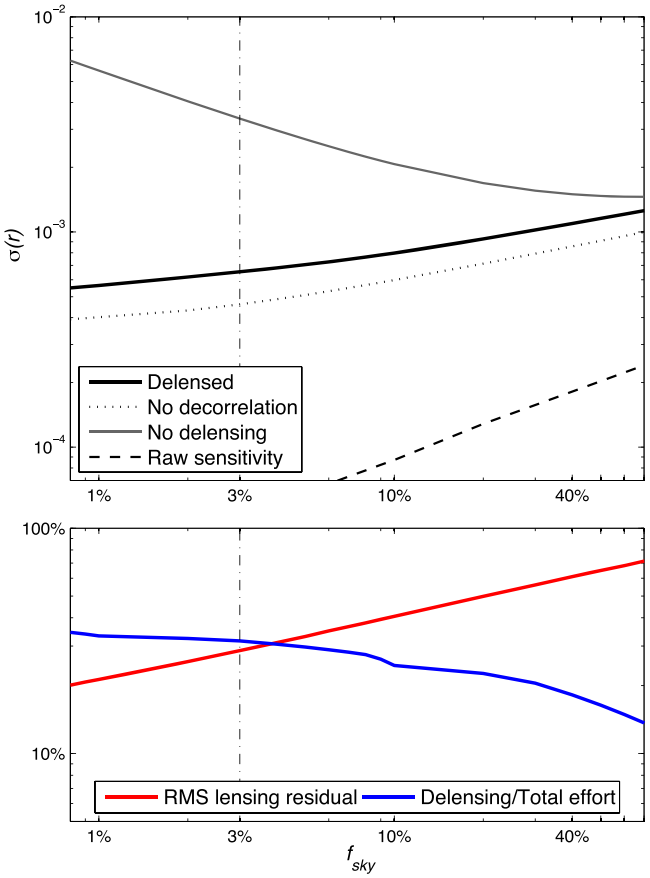


Figure 6. Top: optimized constraints on r as a function of sky fraction, for a fixed effort of 1.2×10^6 150 GHz equivalent detector-years. We include in solid black the case with delensing, allowing for decorrelation of the foregrounds, in solid gray the case without delensing, in dotted gray the case where no decorrelation is allowed in the model (with delensing), and in dashed black the raw sensitivity in the absence of foreground and lensing. Bottom: for the delensed case (with decorrelation), we show the fraction of effort spent on removing the lensing sample variance and the resulting rms lensing residual.

4. Map-based 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 cost of generating and analyzing them. This trade-off includes the choice of the 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 validate semianalytic results, 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 the parameter space for the PGW survey, including map-level noise simulations, sky models, and observation strategy. We also describe our approach to modeling instrumental systematics, the delensing survey, and the analysis methods.

In addition, we use these simulations to validate the spectral domain forecasts for configurations where the approaches are directly comparable, thereby closing the loop presented in Figure 1.

4.1. Map Noise Simulations

We use Equation (7) to obtain the desired noise prescription, for a fixed total effort, and then generate 499 Gaussian noise realizations at each band. As in the Science Book, we have mostly used a circular sky area of 3%. 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 centered at R.A. $= 0^\circ$, decl. $= -45^\circ$ (slightly below the BICEP/Keck patch) with full coverage out to radius $\theta < 12^\circ$ and "relative hits" tapering to zero with a cosine-squared shape for $12^\circ < \theta < 27^\circ$. 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. We also assume an $\ell_{\rm min} = 30$ cutoff, below which we do not recover any information.

For the CMB-S4 Reference Design Report (and Section 7), we include an explicit scan strategy on the sky and produce more realistic sky coverage patterns, but for the moment we regard the above as a reasonable compromise between idealism and reality.

4.2. Foreground Models

To make simulated sky maps, we add realizations of lensed CMB (Planck Collaboration et al. 2020d), both with and without a PGW component, to models of the Galactic foregrounds. So far, we have run simulations with seven foreground models, which we now describe.

1. 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, modified blackbody for dust).
2. The PySM[103] model `a1d1f1s1`, where the letters refer to anomalous microwave emission (AME), dust, free–free, and synchrotron, respectively, and the numbers are the base models described in Thorne et al. (2017). Free-free and AME are assumed to be unpolarized in this model, and thus do not affect the analysis in this paper.
3. The PySM model `a2d4f1s3`, where the models have been updated to variants that are also described in Thorne et al. (2017). Note that these include 2% polarized AME, a curvature of the synchrotron SED, and a two-temperature model for dust.
4. The PySM model `a2d7f1s3`, where the dust model has been updated to a more sophisticated physical characterization of dust grains as described in Hensley (2015). This model is interesting in that it does not necessarily conform to the modified blackbody SED.
5. The dust in Model 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 what is described in that paper to produce a modest amount of decorrelation of the dust emission pattern as a function of frequency motivated by the analysis of Planck data in Planck Collaboration (2017).
6. A toy model where the strong dust decorrelation suggested in Figure 3 of Planck Collaboration (2017) is taken at face value ($\Delta^{217\times353} = 0.85$, 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 (2018) and Planck Collaboration et al. (2020c) have reanalyzed the same Planck data, and while they find that the high level of decorrelation in this model is still consistent with the data, their best fit to that same data has no decorrelation.
7. 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 3 use the 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. Models 4 and 6 are not based on Planck data, but still contain a fixed signal realization. Models 0 and 5 have different seeds for each signal map and include the (Gaussian) sample variance. 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 the more complex models are not necessarily more accurate reflections of reality.

4.3. 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 furthermore, their impact on actual results comes not through the modeled effects but through unmodeled residuals. Many existing CMB experiments have published in-depth studies that use calibration data and simulations to set upper limits on a wide variety of effects, e.g., Keisler et al. (2015), BICEP2 Collaboration III (2015), Essinger-Hileman et al. (2016), Polarbear Collaboration et al. (2020), and BICEP2/Keck Array Collaborations XI (2019). There have also been several publications that examine the effect of specific classes of instrumental systematics on a generic polarization experiment, e.g., Hu et al. (2003), O'Dea et al. (2007), Shimon et al. (2008), Wallis et al. (2014), and Duivenvoorden et al. (2019).

For this study, in the absence of detailed instrument and survey designs, 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 reanalyzing them without knowledge of what was put in. We have experimented with components that are both correlated and uncorrelated across frequency bands, and that have white, $1/\ell$, and white $+1/\ell$ spectra, at varying levels compared to single-frequency map noise, or for correlated cases, combined

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

map noise. The leading-order effects of such components can be mitigated via explicit modeling or filtering, but they may still produce map-level residuals. Examples of mechanisms in this class include bandpass mismatches, beam and pointing variations, calibration variations, cross-talk effects, half-wave plate leakage, ground pickup, and readout irregularities.

To assess the impact of instrumental systematics on measurement requirements, for the purpose of 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 semianalytic forecasts and judge at what levels classes of systematics introduce parameter biases or additional uncertainties that are significant compared to the science targets for those parameters.

For the uncorrelated systematic, we add power independently to each frequency map, scaled by the noise power in that map. This systematic is "noise-like," and therefore the power in the map does not roll off at high multipoles due to beam smoothing. The expression for the auto-spectrum of the systematic is

$$C_l = \sigma^2_{\rm map}\left(A + B\left(\frac{l}{l_{\rm knee}}\right)^{\gamma}\right), \tag{8}$$

where $\sigma_{\rm map}$, $l_{\rm knee}$, and γ are same as found in Section 3; A and B are the amplitudes of the white and $1/\ell$ components, in units of fraction of noise power.

For the correlated systematic, we add a common signal at the same amplitude to all frequency maps. Since this signal is common, it affects both auto- and cross-spectra as

$$C_{l,ij} = \sigma^2_{\rm comb}\left(A + B\left(\frac{l}{l_{\rm knee}}\right)^{\gamma}\right)B_{l,i}B_{l,j}. \tag{9}$$

In place of the per-frequency noise power levels, we define the overall level of this systematic relative to $\sigma^2_{\rm comb}$, which is the inverse quadrature sum of the per-frequency noise power levels. We use a common $l_{\rm knee} = 50$ and $\gamma = -2$. Since this systematic is "CMB signal-like," we calculate the bias assuming that it has been smoothed in each map according to the beam size of that frequency.

Other classes of systematics could be simulated by manipulating the analysis procedure only. Examples of such effects include uncertainties in the bandpasses, polarization angles, calibration, and beam shapes. Such examples are not presented here and are left for future work.

4.4. Delensing

We have generated high-resolution simulated maps on which we intend to run explicit lensing reconstruction and then include that information in the analysis. While we are currently working on this analysis, this process has 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 2.3.1.

5. Simulation Analysis Methods

To make simulated maps, the noise realizations described in Section 4.1 are added to the sky models described in Section 4.2. For each realization, one then has a stack of multifrequency I, Q, and U maps containing nonuniform noise, foregrounds, and signal, and the challenge is to reanalyze them to recover the parameter of interest (in this case, r). This can be done by different teams using different methods, and could be carried out in a blind manner, although we have not done this yet.

So far, we have experimented with two methods. The first is a spectral internal linear combination (ILC) method (e.g., Tegmark et al. 2003), which determines the linear combination of multipole coefficients that minimizes the foreground and noise power without altering the CMB contribution. This method only relies on the frequency dependence of the CMB, and does not rely on assumptions about the spectral dependence of foreground components. Determination of cleaning coefficients in each bin ($\Delta\ell = 31$) leads to the smallest foreground residuals, but yields foreground residuals that are difficult to model. Relying on a single cleaning coefficient across all bins leads to foreground residuals that are easy to model but large. As a consequence, we use the same cleaning coefficients in three neighboring bins, which results in residuals that are acceptable and can still be marginalized in the power spectrum likelihood analysis. For the marginalization, since foreground residuals are typically dominated by dust, we assume an ℓ-dependence of the residuals across the three bins that share a common cleaning coefficient that is consistent with the ℓ-dependence observed by Planck for dust. For models that exhibit decorrelation between different dust components, this model is no longer correct. This could be improved with better understanding of foregrounds, but no attempts were made to do so in this analysis method.

The second method is an evolution of the parametric multicomponent fit to the ensemble of auto- and cross-spectra, as used for the BICEP/Keck analysis to date (BICEP2/Keck Array & Planck Collaborations 2015; BICEP2/Keck Array Collaborations VI 2016; BICEP2/Keck Array Collaborations X 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 ($A_{\rm dust}$ and $A_{\rm sync}$), a spatial spectral parameter ($\alpha_{\rm d}$ and $\alpha_{\rm s}$), and a frequency spectral parameter ($\beta_{\rm d}$ and $\beta_{\rm s}$). We also allow a dust/synchrotron correlation (ϵ), and decorrelation of the foreground patterns over frequency ($\Delta_{\rm d}$ and $\Delta_{\rm s}$). This model is equivalent to the one described in Section 2.4.

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 more complex behavior will likely be required, for example, modeling the frequency spectral indices individually in (large) pixels.

6. Map-based Results

Table 2 summarizes the results of the analysis for simulations of the optimized configuration obtained in Section 3 (1.2×10^6 150 GHz equivalent detector-years) and residual lensing power $A_{\rm L} = 0.1$.[104] The lensing residual is expected for iterative EB delensing according to Smith et al. (2012) for the sensitivity and angular resolution of the

[104] The distributions are close to normal, with no distribution showing significant skewness or kurtosis, and therefore well-characterized by the quoted mean and standard deviation values.

Table 2
Results of Two Analysis Methods Applied to Map-based Simulations Assuming the CMB-S4 CDT Report (Lawrence et al. 2017) Configuration and Our Suite of Sky Models (DC4)

		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	4.4	−0.2	4.4	0.2	5.7	0.3
	1	4.6	0.8	4.7	6.8	6.4	5.2
	2	4.7	0.7	4.8	3.8	6.5	1.9
	3	4.6	1.2	4.7	6.0	6.7	0.7
	4	6.5	4.8	7.9	43	8.3	−7.7
	5[a]	18	17	31	340	15	0.2
	6	4.8	−1.8	4.8	0.6	6.5	1.8
0.003	0	6.6	−0.7	6.2	0.3	8.1	0.4
	1	6.9	0.9	6.5	6.9	8.5	5.4
	2	6.5	−0.1	6.4	3.9	7.9	1.9
	3	7.0	1.4	6.6	6.7	8.7	0.9
	4	11	7.1	10	51	11	−6.2
	5[a]	23	17	34	350	17	0.4
	6	7.5	−0.2	7.1	1.4	8.6	2.5

Notes. All simulations assume an instrument configuration including a (high-resolution) 20 GHz channel, a survey of 3% of the sky with 1.2×10^6 150 GHz equivalent detector-years, and $A_L = 0.1$. (These analyses hold the A_L model parameter fixed to the simulated value. When freeing this parameter in the analysis, we recover similar $\sigma(r)$ values to within ∼1% for the case with decorrelation, and ∼9% for the case without; imposing a Gaussian prior of $\sigma(A_L)/A_L = 0.05$ reduces the differences to ∼0.2% and ∼2% respectively.)
[a] An extreme decorrelation model—see Section 4.2. In the right column, the parametric analysis includes a decorrelation parameter. No attempt is made in the ILC analysis to model the decorrelation. The middle columns shows the parametric analysis when we do not include decorrelation parameters.

Table 3
Results on Detection Significance for the CMB-S4 CDT Report (Lawrence et al. 2017) Configuration, 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 yr	6	1.0×10^{-3}	⋯	1.0×10^{-3}	⋯
0.003	4 yr	6	⋯	4.0	⋯	4.2
	8 yr	6	⋯	5.1	⋯	5.6

Note. For the $r = 0$ model, we report the 95% confidence level upper limit (CL UL).

delensing survey. The results from the parametric analysis naturally depend on whether a marginalization over decorrelation is performed, while the ILC analysis did not attempt to capture the effects of decorrelation on the recovery of r and $\sigma(r)$. This is evidenced by the 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 $\sigma(r) \approx 5 \times 10^{-4}$, exactly as expected from the semianalytic formalism. 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 below 1.0σ for all the models. These simulations are sets of 499 realizations, so the statistical uncertainty on the bias is approximately 0.04σ. However, the strong decorrelation in Model 5, as well as the high-significance detection of decorrelation in the parametric analysis of Model 4, do significantly increase $\sigma(r)$ and the level of bias. 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, a different re-optimization to concentrate the sensitivity at closer-in (less decorrelated) frequencies would be called for.

Table 3 shows results on detection significance for the CDT Report 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 after 4 yr of operation is $r = 0.004$.[105] For a tensor-to-scalar ratio of $r = 0.003$, the median detection significance after 4 yr is expected to be 4σ. If a detection were to be emerging at this point, extending the run time to 8 yr would be justified in order to reach a 5σ detection.

While $\sigma(r)$ can be precisely forecast for given assumptions, the true achieved detection level for r depends on the particular realization of the B-mode field in the observed patch of sky. Therefore, we can only forecast a distribution of detection levels. For a tensor-to-scalar ratio of $r = 0.003$ and 8 yr of observing, we expect to achieve more than a 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

[105] The reason this is larger than five times the quoted $\sigma(r)$ is sample variance.

Table 4
Map-based Simulation Results for Dedicated Simulations Containing Systematics (DC3)

	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	...	7.2	...
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	1.4
Correlated $1/\ell$	0	0	0	11	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	1.0
Both, white + $1/\ell$	0.8	1.7	1.5	2.6	5.6	1.1	7.9	0.98

Notes. Simulations here assume the Science Book Configuration (Abazajian et al. 2016), i.e., an instrument configuration including a (low-resolution) 20 GHz channel, a survey of 3% of the sky with 1.0×10^6 150 GHz equivalent detector-years, and $A_L = 0.1$. We report sky Model 3 and $r = 0$ (no decorrelation), with additive systematic effects in varying combinations, the amplitudes of which are specified as percentages of survey noise, for the white (A) and $1/\ell$ (B) components. The r bias columns list the bias due solely to systematic effects, i.e., the shift relative to the "None" case.

foregrounds are substantially different than any of these cases, then the biases could be larger. To obtain 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 or marginalized over, but simply absorbed into the estimated B-mode power spectrum. Doing so with dedicated simulations based on sky Model 6 increases the magnitude of the bias on r to 4.1×10^{-4}. The dominant contribution to the bias comes from synchrotron residuals, which motivated placing a lower-frequency 20 GHz channel on the LAT (see footnote 96), reducing the magnitude of the bias to 1.8×10^{-4}, as shown in Table 2.

Table 4 summarizes the results of the analysis of simulations including additive systematic effects on top of foreground Model 3 (note: these simulations correspond to DC3). Different combinations of uncorrelated and correlated contamination with varying spectra are considered. The levels of systematic contamination for these simulations were chosen to predict biases on r of 1×10^{-4} in semianalytic forecasts. We can see that the different combinations explored increase the bias on r by amounts that typically vary from 0.5 to 1.5×10^{-4} for the two separate analyses, over the different cases. We find that, to restrict the 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/Keck Array Collaborations VI 2016). Assuming that 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.

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 parameterized models of the frequency spectra of different foregrounds, shows biases on r at the 1×10^{-4} level for uncorrelated random deviations in band-center determination of 0.8%, or for correlated deviations of 2%; we adopt these as reasonable benchmark requirements to accommodate a variety of both blind and astrophysical foreground modeling approaches.

7. Modifications Leading to the Reference Design

In this section, we describe updates to the framework and the reference design that attempt to take into account the impact of realistic observing strategies, realistic focal-plane layouts, and mapping onto optics tubes, as well as a more conservative approach to our delensing forecasts. We also seek to answer the question of siting, with the South Pole and Chile being the two choices considered. The work in this section has led to the forecasts and plans presented in the most recent CMB-S4 document—the CMB-S4 Reference Design Report (Abazajian et al. 2019). While this section is self-contained, we direct the reader to the aforementioned source for more details.

In previous versions of our forecasting, we have had the ability to choose the number of detectors in each frequency band in a continuously variable manner, as shown in Figure 5. For the reference design, a mapping of detectors into dichroic optics tubes has been carried out, while seeking to maintain the band distribution as determined in the optimization calculations. In scaling the achieved performance from the existing monochromatic instruments to dichroic detectors and optics, no degradation of optical performance has been assumed at this stage, but this assumption should be verified with upcoming data. This results in the configuration described in Abazajian et al. (2019) with 18 SAT tubes, observing for 7 yr. We use this configuration to scale the BICEP/Keck noise bandpower covariance matrix in the same way as described in Section 2.5. A further rescaling is applied to account for sky coverage, as explained below.

7.1. Sky Coverage Effects

The semianalytic calculations of Section 2.5 assumed a simplified rescaling for sky area, while the map-based simulations of Section 6 assumed an idealized circular sky patch, which is not actually achievable with a practical instrument from a site at any latitude. Figure 7 compares our prior assumptions to more realistic hit patterns.

From the South Pole, it is possible to concentrate the coverage onto a compact region of sky, but from Chile, the region that can be observed is affected by Earth's rotation, resulting in more extended coverage. The large instantaneous

field of view of the SAT telescopes means that there is a minimum field size that can be achieved, and that there is always a strong edge taper in the coverage pattern.

We have performed a calculation that attempts to optimize simulated SAT observations from Chile to produce the densest possible coverage on a 3% patch of low-foreground sky, resulting in the overall pattern shown in Figure 7 as "Chile full." We segment this into its deepest part, which we call "Chile deep," and the remainder, which we call "Chile shallow."

From the South Pole, one can scan the same patch at all times of the day and year at the same observing elevation, with the size of the observed patch controlled by the length of the scan throw in R.A. A minimal-length scan results in the pattern shown in the figure as "Pole deep." Lengthening the scan while remaining in the low foreground sky results in the pattern "Pole wide." In the results below, "Pole deep" and "Pole wide" are therefore "either/or" options.

Because the noise increases in regions with less observing time, the effective sky area for noise is larger than the effective sky area for signal—and both of these also depend on the weighting applied when analyzing the maps. The patterns shown in Figure 7 have the effective sky fractions reported in Table 5, assuming inverse noise-variance weighting.

We can take into account the above effects by rescaling the BICEP/Keck BPCM in a more sophisticated manner. First, we need to scale the noise due to distributing the effort on a patch of sky larger than the original BICEP/Keck one. The noise is scaled by the effective noise factor

$$f_{\rm eff}^{\rm noise} = \frac{\Omega_{\rm pix}}{4\pi}\frac{\sum_i w_i^2 h_i^{-1}\sum_i h_i}{\sum_i w_i^2}, \tag{10}$$

where $\Omega_{\rm pix}$ is the solid angle of a single pixel, w_i are the weights for pixel i, and h_i are the hit counts.

Second, since we observe a different number of modes, we need to scale the signal, noise, and signal-cross-noise contributions of the BPCM by the factors

$$f_{\rm sky}^{\rm noise} = \frac{\Omega_{\rm pix}}{4\pi}\frac{\left(\sum_i w_i^2 h_i^{-1}\right)^2}{\sum_i w_i^4 h_i^{-2}}, \tag{11}$$

$$f_{\rm sky}^{\rm signal} = \frac{\Omega_{\rm pix}}{4\pi}\frac{\left(\sum_i w_i^2\right)^2}{\sum_i w_i^4}, \tag{12}$$

$$f_{\rm sky}^{\rm cross} = \frac{\Omega_{\rm pix}}{4\pi}\frac{\sum_i w_i^2 \sum_i w_i^2 h_i^{-1}}{\sum_i w_i^4 h_i^{-1}}. \tag{13}$$

We also need to take out the effect of these factors from the original BPCM. In the BICEP/Keck analysis, the weights are the inverse noise variance, i.e., $w_i = h_i$. In the case of CMB-S4, we will never be noise-dominated, either due to an actual primordial signal, or due to the lensing and foreground residuals, so here we use the inverse variance of the total signal and noise to determine the optimal weighting.

Last, while the scanning strategy used here has been optimized to mostly observe the cleanest available sky, some regions with high Galactic emission are also observed. Realistically, we would mask such regions when analyzing the maps. To assess this effect, we used masks based on a smoothed Planck 353 GHz polarized intensity map, keeping the cleanest 30% or 60% of the full sky (28% and 58% after

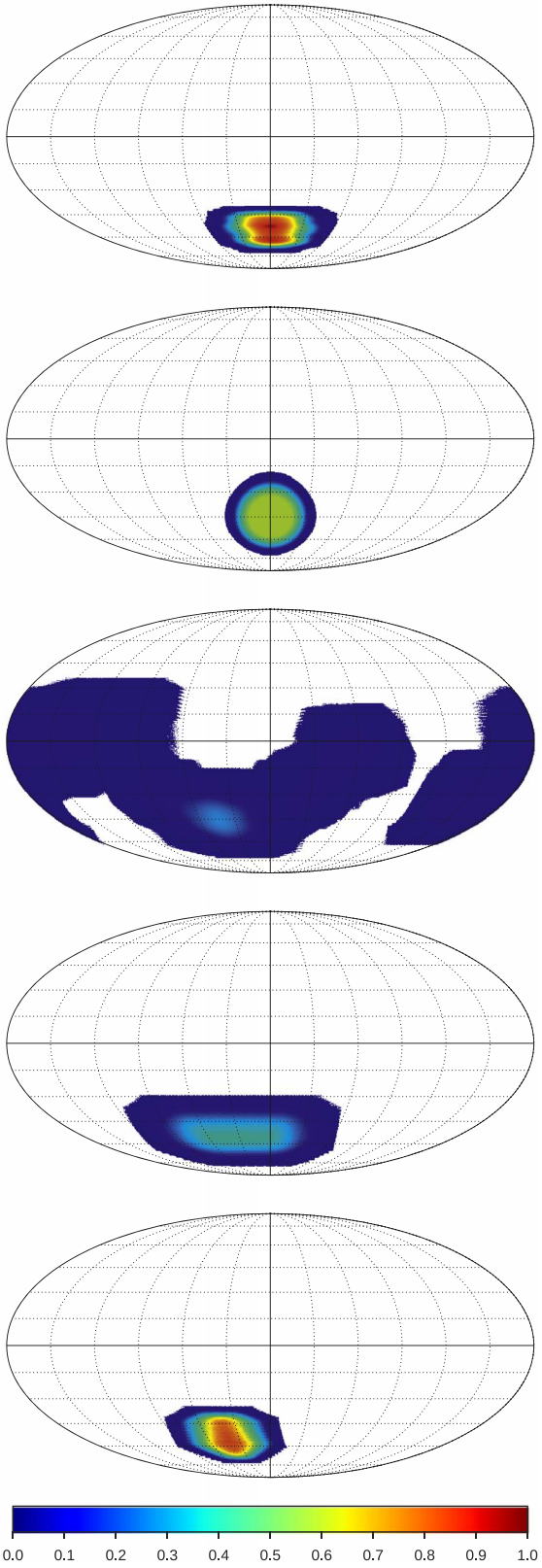


Figure 7. Hit patterns on the sky for small-aperture telescope surveys. Top panel: the actual BICEP3 2017 hit pattern (peak normalized). Second panel: idealized circular pattern as used in Section 4. Third panel: simulated "Chile full" pattern, Fourth panel: simulated "Pole wide" pattern. Fifth panel: simulated "Pole deep" pattern. Each pattern is normalized to the same hit sum as in the top panel, and the color scales are the same. The "Chile deep" and "Chile shallow" regions referred to in the text are subregions of the "Chile full" pattern.

Table 5
Effective Sky Fractions for Signal, Noise, and Signal-cross-noise, as Percentages, for the Observation Patterns Shown in Figure 7, and the Case of Inverse Noise-variance Weighting (i.e., Equations (11)–(13) with $w_i = h_i$)

	SP Deep	SP Wide	CH Deep	CH Shallow	CH Full
$f_{\rm sky}^{\rm sig}$	1.9	4.3	2.4	10	5.9
$f_{\rm sky}^{\rm noi}$	2.9	6.5	3.4	20	18
$f_{\rm sky}^{\rm cross}$	2.5	5.5	3.0	16	12

Notes. Here, "SP" is the South Pole and "CH" is Chile. The CH Shallow numbers appear larger than CH Full due to the effects of the weighting. For comparison, the BICEP2/Keck and BICEP3 values are (1.0, 1.3, and 1.1) and (1.9, 2.7, and 2.3) for signal, noise, and signal-cross-noise respectively.

apodization). We then used these masks to disregard some of the pixels, resulting in a degradation of the constraints on r.

7.2. Delensing Revised

The CMB-S4 PGW science goal can only be achieved if the majority of the lensing B modes can be removed. The optimization in Section 3 assumed a single frequency channel assigned to the higher-resolution delensing observations. The strength of polarized foregrounds at small angular scales is currently poorly constrained by data, and though studies related to the role polarized small-scale galactic foreground residuals play in the delensing process have been performed (e.g., Fabbian et al. 2019; Beck et al. 2020), a complete understanding of their impact is still missing.[106] Separately, while small-scale extragalactic foregrounds limit the temperature reconstruction of the lensing potential, experiments such as CMB-S4 will be dominated by polarization reconstruction, for which these are much less pertinent. To the extent that temperature reconstruction will contribute nontrivial weight to the analysis, Sailer et al. (2020) shows that robust foreground-hardened estimators can be constructed. Furthermore, analyses of delensing with external tracers such as the CIB, for which such foreground-induced biases become yet more important, have also shown that they can be effectively mitigated with multifrequency cleaning techniques (Baleato Lizancos et al. 2021). Balancing these considerations, the high-resolution instrument reference design, therefore, includes some additional coverage at higher and lower frequencies.

To forecast the delensing performance, we proceed in two steps. For a given LAT configuration and sky coverage, we derive the noise levels for an ILC that minimizes the variance of components with a frequency dependence that differs from that of a blackbody (Tegmark & Efstathiou 1996). In this step, we assume that polarized foreground emission is dominated by Galactic synchrotron and thermal dust emission.[107] Using the ILC noise power spectrum, we then forecast the performance expected for iterative EB delensing (Smith et al. 2012).

The LAT reference design, established independently from this work, and described in detail in Section 2.2.2 and Appendix A of Abazajian et al. (2019), includes two LATs in Chile for the wide-field survey targeting science goals other than PGWs, and one LAT at the South Pole for the PGW delensing survey. For the Chile LATs, and a wide-area survey covering 70% of the sky, the two-step procedure predicts that 73% of the lensing power can be removed in the "Chile shallow" region after 7 yr of observation. Similarly, for the single LAT at the South Pole dedicated to delensing of the approximately 3% "Chile deep" and "Pole deep" regions, we expect to be able to remove close to 90% of the lensing power after 7 yr of observation.

The numbers given above assume identical hits maps for the South Pole LATs and SATs; they also assume inverse noise-variance weighting rather than a weighting scheme that accounts for both signal and noise. For the lensing residuals achieved by the reference design, inverse noise-variance weighting for the SAT maps is suboptimal because the signal (e.g., for $r = 0$ lensing residual after foreground removal) is measured with signal-to-noise ratio above unity over a region that extends beyond the region that dominates the analysis in inverse noise-variance weighting. As a consequence, in all the forecasts presented below, we employ weights that account for both signal and noise, which significantly increases the number of modes that contribute to the analysis, especially for small, deep patches. Even though, for the same survey, this leads to slightly higher noise and lensing residuals, the increase in the number of modes overall reduces the lensing sample variance contribution to $\sigma(r)$. In practice, we determine the optimal weights iteratively, accounting for the effect on the SAT and LAT analyses, and find rapid convergence.

7.3. Results

The covariance matrices calculated as described above are used to produce the results given in this section, where the number, siting, and coverage patterns of the SATs are varied. In all cases, a delensing LAT at the South Pole is assumed to concentrate its coverage on a small patch of sky, while delensing over larger sky areas is assumed to be available from the Chilean LATs.

As mentioned earlier, we split the Chilean coverage shown as "Chile full" in Figure 7 into a deep patch, "Chile deep," which overlaps with the "Pole deep" region, and call the remainder "Chile shallow." We then make separate forecasts for each Chilean subregion using the appropriate delensing level for that subregion. To obtain forecasts for the full region, we add the $\sigma(r)$ results in simple inverse quadrature, thereby making the approximation of independence of the measured modes. When we combine these with South Pole observations, we mimic a joint analysis over the overlapping region by taking the sum of the "Pole deep" and "Chile deep" coverage maps and computing the corresponding weights and lensing residuals, and add the "Chile shallow" results in inverse quadrature.

The foreground fiducial model is kept the same as in Section 2.4, and we also explore the possibility of unmodeled foreground residuals contributing residual power to the cleaned maps, assuming a foreground cleaning to 1% at 95 GHz. We do so by adding in quadrature to the r forecasts a foreground bias equal to 1% of the equivalent r of the foreground minimum of the BK15 data at 95 GHz and $\ell = 80$.

Since some parts of the "Chile shallow" coverage lie closer to the Galactic plane, we boost the foreground level in the shallow region, and the foreground bias if it is included, by a factor of 3 with respect to the deep patch. This scaling is based

[106] Of note, Beck et al. (2020) shows that delensing biases due to galactic foregrounds can be mitigated with certain analysis choices.

[107] In the ILC, power from unresolved/unmasked polarized extragalactic point sources is not currently accounted for. With reasonable masking thresholds, we expect the contribution to the ILC residuals from polarized sources to be subdominant compared to instrument noise (Gupta et al. 2019).

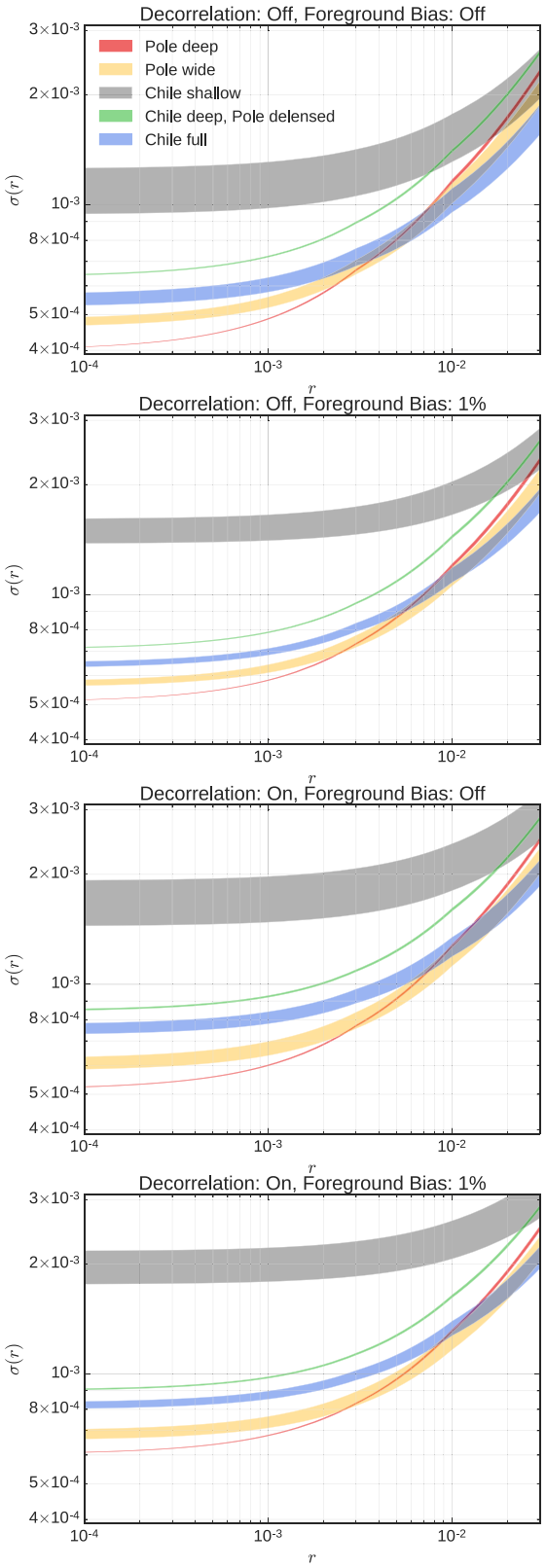


Figure 8. Constraints on r as a function of the value of r. We assume an instrument with 18 SAT optics tubes and an observation time of 7 yr, with the five masks defined in Section 7. Each band shows different Galactic cuts, based on *Planck* polarized foregrounds: the upper edge uses the cleanest 28% of the full sky, whereas the lower edge uses the 58% cleanest. We explore the effect of turning on foreground decorrelation in the forecasting. We also explore adding a foreground bias, in quadrature, with a 1% value of the equivalent r of the current foreground minimum of the BK15 data at $\ell = 80$.

Table 6
Combined $10^4 \times \sigma(r)$ Values (*Smaller* Numbers are Better), Assuming $r = 0$ after 7 yr of Observation, Keeping only the 28% Cleanest Part of the Sky, Assuming no Decorrelation and an Observing Efficiency in Chile the Same as at the South Pole

Chile\Pole	0	6	9	12	18	30
0		6.3	5.0	4.5	**4.0**	3.5
6	12	5.5	4.7	**4.3**	3.9	3.5
9	8.7	5.1	**4.5**	4.1	3.8	3.4
12	7.1	**4.8**	4.3	4.0	3.7	3.4
18	**5.7**	4.4	4.1	3.8	3.6	3.3
30	4.4	3.9	3.7	3.6	3.4	3.2

Note. The bolded cells conserve the 18 SAT optics tube count of the reference design, while nonbolded cells explore other counts at the South Pole and in Chile.

on the dust amplitude measured in the map-based simulations for surveys observed from the South Pole or from Chile, using a map with spatially varying foregrounds based on Planck data.

In Figure 8, we show the dependence of $\sigma(r)$ on r for the different coverage masks. We calculate these constraints for $r = 0$, 0.003, 0.01, and 0.03, for the different Galactic cuts, and show the linear interpolation between these points, sampled on a high-resolution linearly spaced grid. We find that the survey strategy from the South Pole is always favored in the limit of small r, with the crossover point depending on the specific assumptions. This comes as a direct result of the fact that, while a larger fraction of the sky can be observed from Chile, one can concentrate the available sensitivity more deeply from the South Pole (see Figure 7). We note that the delensing requirements are more stringent for the smaller surveys (such as the South Pole one), highlighting the importance of periodically revisiting the assumptions made for the delensing survey of CMB-S4 as new data from Stage-3 experiments become available.

In Tables 6–7, we present a set of $\sigma(r)$ results for seven years of observations and $r = 0$, and in Tables 8–9, we present a set of significance of detection levels for $r = 0.003$, while varying the number of optics tubes at the South Pole and Chile over a wide range of options. We show results for two different variants: (i) with no marginalization over the decorrelation parameters and for the 28% cleanest polarized sky; and (ii) the same as (i), but marginalizing over the foreground decorrelation parameters.

In all these calculations, we assume that the observing efficiency from Chile is 100% equivalent to that from the South Pole. Despite long histories of CMB observations at both sites, it is still quite challenging to make a clean comparison of their observing efficiencies. Results from the BICEP/Keck program are responsible for the leading constraints on r for the last decade, but it is still not currently possible to disentangle the role of the observing site from other factors that have contributed to the success of that program, such as detector performance, instrument design, observing strategy, and operations management. However, note also that even assuming equal observing efficiency at both sites (Table 6), there is still a difference between $\sigma(r)$ obtained for equivalent numbers of optics tubes in Chile versus the South Pole, due to sky fraction, foreground complexity, and delensing. In the CMB-S4 Reference Design Report, we also show forecasting results that

Table 7
Same as Table 6, but Assuming Additional Foreground Decorrelation Parameters

Chile\Pole	0	6	9	12	18	30
0		8.4	6.7	6.0	**5.2**	4.4
6	16	7.3	6.2	**5.6**	5.0	4.3
9	12	6.8	**5.9**	5.4	4.9	4.3
12	9.7	**6.4**	5.7	5.3	4.8	4.2
18	**7.8**	5.8	5.3	5.0	4.6	4.1
30	6.0	5.1	4.8	4.6	4.3	4.0

Table 8
Combined Detection Significance (*Larger* Numbers are Better) for $r = 0.003$ after 7 yr of Observation, Keeping only the 28% Cleanest Part of the Sky, Assuming no Decorrelation and an Observing Efficiency in Chile the Same as at the South Pole

Chile\Pole	0	6	9	12	18	30
0		3.7	4.5	4.9	**5.4**	6.0
6	2.2	4.3	4.8	**5.2**	5.6	6.1
9	3.0	4.6	**5.1**	5.4	5.8	6.3
12	3.6	**4.9**	5.3	5.6	5.9	6.4
18	**4.4**	5.3	5.7	5.9	6.2	6.6
30	5.4	6.0	6.2	6.4	6.6	6.9

Note. The bolded cells conserve the 18 SAT optics tube count of the reference design, while nonbolded cells explore other counts at the South Pole and in Chile.

Table 9
Same as Table 8, but Assuming Additional Foreground Decorrelation Parameters

Chile\Pole	0	6	9	12	18	30
0		3.0	3.6	4.0	**4.5**	5.1
6	1.7	3.5	3.9	**4.3**	4.7	5.2
9	2.3	3.7	**4.1**	4.4	4.8	5.3
12	2.8	**3.9**	4.3	4.6	4.9	5.4
18	**3.4**	4.3	4.6	4.8	5.1	5.6
30	4.2	4.8	5.0	5.2	5.5	5.8

assume an observing efficiency from Chile equivalent to 50% of the South Pole efficiency.

Note that the results where the total number of SAT optics tubes is different from the assumed number of 18 are subject to a caveat: the delensing effort is assumed to be held fixed to the reference design—one LAT at the South Pole and two in Chile. In principle, as the total effort is varied away from the reference design, one should re-optimize the fraction of delensing effort as per Figure 4.

8. Conclusions

In this paper, we have presented the *CMB-S4 r forecasting framework* and its evolution through three major iterations. To ensure forecasting realism, we have chosen to directly ground our forecasting in the end-to-end on-sky achieved performance of Stage 2–3 experiments, rather than rely on *ab initio* assumptions. Additionally, we have checked the semianalytic calculations with map-based simulations of increasing complexity, thereby creating a closed forecasting loop (presented in Figure 1). This approach allows us to confidently incorporate all the effects that impact current real-world experiments, to flexibly trace the end-to-end effects of changes to experimental design or assumptions, and to iterate over the framework as more sophisticated sky and instrument models become available.

The semianalytic calculations described in Sections 2 and 3 indicate that, for a 3% sky fraction, 1.2×10^6 150 GHz equivalent detector-years (or 1.8×10^6 when including marginalization over the decorrelation parameters) are required to reach the science requirement of $\sigma(r) = 5 \times 10^{-4}$, with 30% of this effort assigned to the delensing survey. The resulting optimal distribution of detectors across frequency bands is summarized in Table 3.

The map-based simulations described in Sections 4–6 confirm the $\sigma(r)$ results from the semianalytic calculations. These simulations also indicate that bias in the recovered r value is within 1σ for a suite of different foreground models. However, we note that foregrounds remain a serious issue that must be periodically revisited as the project progresses.

Additional map-based simulations indicate that the systematic bias on r can be controlled to $<1\sigma$, provided that fractional contamination levels similar to those already achieved by small-aperture telescopes can be maintained.

Mapping the requirements from the semianalytic calculations onto realizable instruments (see Section 7) results in the reference design described in Abazajian et al. (2019) for a seven-year survey period, and in the constraints presented in Tables 6–9.

In Section 7, we have also extended the semianalytic calculations to account for realistic observation patterns and probed the dependence of $\sigma(r)$ on r, as a function of experiment siting, finding that the survey strategy from the South Pole is always favored in the limit of small r. We note that surveys with smaller footprints (such as the South Pole survey) depend more tightly on the levels of achieved delensing. Therefore, revising the delensing assumptions will be important as new studies of small-scale foregrounds become available.

Going forward, as the CMB-S4 project matures, the collaboration will need to converge on increasingly specific instrumentation choices. To quantitatively discern between the different options, we intend to continue using and developing the closed-loop framework presented here. Among others, we anticipate adding complexity to the following directions: delensing treatment; foreground simulations; foreground parameterization; survey strategies; and instrumental systematics. Additionally, we expect to apply new analysis methods to our data challenges and incorporate new achieved performance levels at multiple additional frequencies from multiple sites as these data become available. The iterative nature of our framework can easily accommodate these revisions, achieving increased realism with each iteration.

The CMB-S4 collaboration (https://cmb-s4.org/) is working to plan, construct, and operate a next-generation, multisite CMB experiment in the 2020s. The collaboration is led by an elected Governing Board, Spokespeople, Committee Chairs, and Executive Team. Funding for the CMB-S4 Integrated Project Office is provided by the Department of Energy's Office of Science (project level CD-0) and by the National

Science Foundation through the Mid-Scale Research Infrastructure-R1 award OPP-1935892. This research used resources of Argonne National Laboratory, a U.S. Department of Energy (DOE) Office of Science User Facility operated under Contract No. DE-AC02-06CH11357. This document was prepared by the CMB-S4 collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. Work at Lawrence Berkeley National Laboratory was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Work at SLAC National Accelerator Laboratory is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. 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. In the United States, work on CMB-S4 by individual investigators has been supported by the National Science Foundation (awards 1248097, 1255358, 1815887, 1835865, 1852617, 2009469), the Department of Energy (awards DE-SC0009919, DE-SC0009946, DE-SC0010129, DE-SC0011784), and the National Aeronautics and Space Administration (award ATP-80NSSC20K0518). In Australia, the Melbourne authors acknowledge support from an Australian Research Council Future Fellowship (FT150100074). In Canada, R.H. is supported by the Discovery Grants program from NSERC, and acknowledges funding from CIFAR, the Sloan Foundation, and the Dunlap family. In Italy, C.B. acknowledges support under the ASI COSMOS and INFN INDARK programs. In the Netherlands, D.M. acknowledges NWO VIDI award number 639.042.730. In Switzerland, J.C. is supported by an SNSF Eccellenza Professorial Fellowship (No. 186879). In the United Kingdom, A.L., G.F., and J.C. are supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant Agreement No. [616170]. A.L. also acknowledges STFC award ST/P000525/1. S.M. is supported by the research program Innovational Research Incentives Scheme (Vernieuwingsimpuls), which is financed by the Netherlands Organization for Scientific Research through the NWO VIDI grant No. 639.042.612-Nissanke and the Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER, received financial state aid managed by the Agence Nationale de la Recherche, as part of the program Investissements d'avenir under the reference ANR-11-IDEX-0004-02. Some computations in this paper were run on the Odyssey cluster, supported by the FAS Science Division Research Computing Group at Harvard University.

Appendix A
Multicomponent Theory Model

The forecasting framework uses a parametric model to describe the bandpower expectation values as a combination of cosmological and foreground signals. This parameterization follows the one presented in Appendix G of BICEP2/Keck Array Collaborations X (2018).

The model includes signals from the CMB (lensed-scalar and tensor contributions), Galactic dust, and Galactic synchrotron, with the possibility of spatial correlation between dust and synchrotron. Contributions from dust and synchrotron to the *BB* spectrum between maps at frequencies ν_1 and ν_2 can be written in the following form, after computing the expected spectrum as a function of ℓ, and bandpower window functions are applied to calculate binned bandpower expectation values:

$$\begin{aligned}\mathcal{D}_{\ell,BB}^{\nu_1\times\nu_2} &= A_{\rm dust}\Delta'_{\rm d} f_{\rm d}^{\nu_1} f_{\rm d}^{\nu_2}\left(\frac{\ell}{80}\right)^{\alpha_{\rm d}} \\ &+ A_{\rm sync}\Delta'_{\rm s} f_{\rm s}^{\nu_1} f_{\rm s}^{\nu_2}\left(\frac{\ell}{80}\right)^{\alpha_{\rm s}} \\ &+ \epsilon\sqrt{A_{\rm dust}A_{\rm sync}}\,(f_{\rm d}^{\nu_1} f_{\rm s}^{\nu_2} + f_{\rm s}^{\nu_1} f_{\rm d}^{\nu_2})\left(\frac{\ell}{80}\right)^{(\alpha_{\rm d}+\alpha_{\rm s})/2}.\end{aligned} \tag{A1}$$

Parameters $A_{\rm dust}$ and $A_{\rm sync}$ specify the dust and synchrotron power in units of $\mu{\rm K}^2_{\rm CMB}$ at angular scale $\ell = 80$. These are defined at pivot frequencies of 353 GHz for dust and 23 GHz for synchrotron. The dust and synchrotron components scale as power laws in ℓ with slopes $\alpha_{\rm d}$ and $\alpha_{\rm s}$, respectively (note that slope parameters are defined for $\mathcal{D}_\ell \equiv \ell(\ell+1)C_\ell/2\pi$). The parameter ϵ specifies the level of spatial correlation between dust and synchrotron; this correlation coefficient is assumed to be constant across all ℓ. If either $A_{\rm dust}$ or $A_{\rm sync}$ are negative, the contribution of the correlated component to the expectation value flips sign. Negative foreground amplitudes are technically nonphysical, but this analytic continuation becomes important when we explore entire parameter phase spaces—for instance, in the simulation reanalysis steps.

Coefficients $f_{\rm d}^{\nu}$ and $f_{\rm s}^{\nu}$, given by Equations (A5) and (A6), describe the scaling of dust and synchrotron amplitude from the pivot frequencies to the actual bandpasses of the maps at frequencies ν_1 and ν_2. The SED model used for dust emission is a modified blackbody, i.e., a Planck function multiplied by a power law with emissivity spectral index $\beta_{\rm d}$ (Planck Collaboration XXII 2015). We adopt a dust temperature $T_{\rm d} = 19.6$ K. The SED model used for synchrotron is defined as a power law with spectral index $\beta_{\rm s}$ relative to a Rayleigh–Jeans spectrum. The frequency scaling also includes the unit conversion for $\mu{\rm K}_{\rm CMB}$ from the pivot frequency to the target bandpass.

To integrate the SED and unit conversion over the bandpass of the target map, we adopt the convention used by Planck Collaboration IX (2014), in which our bandpass functions describe response as a function of frequency to a beam-filling source with uniform spectral radiance. For emissivity spectral index $\beta_{\rm d}$ and dust temperature $T_{\rm d}$, the modified blackbody scaling (MBBS) from pivot frequency $\nu_{\rm pivot}$ to a map with bandpass $R(\nu)$ is given by

$$f_{\rm MBBS}^{\nu} = \frac{\int d\nu R(\nu)\nu^{3+\beta_{\rm d}}\left(\exp\frac{h\nu}{kT_{\rm d}} - 1\right)^{-1}}{\nu_{\rm pivot}^{3+\beta_{\rm d}}\left(\exp\frac{h\nu_{\rm pivot}}{kT_{\rm d}} - 1\right)^{-1}}. \tag{A2}$$

For a synchrotron power-law scaling (PLS) with spectral index β_s, we calculate the coefficient in a similar way:

$$f^{\nu}_{\rm PLS} = \frac{\int d\nu R(\nu)\nu^{2+\beta_s}}{\nu^{2+\beta_s}_{\rm pivot}}. \quad \text{(A3)}$$

The conversion between $\mu K_{\rm CMB}$ units at the pivot frequency and $\mu K_{\rm CMB}$ units at the target map bandpass is given by

$$f^{\nu}_{T_{\rm CMB}} = \frac{\nu^4_{\rm pivot}\exp\frac{h\nu_{\rm pivot}}{kT_{\rm CMB}}\left(\exp\frac{h\nu_{\rm pivot}}{kT_{\rm CMB}}-1\right)^{-2}}{\int d\nu R(\nu)\nu^4\exp\frac{h\nu}{kT_{\rm CMB}}\left(\exp\frac{h\nu}{kT_{\rm CMB}}-1\right)^{-2}}. \quad \text{(A4)}$$

Combining these factors, we obtain the scalings used in Equation (A1):

$$f^{\nu}_{\rm d} = f^{\nu}_{\rm MBBS} \times f^{\nu}_{T_{\rm CMB}}, \quad \text{(A5)}$$

$$f^{\nu}_{\rm s} = f^{\nu}_{\rm PLS} \times f^{\nu}_{T_{\rm CMB}}. \quad \text{(A6)}$$

We also consider dust and synchrotron frequency decorrelation. The simplest possible model of a polarized foreground component is one with a fixed spatial pattern on the sky that scales with frequency according to a single SED. In this case, the expectation value of the cross-spectrum between any two frequencies is the geometric mean of the respective auto-spectrum expectation values. In reality, the spatial pattern might vary as a function of frequency, leading to the cross-spectra being suppressed with respect to the geometric mean of the auto-spectra (Planck Collaboration 2017). We refer to this phenomenon as decorrelation.

We model decorrelation in the power spectrum domain using a set of simple one-parameter models. We define the correlation ratio of dust between two reference frequencies, 217 and 353 GHz, at pivot scale $\ell = 80$ as

$$\Delta_{\rm d} = \frac{\mathcal{D}_{80}(217\times 353)}{\sqrt{\mathcal{D}_{80}(217\times 217)\mathcal{D}_{80}(353\times 353)}}, \quad \text{(A7)}$$

where $\mathcal{D}_{80}$ is the dust power at $\ell = 80$. Here, $\Delta_{\rm d} < 1$ corresponds to decorrelation. We scale to other frequency combinations using the factor suggested by Planck Collaboration (2017):

$$f(\nu_1, \nu_2) = \frac{(\log(\nu_1/\nu_2))^2}{(\log(217/353))^2}. \quad \text{(A8)}$$

Similarly, based on suggestions from Planck Collaboration (2017), we consider three possible scalings with ℓ:

$$g(\ell) = \begin{cases} 1, & \text{flat case;} \\ (\ell/80), & \text{linear case;} \\ (\ell/80)^2, & \text{quadratic case.} \end{cases} \quad \text{(A9)}$$

The scalings above can produce extreme (and nonphysical) behavior at high ℓ or for frequencies that are widely separated. We therefore remap the scaled correlation coefficient using the following function:

$$\Delta'_{\rm d}(\nu_1, \nu_2, \ell) = \exp\left[\log(\Delta_{\rm d}) f(\nu_1, \nu_2)\, g(\ell)\right]. \quad \text{(A10)}$$

With this remapping, $\Delta'_{\rm d}$ remains in the range 0 (no correlation) to 1 (perfect correlation) for all values of f and g. This combination of frequency scaling and nonlinear remapping has been shown to correspond to a Gaussian spatial variation in the foreground spectral index parameter (Vansyngel et al. 2017).

In a similar vein, we define the parameter $\Delta'_{\rm s}$, which describes decorrelation of the synchrotron pattern. We do not include foreground decorrelation parameters in the dust–synchrotron correlated component. A complete foreground model would include all correlations between dust and synchrotron foregrounds across observing frequencies, but the current data do not provide useful guidance about the form of such correlations.

In addition to foregrounds, we include CMB scalar and tensor contributions. We make the simplifying assumptions that the tensor *BB* spectrum is given by a template scaled by parameter r and the *BB* spectrum from lensed CMB is given by a template scaled by parameter $A_{\rm L}$. Using CMB temperature units, the CMB contribution to the *BB* spectrum is given by

$$D^{\nu_1\times\nu_2}_{\ell,BB} = \frac{r}{0.1}D^{\rm tensor}_{\ell,BB} + A_{\rm L}D^{\rm lensing}_{\ell,BB}, \quad \text{(A11)}$$

where $D^{\rm tensor}_{\ell,BB}$ is the *BB* template for a tensor signal with $r = 0.1$ and $D^{\rm lensing}_{\ell,BB}$ is the expected lensing *BB* spectrum for ΛCDM. These are obtained using the `CAMB`[108] package (Lewis & Challinor 2011).

Appendix B
BPCM Construction and Rescaling

The BPCM construction relies on analytic rescaling of simulations. We calculate and store the bandpower covariance of signal and noise simulations for a particular sky and instrument model; rescaling from these covariance matrices eliminates the computational cost of creating large suites of simulations for each desired model. In the rescaling process, we explicitly set to zero terms that have zero expectation value (under the assumption that signal and noise are uncorrelated, and that physically unrelated signals are also uncorrelated) to reduce Monte Carlo error in the covariance matrix that stems from the modest number (499) of simulation realizations used. This framework is described in detail in Buza (2019), but we review it briefly here.

In a general case, let m_a, m_b, m_c, and m_d denote the four maps included in our analysis, with $m_a \times m_b$ denoting the cross-spectrum between those maps. Each map is the sum of independent signal components (CMB and foregrounds) plus a noise contribution, i.e., $m_a = \sum_i s_{ai} + n_a$. Since the synchrotron and dust foregrounds could be spatially correlated, we choose to divide them into three mutually independent components—uncorrelated synchrotron, uncorrelated dust, and the correlated part of synchrotron and dust. Using the properties that these signal and noise fields are independent and have zero mean, the expectation values for many spectra can be set to zero: $\langle s_{xi} \times s_{yj}\rangle = 0$ for $i \neq j$ (independent signal fields) and $\langle s_{xi} \times n_y\rangle = 0$ (signal is independent of noise). Then, a generic bandpower covariance term can be written as

$$\mathrm{Cov}(m_a \times m_b, m_c \times m_d) = \langle (m_a \times m_b)(m_c \times m_d)\rangle - \langle m_a \times m_b\rangle\langle m_c \times m_d\rangle \quad \text{(B1)}$$

[108] https://camb.info/

$$
\begin{aligned}
&= \textstyle\sum_i \mathrm{Cov}(s_{ai} \times s_{bi},\, s_{ci} \times s_{di}) \\
&\quad + \textstyle\sum_i \sum_{j\neq i} \mathrm{Cov}(s_{ai} \times s_{bj},\, s_{ci} \times s_{dj}) \\
&\quad + \textstyle\sum_i \sum_{j\neq i} \mathrm{Cov}(s_{ai} \times s_{bj},\, s_{cj} \times s_{dj}) \\
&\quad + \textstyle\sum_i \mathrm{Cov}(s_{ai} \times n_b,\, s_{ci} \times n_d) \\
&\quad + \textstyle\sum_i \mathrm{Cov}(s_{ai} \times n_b,\, n_c \times s_{di}) \\
&\quad + \textstyle\sum_i \mathrm{Cov}(n_a \times s_{bi},\, s_{ci} \times n_d) \\
&\quad + \textstyle\sum_i \mathrm{Cov}(n_a \times s_{bi},\, n_c \times s_{di}) \\
&\quad + \mathrm{Cov}(n_a \times n_b,\, n_c \times n_d), \qquad \text{(B2)}
\end{aligned}
$$

where the first three terms are signal-only covariances, the next four terms are covariances between select signal-noise cross-spectra, and the last term is a noise-only covariance. Additional terms can be set to zero if we make the further assumption that the noise is independent in each map, i.e., $\langle n_x \times n_y \rangle = 0$ for $x \neq y$. We also set to zero the covariance between any bandpowers that are separated by more than one bin in ℓ (for bins with $\Delta\ell = 35$), since these correlations are very small and not well-measured.

Rather than running simulations with a complicated combination of CMB and foreground signals, we calculate the above covariance terms for a simple signal model and then rescale to obtain the bandpower covariance matrix for an arbitrary model. In practice, the signal simulations are CMB realizations for a lensed ΛCDM model. Using s_{x0} to denote the simulated signal component of map m_x, the rescaled bandpower covariance matrix term is given by

$$
\begin{aligned}
&\mathrm{Cov}(m_a \times m_b,\, m_c \times m_d) \\
&\quad = \mathrm{Cov}(s_{a0} \times s_{b0},\, s_{c0} \times s_{d0}) \\
&\quad \times \textstyle\sum_i \left[\dfrac{\langle s_{ai} \times s_{ai}\rangle \langle s_{bi} \times s_{bi}\rangle \langle s_{ci} \times s_{ci}\rangle \langle s_{di} \times s_{di}\rangle}{\langle s_{a0} \times s_{a0}\rangle \langle s_{b0} \times s_{b0}\rangle \langle s_{c0} \times s_{c0}\rangle \langle s_{d0} \times s_{d0}\rangle} \right]^{1/2} \\
&\quad + \frac{1}{2}\mathrm{Cov}(s_{a0} \times s_{b0},\, s_{c0} \times s_{d0}) \\
&\quad \times \textstyle\sum_i \sum_{j\neq i} \left[\dfrac{\langle s_{ai} \times s_{ai}\rangle \langle s_{bj} \times s_{bj}\rangle \langle s_{ci} \times s_{ci}\rangle \langle s_{dj} \times s_{dj}\rangle}{\langle s_{a0} \times s_{a0}\rangle \langle s_{b0} \times s_{b0}\rangle \langle s_{c0} \times s_{c0}\rangle \langle s_{d0} \times s_{d0}\rangle} \right]^{1/2} \\
&\quad + \frac{1}{2}\mathrm{Cov}(s_{a0} \times s_{b0},\, s_{c0} \times s_{d0}) \\
&\quad \times \textstyle\sum_i \sum_{j\neq i} \left[\dfrac{\langle s_{ai} \times s_{ai}\rangle \langle s_{bj} \times s_{bj}\rangle \langle s_{cj} \times s_{cj}\rangle \langle s_{di} \times s_{di}\rangle}{\langle s_{a0} \times s_{a0}\rangle \langle s_{b0} \times s_{b0}\rangle \langle s_{c0} \times s_{c0}\rangle \langle s_{d0} \times s_{d0}\rangle} \right]^{1/2} \\
&\quad + \mathrm{Cov}(s_{a0} \times n_b,\, s_{c0} \times n_d) \\
&\quad \times \textstyle\sum_i \left[\dfrac{\langle s_{ai} \times s_{ai}\rangle \langle s_{ci} \times s_{ci}\rangle}{\langle s_{a0} \times s_{a0}\rangle \langle s_{c0} \times s_{c0}\rangle} \right]^{1/2} \\
&\quad + \mathrm{Cov}(s_{a0} \times n_b,\, n_c \times s_{d0}) \\
&\quad \times \textstyle\sum_i \left[\dfrac{\langle s_{ai} \times s_{ai}\rangle \langle s_{di} \times s_{di}\rangle}{\langle s_{a0} \times s_{a0}\rangle \langle s_{d0} \times s_{d0}\rangle} \right]^{1/2} \\
&\quad + \mathrm{Cov}(n_a \times s_{b0},\, s_{c0} \times n_d) \\
&\quad \times \textstyle\sum_i \left[\dfrac{\langle s_{bi} \times s_{bi}\rangle \langle s_{ci} \times s_{ci}\rangle}{\langle s_{b0} \times s_{b0}\rangle \langle s_{c0} \times s_{c0}\rangle} \right]^{1/2} \\
&\quad + \mathrm{Cov}(n_a \times s_{b0},\, n_c \times s_{d0}) \\
&\quad \times \textstyle\sum_i \left[\dfrac{\langle s_{bi} \times s_{bi}\rangle \langle s_{di} \times s_{di}\rangle}{\langle s_{b0} \times s_{b0}\rangle \langle s_{d0} \times s_{d0}\rangle} \right]^{1/2} \\
&\quad + \mathrm{Cov}(n_a \times n_b,\, n_c \times n_d). \qquad \text{(B3)}
\end{aligned}
$$

The factors of $1/2$ in the second and third terms are necessary to account for the factor of 2 difference between the variance of an auto-spectrum and the variance of the cross-spectrum between two uncorrelated fields.

An important feature of this rescaling method is that it considers signal and noise separately, rather than rescaling a combined signal-plus-noise covariance matrix. This is important because, for the highly nonuniform hit patterns (see Figure 7) that result from a large field-of-view telescope making deep maps, the spatial distribution of signal and noise in the map are very different. This leads to a significant difference in the number of signal versus noise degrees of freedom, and therefore different amounts of bandpower variance for the same overall power. This rescaling method does not attempt to account for different degrees of freedom between the different signal types, but that is a much smaller effect.

References

BICEP2 Collaboration III 2015, ApJ, 814, 110
BICEP2/Keck Array and Planck Collaborations 2015, PhRvL, 114, 101301
BICEP2/Keck Array Collaborations VI 2016, PhRvL, 116, 031302
BICEP2/Keck Array Collaborations VII 2016, ApJ, 825, 66
BICEP2/Keck Array Collaborations X 2018, PhRvL, 121, 221301
BICEP2/Keck Array Collaborations XI 2019, ApJ, 884, 114
Abazajian, K. N., Adshead, P., Ahmed, Z., et al. 2016, arXiv:1610.02743
Abazajian, K., Addison, G., Adshead, P., et al. 2019, arXiv:1907.04473
Abitbol, M., Aboobaker, A. M., Ade, P., et al. 2018, ApJS, 239, 8
Ade, P., Aguirre, J., Ahmed, Z., et al. 2019, JCAP, 2019, 056
Aiola, S., Calabrese, E., Maurin, L., et al. 2020, JCAP, 2020, 047
Albrecht, A., & Steinhardt, P. J. 1982, PhRvL, 48, 1220
Araujo, D., Bischoff, C., Brizius, A., et al. 2012, ApJ, 760, 145
Austermann, J. E., Aird, K. A., Beall, J. A., et al. 2012, Proc. SPIE, 8452, 393
Baleato Lizancos, A., Challinor, A., Sherwin, B. D., & Namikawa, T. 2021, arXiv:2102.01045
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Barron, D., Chinone, Y., Kusaka, A., et al. 2018, JCAP, 2018, 009
Beck, D., Errard, J., & Stompor, R. 2020, JCAP, 2020, 030
Bender, A. N., Ade, P. A. R., Ahmed, Z., et al. 2018, Proc. SPIE, 10708, 1070803
Bunn, E. F., Zaldarriaga, M., Tegmark, M., & de Oliveira-Costa, A. 2003, PhRvD, 67, 023501
Buza, V. 2019, PhD thesis, Harvard Univ., https://dash.harvard.edu/handle/1/42029563
Carron, J. 2019, PhRvD, 99, 043518
Carron, J., & Lewis, A. 2017, PhRvD, 96, 063510
Carron, J., Lewis, A., & Challinor, A. 2017, JCAP, 2017, 035
Choi, S. K., & Page, L. A. 2015, JCAP, 2015, 020
Cramér, H. 1946, Mathematical Methods of Statistics (1st ed.; Princeton, NJ: Princeton Univ. Press)
Duivenvoorden, A. J., Gudmundsson, J. E., & Rahlin, A. r. S. 2019, MNRAS, 486, 5448
Errard, J., Feeney, S. M., Peiris, H. V., & Jaffe, A. H. 2016, JCAP, 1603, 052
Essinger-Hileman, T. 2011, PhD thesis, Princeton Univ.
Essinger-Hileman, T., Kusaka, A., Appel, J. W., et al. 2016, RScI, 87, 094503
Fabbian, G., Lewis, A., & Beck, D. 2019, JCAP, 2019, 057
Fraisse, A. A., Ade, P. A. R., Amiri, M., et al. 2013, JCAP, 2013, 047
Fuskeland, U., Wehus, I. K., Eriksen, H. K., & Næss, S. K. 2014, ApJ, 790, 104
Gelaro, R., McCarty, W., Suarez, M. J., et al. 2017, JCli, 30, 5419
Ghosh, T., Boulanger, F., Martin, P. G., et al. 2017, A&A, 601, A71
Gualtieri, R., Filippini, J. P., Ade, P. A. R., et al. 2018, JLTP, 193, 1112
Gupta, N., Reichardt, C. L., Ade, P. A. R., et al. 2019, MNRAS, 490, 5712
Guth, A. H. 1981, PhRvD, D23, 347
Guth, A. H., & Pi, S. Y. 1982, PhRvL, 49, 1110
Harrington, K., Marriage, T., Ali, A., et al. 2016, Proc. SPIE, 9914, 99141K
Hasegawa, M., Ade, P., Aguilar, M., et al. 2018, Proc. SPIE, 10708, 1070802
Hawking, S. 1982, PhLB, 115, 295
Hazumi, M., Ade, P. A. R., Akiba, Y., et al. 2019, JLTP, 194, 443
Henning, J. W., Sayre, J. T., Reichardt, C. L., et al. 2018, ApJ, 852, 97
Hensley, B. 2015, PhD thesis, Princeton Univ.

Hivon, E., & Kamionkowski, M. 2002, Sci, 298, 1349
Hu, W., Hedman, M. M., & Zaldarriaga, M. 2003, PhRvD, 67, 043004
Hu, W., & White, M. 1997, NewA, 2, 323
Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, PhRvL, 78, 2058
Kamionkowski, M., & Kovetz, E. D. 2016, ARA&A, 54, 227
Kang, J. H., Ade, P. A. R., Ahmed, Z., et al. 2018, Proc. SPIE, 10708, 391
Kazanas, D. 1980, ApJL, 241, L59
Keisler, R., Hoover, S., Harrington, N., et al. 2015, ApJ, 807, 151
Kendall, M., & Stuart, A. 1979, Advanced Theory of Statistics (4th ed.; Oxford: Oxford Univ. Press)
Kermish, Z. D., Ade, P., Anthony, A., et al. 2012, Proc. SPIE, 8452, 366
Knox, L., & Song, Y.-S. 2002, PhRvL, 89, 011303
Krachmalnicoff, N., Carretti, N., Baccigalupi, C., et al. 2018, A&A, 618, A166
Krauss, L. M., & Wilczek, F. 2014, PhRvD, 89, 047501
Kritsuk, A. G., Ustyugov, S. D., & Norman, M. L. 2017, NJPh, 19, 065003
Kusaka, A., Apple, J., Essinger-Hileman, T., et al. 2018, JCAP, 2018, 005
Lawrence, C. R., Borrill, J., Carlstrom, J., et al. 2017, CMB-S4 Concept Definition Task Force Report, https://www.nsf.gov/mps/ast/aaac/cmbs4cdt.jsp
Lawrence, C. R., Gaier, T. C., & Seiffert, M. 2004, Proc. SPIE, 5498, 220
Lee, A. T., Tran, H., Ade, P., et al. 2008, in AIP Conf. Proc. 1040, KEK Cosmophysics Group Inaugural Conf. "Accelerators in the Universe": Interplay between High Energy Physics and Cosmophysics (Melville, NY: AIP), 66
Lewis, A., & Challinor, A. 2006, PhR, 429, 1
Lewis, A., & Challinor, A. 2011, CAMB: Code for Anisotropies in the Microwave Background, Astrophysics Source Code Library, ascl:1102.026
Linde, A. D. 1982, PhLB, 108, 389
Linde, A. D. 1983, PhLB, 129, 177
Louis, T., Grace, E., Hasselfield, M., et al. 2017, JCAP, 2017, 031
McMahon, J. J., Aird, K. A., Benson, B. A., et al. 2009, in AIP Conf. Proc. 1185, The Thirteenth Int. Workshop on Low Temperature Detectors (Melville, NY: AIP), 511
Millea, M., Anderes, E., & Wandelt, B. D. 2020, PhRvD, 102, 123542
Montroy, T. E., et al. 2006, Proc. SPIE, 6267, 239
Mukhanov, V. F. 1985, JETPL, 41, 493
Mukhanov, V. F., & Chibisov, G. V. 1981, JETPL, 33, 532
Mukhanov, V. F., & Chibisov, G. V. 1982, JETP, 56, 258
Nguyen, H. T., Kovac, J., Ade, P., et al. 2008, Proc. SPIE, 7020, 395
O'Dea, D., Challinor, A., & Johnson, B. R. 2007, MNRAS, 376, 1767
Paine, S. 2017, The am Atmospheric Model (v9.2), Zenodo, doi:10.5281/zenodo.438726
Planck Collaboration XXII 2015, A&A, 576, A107
Planck Collaboration IX 2014, A&A, 571, A9
Planck Collaboration XI 2016, A&A, 594, A11
Planck Collaboration XV 2014, A&A, 571, A15
Planck Collaboration 2006, arXiv:astroph/0604069
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A5
Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020b, A&A, 641, A10
Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020c, A&A, 641, A11
Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020d, A&A, 641, A7
Planck Collaboration L 2017, A&A, 599, A51
Polarbear Collaboration, Adachi, S., Aguilar Faúndez, M. A. O., et al. 2020, ApJ, 897, 55
Reichborn-Kjennerud, B., Aboobaker, A. M., Ade, P., et al. 2010, Proc. SPIE, 7741, 381
Sailer, N., Schaan, E., & Ferraro, S. 2020, PhRvD, 102, 063517
Sato, K. 1981, MNRAS, 195, 467
Sayre, J. T., Reichardt, C. L., Henning, J. W., et al. 2020, PhRvD, 101, 122003
Seljak, U., & Hirata, C. M. 2004, PhRvD, 69, 043005
Seljak, U., & Zaldarriaga, M. 1997, PhRvL, 78, 2054
Sheehy, C., & Slosar, A. 2018, PhRvD, 97, 043522
Shimon, M., Keating, B., Ponthieu, N., & Hivon, E. 2008, PhRvD, 77, 083003
Smith, K. M., Hanson, D., LoVerde, M., Hirata, C. M., & Zahn, O. 2012, JCAP, 1206, 014
Smith, K. M., & Zaldarriaga, M. 2007, PhRvD, 76, 043001
Starobinsky, A. 1982, PhLB, 117, 175
Starobinsky, A. A. 1980, PhLB, 91, 99
Suzuki, A., Ade, P., Akiba, Y., et al. 2016, JLTP, 184, 805
Tegmark, M., de Oliveira-Costa, A., & Hamilton, A. 2003, PhRvD, 68, 123523
Tegmark, M., & Efstathiou, G. 1996, MNRAS, 281, 1297
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2017, MNRAS, 469, 2821
Vansyngel, F., Boulanger, F., Ghosh, T., et al. 2017, A&A, 603, A62
Wallis, C. G. R., Brown, M. L., Battye, R. A., Pisano, G., & Lamagna, L. 2014, MNRAS, 442, 1963
Wu, W. L. K., Errard, J., Dvorkin, C., et al. 2014, ApJ, 788, 138
Yoon, K. W., Ade, P. A. R., Barkats, D., et al. 2006, Proc. SPIE, 6275, 508