The Astrophysical Journal, 927:174 (26pp), 2022 March 10 https://doi.org/10.3847/1538-4357/ac20df

A Constraint on Primordial *B*-modes from the First Flight of the SPIDER Balloon-borne Telescope

P. A. R. Ade[1], M. Amiri[2], S. J. Benton[3], A. S. Bergman[3], R. Bihary[4], J. J. Bock[5,6], J. R. Bond[7], J. A. Bonetti[6], S. A. Bryan[8], H. C. Chiang[9,10], C. R. Contaldi[11], O. Doré[5,6], A. J. Duivenvoorden[3,12], H. K. Eriksen[13], M. Farhang[7,14,15], J. P. Filippini[16,17], A. A. Fraisse[3], K. Freese[12,18], M. Galloway[13], A. E. Gambrel[19], N. N. Gandilo[20], K. Ganga[21], R. Gualtieri[22], J. E. Gudmundsson[12], M. Halpern[2], J. Hartley[23], M. Hasselfield[24], G. Hilton[25], W. Holmes[6], V. V. Hristov[5], Z. Huang[7], K. D. Irwin[26,27], W. C. Jones[3], A. Karakci[13], C. L. Kuo[26], Z. D. Kermish[3], J. S.-Y. Leung[15,28], S. Li[3,29], D. S. Y. Mak[11], P. V. Mason[5], K. Megerian[6], L. Moncelsi[5], T. A. Morford[5], J. M. Nagy[30,31], C. B. Netterfield[15,23], M. Nolta[7], R. O'Brient[6], B. Osherson[16], I. L. Padilla[15,32], B. Racine[13], A. S. Rahlin[19,33], C. Reintsema[25], J. E. Ruhl[4], M. C. Runyan[5], T. M. Ruud[13], J. A. Shariff[7], E. C. Shaw[16], C. Shiu[3], J. D. Soler[34], X. Song[3], A. Trangsrud[5,6], C. Tucker[1], R. S. Tucker[5], A. D. Turner[6], J. F. van der List[3], A. C. Weber[6], I. K. Wehus[13], S. Wen[4], D. V. Wiebe[2], and E. Y. Young[26,27]

SPIDER Collaboration

[1] School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, UK
[2] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[3] Department of Physics, Princeton University, Jadwin Hall, Princeton, NJ 08544, USA; wcjones@princeton.edu
[4] Physics Department, Case Western Reserve University, 10900 Euclid Ave, Rockefeller Building, Cleveland, OH 44106, USA
[5] Division of Physics, Mathematics and Astronomy, California Institute of Technology, MS 367-17, 1200 E. California Boulevard, Pasadena, CA 91125, USA
[6] Jet Propulsion Laboratory, Pasadena, CA 91109, USA
[7] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[8] School of Electrical, Computer, and Energy Engineering, Arizona State University, 650 E Tyler Mall, Tempe, AZ 85281, USA
[9] Department of Physics, McGill University, 3600 Rue University, Montreal, QC, H3A 2T8, Canada
[10] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa
[11] Blackett Laboratory, Imperial College London, SW7 2AZ, London, UK
[12] The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden
[13] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
[14] Department of Physics, Shahid Beheshti University, 1983969411, Tehran, Iran
[15] Department of Astronomy and Astrophysics, University of Toronto, 50 Saint George Street, Toronto, ON M5S 3H4, Canada
[16] Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green Street, Urbana, IL 61801, USA
[17] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
[18] Department of Physics, University of Texas, 2515 Speedway, C1600, Austin, TX 78712, USA
[19] Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[20] Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[21] Université de Paris, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France
[22] High Energy Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA
[23] Department of Physics, University of Toronto, 60 Saint George Street, Toronto, ON M5S 3H4, Canada
[24] Department of Astronomy and Astrophysics, Pennsylvania State University, 520 Davey Laboratory, University Park, PA 16802, USA
[25] National Institute of Standards and Technology, 325 Broadway Mailcode 817.03, Boulder, CO 80305, USA
[26] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[27] SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[28] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 Saint George Street, Toronto, ON M5S 3H4, Canada
[29] Department of Mechanical and Aerospace Engineering, Princeton University, Engineering Quadrangle, Princeton, NJ 08544, USA
[30] Department of Physics, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA
[31] McDonnell Center for the Space Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA
[32] Department of Physics and Astronomy, Johns Hopkins University, 3701 San Martin Drive, Baltimore, MD 21218, USA
[33] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510-5011, USA
[34] Max-Planck-Institute for Astronomy, Konigstuhl 17, D-69117, Heidelberg, Germany

Received 2021 April 6; revised 2021 August 5; accepted 2021 August 15; published 2022 March 14

Abstract

We present the first linear polarization measurements from the 2015 long-duration balloon flight of SPIDER, which is an experiment that is designed to map the polarization of the cosmic microwave background (CMB) on degree angular scales. The results from these measurements include maps and angular power spectra from observations of 4.8% of the sky at 95 and 150 GHz, along with the results of internal consistency tests on these data. While the polarized CMB anisotropy from primordial density perturbations is the dominant signal in this region of sky, Galactic dust emission is also detected with high significance. Galactic synchrotron emission is found to be negligible in the SPIDER bands. We employ two independent foreground-removal techniques to explore the sensitivity of the cosmological result to the assumptions made by each. The primary method uses a dust template derived from *Planck* data to subtract the Galactic dust signal. A second approach, which constitutes a joint analysis of SPIDER and *Planck* data in the harmonic domain, assumes a modified-blackbody model for the spectral energy distribution of the dust with no constraint on its spatial

morphology. Using a likelihood that jointly samples the template amplitude and r parameter space, we derive 95% upper limits on the primordial tensor-to-scalar ratio from Feldman–Cousins and Bayesian constructions, finding $r < 0.11$ and $r < 0.19$, respectively. Roughly half the uncertainty in r derives from noise associated with the template subtraction. New data at 280 GHz from SPIDER's second flight will complement the *Planck* polarization maps, providing powerful measurements of the polarized Galactic dust emission.

Unified Astronomy Thesaurus concepts: Cosmic microwave background radiation (322); Observational cosmology (1146); Cosmological parameters (339); Interstellar emissions (840)

1. Introduction

In the standard cosmological model (ΛCDM), the universe consists of a blend of radiation, baryonic matter, cold dark matter, and a vacuum energy density consistent with a cosmological constant. The observed structure in the universe originates from primordial fluctuations of matter and energy that grow through gravitational instability. These perturbations evolve within a spacetime geometry that is spatially flat on the largest observed scales. This simple paradigm has proven to be in remarkable agreement with the overwhelming majority of all observational tests (Peebles 2012; *Planck* Collaboration et al. 2020a, 2020b).

Observational data place stringent constraints on the properties of these primordial density fluctuations: they must be predominantly adiabatic in nature, Gaussian-distributed, follow a nearly—but not quite—scale-invariant spectrum, and encode correlations on scales larger than the horizon during recombination. Mechanisms to generate such fluctuations have been proposed within the context of inflationary, bouncing, and cyclic models (Guth & Pi 1982; Hawking 1982; Mukhanov & Chibisov 1982; Starobinsky 1982; Bardeen et al. 1983; Ijjas & Steinhardt 2018, 2019; Tanabashi et al. 2018; Shandera et al. 2019; Cook et al. 2020).

In addition to the well-studied scalar perturbations, some early-universe models—particularly inflationary models—predict a spectrum of tensor perturbations, or primordial gravitational waves. Their amplitude is characterized by the dimensionless tensor-to-scalar ratio, r.[35] The *Planck* collaboration combines precision measurements of the matter power spectrum and the largest-scale cosmic microwave background (CMB) intensity fluctuations to constrain r to be less than $r < 0.10$ (*Planck* Collaboration et al. 2020c).[36]

Local quadrupole anisotropies sourced by tensor fluctuations can also imprint a unique "*B*-mode" (curl) component to the polarization of the CMB at degree angular scales (Seljak & Zaldarriaga 1997; Kamionkowski & Jaffe 2001). Though challenging to measure, this signature is relatively free of sample variance from the brighter scalar modes and thus allows observational access to much smaller values of r. The detection of the signature of tensor fluctuations would bring remarkable new insights into early-universe physics. This scientific potential has motivated an ambitious observational effort to search for the signature of primordial gravitational waves in the polarization of the CMB (Abazajian et al. 2016; Kamionkowski & Kovetz 2016).

The *Planck* polarization data, which span more than half of the full sky, constrain $r < 0.158$ using limits on the *B*-mode contribution alone (Tristram et al. 2021). Using *BB* limits derived from observations of less than 1% of the full sky, the Keck team reports $r < 0.072$ (BICEP2/Keck Array Collaboration 2018). *Planck* measurements of the CMB intensity, the *E*-mode polarization, and lensing over more than half the full sky, together with the Keck *BB* limits, improve the constraint to $r < 0.056$ (*Planck* Collaboration et al. 2020c). In Tristram et al. (2021), this same constraint is obtained using only *Planck* temperature and polarization data. Combining the *B*-mode results from the Keck experiment with this re-analysis of the *Planck* polarization data, the same team reports a somewhat tighter constraint, $r < 0.044$ (Tristram et al. 2021).

As anticipated even prior to the *Planck* results, any cosmological *B*-mode signal is subdominant to the diffuse polarized emission from our Galaxy along any line of sight (Fraisse et al. 2013). Current CMB observations must thus contend with modeling uncertainties associated with diffuse Galactic emission. To date, the *Planck* polarization data provide the most accurate estimate of polarized Galactic emission across the full sky (*Planck* Collaboration et al. 2020d).

In this paper, we report results from the first flight of SPIDER, which is a balloon-borne instrument that is designed to measure the polarization of the CMB on degree angular scales. This paper is organized as follows. After a brief description of the SPIDER instrument in Section 2 and observation strategy in Section 3, we discuss the low-level data processing leading up to maps of the sky in Section 4. Section 5 presents two complementary angular power spectrum estimators, while Section 6 discusses the consistency tests performed with each of these estimators, and Section 7 addresses sources of systematic error. Results from several distinct methods of component separation are presented in Section 8, while Section 9 provides constraints on cosmological parameters for each method. The main conclusions and SPIDER's future prospects are summarized in Section 10.

2. The SPIDER Instrument

The SPIDER payload consists of six monochromatic refracting telescopes that are housed within a single liquid helium cryostat, which is supported and pointed by a lightweight carbon fiber gondola. Here we provide a brief overview of the payload design, a more detailed description can be found in Runyan et al. (2010), Filippini et al. (2010), Rahlin et al. (2014), and Gualtieri et al. (2018).

2.1. Receivers

Each SPIDER receiver is an axisymmetric two-lens cryogenic refractor with a 270 mm cold stop, which is designed to minimize polarized systematics. In each receiver, two high-density polyethylene lenses cooled to 4 K focus light onto a 300 mK focal plane. The blackened cold stop and internal baffles surrounding the optics are cooled to 1.6 K to reduce stray photon loading on the detectors. A sapphire half-wave plate (HWP) mounted to a 4 K flange skyward of each receiver's stop is rotated to a new fixed orientation angle twice

[35] Throughout we specify r at a scale of $k_0 = 0.05$ Mpc^{-1}, and further assume a scale-invariant tensor spectrum ($n_t = 0$). The six ΛCDM parameters are fixed to those of *Planck* Collaboration et al. (2020b).

[36] This constraint relaxes to $r < 0.16$ when excluding the low-ℓ data ($2 \leqslant \ell \leqslant 29$) that include the temperature deficit.

Table 1
Summary of Instrumental Parameters for the Data Used in this Analysis

Band	Center (GHz)	Width (%)	FWHM (arcmin)	# Det. Used	NET$_{tot}$ (μK$\sqrt{s}$)	Data Used (days)	Map Depth (μK arcmin)
95 GHz	94.7	26.4	41.4	675	7.1	6.5	22.5
150 GHz	151.0	25.7	28.8	815	6.0	5.6	20.4

Note. Band center and width are averages of per-detector measurements. Beam full-width at half-maximum is derived from a combined fit to all detectors in a given band. Noise-equivalent temperature is the quadrature average over all detectors used. Data used is the NET-weighted average of unflagged data in each channel, and is restricted to samples inside our sky mask (Section 5) with hits-weighted $f_{\rm sky}$ of 3.9%. Approximate map depths do not account for effects of filtering on signal-to-noise. All sensitivities are reported in CMB temperature units.

daily to provide polarization modulation (Bryan et al. 2010a, 2016). Each receiver views the sky through a series of reflective metal-mesh (Ade et al. 2006) and lossy nylon filters to reduce infrared loading on the cryogenic system and detectors, as well as a thin ($\sim$3 mm) ultra-high-molecular-weight polyethylene (UHMWPE) vacuum window. An appropriate single-layer anti-reflection coating, which is matched to the receiver's band (95 or 150 GHz), is attached to each side of the HWPs, lenses, vacuum windows, and relevant filters.

Each telescope focuses radiation onto four wafers ("tiles") of antenna-coupled transition-edge sensors (TESs), which were fabricated at JPL (Ade et al. 2015). Each wafer is patterned with an array of polarimeter pixels, consisting of two inter-penetrating arrays of slot antennas (one for each perpendicular polarization mode). This arrangement provides for an instantaneous measurement of total intensity and one of two linear polarization components. A complete measurement of partial linear polarization—Stokes I, Q and U parameters—is obtained for each pixel through rotations of the HWP and the sky, which modulate the polarization angle (Jones et al. 2007). A microstrip feed network coherently couples optical power from these synthesized antennas through a band-defining lumped-element filter before dissipating the power incoherently on a thermally isolated island. Each island supports two TESs with different critical temperatures, T_c, wired in series: a Ti sensor ($T_c \sim 500$ mK) for science observations and an Al sensor ($T_c \sim 1.3$ K) for laboratory testing. The 512 (288) TESs of each 150 GHz (95 GHz) focal plane are read out using a time-division SQUID multiplexing system (de Korte et al. 2003; Battistelli et al. 2008; Stiehl et al. 2011). The TESs and SQUIDs are housed within extensive magnetic shielding (Runyan et al.2010).

Table 1 summarizes the properties of all of the detectors used in the analysis presented in this paper.[37] This flight of SPIDER deployed a total of 2400 TESs. The channel counts in Table 1 account for intentionally dark (non-optical) TES channels, losses due to detector and readout performance, and the conservative channel cuts used in the present analysis. Notably, one of the three 150 GHz receivers was excluded late in the analysis due to a null test failure (Section 6.1.3), but should be recoverable with future work. Across the remaining five receivers, $\sim$80% of TESs are used in this analysis.

2.2. Cryogenics

SPIDER's cryogenic system (Gudmundsson et al. 2015), which is the largest yet deployed on a long-duration balloon flight, consists of two liquid helium reservoirs: a 1284 l main tank and a 16 l superfluid tank. The main tank is maintained at a pressure of roughly 1 bar during the flight, providing cooling power at $\sim$4 K for the receiver optics and the ^{3}He sorption coolers. The boil-off from the main tank flows through heat exchangers on each of two vapor-cooled shields, which intercept the radiative and conductive parasitic loads on the cryogenic system and cool the infrared filter stack. The superfluid system provides cooling power at 1.6 K to each telescope's ^{3}He sorption cooler and internal optical baffles. The superfluid tank fills continuously from the main tank through a capillary assembly, and is maintained at the ambient pressure of the altitude at float (about 6 mbar). The superfluid system is pumped down on the ground, and maintained at low pressure during launch and ascent with a small diaphragm pump on the gondola. The focal planes themselves are cooled to $\sim$300 mK by a dedicated ^{3}He sorption cooler within each telescope.

2.3. Gondola and Pointing System

The cryostat is supported within a lightweight carbon fiber gondola (Soler et al. 2014). A reaction-wheel and motorized pivot scan the gondola in azimuth, while a linear drive steps the cryostat in elevation (Shariff et al. 2014). Absolute referencing of the payload orientation is provided by a suite of three star cameras: one attached to the cryostat and oriented along the boresight axis, the other two mounted to the outer gondola frame on a rotating table that allows them to track the sky during azimuthal scans. Information from the star cameras is combined with that from GPS receivers, Sun sensors, encoders, and gyroscopes to enable in-flight pointing and post-flight pointing reconstruction (Gandilo et al. 2014). Control and monitoring of the pointing and cryogenic systems is performed by a pair of redundant flight computers interfaced with the custom BLASTbus electronics (Benton et al. 2014). A Sun shield protects the instrument and optics during the 24 hr Antarctic summer daylight. Continuous electric power is provided by a 2 kW solar panel system, while various antenna arrays provide commanding, telemetry, and location information during the flight.

3. Science Observations

SPIDER was launched on 2015 January 1, from the NASA/NSF Long-Duration Balloon (LDB) facility near McMurdo Station, Antarctica. All of the payload systems performed well throughout the flight, with the exception of a differential GPS unit failure that had no significant impact on flight operations or pointing reconstruction. SPIDER's flight lasted 16.5 days at an average altitude of 35 km. The flight was terminated when cryogens were exhausted and the circumpolar wind system began to fail. The payload touched down in a remote region of

[37] In this paper, all of the temperatures used in reference to signal or noise are in units of $\Delta T_{\rm CMB}$, the equivalent CMB fluctuation, in which the data are natively calibrated.

Ellsworth Land of West Antarctica. Data drives and key flight hardware were recovered in February by personnel from the British Antarctic Survey, while a second team recovered the remainder of the instrument in November. The payload optics and focal planes have subsequently been refurbished and upgraded in preparation for a second flight (Shaw et al. 2020).

During an Antarctic LDB flight, the Sun remains above the horizon at all times. The accessible region of sky is therefore constrained by the need for the field center to remain roughly anti-solar and, for CMB observations, to avoid the Galactic plane. This favors a launch as early in the season as possible because the anti-solar direction progresses to lower Galactic latitude over time. SPIDER's launch opportunity came relatively late in the Antarctic LDB season, which pushed the field center toward the lower range of possible field centers.

SPIDER scanned in azimuth throughout the flight, with a sinusoidal speed profile peaking as high as $4^\circ\,s^{-1}$. The sinusoidal speed profile allowed smooth torque variations in the pivot and reaction-wheel motors, without sustaining peak torque for long. For the first two-thirds of the flight, SPIDER scanned a $\sim 75^\circ$ azimuthal range that was limited on either side by the Galaxy and the Sun. To obtain more uniform coverage, the azimuthal range was reduced for the final third of the flight to cover the middle half of this range. Scan turnarounds are separated by as long as 36 s, shortening to as little as 22 s for the narrower region later in the flight. Small steps in elevation were made at every third scan turnaround, covering the full 22°–50° range upwards and downwards once per day. A brief scan over the bright Galactic source RCW38 was used to confirm pointing in flight.

The in-flight pointing solution using only coarse sensors has an error of $22'$ rms, which is less than a SPIDER beamwidth and adequate for scan control. The post-flight pointing reconstruction integrates gyroscopes between star camera solutions, and matches raw solutions of the boresight star camera to within $0.'9$ rms. The relative pointing between the boresight camera and microwave detectors is calibrated with cross-correlation and deprojection methods, which will be described in Section 4.2.4.

The half-wave plate is stepped in angle twice per sidereal day (Bryan et al. 2010a, 2016). The nominal HWP angles are chosen to rotate each receiver between Stokes Q and U sensitivity every half day, and to cover each rising and setting raster in Q and U with every detector on alternate days. A total of eight discrete HWP angles, separated by $22.^\circ5$ over a range of 180° are cycled through an eight-day pattern to reduce sensitivity to beam and HWP systematics (Bryan et al. 2010b; Nagy et al. 2017). The HWP angles are measured with a combination of absolute and relative encoders, which provides an accuracy of $0.^\circ1$.

SPIDER's flight control system implements a number of autonomous watchdog routines that monitor the quality of data returned from the detectors, perform limited corrective actions, and package compressed summary data packets for return to the ground system. Of particular note are the detector monitoring systems, which are described more fully in Rahlin (2016). These use regular measurements of the TES differential resistance (dV/dI) from small 2 Hz square waves imposed on the TES bias lines, which are carried out for 2 s intervals every fifth scan turnaround. These values are used to automatically identify channels that are superconducting or normal, have accumulated a large DC offset, or have drifted significantly in TES resistance. When the count of such anomalous channels grows large enough, the system initiates a reset of the TES feedback loop or an adjustment of the TES bias. Due to an unforeseen software race condition, this monitoring system did not function for most receivers during the latter portion of the flight; however, in practice this had little meaningful effect, given SPIDER's excellent detector performance stability. Using both electrical and optical measurements of the temporal gain variations, the excursions on all timescales are found to be less than 5%, and are not strongly correlated (Section 4.2.3).

4. Data Processing and Map Making

In this section, we will present an abbreviated discussion of SPIDER's low-level processing from raw data to calibrated maps and simulations of the sky. More details can be found in Rahlin (2016), Gambrel (2018), Young (2018).

SPIDER's raw data consist of 2.1 TB of time-ordered samples. Bolometer and pointing data were recorded at 119 Hz, while a variety of gondola and cryogenic performance parameters were recorded at reduced sample rates. All data were recorded in-flight across multiple redundant drives. Data from the six bolometer arrays and the flight system were synchronized using data-valid clock signals and sequential counter values distributed from a single crystal clock system.

For a number of data processing operations, samples are grouped into contiguous "chunks" that are approximately 10 minutes in length. These evenly partition the periods between HWP angle steps and divide only at turnarounds of the azimuthal scan. Chunk length is a compromise between containing a sufficient number of samples for analysis tasks such as estimating low-frequency noise, while remaining short enough to ensure that neighboring chunks have similar observing conditions and sky signal. Ten minutes is long compared to the azimuthal scan period, but short relative to the timescale for changes in telescope elevation or cryogenic temperatures. Similar chunk partitioning is also used to construct data subsets for power spectrum estimation (Section 5).

4.1. Timestream Flagging, Cleaning, and Filtering

In addition to the expected Gaussian uncorrelated noise, we observe two broad classes of correlated noise in the SPIDER timestream data: intermittent and quasi-stationary. *Intermittent* noise encompasses noise sources that appear to be discretely on or off at any given time. The primary sources of intermittent noise are the telemetry transmitters on board the payload. In particular, the three Iridium transmitters are active for about two seconds at a time, operating asynchronously with periods between 1 and 15 minutes. Other sources of intermittent noise include cosmic ray interactions in the detectors—which are discussed further in Osherson et al. (2020)—and various glitches or step discontinuities due to the multiplexing readout. *Quasi-stationary* noise consists of non-astrophysical signals that are partially correlated across the field of view and change very little over multiple azimuthal scans. These have a peak-to-peak amplitude that is typically less than 3 mK_{CMB}, and vary slowly over time. It is believed that the majority of this contamination is sourced by sidelobe pickup, primarily from the Earth's limb; its variation with elevation is not consistent with signal from any residual atmospheric emission. Additionally, RF-coupled interference was observed in some of the detector channels. One consequence of this is that a subset of

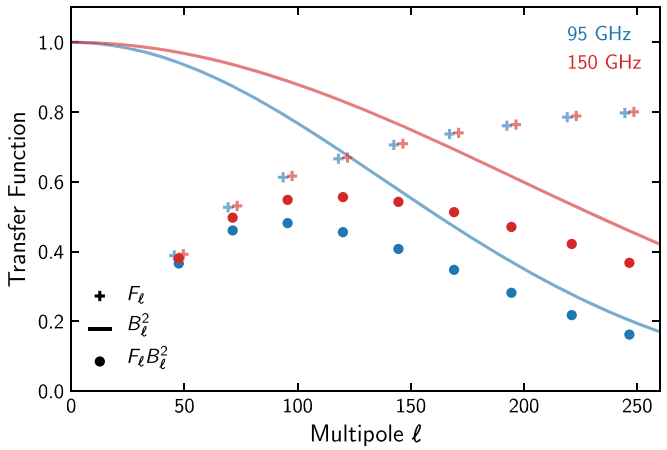

Figure 1. SPIDER's filter transfer function (F_ℓ), beam window function (B_ℓ^2), and total transfer function for 95 and 150 GHz. Quantities shown are the average of the *EE* and *BB* transfer functions, which are similar but are not assumed to be identical. In this work, signal estimation is achieved using a simple binning of the data, which necessitates the filtering applied here. This simple approach is driven by the desire to perform relatively efficient analysis of the signal and null tests in the time domain, but is unrelated to the atmosphere and does not represent a limitation of the stratospheric balloon platform.

detectors evidence a signal that is well-correlated with the orientation of the reaction wheel, with couplings that vary in strength both between warm readout electronics racks, and within them channel by channel. During both pre-launch testing and the flight, one of the three readouts serving the 150 GHz focal planes proved to be substantially more susceptible to these effects than the others. As further discussed in Section 6.1.3, the data from this focal plane contribute to null test failures and are not included in any other part of this analysis.

Intermittent noise is mitigated by flagging the affected detector samples. Such samples are tagged, replaced with constrained noise realizations, and excluded from map-making. Step discontinuities arise intermittently in SPIDER's data, due primarily to transmitter interference and large cosmic ray interactions. In addition to flagging the discontinuity itself, we adjust the data to eliminate the discontinuity using a linear fit to data before and after the event. This procedure accounts empirically for cross-talk of the discontinuity among channels. This "stitching" operation improves low-frequency noise significantly, and simulations show that it has negligible effect on signal response.

Quasi-stationary noise is mitigated by time-domain filtering of this flagged data set, conducted at the full detector sample rate. To reduce noise correlated with the reaction wheel, detector timestreams are binned according to the angle of the reaction wheel to form templates that are then subtracted. The impact of this operation was checked using the full-flight time-domain simulations described in Section 4.4, and found to have negligible effect on the astrophysical signal. This fit is thus performed only on data timestreams and is not performed on the large simulation ensembles. To reduce low-frequency noise and pickup more broadly, the data are filtered between each scan turnaround by subtracting a fifth-order polynomial fit to each detector's data as a function of azimuth.

The effects of scanning, filtering, and flagging are determined by applying the entire analysis pipeline to an ensemble of time-domain signal simulations. The transfer functions due to filtering and beams, which derive from these simulations, are shown in Figure 1 and are discussed further in Section 5. The primary effect of filtering is a suppression of power on large angular scales. Because the typical scan speed and direction vary across the sky, the effect of filtering is both anisotropic and inhomogeneous, with greater suppression on the edges of the field of view. This latter effect is not visibly evident in the temperature or polarization maps (Figures 2 and 3), and the net impact of the filtering has been shown to be adequately modeled with a simple multipole domain transfer function.

Detectors with consistently high noise after cleaning are completely cut from the analysis. A small fraction ($\ll$1%) of entire azimuth scans are also flagged for having too much residual noise after cleaning. In total, when weighting data by their estimated noise level, complete detector cuts remove 5% of the data, and 28% of samples on remaining detectors are flagged, including periods of cryogenic recycling. All detectors in the 150 GHz receiver most strongly affected by RF-coupled interference are discarded, resulting in a further reduction of 22% at that frequency (10% overall). Finally, 27% of unflagged samples lie outside the sky mask of the present analysis (Section 5).

4.2. Detector Characterization

Both pre-launch and in-flight data are used to characterize instrumental parameters needed to construct accurate temperature and polarization maps. The pre-launch data include spectroscopic and polarimetric measurements. In-flight data are used for the absolute calibration, to monitor gain fluctuations, and to refine pre-launch estimates of beam response and pointing offsets.

4.2.1. Polarization Angle

The individual detector polarization angles were measured prior to launch with a rotating polarized thermal source in the near-field of each receiver, as described in Nagy (2017). The uncertainty on each measured detector angle is approximately $0\overset{\circ}{.}5$, which is better than SPIDER's target of 1° (Fraisse et al. 2013). These measured angles are used directly by the map maker, with no correction applied based on the flight data.

4.2.2. Frequency Response

The frequency response of each detector was measured prior to launch with a custom high-throughput Fourier Transform Spectrometer, which was mounted on top of the cryostat. Since the hot thermal source did not illuminate the full telescope solid angle, an actuated mirror steered the output over the full field of view. In total, 95% of all of the detectors used in the science analysis were measured with band center and bandwidth accurate to 1 GHz. Measurements of the band centers and widths at different HWP angles and output mirror positions are consistent within errors. Further details are provided in Gambrel (2018).

These per-detector measurements are not used directly in making maps, which could result in leakage of spectrally mismatched temperature signals into polarization. Instead, a null test is constructed by splitting detectors with high and low band centers (Section 6.1). Since no difference is detected, we conclude that the SPIDER data, including mitigation from HWP and sky rotation, have negligible leakage from bandpass mismatch.

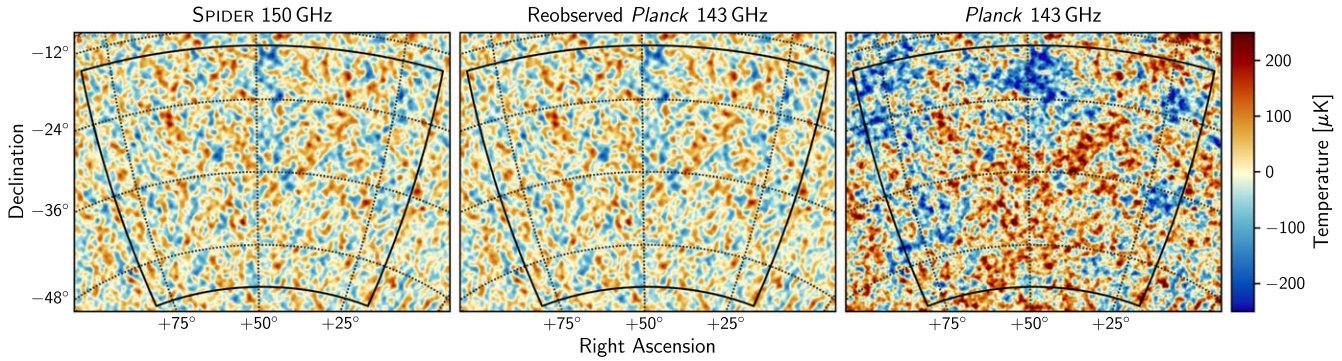

Figure 2. Left-hand panel: The total intensity map as observed by the SPIDER 150 GHz receivers. No additional filtering is applied to the maps beyond that in the timestream processing. The black outline indicates the sky region used to compute power spectra, though the additional point-source mask is not shown. Middle panel: The *Planck* 143 GHz map as re-observed using the SPIDER scan strategy and filtering, indicating strong agreement in the temperature signal. Right-hand panel: The raw *Planck* 143 GHz map, shown to illustrate the impact of SPIDER's scan strategy and filtering, which suppresses power at large angular scales.

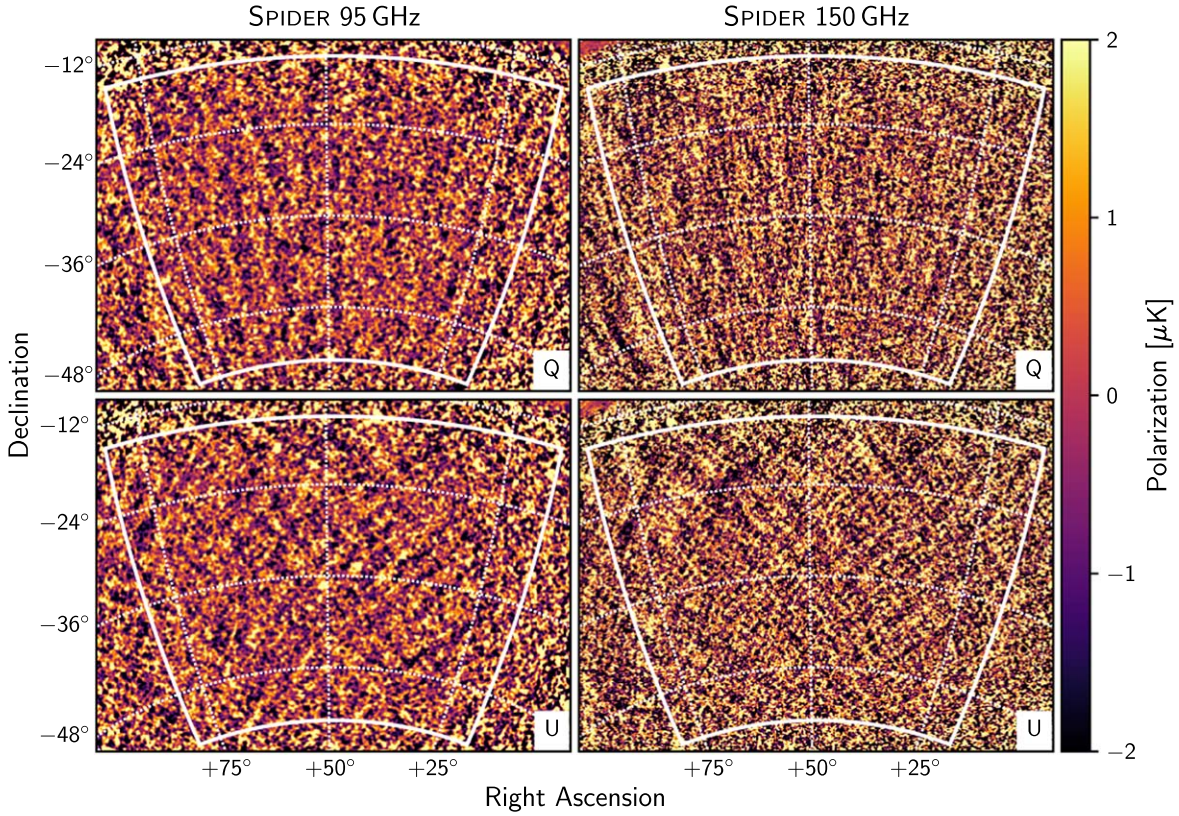

Figure 3. Q and U polarization maps as observed by SPIDER's 95 and 150 GHz receivers. The maps have been smoothed with a 10′ Gaussian for clarity. The temperature-to-polarization leakage from the map maker is subtracted (Section 4.4), although this effect is not visible by eye. The dominant E-mode pattern of the cosmological signature is evident in the maps, although it is diluted by the Galactic signal. The white outline indicates the sky region used to compute power spectra, though the additional point-source mask is not shown.

4.2.3. Calibration and Beam

SPIDER's absolute calibration is derived by cross-calibrating degree-scale power with *Planck* temperature anisotropy data at 100 and 143 GHz.[38] This procedure finds the absolute calibration factor and parameterized beam model that minimizes the difference with the *Planck* temperature spectra at a per-detector level in the range $100 < \ell < 275$ $(100 < \ell < 375)$ for the 95(150) GHz frequency band. The absolute calibration is obtained by finding the scalar, c, that minimizes

$$\sum_{\ell=\ell_1}^{\ell_2} R_\ell \equiv \sum_{\ell=\ell_1}^{\ell_2} \left| c \frac{\widehat{C}_\ell^{TT}}{\widehat{C}_{\ell,\mathrm{ref}}^{TT}} \frac{b_\ell^{\mathrm{Planck}}}{b_\ell^{\mathrm{SPIDER}}} - 1 \right|, \quad (1)$$

where $\ell_1 = 100$ and $\ell_2 = 275\,(375)$ for the 95(150) GHz frequency band. We use $\widehat{C}_{\ell,\mathrm{ref}}^{TT}$ to represent a temperature power spectrum calculated using maps obtained from re-scanning the *Planck* half-mission reference maps, while $\widehat{C}_\ell^{TT}$ is calculated from single-detector maps cross-correlated with a *Planck* half-mission map. The beam transfer functions, b_ℓ^{Planck} and b_ℓ^{SPIDER},

[38] Throughout this paper, we use release 3.01 of the *Planck* HFI maps (*Planck* Collaboration et al. 2020e).

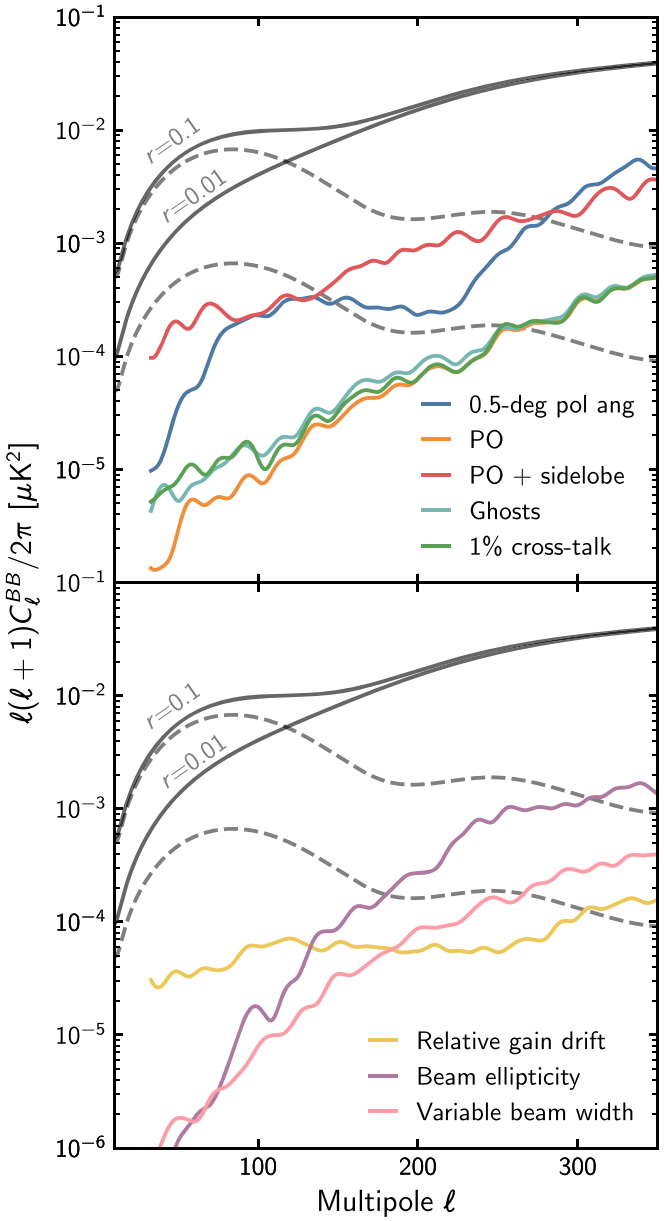

Figure 4. Simulated residual *B*-mode power from several systematic effects at 150 GHz. The top panel shows the residuals from offset detector orientation angles and beam model extensions simulated using the `beamconv` algorithm. The bottom panel shows residuals determined from time-domain deprojection templates. In the legend, "PO" refers to physical optics simulations; the entries are described further in Section 7. Also shown for comparison are total *BB* spectra with lensing (solid gray) and without (dashed gray) for two benchmark values of r.

quantify the relative sensitivity the *Planck* and SPIDER spatial response as a function of multipole.

We use a simple Gaussian beam model, $b_\ell^{\rm SPIDER}$, to extend this calibration to other angular scales included in our analysis; note that this extrapolation is small, and primarily to larger angular scales. Each SPIDER telescope is fitted with a single common beam model, which is then used to determine an independent calibration factor for each individual detector.

Various consistency tests show that our analysis is not sensitive to a more physically motivated beam model because significant deviations from a simple Gaussian are only evident below multipoles used in constructing bandpowers. By using beam models informed by physical optics simulations, we have quantified the potential bias in our absolute calibration on the largest angular scales caused by the Gaussian beam model assumption. At most, this results in a 5% bias in the beam transfer function in the lowest bin ($33 \leqslant \ell \leqslant 57$), which has a negligible impact on the results. The "Inner/Outer Focal Plane Radius" null test (Section 6.1) shows no detectable difference between the detectors expected to be the best and worst matches to our beam model.

We further explore the possibility of time-varying detector calibrations in several ways. We use our regular TES resistance measurements (Section 3) to generate a rough proxy for small changes in the TES bias state, and hence responsivity. When TES monitoring is not available late in the flight, we use the average level of the TES current to calibrate a similar proxy. We find that these estimates are consistent with one another. Additionally, we find that gain excursions on all timescales are less than 5%, and are not strongly correlated. While we correct our timestreams for an interpolated version of our TES-monitoring gain, simulations and null tests show that this has negligible impact on our analysis (see Figure 4).

Time domain simulations are used to quantify errors in the absolute calibration and effective beamwidth on both per-detector and full focal plane bases. Statistical error in the determination of those parameters is caused by noise in the SPIDER and *Planck* data, both of which are incorporated in our simulations with appropriate noise models. At the telescope level, statistical error is relatively small because of the data's high signal-to-noise ratio. For example, the fractional error in the per-telescope beam transfer function, b_ℓ, at degree angular scales, is approximately 0.1%. The likelihood analysis described in Section 9 incorporates a model of the statistical beam error. Some potential systematic effects are also investigated and found to be subdominant at our sensitivity level (Section 7).

4.2.4. Pointing Offset

Each detector's pointing relative to the boresight star camera solution, averaged over the full flight, is initially characterized by maximizing the cross-correlation between single-detector SPIDER temperature maps and *Planck* maps. To reduce error, a model for each detector tile—allowing free translation, rotation, and plate scale—is fitted to the individual detector offsets.

These initial pointing offsets are refined using a time-domain "deprojection" technique based upon Bicep2 Collaboration (2015). The deprojection method involves fitting for perturbations in leading-order beam systematics—calibration, pointing offset, width, and ellipticity—using time-domain templates generated from *Planck* temperature maps and their derivatives. Unlike the BICEP2 implementation, we fit for perturbations not between paired detectors but between each detector's data and a simulation thereof, which is generated by re-observing *Planck* maps using the Gaussian beam model and initial pointing estimates from cross-correlation. The per-detector pointing offsets measured this way are consistent with the cross-correlation results but have greater precision. In addition to measuring the average pointing offset of each detector over the full flight, the average offset of all detectors in each 10-minute chunk is used to measure and correct a slow thermal/mechanical drift over the course of the flight relative to the boresight position estimated by the pointing sensors.

Deprojection fits are also used to measure per-detector calibration, beamwidth, and ellipticity. The estimated calibrations are in good agreement with those found in Section 4.2.3

but are less precise and are therefore not used. The beamwidth and ellipticity parameters are used in Section 7 to simulate the effect of systematics not accounted for in the focal plane average beam model.

4.3. Map Making

In the next stage of the analysis, the processed data are binned into two-dimensional maps of the microwave sky. The flagged and cleaned (Section 4.1), and calibrated (Section 4.2.3) data (d) can be expressed as a sum of signal and noise (n):

$$d = \boldsymbol{P}s + n, \tag{2}$$

where the signal vector s is a pixelized representation of the true Stokes parameters observed on the sky. The sparse "pointing matrix" $\boldsymbol{P}$ has non-zero elements that correspond to the Stokes pixel observed in each time sample, which is given by the boresight pointing (Section 3) and per-detector offsets (Section 4.2.4). Each non-zero element has an amplitude that varies with the per-detector polarization efficiencies and angles (Section 4.2.1) as modulated by the HWP and sky rotation. See Jones et al. (2007) for more detail on the pointing matrix elements.

To reconstruct the signal, the least-squares solution to Equation (2) is

$$\hat{s} = (\boldsymbol{P}^T\boldsymbol{N}^{-1}\boldsymbol{P})^{-1}\boldsymbol{P}^T\boldsymbol{N}^{-1}d, \tag{3}$$

where $\boldsymbol{N} = \langle nn^T \rangle$ is the covariance matrix of the noise. Because $\boldsymbol{N}$ is very large and dense, this cannot be evaluated directly without simplifying assumptions. For this result, we have applied a filter $\boldsymbol{F}$ to the data (Section 4.1), which makes the noise mostly uncorrelated between detectors and over time. Therefore, the filtered noise covariance matrix $\tilde{\boldsymbol{N}} = \langle (\boldsymbol{F}n)(\boldsymbol{F}n)^T \rangle$ is approximately diagonal. The diagonal noise variances are estimated for each detector in each 10-minute chunk of data.

The full equation used to estimate maps is

$$\hat{s} = (\boldsymbol{P}^T\tilde{\boldsymbol{N}}^{-1}\boldsymbol{P})^{-1}\boldsymbol{P}^T\tilde{\boldsymbol{N}}^{-1}\boldsymbol{F}d. \tag{4}$$

This can be evaluated very quickly because for diagonal $\tilde{\boldsymbol{N}}$, the elements of $(\boldsymbol{P}^T\tilde{\boldsymbol{N}}^{-1}\boldsymbol{P})$ and $(\boldsymbol{P}^T\tilde{\boldsymbol{N}}^{-1}\boldsymbol{F}d)$ simplify to per-pixel weighted averages. This simplification computationally enables large ensembles of time-domain simulations that model all relevant aspects of the experiment, including the effect of the filter $\boldsymbol{F}$ on information in the map (Tegmark 1997). The resulting maps use `HEALPix` pixelization[39] with $N_{\rm side} = 512$ ($\sim 6\farcm9$ resolution).

4.4. Simulated Maps

In addition to processing SPIDER data, the map maker can be run on simulated data using the same flagging, filtering, beams, pointing, and polarization angles as SPIDER. The simulated data can include signal from an input sky map, random noise generated from a power spectral density, and/or various injected glitches and systematics.

Noise simulations are generated separately for each detector, and are derived from the power spectral density of signal subtracted timestreams, averaged over all 10-minute chunks of data. As such, the fiducial noise model assumes the detector noise is stationary and uncorrelated over the course of flight. This model has been found to overestimate the true map noise in the data by $\sim$15%, due primarily to the asymmetric impact of high outliers in the sample. It is thus empirically recalibrated by pipelines that use the noise simulations (Sections 5.1 and 8.2).

The input signal for simulated skies can use a power spectrum (typically ΛCDM) as a source, where Gaussian random realizations are made with the `synfast` utility from `HEALPix`. In simulations where the input signal should represent the true microwave sky, *Planck* maps are used as an imperfect estimate, matching the SPIDER bands at 95 and 150 GHz with the similar *Planck* frequency maps at 100 and 143 GHz. Where these re-observed *Planck* maps appear in cross-spectra, different half-mission or half-ring data splits are used to avoid noise bias in the power spectra. To simulate the effects of various instrumental properties and systematics, channel parameter values (pointing, calibration, etc.) may be applied differently when simulating timestreams than when binning those simulated timestreams into a map.

Temperature-to-polarization leakage generated by the map-making pipeline is estimated by simulated observation of a *Planck* temperature-only map. The resulting polarization in this simulation, which is primarily caused by filtering of the temperature signal, constitutes a bias and is subtracted from data in the map domain. In the harmonic domain, this is an approximately 0.05 μK_2 correction in both *EE* and *BB* in units of $\ell(\ell+1)C_\ell/(2\pi)$. The frequency mismatch between the SPIDER and *Planck* bands is neglected and thus introduces a small error in this bias subtraction. We have used `Commander` (*Planck* Collaboration 2016a) temperature foreground estimates to verify that this approximation results in a relative error in the final polarization power spectra that is below 1% for each of our multipole bins. Additionally, simulations show that *E*-to-*B* leakage is negligible, measuring at most 3% of the *B*-mode error. It is nonetheless corrected for in the NSI pipeline, which is described in Section 5.2.

4.5. Maps

Figure 2 shows the temperature map observed by SPIDER's 150 GHz channels; the 95 GHz map (not shown) is visually very similar because the dominant structure is fully resolved at both frequencies. The corresponding polarization maps (Stokes Q and U) are shown in Figure 3. The rectangular outline shown over the maps encloses the region used for estimating angular power spectra, which covers 4.8% of the sky.

5. Power Spectrum Estimation

To estimate the underlying power spectrum of the sky from SPIDER's maps, we must efficiently account for the effects of finite and uneven sky coverage, distortions from TOD processing and the map maker, and the complex impact of instrumental noise. The latter is particularly challenging to model or measure with the required precision and accuracy for space- and balloon-borne experiments (Jones et al. 2007). The relatively short duration of the observations provides limited data redundancy, which poses a challenge for fully empirical noise models (e.g., BICEP2 Collaboration et al. 2014a).

We have developed two parallel power spectrum estimation pipelines for processing SPIDER maps: XFaster, which is a

[39] https://healpix.sourceforge.io

maximum-likelihood estimator; and the simpler Noise Simulation Independent (NSI) pipeline. Each pipeline begins with a set of maps that are constructed from independent subsets of SPIDER's data, from which we construct a set of cross-spectra. XFaster uses four data subsets, each combining every fourth 10-minute chunk of data (the same chunks used for low-level processing in Section 4). The NSI algorithm benefits from having a larger number of cross-spectra. This pipeline thus works with 14 data subsets composed from interleaved 3-minute chunks, with the shorter chunk length chosen to prevent gaps in sky coverage.

A common sky mask is used for all of the results in this paper. The mask covers 1964 square degrees with uniform weighting, consisting of the 1992 square degree rectangle shown in Figures 2 and 3 with point sources removed. The point-source mask excludes 1° diameter circular regions around objects from the *Planck* compact object catalog (*Planck* Collaboration et al. 2016a), plus a 2° region around the bright radio galaxy NGC 1316. The 50 brightest sources in SPIDER's observation region are masked this way, though not all lie within the chosen rectangle. Among a handful of simple mask options, this mask was the largest subset of the data that was well-conditioned and passed null tests. This mask was established prior to the calculation of the signal power spectra to avoid potential bias. Apodization of the mask was found to negligibly affect signal and null spectra for both simulations and data, and thus the mask is unapodized to maximize sky coverage. Null test and signal power spectra are computed with both pipelines, and the results are shown in Section 6.

Each pipeline ultimately produces a spectrum and covariance matrix for $33 \leqslant \ell \leqslant 257$, binned into nine "science" bandpowers with an ℓ width of 25. One lower ($8 \leqslant \ell \leqslant 32$) and two higher bandpowers are also computed for each pipeline to accurately account for their leakage into the nine bins used for cosmological analysis. The bin starting at $\ell = 8$ was found to contain residual systematic signal, and the bins above $\ell = 257$ contribute little to the cosmological and foreground constraints; thus, they are excluded from the science bins. Throughout the following, unless explicitly stated otherwise, the lowest or first bin refers to the first science bin; i.e., that starting at $\ell = 33$. In all plots shown, the XFaster bin centers are a weighted average of the multipoles in each bin, with weights given by the bandpower window functions (Gambrel et al. 2021). The NSI band centers are an unweighted average of the multipoles in the bin. The details that distinguish the two pipelines are provided below.

5.1. XFaster Pipeline

XFaster is a maximum-likelihood estimator that is built as a hybrid of Monte Carlo estimators, such as MASTER (Hivon et al. 2002) and PolSpice (Chon et al. 2004), and iterative quadratic estimators (Tegmark & de Oliveira-Costa 2001; Rocha et al. 2011). It was developed to allow the application of a maximum-likelihood estimator to maps made with disjoint masks. XFaster is based on an algorithm that was originally written for analysis of the BOOMERanG data set (Netterfield et al. 2002; Jones et al. 2006), and has since been employed as one of the estimators used in the *Planck* analysis (Rocha et al. 2011). A number of new features were implemented for use in the SPIDER pipeline, notably the ability to calculate null spectra. The main features are summarized here, with further details left to a dedicated paper (Gambrel et al. 2021).

XFaster iteratively solves for bandpower deviations from a fiducial full-sky signal model using an approximation for the likelihood of cut-sky $a_{\ell m}$ modes. This signal model is constructed using the MASTER formalism (Hivon et al. 2002), in which the mode-mixing from the mask is computed analytically (including an *E-B* mixing component), the beams are precomputed as described in Section 4.2.3, and the filter transfer functions are estimated from an ensemble of 1000 ΛCDM simulations that are run through the full map-making pipeline (Section 4.4). In addition to the signal power, XFaster also estimates the instrumental noise from the auto- and cross-spectra of the input maps. An ensemble of 1000 time-domain noise simulations is input to the pipeline to provide a fiducial noise model. The noise model is itself iteratively recalibrated by including deviations from the fiducial model as parameters in the likelihood maximization alongside the signal bandpowers. The noise residual parameters are modeled as the same for *EE* and *BB*.

In this pseudo-$a_{\ell m}$ space, the full likelihood of the observed data modes $\widetilde{\boldsymbol{d}}$, given the signal model $\widetilde{\boldsymbol{S}}$ and noise, is approximated, for a single map and for a single spectrum, as

$$-2\ln L(\widetilde{\boldsymbol{d}}|\widetilde{\boldsymbol{S}}) = \sum_{\ell} g_{\ell}(2\ell+1)\left[\frac{\widehat{C}_{\ell}}{\widetilde{S}_{\ell}+\widetilde{N}_{\ell}} + \ln(\widetilde{S}_{\ell}+\widetilde{N}_{\ell})\right], \tag{5}$$

where $\widehat{C}_{\ell}$ is the data pseudo-spectrum, g_{ℓ} is an effective mode-counting factor, $\widetilde{S}_{\ell}$ is the estimated signal, and $\widetilde{N}_{\ell}$ is the noise bias estimated from the mean of noise-only simulation ensembles. This approximation assumes that the signal and noise components are uncorrelated. The likelihood is diagonalized by assuming that by binning power into bandpowers of sufficient width, the effect of correlations between multipoles on the estimate is greatly reduced. In practice, the components of Equation (5) are matrices of all cross-spectra among all maps used for the analysis. The structure of the generalized likelihood is band-diagonal to account for correlations between spectral combinations and overlapping maps.

The vector g_{ℓ} is a recalibration of the effective mode count. This corrects the likelihood, and the Fisher matrix obtained from it, for the effects of masking, filtering, and diagonalizations. The recalibration is computed using an ensemble of simulations. Because this factor differs based on whether the spectrum is signal- or noise-dominated, we compute it separately for signal and null spectra. For signal spectra, we use signal-only simulations, while for null spectra we use simulations that include signal and noise. The g_{ℓ} factor is described in more detail in Gambrel et al. (2021).

The iterative estimate of bandpower deviations on the signal and noise model automatically produces the Fisher information matrix, whose inverse is the bandpower covariance. Extensive simulations have been performed to ensure the XFaster estimator is unbiased and that the resulting covariance matrix is accurate. These are discussed further in Gambrel et al. (2021).

5.2. NSI Pipeline

The NSI pipeline was developed to estimate statistical bandpower errors directly from the data. It uses the cross-spectra of 14 temporally independent maps at each observing frequency, generated from interleaved 3-minute data chunks. All possible cross-spectra are constructed from the map

ensemble (neglecting the auto-spectra), providing 91 at each single frequency (95×95 GHz or 150×150 GHz) and 196 with one map at each frequency (95×150 GHz), for a combined total of 378 cross-spectra. The bandpowers are estimated from the noise-weighted mean of all cross-spectra. The associated statistical uncertainties are estimated from the distributions of these cross-spectra by computing the standard error on the mean with jackknife resampling. By using many cross-spectra of uncorrelated maps, the sensitivity of the NSI pipeline approaches that of an auto-spectrum analysis, but the resulting bandpower estimates are not biased by any mischaracterization of the noise auto-spectrum. This methodology is similar to Xspect (Tristram et al. 2005) and Xpol (Tristram 2006), and similar approaches have been used by several experiments including SPT (Lueker et al. 2010) and CLASS (Padilla et al. 2020) as well as for SPIDER's circular polarization analysis (Nagy et al. 2017).

The cross-spectra are computed with PolSpice (Chon et al. 2004), using cosine apodization with $\sigma = 10°$, $\theta_{\rm max} = 50°$, $\ell_{\rm max} = 700$, and `symmetric_cl`. These choices were verified with simulations to produce unbiased and quick results for our sky region, with little coupling between our binned bandpowers.

The NSI pipeline uses a two-dimensional "transfer matrix" to correct the power spectra for mode mixing and power attenuation from filtering. This approach considers the leakage from a given multipole bin to all others, both within the same spectrum and between spectral types (*TT*, *EE*, *BB*). By design, this procedure also includes a correction for the instrumental beam, as well as any residual leakage effects induced by the cut-sky mask that are not corrected by the spectral estimator. Though found to be negligible for SPIDER, it also includes a correction for *E*-to-*B* leakage. The transfer matrix is constructed from a simulation ensemble in which each simulated map (Section 4.4) has a source spectrum with only one non-zero multipole bin, which is set to the value of the appropriate fiducial ΛCDM spectrum. When the maps are processed with the NSI pipeline, the ratio of the output and input spectra encodes the leakage from that bin to all others. Further discussion of this method and its impact on the recovered spectra is provided in J. S.-Y. Leung et al. (2021, in preparation).

6. Consistency Tests

SPIDER's data processing is designed to flag or filter out the dominant sources of systematic contamination in the time-ordered data, such as intermittent pickup and quasi-stationary noise. To ensure that low-level residuals do not remain at the level of our sensitivity, we conduct two types of tests: null tests (described in Section 6.1), and simulated injections of modelable systematics (described in Section 7). We also discuss the consistency between the two power spectrum estimation pipelines in Section 6.2.

6.1. Null Tests

Null tests check for systematic noise residuals in the differences between pairs of maps, which are constructed from various splits of SPIDER's data by time period or detector set. The pairs of maps are chosen to share a common signal but to have independent noise, and to maximize the residuals from possible systematic effects within the data. If the power spectra of these differences are consistent with statistical noise, then we have evidence that systematic errors probed by the splits do not significantly contaminate the maps. These tests are performed separately for both the NSI and XFaster pipelines.

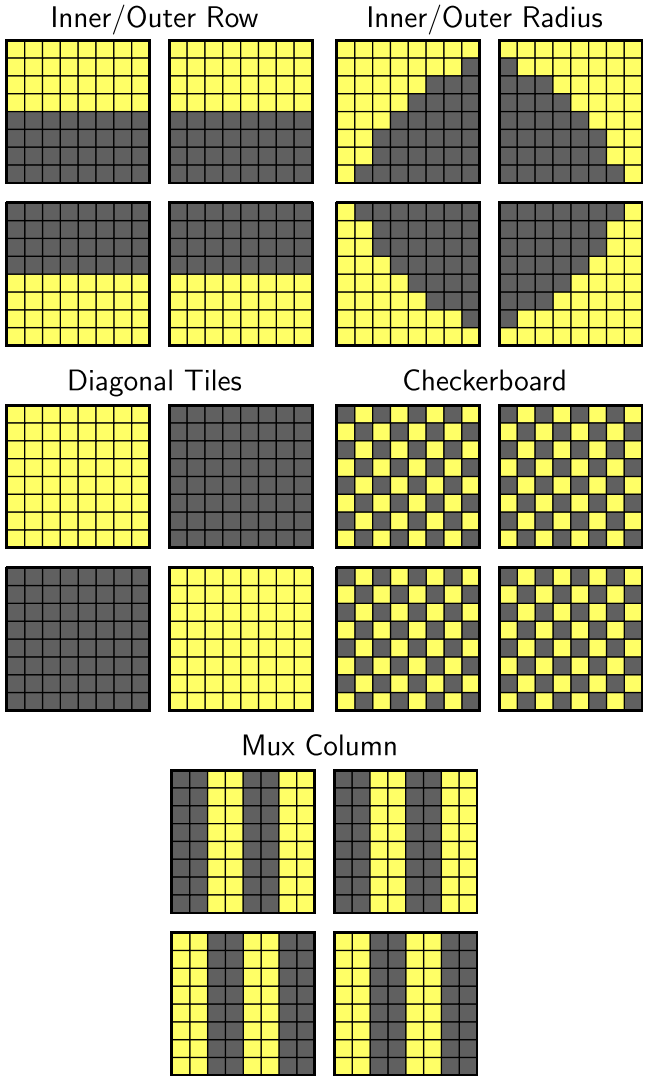

Figure 5. Physical representation of coordinate-based detector splits used for null tests. Each square represents a pixel, each of which contains two orthogonally polarized detectors. Focal plane maps shown are for 150 GHz; 95 GHz splits look similar but with 6×6 grids of pixels per tile.

6.1.1. Null Split Definitions

The suite of null splits is listed below. The first five splits are based on channel location within the focal plane, as illustrated in Figure 5. Two further splits employ alternate divisions of the detectors: by pointing relative to payload azimuth (dependent on the orientation of each receiver about its boresight axis) and by frequency response. The final three splits are by time throughout the mission.

1. *Inner/Outer Focal Plane Rows:* Split by physical detector row on the focal plane. This split also divides detectors according to their location in the multiplexed readout. We have observed RF-coupled interference (including the reaction-wheel synchronous noise) that is more prevalent in the inner focal plane rows, which makes this split sensitive to any related residuals.

2. *Inner/Outer Focal Plane Radius:* Split by detector distance from the center of the focal plane. This probes beam shape, which becomes increasingly elliptical toward the focal plane edges.
3. *Diagonal Tiles:* Split focal planes into sets of two tiles located diagonally across from each other. Since each tile is fabricated independently, this tests for detector non-uniformity in fabrication.
4. *Checkerboard:* Split the square grid of detectors in a checkerboard pattern (not splitting orthogonally polarized pairs). This test probes the noise model because we do not expect any instrumental systematics to vary on this basis.
5. *Alternating Mux Column:* Split every other readout column. This probes differences in the detector bias and SQUID readout among columns.
6. *Port/Starboard Detectors:* Split by detector pointing azimuth to difference detectors located on the port and starboard side of a given receiver. This is sensitive to sidelobe pickup from the Galaxy and Sun, which are on opposite sides of the azimuthal scan.
7. *Band Center:* Split all detectors at a given observing frequency by their measured band center. The sets have mean band center differences of 2 and 4 GHz at observing frequencies 95 and 150 GHz, respectively. This probes SPIDER's sensitivity to differential responsivity to Galactic dust between orthogonally polarized detectors, which could bias the foreground and cosmological results.
8. *Left/Right Scan:* Split azimuthal scans into left-going and right-going. This probes time constant effects.
9. *Alternating Days:* Split into every other day. This probes HWP systematics because full polarization angle coverage for a given detector requires four independent HWP angles, which corresponds to two days of observing time.
10. *Early/Late Flight:* Split each of the two scan strategies into early and late halves. This probes longer trends, such as effects from cryogen loss.

6.1.2. Processing

For each half of a null split, a data map is made with the standard processing pipeline described in Section 4. A simulated re-observed *Planck* map is also made for each SPIDER half-data map, using the same detector/time split. The *Planck* maps provide estimates of the expected null signal residual because they capture both CMB and foreground power; the latter is consistent with being the dominant source of signal residual power at large scales in the SPIDER null maps. For most null splits, the foreground and CMB signal residuals are small compared to the noise. However, they are particularly significant for the Port/Starboard 150 GHz null split, in which the lowest bin's residual null power is reduced by a factor of two when accounting for foregrounds. To perform a null test, we first subtract the simulated *Planck* maps from the SPIDER half-data maps, and we then use the difference between the two halves to form a null map. The power spectra and covariances are estimated from the null maps with the XFaster and NSI pipelines.

XFaster uses 500 signal and noise simulations per null split, unlike the 1000 simulations that are used for signal power spectra. Additionally, the mode-loss factor, g_ℓ, is determined differently for null spectra. It is computed for each null using signal and noise simulations, rather than signal-only simulations (as described in Section 5.1). This procedure has been validated with simulations. NSI null spectra are computed in the same way as signal power spectra (as described in Section 5.2).

6.1.3. Null Test Results

Both pipelines are used to construct *EE*, *BB*, and *EB* spectra. χ^2 values are computed for each test over the nine ℓ-bins and expressed as probability-to-exceed (PTE) values in Table 2. Example null spectra are shown in Figure 6.

For a large ensemble of uncorrelated, noise-dominated spectra, the PTE values are expected to be uniformly distributed between 0 and 1, with no extreme outliers at either end of the distribution. However, non-negligible correlations among the null splits are known to exist because they share detector samples. Thus, the distributions of the null spectra are evaluated using simulations incorporating these correlations. For XFaster, null bandpowers and covariance matrices are computed from 500 simulated signal and noise maps, each seed of which naturally incorporates correlations across null splits. For NSI, simulated null bandpowers are instead generated from realizations of the covariance matrix among the null test bandpowers, which is itself estimated from the ensemble of chunk cross-spectra. Two tests are conducted on the resulting null statistic distributions: the first probing the distributions' outliers, and the second probing their shapes. We require that each test results in a PTE for the observed data of at least 1%.

In the outlier test, we count how many simulations have a largest χ^2 at least as high as the largest χ^2 measured for the data. The results are shown in Table 3 in the Outlier Test column, with both pipelines passing this test.

The NSI distribution shape test is computed as a single "combined χ^2" from a single covariance matrix including all of the null tests. The resulting value is compared to simulations drawn from the combined covariance matrix, with a p-value calculated as the fraction of simulations with a higher combined χ^2 than the data. Because the combined covariance matrix is 270×270 ($270 = 10$ splits $\times$ 3 spectra $\times$ 9 bins), only the combined frequency case with 378 cross-spectra is sufficient to compute it. Thus, the individual frequency χ^2 distribution tests rely on the XFaster result.

The XFaster distribution shape test is performed using Kolmogorov–Smirnov (KS) tests. A KS-test p-value for the data is computed by comparing the 30 data χ^2 values (10 null splits $\times$ 3 spectra) to the distribution of 150,000 simulated χ^2 values (5000 draws $\times$ 10 null splits $\times$ 3 spectra) from random bandpower draws from the data covariance matrix. This exercise is then repeated for each of the 500 simulations. This gives a p-value per simulation, each of which intrinsically includes the effect of correlations between null splits. We then determine the number of simulations with a lower p-value than the data, and find that all frequency combinations pass our threshold of 1%, as shown in the Distribution Test column of Table 3.

We note that the data collected by one of the 150 GHz receivers are excluded from this null test analysis, and from all results shown in this paper. Their inclusion leads to an unlikely χ^2 distribution, with a Distribution Test PTE of 0.002 for both XFaster 150 GHz and NSI Combined. While the NSI pipeline cannot compute the full null test covariance matrix for a single frequency, the excluded 150 GHz data has enough impact that even when combined with 95 GHz data, the PTE fails to exceed

Table 2
Null Test Results Expressed as Probability-to-exceed (PTE) Values, Assuming a χ^2 Distribution for Each

	XFaster			NSI		
	95 GHz	150 GHz	Comb	95 GHz	150 GHz	Comb
Inner/Outer Focal Plane Rows						
EE	0.06	0.52	0.14	0.18	0.52	0.09
BB	0.007	0.14	0.10	0.08	0.05	0.02
EB	0.41	0.77	0.85	0.59	0.79	0.72
Inner/Outer Focal Plane Radius						
EE	0.67	0.86	0.98	0.65	0.94	0.99
BB	0.02	0.92	0.41	0.11	0.29	0.22
EB	0.15	0.34	0.82	0.33	0.12	0.95
Diagonal Tiles						
EE	0.37	0.16	0.18	0.35	0.10	0.07
BB	0.80	0.90	0.82	0.28	0.53	0.20
EB	0.05	0.83	0.33	0.09	0.27	0.15
Checkerboard Detectors						
EE	0.15	0.64	0.42	0.26	0.54	0.24
BB	0.02	0.32	0.49	0.05	0.16	0.60
EB	0.01	0.39	0.66	0.03	0.38	0.35
Alternating Mux Columns						
EE	0.62	0.06	0.61	0.75	0.17	0.40
BB	0.33	0.33	0.03	0.92	0.13	0.12
EB	0.48	0.62	0.72	0.38	0.43	0.67
Port/Starboard Pointing Detectors						
EE	0.82	0.15	0.66	0.50	0.003	0.46
BB	0.53	0.17	0.20	0.51	0.004	0.009
EB	0.57	0.35	0.32	0.61	0.20	0.25
Band Center						
EE	0.76	0.39	0.92	0.98	0.24	0.93
BB	0.19	0.18	0.88	0.23	0.16	0.43
EB	0.51	0.72	0.73	0.68	0.62	0.61
Left/Right Scan						
EE	0.66	0.61	0.62	0.51	0.57	0.37
BB	0.54	0.02	0.08	0.70	0.009	0.03
EB	0.94	0.08	0.62	0.71	0.06	0.57
Alternating Days						
EE	0.04	0.24	0.68	0.07	0.24	0.84
BB	0.04	0.66	0.27	0.24	0.59	0.37
EB	0.94	0.05	0.83	0.47	0.24	0.92
Early/Late Flight						
EE	0.78	0.008	0.69	0.73	0.15	0.88
BB	0.08	0.14	0.14	0.29	0.34	0.50
EB	0.07	0.43	0.42	0.06	0.04	0.21

Note. Results are shown for each polarization spectrum and pipeline, as well as for each frequency band and for the combined data set ("Comb"). While these PTE values are useful for evaluating individual tests, correlations affect their interpretation across multiple tests (see Section 6.1.3 and Table 3). The color scale extends from 0 (dark) to 1 (light) to draw attention to more unlikely PTEs at either end of the range.

our a priori 1% requirement. These failures do not originate from a single null test. Rather, many χ^2 values increase significantly. Although this failure mode makes it difficult to assign blame to a single systematic effect, the excluded receiver was uniquely susceptible to noise correlated with the reaction-wheel angle.

6.2. Raw Spectrum Comparison

While the null tests provide important consistency checks on the SPIDER data, the two independent power spectrum estimation pipelines also provide an important consistency check on the methodology. Figure 7 shows the power spectra from both pipelines for the full SPIDER data set. The bandpower error bars for both pipelines include only the instrumental noise contribution (no sample variance) for ease of comparison. Since no foreground cleaning has been applied to these raw power spectra, excess power over the ΛCDM model is expected, particularly on large angular scales. Small differences are observed in the bandpowers from each pipeline, particularly in the lowest multipole bins of the *TT* and *TE* spectra. These derive from differences in the estimators, notably effects of the different bandpower window functions and transfer functions. These are handled consistently in the likelihood analysis described in Section 9. The amplitudes of the *TT* differences are consistent with those seen in signal simulations.

7. Systematic Error Budget

We use an ensemble of time-domain simulations to study the impact of various instrumental systematic effects on our ability to constrain a *B*-mode signal. We consider eight classes of systematic effect that were previously identified as relevant for SPIDER (Fraisse et al. 2013), spanning optical non-idealities, calibration errors, and electrical cross-talk among detectors. For each of these, we re-observe simulated sky maps with the full filtering and map-making pipeline (Section 4.4) while injecting systematic signal into the time-ordered data. The simulated sky maps are generated from the *Planck* best-fit ΛCDM power spectrum (*Planck* Collaboration 2016b), but with no input *B*-mode power. Any output *BB* power is thus ascribed to systematic effects. Figure 4 shows the impact of these effects on the *B*-mode power spectrum at 150 GHz. The result for 95 GHz is qualitatively similar. Of the eight systematic effects considered, none are large enough to meaningfully impact SPIDER's measured *B*-mode power spectrum. We rely on our null tests to provide limits on the contribution of known, and unknown, sources of systematic error, including those that cannot be reliably simulated.

We start by considering an offset in the detectors' polarization orientation angles, which causes *E*-mode sky signal to be misinterpreted as *B*-modes. We simulate this by introducing a 0°.5 common (shared among all detectors) offset between the polarization angles used during re-observation and map-making. This simulated offset is taken from the per-detector error determined in preflight characterization (Section 4.2.1). Note that a polarization error that varies among detectors would generally average down to less net effect, so the use of a common offset is conservative. The simulations indicate no significant contamination, even for this pessimistic case.

To investigate deviations from the idealized Gaussian beam model, we consider beams derived from Physical Optics (PO) simulations of the full SPIDER telescopes. This model was developed using the GRASP software package (Ticra Tools 2020), and includes variations in beamwidth, non-Gaussian shape, and cross-polarization response across the

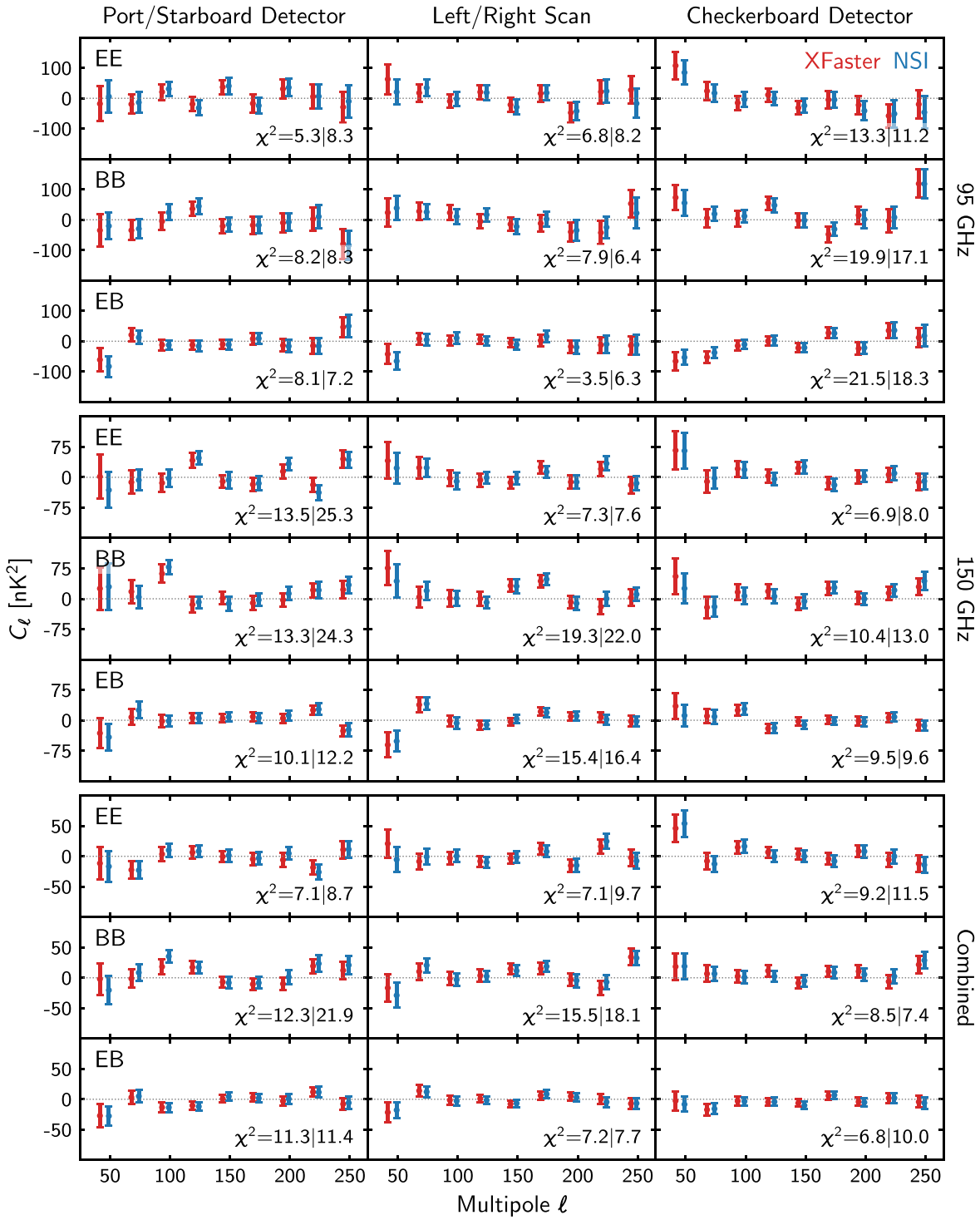

Figure 6. Three null tests, showing the comparison between the XFaster and NSI pipelines for the different frequency combinations. For visual clarity, the NSI band centers are horizontally offset by $\Delta\ell = 4$, while XFaster is left unchanged. The same ℓ offsets are applied to Figures 7 and 8. χ^2 values are computed independently for each pipeline and spectrum. See Table 2 for the corresponding probability-to-exceed (PTE) value for each test.

focal plane. These PO beam models are truncated to an angular extent of 2.°5; therefore, they do not constrain the impact of extended sidelobes. We further explore the impact of large-amplitude sidelobe response using a conservative model of the SPIDER beam derived from advanced optical simulations. These beam models are convolved with simulated skies using the `beamconv` algorithm (Duivenvoorden et al. 2019), which allows efficient generation of detector timestreams for these more general beam models. To probe far-sidelobe coupling to the Galaxy, we add `Commander` (*Planck* Collaboration 2016a) dust and synchrotron foreground templates to the CMB signal outside the nominal SPIDER observation region.

We further use `beamconv` to simulate the instrument's response to two more complex effects: reflective ghosts within the optical system (which are modeled as 1% of each main beam's magnitude, comparable to laboratory tests and in-flight

Table 3
Null Test Ensemble Results, Given as the Proportion of Simulations with Worse Test Statistics than the Data

	Outlier Test PTE		Distribution Test PTE	
Band	XFaster	NSI	XFaster	NSI
95 GHz	0.38	0.80	0.07	N/A
150 GHz	0.36	0.20	0.25	N/A
Combined	0.78	0.34	0.55	0.50

Note. The outlier test probes the number of simulations with at least as large of a χ^2 as the largest data χ^2. The distribution test probes the shape of the null statistics across all tests, accounting for correlations among similar null splits. For the XFaster pipeline, this is computed as a KS-test using simulations to calibrate the bias due to correlations, and for NSI it is computed as a single "combined χ^2".

estimates) and cross-talk between detectors (which are modeled as 1% between channels in adjacent multiplexing rows; lab testing is consistent with <0.5%, as expected from the readout system de Korte et al. 2003; BICEP2 Collaboration et al. 2014b). The results of these simulations are shown in the upper panel of Figure 4—they all have negligible impact.

In addition to the PO beam investigation, we simulate other residuals estimated as part of the time-domain analysis. Detector gain drifts over time are injected based on the measurements from TES resistance (Section 4.2.3). Per-detector deviations from the simple beam model are injected with best-fit deprojection templates (Section 4.2.4) for beamwidth and ellipticity. These simulations inject the full estimated effects and assume that no attempt is made to correct for them. This assumption is conservative for gain drifts, where the estimated amplitude is corrected for in the data analysis. Still, as shown in the lower panel of Figure 4, these beam and gain systematics are of negligible importance.

8. Component Separation

Measurements of CMB polarization, particularly on large angular scales, are complicated by the need to model and remove diffuse Galactic emission. Modeling the Galactic signal represents one of the most challenging obstacles in characterizing the polarization of the CMB. We implement a variety of methods to disentangle the Galactic and cosmological signals, each subject to different assumptions, and assess their consistency. This paper focuses on the CMB component estimate, while detailed discussion of foreground components is left for future work.

In Section 8.1 we present the template-subtraction method that serves as the nominal foreground-removal technique for the cosmological results in this paper. Section 8.2 describes an implementation of the SMICA component separation method (Cardoso et al. 2008) on the SPIDER data. These two methods —the former map-based, the latter operating in the harmonic domain—are both used to compute foreground-cleaned spectra, and are propagated to r constraints in Section 9. Both analyses assume that the emission from interstellar dust is the only polarized foreground in the SPIDER data. We use another harmonic-space method, which is described in Section 8.4 and builds upon the Choi & Page (2015) analysis, to assess the contribution of the Galactic synchrotron emission to the polarized signal measured by SPIDER, finding it negligible for the present purpose.

8.1. Template Subtraction

Under the assumption that the spatial morphology of the polarized emission from interstellar dust is frequency-independent, this Galactic component can be projected out of a map at a given frequency by fitting a scalar amplitude to a morphological template of the emission. This approach has been successfully applied in the analysis of data from WMAP and *Planck*, and it has been studied in the context of future orbital CMB missions (see the review by Delabrouille & Cardoso 2007 as well as Dunkley et al. 2009; Katayama & Komatsu 2011; Efstathiou & Gratton 2019).

8.1.1. Implementation

We model the polarized intensity measured by SPIDER in a given pixel as

$$S_\nu = S^{\rm CMB} + A_{\nu,\nu_0} S_{\nu_0}^{\rm dust} + n_\nu, \tag{6}$$

where S is a Stokes parameter, ν is the frequency of the map, ν_0 is the frequency of the dust-dominated map used to build the template, A_{ν,ν_0} is a scalar amplitude, and n_ν is the map noise. We construct dust-template maps, $S_{\nu_0}^{\rm t}$, from the *Planck* data by subtracting the 100 GHz map from a map at higher frequency $\nu_0 = 217$ or 353 GHz, where dust emission dominates. With the notation in Equation (6),

$$S_{\nu_0}^{\rm t} = S_{\nu_0} - S_{100} = (1 - A_{100,\nu_0}) S_{\nu_0}^{\rm dust} + n_{\nu_0}^{\rm t}, \tag{7}$$

where $n_{\nu_0}^{\rm t} \equiv n_{\nu_0} - n_{100}$ is the template noise.

A SPIDER map S_ν is cleaned by subtracting from it a dust template $S_{\nu_0}^{\rm t}$ multiplied by a scalar α. With the notation above,

$$\begin{aligned} S_\nu^{\rm cleaned} &= S_\nu - \alpha S_{\nu_0}^{\rm t}, \\ &= S^{\rm CMB} + (A_{\nu,\nu_0} - \alpha[1 - A_{100,\nu_0}]) S_{\nu_0}^{\rm dust} \\ &\quad + n_\nu - \alpha n_{\nu_0}^{\rm t}. \end{aligned} \tag{8}$$

For each SPIDER map frequency and choice of dust template, the NSI pipeline finds the value of α that minimizes the summed power in the lowest three multipole bins ($33 \leqslant \ell \leqslant 107$) of the cleaned *EE* spectrum. In the XFaster pipeline we fit α and r simultaneously in the likelihood using all nine multipole bins of the *EE* and *BB* spectra.

Finally, we note that two versions of each template are constructed, each using data from only one *Planck* half-mission. This allows both the NSI and XFaster pipelines to compute the required template-subtracted spectra as cross-spectra between two maps with independent template noise. This eliminates the significant noise bias that would come from the noise auto-spectrum of a full-mission dust template, albeit at the cost of an increase in template noise.

8.1.2. Template-subtraction Results

Table 4 gathers the values of the fitting parameter α measured by the NSI and XFaster pipelines at 95 and 150 GHz for two independent dust templates that are derived from the *Planck* 217 and 353 GHz maps (Equation (7)).[40] For NSI, the

[40] We use effective band centers for the relevant *Planck* maps of 101.3, 220.6, and 359.7 GHz, which are computed from the spectral response functions of *Planck*'s polarization sensitive bolometers for a flat-spectrum source (*Planck* Collaboration et al. 2014).

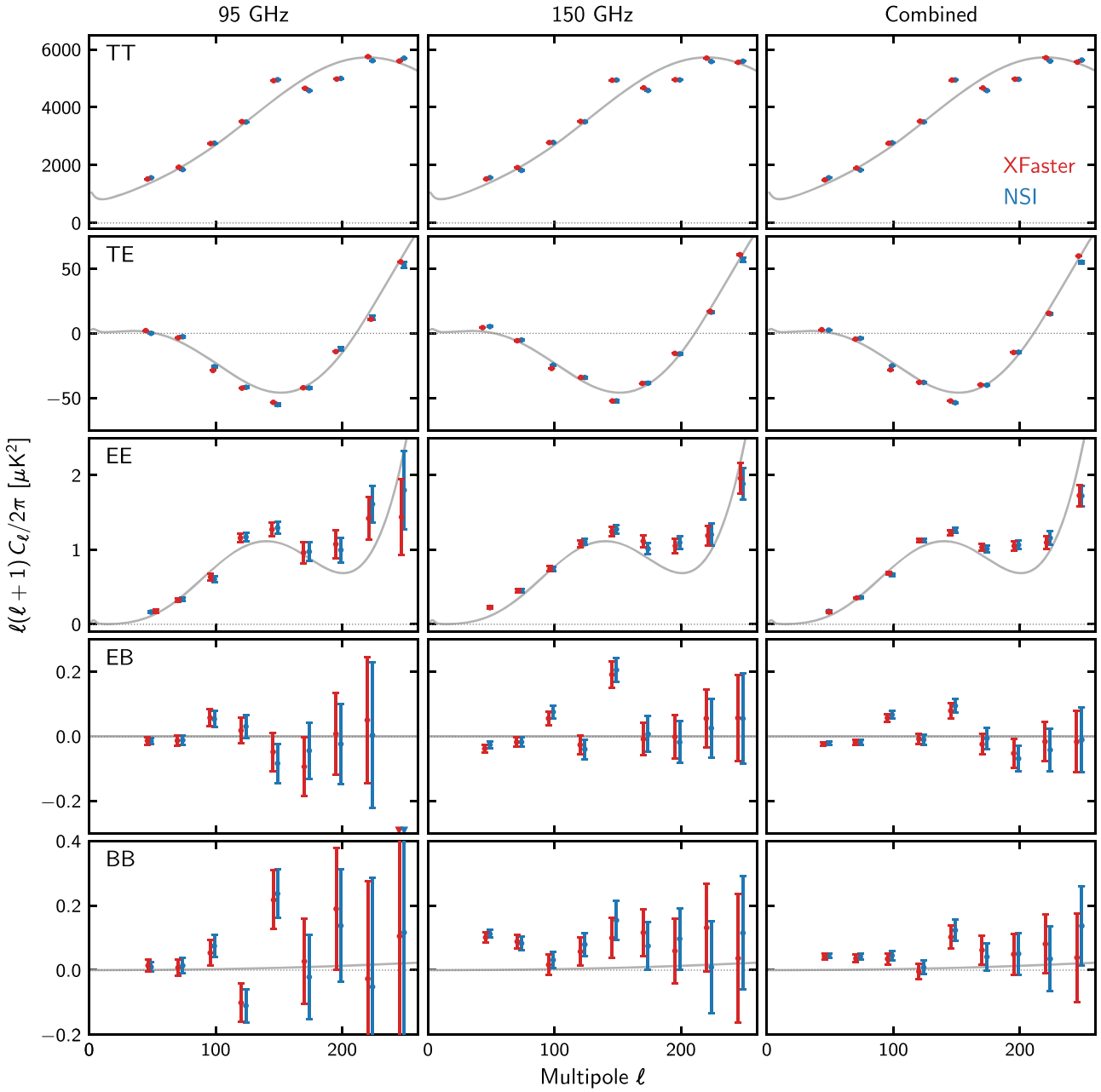

Figure 7. Raw power spectra from the XFaster and NSI pipelines. Spectra are computed for each of the two frequency bands individually in the left-hand and middle columns, and the combined best estimate spectrum from both frequencies is shown in the right-hand column. The best-fit *Planck* ΛCDM power spectrum from *Planck* Collaboration et al. (2016b) is shown in gray. The error bars do not include sample variance because this allows us to better compare the instrumental noise estimates between the two power spectrum pipelines. The *TB* spectrum is not used in this cosmological analysis and is thus omitted.

error in α for the 353–100 GHz template is dominated by the contribution from chance correlations between the dust and CMB *EE* components. Because XFaster includes both *EE* and *BB* spectra in the α fits, this chance correlation is subdominant. Instead, the measurement of α_{95} (the value of α appropriate to cleaning a 95 GHz map) is limited by SPIDER noise, while the error on α_{150} is contributed in equal parts by SPIDER noise and *Planck* noise in the template. For the 217–100 GHz template, the template noise is a larger contributor to the α error at both frequencies; it nearly equals the SPIDER noise contribution at 95 GHz and is 3 times more significant than SPIDER noise at 150 GHz. The use of both *EE* and *BB* spectra over the full multipole range accounts for XFaster's significantly smaller uncertainty in the determination of α (by a factor of 2 for $\nu_0 = 353$ GHz, and 3.5–4 for $\nu_0 = 217$ GHz) compared to NSI.

As can be inferred from Table 4, we detect polarized dust emission at high significance at both 95 and 150 GHz. The NSI and XFaster pipelines provide consistent α values, which are also broadly consistent with *Planck* expectations. Assuming a modified-blackbody dust spectral energy distribution (SED) with temperature $T_d = 19.6$ K, *Planck* Collaboration et al. (2020d) find the polarized emission from interstellar dust over 71% of the high-Galactic-latitude sky to be consistent with a dust spectral index $\beta_d = 1.53 \pm 0.02$, corresponding to the α values reported in the *Planck* columns of Table 4. All SPIDER values are within 2σ of those estimated by *Planck*.

Table 4
XFaster- and NSI-fitted Values of the Dust-template Frequency Scaling Factor at 95 GHz (α_{95}) and 150 GHz (α_{150}) for $\nu_0 = 353$ GHz and $\nu_0 = 217$ GHz (Equation (8))

	$10^3\ \alpha_{95}$	$10^3\ \alpha_{150}$	β_d^{95}	β_d^{150}
Template: $\nu_0 = 353$ GHz				
Planck	16.8±0.5	44.4±0.8	1.53±0.02	
XFaster	18±2	45±2	$1.49^{+0.07}_{-0.09}$	1.52±0.05
NSI	19±5	45±4	$1.44^{+0.22}_{-0.17}$	1.51±0.10
Template: $\nu_0 = 217$ GHz				
Planck	153±3	404±4	1.53±0.02	
XFaster	159±17	377±16	$1.51^{+0.10}_{-0.12}$	$1.68^{+0.08}_{-0.09}$
NSI	140±50	350±58	$1.63^{+0.46}_{-0.31}$	$1.81^{+0.38}_{-0.31}$
SMICA				
FFP10	⋯	⋯	1.43±0.04	
Auto-Cross	⋯	⋯	1.50±0.04	

Note. Assuming a modified-blackbody dust SED with temperature $T_d = 19.6$ K (*Planck* Collaboration et al. 2020d), we derive from each α value the dust spectral index β_d, which can be compared directly to the SMICA measurement of this parameter and to its value derived by *Planck* over 71% of the high-Galactic-latitude sky. For ease of comparison, we also report the α values at SPIDER frequencies expected from the dust spectral index derived by *Planck*. The SMICA recovered value depends on the noise model at 353 GHz. We explore two options: the first assumes a noise model that is the ensemble average of FFP10 simulations, while the second assumes a noise model from the difference between the auto-spectrum of the full-mission map and the cross-spectrum of the two half-mission maps.

8.2. SMICA

Spectral matching independent component analysis (SMICA; Delabrouille et al. 2003; Cardoso et al. 2008) is a harmonic-space component separation technique. In brief, this technique involves the calculation of the cross-spectra that preserve the joint correlation structure between the input maps. The power in these spectra is then partitioned among individual components based on their spectral shape. The fitted spectral components uniquely determine the weight assigned to each map and allow recovery of component-separated maps. The formalism behind SMICA has been discussed in other publications, therefore we summarize the method as implemented for SPIDER.

The SMICA pipeline is highly complementary to the template methods that were used earlier. As implemented in this work, SMICA adopts a rigid model for the spectral energy density of the dust foreground but, unlike the template methods, it assumes relatively little about its spatial morphology. This modeling flexibility comes at the cost of a larger number of fit parameters for a given set of input data. The SMICA pipeline also enables a fully consistent joint analysis of the *Planck* and SPIDER data.

8.2.1. Implementation

SPIDER's implementation of SMICA takes as input $N_{\rm chan}$ polarized maps and uses PolSpice to compute the spectral covariance matrix $\widehat{\boldsymbol{R}}_b$: a $2N_{\rm chan} \times 2N_{\rm chan}$ matrix for each bandpower b, gathering all possible combinations of binned *EE* and *BB* auto and cross pseudo-spectra. We then construct a parameterized model covariance $\widetilde{\boldsymbol{R}}_b$, also in pseudo-spectrum space, that accurately describes the data $\widehat{\boldsymbol{R}}_b$.

While not required by the approach, in this work we assume that the dust polarization amplitude follows a modified-blackbody frequency scaling whose index is scale-independent and identical in *EE* and *BB*. The model for polarization spectrum X ($X \in \{EE, BB\}$) is as follows:

$$\widetilde{\boldsymbol{R}}_b^X(\theta) = \widetilde{\boldsymbol{N}}_b^X + \sum_{b'} \boldsymbol{J}_{b,b'}[f_{b'}^X(\beta_d) P_{b'}^X f_{b'}^X(\beta_d)^T + C_{b'}^X], \quad (9)$$

where $f_b^X(\beta_d)$ is a vector of size $N_{\rm chan}$ that captures the dust amplitude scaling in the map domain, P_b^X the full-sky bandpowers of the dust at a fixed reference frequency, C_b^X the full-sky CMB bandpowers, and $\widetilde{\boldsymbol{N}}_b^X$ is a $N_{\rm chan} \times N_{\rm chan}$ matrix representing the auto-correlated noise terms for all inputs. These are all free parameters fit in the model, which are notated together by θ for brevity. The transfer matrix, as introduced in Section 5.2, is represented by $\boldsymbol{J}_{b,b'}$ and is applied only to the terms that contribute to the sky signal. This matrix includes effects from filtering, beam smoothing, and the mode-coupling kernel from the mask.

The model is fit by finding optimal parameters θ that minimize the spectral mismatch between the data $\widehat{\boldsymbol{R}}_b$ and the model $\widetilde{\boldsymbol{R}}_b(\theta)$. This optimization fits all parameters simultaneously to account for covariance between bins and different components. The test statistic is the Kullback–Leibler divergence between data and model:

$$-2\ln L = \sum_b w_b\, \mathrm{Tr}\,[\widehat{\boldsymbol{R}}_b \widetilde{\boldsymbol{R}}_b^{-1}(\theta) - \ln(\widehat{\boldsymbol{R}}_b \widetilde{\boldsymbol{R}}_b^{-1}(\theta))], \quad (10)$$

$$w_b = \sum_{\ell \in b} (2\ell + 1) f_{sky}. \quad (11)$$

The component-separated bandpowers are recovered by maximizing Equation (10) using the Markov Chain Monte Carlo (MCMC) solver `emcee` (Foreman-Mackey et al. 2013).

This particular choice of model parameterization and likelihood presents some subtleties in implementation. The likelihood presented in Equation (10) does not account for increased uncertainty of the bandpower estimates due to time-domain filtering. This factor is difficult to compute analytically, so this correction is instead applied to the chains after the fact. The covariance $\boldsymbol{\Sigma}'$ of the component-separated bandpowers is modified to increase the uncertainty of the CMB signal bandpowers: $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}'\boldsymbol{B}^T$. Here $\boldsymbol{B}$ is a diagonal matrix that is determined through simulations and is insensitive to the input sky model. Both the cleaned spectra points in Figure 8 and the SMICA likelihood in Section 9.3 include this correction factor.

An additional challenge arises from noise fitting. SMICA performs component separation based upon spectral shape. Unfortunately, the dust component and noise have similar scalings, each increasing with frequency. This introduces a degeneracy within the model (Equation (9)) between the dust index β_d and the noise parameters. To break this degeneracy, the noise is pinned at the highest-frequency map (generally *Planck* 353 GHz) using a noise model. This can be constructed in one of two ways: by taking the ensemble average of FFP10 noise simulations,[41] or from the difference between the auto-spectrum of the full-mission map and the cross-spectrum of two half-mission maps. Both noise options are propagated through the full analysis. CMB components are largely unaffected by

[41] *Planck* end-to-end "full focal plane" simulations (*Planck* Collaboration et al. 2020a).

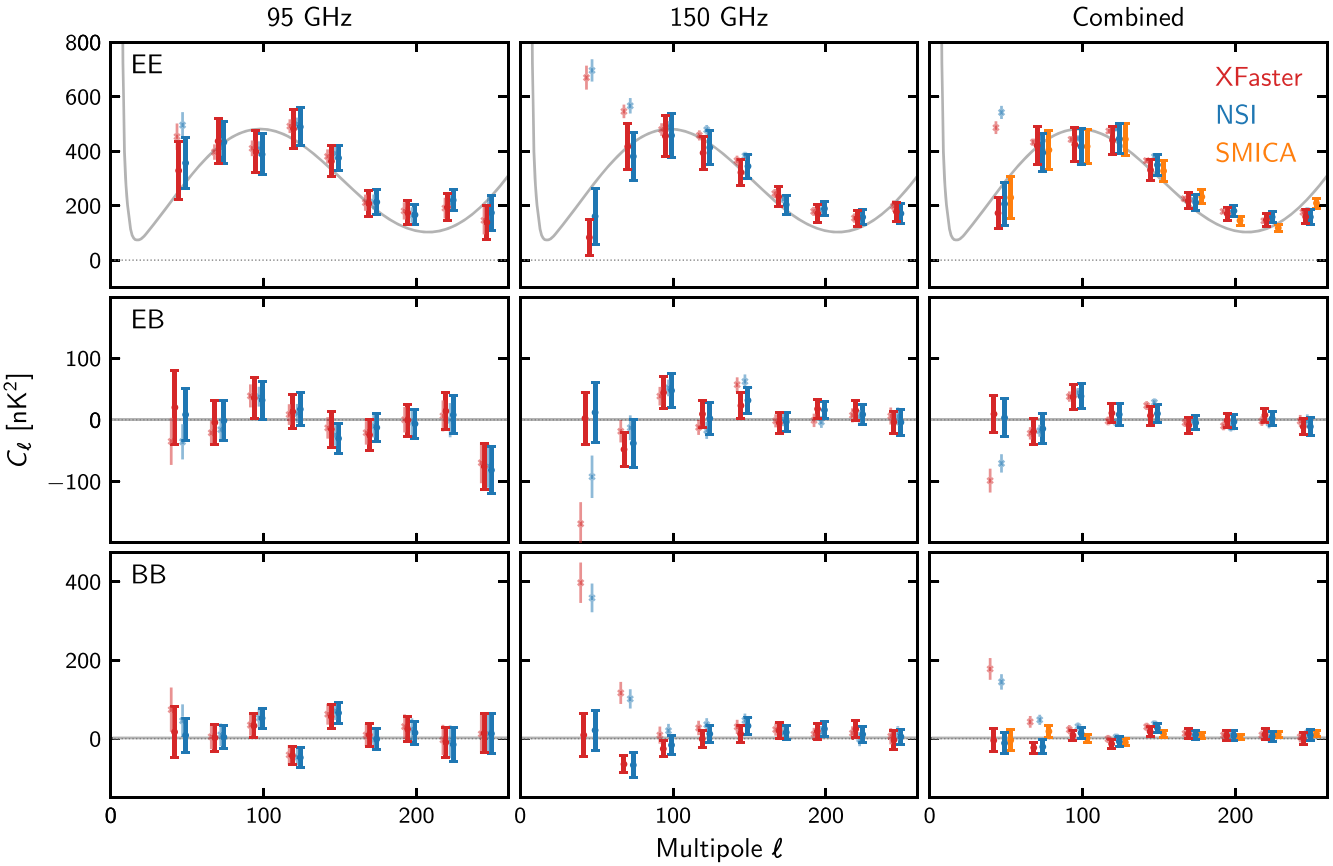

Figure 8. Foreground-cleaned power spectra from the XFaster, NSI, and SMICA pipelines. Foreground templates for XFaster and NSI are constructed as described in Section 8.1 assuming $\nu_0 = 353$ GHz, scaled by best-fit factors listed in Table 4, and subtracted from the SPIDER maps prior to computing spectra with the two pipelines. The raw spectra from Figure 7 are also shown (×), along with the best-fit *Planck* ΛCDM power spectrum from *Planck* Collaboration et al. (2016b) (gray line). The SMICA points are the best-fit component-separated CMB spectrum as described in Section 8.2. Error bars for the foreground-cleaned data include sample variance for all pipelines.

any particular choice of a noise model because their spectral shape is very different from either foregrounds or noise.

The SMICA inputs are the SPIDER maps in four chunk sets for each of the two frequencies (95 and 150 GHz), and full-mission *Planck* HFI polarized maps (100, 143, 217, and 353 GHz). WMAP and LFI channels can be incorporated to add sensitivity at lower frequencies, but were omitted from this analysis because the SPIDER data are consistent with being dominated by dust foregrounds (Section 8.4). Excluding these data allows for a simplified fit to a single-component foreground model. The *Planck* maps are masked to the SPIDER observation region, re-observed, and smoothed to a common resolution given by the SPIDER 150 beams.

This preprocessing ensures that foreground modes that are filtered out in the SPIDER low-level pipeline are also filtered out in the *Planck* maps, which is important when constructing signal or dust component-separated maps.

8.2.2. SMICA Results

SMICA-derived fits for the dust spectral index are shown in Table 4. As expected, recovery of the dust spectral index is somewhat sensitive to the choice of the noise model—indices derived from the two choices of 353 GHz noise differ at a mutual $\sim 1\sigma$ level. While important to SMICA's ability to determine dust properties, the practical effect of this on the CMB component is mitigated by the correlation within the fit between β_d and the dust amplitude P_b.

Because of this correlation, the dust power propagated to lower frequencies is relatively insensitive to the choice of the noise model—the dust bandpowers at SPIDER frequencies agree to within a mutual $\sim 0.25\sigma$ between the two models.

The SMICA-fitted β_d is also generally consistent with that from the template methods. NSI has a larger α error, which allows for good agreement with either of SMICA's noise configurations. XFaster's 353–100 GHz template value shows good agreement with the SMICA auto–cross noise model (within a mutual 0.4σ) but the consistency decreases with the FFP10 noise model (mutual 1.5σ). While we have no a priori reason to favor one SMICA noise model over the other, for reasons unrelated to the foreground estimate (discussed in Section 9.3), auto–cross was chosen as the baseline noise configuration.

8.3. Cleaned Power Spectra

The power spectra cleaned with the map-based (XFaster and NSI) and harmonic-space (SMICA) methods are presented in Figure 8, alongside the raw spectra from Figure 7 for comparison. The foreground-cleaned data points from all three pipelines include sample variance and the noise estimated by each pipeline.

All of the pipelines remove significant foreground power in the low ℓ bins, yielding *EE* spectra in good agreement with one another and with the ΛCDM model. The biggest deviation occurs for the $\ell \sim 200$ bin, where all three pipelines detect more

power than the ΛCDM model. The NSI bandpower is the most discrepant, and is roughly 2σ higher than the equivalent *Planck* bandpower in the same sky region. While the *Planck* release 3.01 data products (*Planck* Collaboration et al. 2020e) are used by default throughout this paper, other public *Planck* variants offer an additional consistency check. For this particular bandpower and sky region, we find that the release 3.01 measurement is consistent with SRoll2 (Delouis et al. 2019) but differs from NPIPE (Tristram et al. 2021). The equivalent NPIPE bandpower is even higher than the NSI measurement, though at less than 1σ. As an additional check, the *Planck* 100 and 143 GHz information (from release 3.01) can be omitted from the SMICA pipeline. As expected, this causes the bandpower to drift up and indicates a larger power contribution from SPIDER than *Planck*.

A comparison of the cleaned *BB* spectra to a lensed ΛCDM model derived from the best-fit *Planck* parameters (*Planck* Collaboration et al. 2016b) yields $\chi^2_{\rm lensed} = 9.2\,(9.3)$ for the XFaster (NSI) bandpowers. This is an improvement of $\Delta\chi^2 = 0.7\,(1.4)$ over a model without lensing. The SMICA CMB *BB* spectrum also prefers lensing, with $\chi^2_{\rm lensed} = 9.2$ and an improvement of $\Delta\chi^2 = 4.0$ over an unlensed model. The biggest difference between the template-based pipelines and SMICA occurs in the second multipole bin of the *BB* spectrum, where the template methods fluctuate low while SMICA fluctuates high. The origin and impact of this difference on the cosmological results are discussed in Section 9.

8.4. Polarized Synchrotron Emission

Both the template-subtraction method (Section 8.1) and SMICA (Section 8.2) assume that the sky signal contains only one polarized foreground component: interstellar dust. Polarized synchrotron emission is known to be significant, even at high-Galactic latitudes, at the lower frequencies mapped by CMB experiments (see, e.g., Page et al. 2007, for an early measurement). However, for SPIDER's region and frequencies of interest, this emission is expected to be subdominant to that from Galactic dust. In this section, we present the results of a harmonic-domain foreground-separation analysis that assumes the presence of polarized Galactic dust *and* synchrotron emission in the SPIDER data, and constrains their relative power.

Following Choi & Page (2015), we construct the ensemble of *EE* and *BB* power spectra made up of all possible cross- and auto-spectra, computed with PolSpice, between WMAP, *Planck* HFI, and SPIDER maps. To avoid noise bias, each "auto"-spectrum is computed as the cross-spectrum between two maps constructed from years 1–5/6–9 (WMAP), half-missions (*Planck*), and interleaved sets of 10-minute data chunks (SPIDER; see Section 5). Before spectrum estimation, all *Planck* and WMAP maps are re-observed and all maps are corrected for T-to-P leakage and smoothed to a common 1° resolution, corresponding to the resolution of the WMAP K-band map. When correcting power spectra, we use the unbinned filter transfer function because the effects other than filter attenuation are highly subdominant to other sources of error.

Under the assumption that the foreground signal in each map is made up of Galactic dust and synchrotron emission, each spectrum in the ensemble is the sum of three physical components (CMB, dust, and synchrotron) and their correlations. We model the SEDs of each at a given frequency ν as follows:

$$
\begin{aligned}
I_{\rm CMB} &= A_{\rm CMB}\frac{\delta B_\nu}{\delta T}\Big|_{T_{\rm CMB}},\\
I_{\rm s}(\nu) &= A_{\rm s}\left(\frac{\nu}{23\ {\rm GHz}}\right)^{\beta_{\rm s}},\\
I_{\rm d}(\nu) &= A_{\rm d}\left(\frac{\nu}{353\ {\rm GHz}}\right)^{\beta_{\rm d}}\frac{B_\nu(T_{\rm d})}{B_{353\ {\rm GHz}}(T_{\rm d})}.
\end{aligned}
\tag{12}
$$

$A_{\rm CMB}$, $A_{\rm s}$, and $A_{\rm d}$ are the amplitudes of the three components (in $\mu{\rm K}_{\rm CMB}$ for $A_{\rm CMB}$ and MJy for $A_{\rm s}$ and $A_{\rm d}$) referenced at 23 and 353 GHz for synchrotron and dust, respectively. $\beta_{\rm s}$ and $\beta_{\rm d}$ are the synchrotron and dust spectral indices, $B_\nu(T_{\rm d})$ is the *Planck* function computed at the dust temperature $T_{\rm d} = 19.6$ K, and $\frac{\delta B_\nu}{\delta T}|_{T_{\rm CMB}}$ is its derivative with respect to T computed at the FIRAS temperature $T_{\rm CMB} = 2.725$ K (Mather et al. 1999). In this formalism, the contribution of a given component to the cross-spectrum between two frequencies ν_1 and ν_2 is simply the product $I_{\rm c}(\nu_1)I_{\rm c}(\nu_2)\Sigma_{\rm c}$, where the index c labels the component and $\Sigma_{\rm c}$ is the cross-spectrum of the associated (Q, U) component spatial templates multiplying the SEDs. Similarly, a correlation between two components $\rm c_1$ and $\rm c_2$ yields a contribution $I_{\rm c_1}(\nu_1)I_{\rm c_2}(\nu_2)\Sigma_{\rm c_1\times c_2}$, where the cross-spectrum $\Sigma_{\rm c_1\times c_2}$ can be interpreted as a scale-dependent spatial correlation coefficient between the two components.

We perform an MCMC analysis to fit, independently in each multipole bin and for each polarization (*EE* or *BB*), a model to this cross-spectrum ensemble consisting of the spectral indices $\beta_{\rm s}$ and $\beta_{\rm d}$, the three parameters $\tilde{A}_{\rm c} \equiv A_{\rm c}\sqrt{\Sigma_{\rm c}}$, and the correlation coefficients $\delta \equiv \Sigma_{\rm d\times CMB}$ and $\rho \equiv \Sigma_{\rm d\times s}$. We use broad, uniform priors on nearly all components, excepting $\beta_{\rm s}$, where we use the *Planck* posterior of $\beta_{\rm s} = -1.15 \pm 0.17$ (*Planck* Collaboration et al. 2020d), and δ, where we use the distribution of correlations observed between a large number of CMB- and dust-only sky simulations. Note that the value of $\beta_{\rm s}$ in this text is equivalent to the *Planck* value of $\beta_{\rm s} + 2$, due to a difference in the calibration of the model. See Appendix for a detailed discussion. There is no a priori reason to expect more than chance correlation between the dust and CMB components, and so it is reasonable to measure δ in this fashion. Unlike the CMB, the correlation between the dust and synchrotron emission (parameterized by ρ) is expected to be non-zero. We adopt a uniform prior on this parameter. Note that we do not fit for the $\Sigma_{\rm s\times CMB}$ correlation coefficient, which is expected to be the smallest of the three, because the validation of this method on simulations did not indicate a need for it. Errors on the input spectra are computed from a distribution of signal and noise simulations, with each map having a noise model appropriate for its associated instrument: SPIDER's stationary noise model (without the noise model scaling factor), *Planck* FFP10 simulations, and Gaussian pixel noise for WMAP.

Drawing from the posteriors of the fitted parameters, we compute the *EE* and *BB* power spectra of the dust and synchrotron emission at SPIDER frequencies. The 95 GHz bandpowers in Figure 9 are shown as the maximum-likelihood point of the posteriors, with error bars bounding the narrowest 95% of the posterior about the maximum likelihood. For those bins in which 0 is within that 95%, only the upper limit is shown. In each multipole bin with a reported maximum

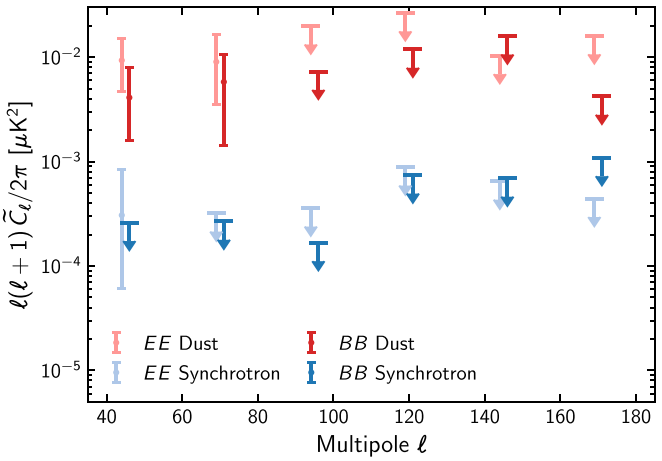

Figure 9. 95 GHz *EE* (light) and *BB* (dark) angular power spectra of the synchrotron (blue) and dust (red) emission in the SPIDERregion as estimated by the foreground-separation method described in Section 8.4. These power spectra are *not* corrected for the SPIDER beam or filtering. In each multipole bin, bandpower points are reported as the maximum-likelihood point of the bandpower posterior, with the narrowest 95% of samples about the maximum likelihood defining error bars. For bandpowers that contain 0 in the narrowest 95% of points, we report 95% upper limits.

likelihood for dust, the 95% upper limit on the polarized synchrotron emission is an order of magnitude or more below the maximum likelihood of the dust bandpower distribution. The foreground analyses presented in Sections 8.1 and 8.2, which measure dust levels that are comparable to those shown in Figure 9, therefore assume the presence of only one polarized Galactic emission in the SPIDER and *Planck* HFI maps—that from interstellar dust. Note that the power reported in Figure 9 is *not* corrected for the SPIDER beam or filter transfer function; the on-sky foreground power will be discussed in a forthcoming publication.

8.5. Discussion

The map-based (template-fitting) and harmonic-domain (SMICA) component separation techniques agree well in their estimation of dust model parameters and CMB spectra. The template-fitting method imposes no model for dust's spectral scaling, and the SMICA method makes no assumptions about the spatial distribution of dust. Thus, this agreement is evidence that their differing assumptions are valid for the SPIDER data set in combination with *Planck*. Furthermore, the level of synchrotron emission is constrained to be well below that of dust in this region of sky at SPIDER frequencies. This justifies the assumption of dust-dominant foregrounds made in the template-fitting and SMICA pipelines. Having established confidence in and consistency among the methods for calculating CMB component power spectra, we then use each to construct likelihoods for the tensor-to-scalar ratio, r.

9. Likelihoods

In this section, we use the power spectrum estimates to construct likelihoods for the tensor-to-scalar ratio, r. Two separate approaches are taken to construct and sample from the parameter likelihood. The XFaster likelihood construction is Gaussian in the $a_{\ell m}$ coefficients, and the algorithm can naturally be adapted to sample that likelihood as a function of parameters other than bandpowers. This approach relies solely on the assumption that the cleaned maps are dominated by CMB and noise, both well-approximated as Gaussian random fields. The NSI and SMICA methods proceed instead from the bandpowers computed above, assuming Gaussian likelihoods for the CMB bandpowers to sample additional parameters. This approach is susceptible to sample variance, which limits the validity of the Gaussian approximation in the presence of significant E-mode or B-mode power (Bond et al. 2000).

The two approaches also differ significantly in the way that parameter covariance is propagated. The XFaster approach samples a likelihood of the $a_{\ell m}$s as a function of the three-dimensional parameter space of r and α, which is then marginalized into a final posterior for r. The NSI and SMICA approach samples a profile likelihood in r, which optimizes over foreground parameter dependence in a separate step. All of the other ΛCDM parameters (notably A_s and τ) are held fixed.

Within the template-based methods, we do not attempt to fit out any foreground structure not captured by the dust template. Since our knowledge of these foreground residuals is by definition limited, we opted not to introduce further modeling in an attempt to capture them. Instead, we use the orthogonal SMICA approach to assess the robustness of the cleaning process used by XFaster and NSI. Within SMICA, we make no assumption as to the spatial structure of the dust emission, but model its frequency dependence. We compare the results from the independent SMICA and template-based approaches in Section 9.4.

Each of the methods described in this work has been extensively validated on simulations that include the cosmological signal, a model of Galactic foregrounds, and time-domain instrumental effects, including the noise and in-flight pointing. Subject to the assumptions made regarding these inputs, we find all estimators to be free from bias. We adopt the XFaster pipeline as our baseline, given its generality and self-consistency.

9.1. XFaster

The XFaster power spectrum estimator naturally provides a likelihood for a parameterized model, rather than fitting for the maximum-likelihood bandpower deviations. The generalized XFaster parameter likelihood has a form similar to that in Equation (5):

$$-2\ln L(\theta|\tilde{\boldsymbol{d}}) = \sum_{\ell,k}(2\ell+1)\,g_\ell^k\,[\widetilde{\boldsymbol{C}}_\ell^{-1}(\theta)\widehat{\boldsymbol{C}}_\ell + \ln\widetilde{\boldsymbol{C}}_\ell(\theta)]_{kk}\,, \tag{13}$$

where θ is a set of parameters, $\widetilde{\boldsymbol{C}}_\ell$ is the model pseudo-C_ℓ matrix, $\widehat{\boldsymbol{C}}_\ell$ is the data pseudo-C_ℓ matrix, g_ℓ is the mode count recalibration factor, and the index k labels the maps used (for this paper, four 95 GHz and four 150 GHz maps). The likelihood is based on the XFaster approximation of the likelihood for the observed pseudo-$a_{\ell m}$s and is therefore also Gaussian without loss of generality. The model pseudo-C_ℓs are computed using the same transfer functions, beam window functions, and mode-coupling kernels used for the bandpower computation, with the tensor contributions to the *EE* and *BB* power spectra modeled as a function of r.

Foreground fitting could, in principle, be accomplished by adding a scaled dust-template spectrum to the model and fitting for the scale factor α at each frequency. However, because all terms in the signal model are treated as Gaussian random fields,

XFaster's estimated error would include sample variance proportional to the foreground amplitudes, which is not appropriate for the fit to a non-Gaussian template. Therefore, we instead subtract the scaled template from the data at the map level and model the residuals as CMB and noise. Because the template subtraction is not accounted for in the covariance, the additional error from foreground cleaning must be calibrated using an ensemble of CMB, noise, and template simulations.

The data terms after template subtraction are

$$\widehat{\boldsymbol{C}}^{i\times j} = \langle (\boldsymbol{d}_i - \alpha_i \boldsymbol{t}_i) \times (\boldsymbol{d}_j - \alpha_j \boldsymbol{t}_j) \rangle, \tag{14}$$

where $\boldsymbol{d}$ is a data map, $\boldsymbol{t}$ is a template map, α is the frequency-dependent template scaling, and i and j are indices of the eight maps. The template maps used for each cross-spectrum are separate *Planck* half-missions to avoid contributions from *Planck* noise auto-spectra. To compute the maximum-likelihood parameter estimates, an MCMC sampler steps through values of r, α_{95}, and α_{150}, recomputing the data and model terms in the likelihood at each step.

Monte Carlo simulations of this process are used to propagate the uncertainty in template fitting to the likelihood derived from the data. The same process is repeated for 300 simulations using different realizations of CMB, SPIDER noise, and *Planck* noise from the FFP10 simulation ensemble. The parameter distributions recovered from these simulations are used to estimate the added covariance from uncertainty in the template fit. We found that the magnitude of the additional covariance did not change significantly within reasonable ranges of r or α, or with different morphologies of the simulated foregrounds. We incorporate this added uncertainty to the data's parameter likelihoods by adding it to each Monte Carlo sample. The terms contributing to the additional covariance are noise in the template (45%), chance correlations between SPIDER noise and the template (45%), and chance correlations between the CMB and the template (10%), where the total quadrature 1σ error added for r is 0.10. This accounts for approximately half of the total error. Uncertainties in the beam window functions and in the corrections to the SPIDER noise model are also parameterized and marginalized over in the final result—we find that their effects are negligible.

Figure 10 shows the complete XFaster likelihood result using both choices of dust template (353–100 GHz and 217–100 GHz), incorporating error contributions from the template subtraction and with no priors imposed on these parameters. The scaling of the 217–100 GHz template αs to plot on common axes relies on the assumption of a modified-blackbody dust model. The two templates yield consistent results for all parameters. Due to its greater constraining power, we use the 353–100 GHz template result for our final constraint, yielding a maximum-likelihood estimate for the tensor-to-scalar ratio of $r_{\rm mle} = -0.21$. We find that 6% of simulations with input $r = 0$ yield $r_{\rm mle} < -0.21$, hence this value is not inconsistent with expected noise fluctuations.

We can compute an upper limit on r from this likelihood, subject to the physical constraint that $r \geqslant 0$. Imposing a flat prior on r, truncated for $r < 0$ to implement this physical constraint, we obtain a 95% Bayesian upper limit of $r < 0.19$.

We also construct a classical confidence interval for r, following the approach discussed in Feldman & Cousins (1998). In this approach, simulations are conducted for a range of values of input $r_{\rm in}$, each carried through to a value of $r_{\rm mle}$.

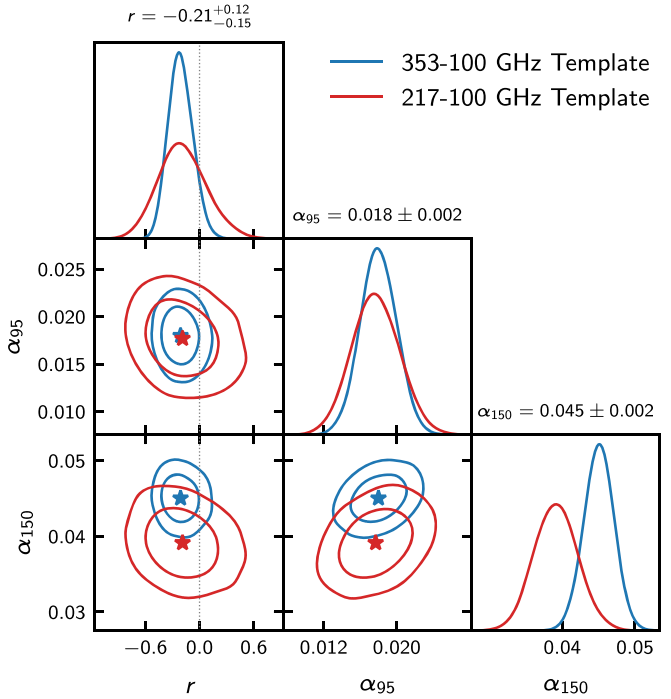

Figure 10. The combined XFaster likelihood for r and α, imposing no priors on these parameters. Blue curves show likelihoods computed using a 353–100 GHz template. Red shows the same for a 217–100 GHz template, where α values have been scaled assuming a modified blackbody for dust ($T_d = 19.6$ K, β_d determined from each α sample) to the corresponding values for a 353–100 GHz template. 1σ constraints for the 353–100 GHz template are shown in the panel titles.

For each $r_{\rm in}$, an interval of $r_{\rm mle}$ is defined containing 95% of simulations—those with the largest values of the likelihood ratio $R \equiv \mathcal{L}(r_{\rm mle}|r_{\rm in})/\mathcal{L}(r_{\rm mle}|r^*)$, where r^* is the value of $r_{\rm in}$ that maximizes $\mathcal{L}(r_{\rm mle}|r_{\rm in})$. Note that $r^* = 0$ for $r_{\rm mle} \leqslant 0$. Figure 11 shows this confidence interval as black-dashed lines, which transition smoothly between detection and upper limit while maintaining correct coverage for $r_{\rm mle}$ near or beyond the physical boundary $r \geqslant 0$. The observed $r_{\rm mle}$ yields an upper limit of $r < 0.11$ (95% CL). The difference between this and the Bayesian limit reflects their disparate definitions and interpretations, as well as the modest over-coverage (conservatism) of the Bayesian limit near the physical boundary.

An ensemble of CMB, SPIDER noise, and template noise simulations are used to determine relative contributions to the error budget. Simulated maps are constructed by creating an ensemble in which only one of these components is allowed to vary (e.g., 300 maps made by joining a single CMB realization, a single SPIDER noise realization, and 300 template realizations). By comparing the scatter in the estimated r values, we can estimate the relative contributions to the total error; i.e., the scatter when all three components are varied together. Assuming that the template is a perfect representation of the dust morphology, the largest contributor to σ_r is SPIDER noise, including its chance correlations with the template, at ∼60%. CMB sample variance and chance correlations contribute ∼25%, and template noise adds ∼15%. For the 217–100 GHz template, the statistical error is instead dominated by the noise in the template.

9.2. NSI

The foreground-cleaned power spectra from the NSI pipeline are also propagated to an r-likelihood. This likelihood proceeds

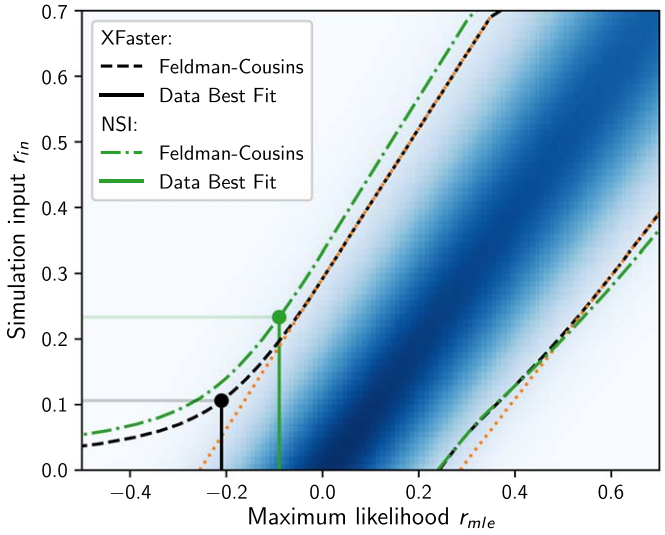

Figure 11. Feldman–Cousins 95% confidence interval (dashed black) on r as a function of $r_{\rm mle}$, derived from template-subtracted XFaster likelihoods. The observed $r_{\rm mle}$ is indicated by the black vertical line; our upper limits are the intersections of the curves above with this line, projected onto the vertical axis. Blue shading indicates the distribution of $r_{\rm mle}$ as a function of input r, which is used to construct these curves. For each input r, we conduct 300 XFaster simulations to produce a histogram of $r_{\rm mle}$, which we smooth by fitting a Gaussian model (a good fit). At each input r, 95% of simulation results lie between the dotted orange lines. These simulations use noise maps rescaled in each bandpower bin by the associated XFaster estimate of the noise model scaling factor. Similarly, the Feldman–Cousins interval (green dashed–dotted) and data best fit (green solid) are shown for the NSI pipeline; the associated simulations are not shown.

in separate steps: the bandpowers are first estimated, then fit for the foreground template amplitudes α (as in Section 8.1), and finally an r-likelihood is constructed from the cleaned spectra. To mitigate potential bias from fitting α before r, α is fit to only EE, while r is fit to only BB. Fits for α provide consistent results whether using EE only, BB only, or both together. The fit for r uses the simple Gaussian likelihood:

$$-2\ln L = (\widehat{C}_b^{BB} - C_b^{BB}(r))^T \boldsymbol{M}^{-1} (\widehat{C}_b^{BB} - C_b^{BB}(r)) + \ln|\boldsymbol{M}|, \quad (15)$$

where $\widehat{C}_b^{BB}$ is the cleaned B-mode spectrum measured by SPIDER and $C_b^{BB}(r)$ is a ΛCDM model using *Planck* parameters, lensing, and allowing r to vary. The bandpower covariance matrix $\boldsymbol{M}$ is a sum of three contributions: SPIDER's statistical noise, estimated from the distribution of 378 NSI cross-spectra; sample variance, estimated from an ensemble of re-observed signal-only ΛCDM simulations; and propagated error on α, to capture the statistical error on the foreground template fitting. The cross-spectra between terms for signal, noise, and foreground template, while uncorrelated in the mean, also contribute to the total covariance. These extra contributions are estimated together from an ensemble of full signal plus noise simulations. Note that, unlike NSI results for raw power spectra and α fits, this result depends on simulations of SPIDER noise. While bandpowers do not in general follow a Gaussian likelihood (Bond et al. 2000; Gerbino et al. 2020), simulations show that this approximation is adequate for SPIDER's sky coverage and ℓ bins.

The errors on α_{95}, α_{150}, and the associated covariance are propagated to both the bandpower covariance matrix and the r-likelihood using a Monte Carlo method. Starting with the two-dimensional Gaussian distribution described by the best-fit parameters in Table 4, random α values are drawn and new template-subtracted bandpowers are computed. A set of 1000 such randomly cleaned bandpowers are used to estimate the bandpower covariance due to α error, which is added to $\boldsymbol{M}$ as above. To estimate r, Equation (15) is evaluated 4096 times for another set of 4096 randomly cleaned bandpowers (as $\widehat{C}_\ell^{BB}$). This step allows α to shift slightly from its best-fit value when the r-likelihood prefers it. More draws are required than for the previous step (4096 versus 1000) to ensure that the random seed does not significantly impact the results. The final r-likelihood is the average of the likelihoods evaluated for each random draw from the α distribution.

Finally, as for XFaster, we derive upper limits on r from the NSI likelihood under the physical constraint $r \geqslant 0$. A Feldman–Cousins approach (Figure 11) yields $r < 0.23$, while a Bayesian calculation yields $r < 0.27$, both at 95% confidence. Section 9.4 further discusses how the NSI and XFaster results compare.

9.3. SMICA

As for NSI, we construct an r-likelihood for SMICA under a simple Gaussian approximation for the bandpower likelihood (Equation (15)). We note that this is a suboptimal approximation for the r-likelihood, due to measured non-Gaussianity of the BB bandpowers in the lowest bins. For SMICA, the covariance $\boldsymbol{M}$ is empirically determined from MCMC chains. While this covariance does not capture the full r-dependence of sample variance, this contribution to the total covariance is expected to be small. Furthermore, because the SMICA likelihood maximization that produced these bandpowers jointly fits for CMB signal, instrumental noise, and dust foregrounds, the statistical distribution of the fitted CMB bandpowers, C_b, naturally includes noise and foreground uncertainty.

As discussed in Section 8.2, this implementation of SMICA requires a choice of noise model at 353 GHz. Lacking strong justification in preferring one over the other, and in the spirit of reporting a conservative upper limit, we run the analysis with both choices and report the less stringent result. Consequently, the reported SMICA results come from the auto–cross noise model.

The resulting SMICA r-likelihood is shown in Figure 12. The nominal configuration with all SPIDER and *Planck* data yields a maximum-likelihood estimate of $r_{\rm mle} = 0.06 \pm 0.11$. Subject to a physical prior that $r \geqslant 0$, this corresponds to a 95% Bayesian upper limit of $r < 0.24$. A Feldman–Cousins constraint is computationally impractical in the SMICA framework.

In Figure 12 we further explore the effect on the SMICA likelihood of incorporating different subsets of the Planck data. The *Planck* 100 and 143 GHz data are particularly interesting in this regard, as they contribute appreciably to the SMICA CMB component but not directly to the template analysis. At the angular scales of interest, the SMICA algorithm assigns weights to the SPIDER data at 95 (150) GHz that are approximately five (three) times those applied to the *Planck* data at the nearest corresponding frequency. Omitting the *Planck* data at 100 and 143 GHz shifts the r-estimate downward to $r_{\rm mle} = -0.03 \pm 0.12$, in closer agreement with the template results. A configuration similar to the template methods (*Planck* 143 and 217 GHz data omitted) results in a similar value ($r_{\rm mle} = -0.02^{+0.12}_{-0.11}$). Even in this configuration, however, *Planck* 100 GHz still has substantial influence on CMB recovery, with a weight approximately four times that in

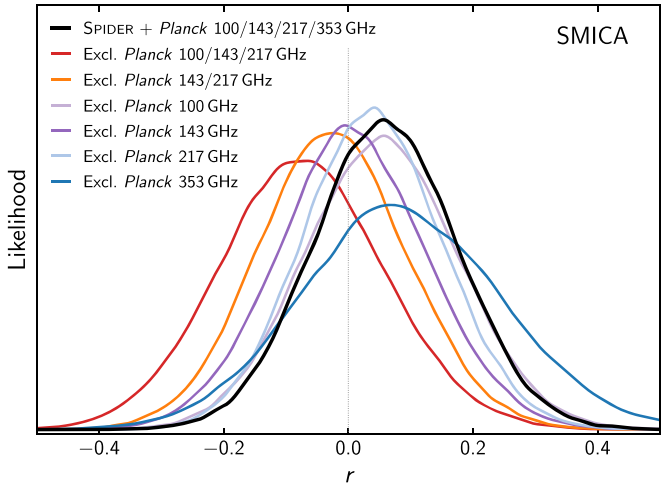

Figure 12. The SMICA r-likelihood, showing the impact of varying the inputs into the pipeline. Removing any of the non-353 *Planck* bands lowers $r_{\rm mle}$, consistent with SPIDER data driving r low. This suggests that the SPIDER and *Planck* noise, or their chance correlation with the foreground signal, must fluctuate with opposite sign.

the template methods. When omitting *Planck* 100, 143 and 217 GHz data, SMICA recovers an $r_{\rm mle}$ closest to that from the template methods: $r_{\rm mle} = -0.07 \pm 0.13$.

In each case, we find that the shift in $r_{\rm mle}$ is primarily driven by the first two multipole bins of the *BB* spectrum. This suggests that either the SPIDER noise or chance correlations between noise and foregrounds result in a negative fluctuation in *BB* relative to the *Planck* data, irrespective of the method of foreground removal. Variation among the r estimates may also arise from the differing assumptions made regarding the modeling of the foregrounds in each pipeline.

9.4. Discussion

Table 5 presents the maximum likelihood r for each of the three pipelines, both in their standard configurations and in various modified configurations that are chosen to explore the impact of their structural differences (which will be discussed in more detail later on). The nominal configuration of each pipeline was chosen prior to running the estimator on data and includes its maximal data set—all *EE* and *BB* science bins for XFaster, and the full set of SPIDER and *Planck* maps for SMICA. The same table reports the 95% upper limit for each pipeline in its the nominal configuration. In all cases, XFaster, NSI, and SMICA are found to return unbiased r posteriors that are broadly consistent with one another when run on time-domain simulations. We adopt the XFaster pipeline as our primary result, due to the more formally correct construction of its likelihood.

Insofar as each of the above pipelines is unbiased on simulations and makes relatively simple (and non-contradictory) assumptions, the observed difference in $r_{\rm mle}$ when restricted to a closely comparable subset of data merits investigation. An important question is whether the observed discrepancy between methods is consistent with expected variation given the difference in methodologies alone. We address this question in two ways: by observing the effects on the data's $r_{\rm mle}$ from slight modifications of each pipeline, and (where feasible) by comparing the results of each pipeline when applied to identical simulated maps.

Table 5
Summary of r-likelihood Values from Various Pipelines, with Nominal Upper Limits in Bold

Pipeline	Description	$r_{\rm mle}$	$r \leqslant 95\%$
XFaster	Nominal, Feldman–Cousins	−0.21	**0.11**
	Nominal, Bayesian	−0.21	**0.19**
	NSI-like:		
	(*a*) r from *BB* only	−0.19	⋯
	(*b*) Independent *EE* and *BB* noise	−0.19	⋯
	(*a*) + (*b*)	−0.15	⋯
NSI	Nominal, Feldman–Cousins	−0.09	0.23
	Nominal, Bayesian	−0.09	0.27
SMICA	Nominal, Bayesian	0.06	0.24
	Template-like:		
	Excl. *Planck* inputs < 353 GHz	−0.07	⋯

We first compare the two template-subtraction methods, XFaster and NSI. XFaster's tighter upper limit results primarily from a lower $r_{\rm mle}$, as illustrated in Figure 11. The NSI limit is also increased slightly by having a broader distribution than XFaster (by about 18% for $r = 0$), which is a result of less-optimal weighting of the available data. Table 5 highlights the effects of modifying some of the assumptions that differ between the two methods. When XFaster is run in a more NSI-like configuration—fitting r from *BB* only, with independently estimated *EE* and *BB* noise—its nominal $r_{\rm mle} = -0.21$ shifts to -0.15, in better agreement with NSI's $r_{\rm mle} = -0.09$. This shift of 0.06 in $r_{\rm mle}$ is consistent with expectations from simulations, which show random fluctuations with $\sigma = 0.06$ between the two different XFaster configurations when applied to the same simulated maps. Since XFaster, unlike NSI, is able to fit r and α simultaneously, thereby leveraging the information in both the *EE* and *BB* spectra, and fully exploring the degeneracy between these parameters, we retain XFaster's nominal configuration. (As shown in Table 5, $r_{\rm mle}$ does not sensitively depend on how noise is fitted when this change alone is implemented.)

SMICA differs from the template-subtraction methods in the quantity of *Planck* data that are incorporated in the analysis. As shown in Table 5 and Figure 12, changing the selection of *Planck* data to better match the template methods results in a downward shift of $r_{\rm mle}$ by 0.13, in the direction of the template results. This implies that fluctuations in the SPIDER and *Planck* data drive the *BB* spectrum in opposite directions.

The difference in $r_{\rm mle}$ between SMICA and XFaster is not entirely resolved through the inclusion of common data products. This is shown most clearly in Figure 13. To quantify the significance of the remaining difference ($\delta r_{\rm mle} = 0.14$), we compare XFaster and SMICA in simulation by applying them to a nearly identical set of CMB, noise, and foreground simulation maps.

For SMICA, we simulate the configuration in which the only *Planck* map used is 353 GHz. For XFaster, the simulated template includes both 353 and 100 GHz simulated *Planck* noise. NSI results are not similarly simulated because SPIDER noise is seeded differently for NSI's shorter data chunks. Both SMICA and XFaster recover $r_{\rm mle}$ without bias and with partially correlated variance. The covariance between the two

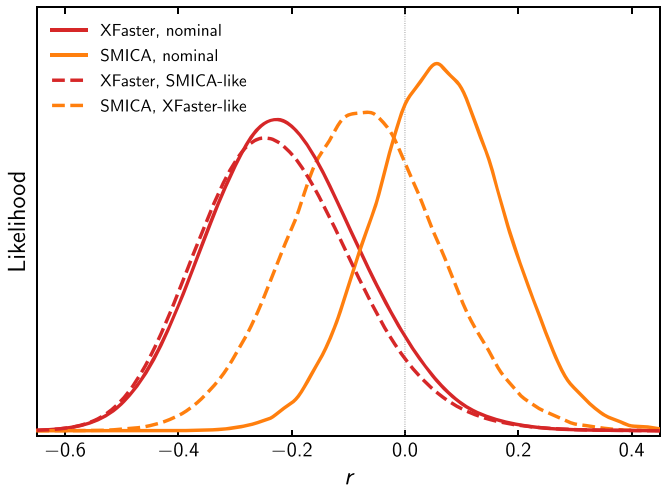

Figure 13. Comparison of the SMICA and XFaster r-likelihoods. The nominal cases correspond to the "Nominal" row for each in Table 5, using the maximal data set for each pipeline. Posteriors are also shown for configurations in which the pipelines use approximately the same data inputs. For XFaster, this corresponds to case *a* in Table 5, limiting the r fit to the *BB* spectrum. For SMICA, this corresponds to the final case in Table 5, removing all *Planck* data except at 353 GHz (reproduced from the red line in Figure 12).

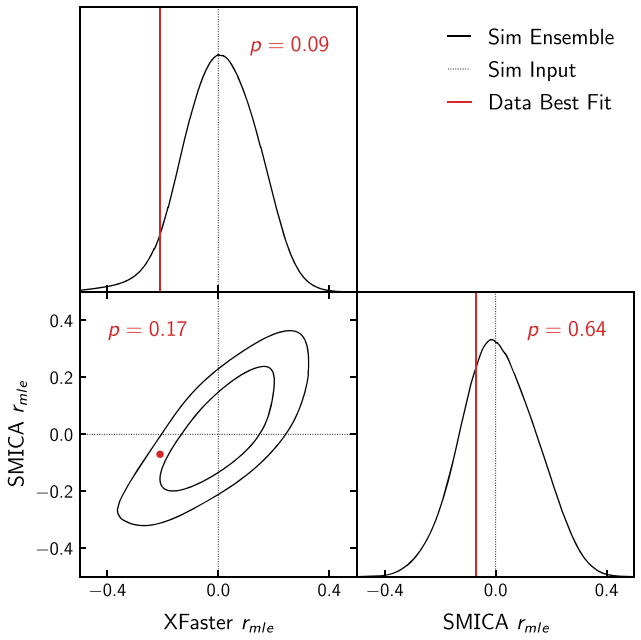

Figure 14. Comparison of XFaster and template-like SMICA $r_{\rm mle}$ over a 200 sim ensemble with identical simulation seeds and input $r = 0$. The foreground model present in these simulations agrees with both pipelines' modeling assumptions. Both pipelines return unbiased estimates of $r_{\rm mle}$, but with substantial scatter between the two estimators. These distributions are consistent with the values of $r_{\rm mle}$ recovered from the data. Each probability shown is the fraction of the integral over the distribution that comes from points with $r_{\rm mle}$ values that occur less often than $r_{\rm mle}$ from data.

estimators can be written as:

$$\Sigma_{\rm XF,SMICA} = \begin{pmatrix} \sigma^2_{\rm XF} & \rho\ \sigma_{\rm XF}\ \sigma_{\rm SMICA} \\ \rho\ \sigma_{\rm XF}\ \sigma_{\rm SMICA} & \sigma^2_{\rm SMICA} \end{pmatrix}. \quad (16)$$

For a 200-simulation ensemble with input $r = 0$, shown in Figure 14, we find $\sigma_{\rm XF} = 0.13$, $\sigma_{\rm SMICA} = 0.13$, and $\rho = 0.74$. The uncorrelated variance between the two estimators ($\rho < 1$) captures the degree to which each is sensitive to a different projection of the data when estimating $r_{\rm mle}$, leading to statistical variation between the methods even when given nearly identical input data. It also suggests some degree of non-optimality in the estimators. Comparing this ensemble to XFaster's nominal result and SMICA's template-like result in Table 5, we find that the observed difference between these estimators (0.14) is consistent with the range of differences seen in simulations ($\sigma = 0.1$). Furthermore, if we compare the pair of observed estimator values to the simulation ensemble in two-dimensions, then we find that about one in six of the simulated pairs result in a difference that is equal to or greater than that obtained on the data.

10. Conclusion

The data from SPIDER's first flight have returned maps of the intensity and polarization at 95 and 150 GHz that are substantially deeper than the *Planck* data in the same region of sky. A rigorous suite of consistency tests have been used to define a subset of these data that can be reliably used for cosmological analysis. These maps, in concert with data from *Planck*, are used to constrain the amplitude of any cosmological *B*-mode signal in the CMB.

As anticipated, polarized Galactic dust emission is observed with high signal-to-noise. In SPIDER's sky region, the Galactic *E*-mode component has roughly twice the power of the *B*-mode component, and is found to be dominated by thermal dust emission at 95 GHz and above at all angular scales probed, while Galactic synchrotron radiation is found to be strongly subdominant.

Separating the dust component from the cosmological signal is the principal challenge of the present analysis. To this end, two basic approaches are employed: map-based template subtraction and SMICA, which is an internal linear combination applied in the harmonic domain. While the SPIDER and *Planck* *B*-mode data are found to push the constraints in opposite directions, the $r_{\rm mle}$ derived from the template methods and SMICA are found to be consistent, subject to the assumptions made in each. Of the two template-based methods, XFaster provides significantly more optimal constraints on r, which was discussed in Section 9.4. As a result, we select this pipeline as our baseline template-based method. Under the assumption that our *Planck*-derived template accurately captures the morphology of the dust, we derive the 95% upper limit on the primordial tensor-to-scalar ratio as $r < 0.11$ and $r < 0.19$ using Feldman–Cousins and Bayesian approaches, respectively.

While relaxing assumptions regarding the morphology of the dust component and assuming a dust SED that is both independent of angular scale and well characterized by a modified-blackbody spectrum, SMICA gives a somewhat higher upper limit of $r < 0.24$. Unlike the template-based method, this constraint is derived from a joint analysis of SPIDER and *Planck* 100–353 GHz data. Further characterization of the dominant Galactic foreground emission is the subject of a forthcoming paper.

SPIDER's upcoming flight, which will feature a suite of three new 280 GHz receivers (Bergman et al. 2018; Shaw et al. 2020), will focus on an improved characterization of the foreground emission. These data will both complement the *Planck* data at 217 and 353 GHz, and achieve significantly

higher sensitivity. At the same time, the availability of an independent data set over a substantial portion of the full sky facilitates qualitatively new measures of the robustness of foreground-separation techniques to choices made in the analysis and the selection of data.

SPIDER is supported in the U.S. by the National Aeronautics and Space Administration under grants NNX07AL64G, NNX12AE95G, and NNX17AC55G issued through the Science Mission Directorate and by the National Science Foundation through PLR-1043515. Logistical support for the Antarctic deployment and operations is provided by the NSF through the U.S. Antarctic Program. Support in Canada is provided by the Natural Sciences and Engineering Research Council and the Canadian Space Agency. Support in Norway is provided by the Research Council of Norway. Support in Sweden is provided by the Swedish Research Council through the Oskar Klein Centre (Contract No. 638-2013-8993) as well as a grant from the Swedish Research Council (dnr. 2019-93959) and a grant from the Swedish Space Agency (dnr. 139/17). The Dunlap Institute is funded through an endowment established by the David Dunlap family and the University of Toronto. The multiplexing readout electronics were developed with support from the Canada Foundation for Innovation and the British Columbia Knowledge Development Fund. KF holds the Jeff & Gail Kodosky Endowed Chair at UT Austin and is grateful for that support. W.C.J. acknowledges the generous support of the David and Lucile Packard Foundation, which has been crucial to the success of the project. C.R.C. was supported by UKRI Consolidated Grants, ST/P000762/1, ST/N000838/1, and ST/T000791/1.

Some of the results in this paper have been derived using the `HEALPix` package (Gorski et al. 2005). The computations described in this paper were performed on four computing clusters: Hippo at the University of KwaZulu-Natal, Feynman at Princeton University, and the GPC and Niagara supercomputers at the SciNet HPC Consortium (Loken et al. 2010; Ponce et al. 2019). SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, Ontario Research Fund—Research Excellence, and the University of Toronto.

This collaboration is grateful to the British Antarctic Survey, particularly Sam Burrell, and to the Alfred Wegener Institute and the crew of R.V. *Polarstern* for invaluable assistance with the recovery of the data and payload after the 2015 flight. Brendan Crill and Tom Montroy made significant contributions to SPIDER's development. Paul Steinhardt provided very helpful comments regarding the status of early-universe models. This project, like so many others that he founded and supported, owes much to the vision and leadership of the late Professor Andrew E. Lange.

Appendix
SED Component Spectral Indices

The signal response d of a SPIDER or *Planck* detector to a beam-filling source on the sky can be written as

$$d = g \int \lambda^2 F(\nu) I(\nu)\, d\nu, \tag{A1}$$

where g incorporates frequency-independent responsivity terms, $I(\nu)$ is the spectral radiance of the source, $0 \leqslant F(\nu) \leqslant 1$ is the dimensionless spectral transmission of the instrument, and $\lambda = c/\nu$ is the wavelength.

For sufficiently narrow-band detectors, we may approximate $F(\nu)$ as a delta function around an effective band center ν, where the appropriate ν may vary with source spectrum. We can then absorb F into the value of g to obtain

$$d \approx g \frac{c^2}{\nu^2} I(\nu). \tag{A2}$$

As discussed in Section 4.2.3, SPIDER uses degree-scale CMB fluctuations to provide a native calibration in units of $K_{\rm CMB}$—small variations in the CMB temperature. It is thus particularly useful to consider a blackbody source with spectral radiance B, for which we have

$$d \approx g \frac{c^2}{\nu^2} \Delta T \frac{\partial B(\nu, T)}{\partial T}, \tag{A3}$$

where we have expanded $B(\nu, T)$ to first order in T at fixed band center ν and expressed d in terms of a temperature fluctuation ΔT. Using the measured instrumental responses, $F(\nu)$, we find this approximation to be accurate within a few percent when using effective band centers computed for a spectrum typical of dust or the CMB.

Under the above approximation, the conversion factor $h_{\rm CMB}$ from detected power to CMB units is computed from the ratio of temperature fluctuation to detector signal:

$$h_{\rm CMB} \equiv \frac{\Delta T_{\rm CMB}}{d} = \frac{\nu^2}{c^2} \left(g \frac{\partial B}{\partial T} \Big|_{\nu,\, T_{\rm cmb}} \right)^{-1}. \tag{A4}$$

When working in brightness, or Rayleigh–Jeans, units, defined by

$$I(\nu) = 2 \frac{\nu^2}{c^2} k_B T_b, \tag{A5}$$

the analogous conversion factor h_b is given by

$$h_b \equiv \frac{\Delta T_b}{d} = \frac{1}{2 g k_B}. \tag{A6}$$

In modeling the SED of foreground emission, we choose to parameterize the spectral radiance of both synchrotron and dust using consistent units of the spectral radiance, as in (Brandt et al. 1994). We model the spectral radiance of synchrotron emission as a power-law in frequency, with an index β_s,

$$I_s(\nu) = A_s \left(\frac{\nu}{\nu_0} \right)^{\beta_s}. \tag{A7}$$

Expressed in CMB units, the contribution of synchrotron emission to the SPIDER model is

$$\begin{aligned} h_{\rm CMB}\, d_s &= \frac{\nu^2}{c^2} \left(g \frac{\partial B}{\partial T} \Big|_{\nu,\, T_{\rm cmb}} \right)^{-1} g \frac{c^2}{\nu^2} A_s \left(\frac{\nu}{\nu_0} \right)^{\beta_s} \\ &= A_s \left(\frac{\nu}{\nu_0} \right)^{\beta_s} \left(\frac{\partial B}{\partial T} \Big|_{\nu,\, T_{\rm cmb}} \right)^{-1}. \end{aligned} \tag{A8}$$

Note that this convention differs from that used in, for example, *Planck* Collaboration et al. (2020d), where synchrotron emission is modeled as a power law in the brightness temperature (see Section 4.2 of *Planck* Collaboration et al. 2020d and

Section 3.2.2 of *Planck* Collaboration et al. 2014), such that

$$T_b(\nu) \propto \left(\frac{\nu}{\nu_0}\right)^{\beta_s'}. \tag{A9}$$

Combining Equations (A9) and (A5), we see that the spectral radiance corresponding to the *Planck* model is $I_s(\nu) \propto \nu^{\beta_s'+2}$, in contrast to the model of Equation (A7); the two power-law indices are related to one another as $\beta_s = \beta_s' + 2$. Therefore, when adopting the *Planck* β_s' posterior as our prior in Section 8.4, we use a value of -1.15 ± 0.17 rather than -3.15 ± 0.17. It is important to note that this index difference applies to both synchrotron and dust but the *Planck* model scales the dust component by $\beta_d - 2$ rather than simply by β_d, so the value of β_d in the two models is equivalent.

ORCID iDs

P. A. R. Ade https://orcid.org/0000-0002-5127-0401
S. J. Benton https://orcid.org/0000-0002-4214-9298
J. R. Bond https://orcid.org/0000-0003-2358-9949
H. C. Chiang https://orcid.org/0000-0002-4098-9533
C. R. Contaldi https://orcid.org/0000-0001-7285-0707
O. Doré https://orcid.org/0000-0001-7432-2932
A. J. Duivenvoorden https://orcid.org/0000-0003-2856-2382
J. P. Filippini https://orcid.org/0000-0001-8217-6832
K. Ganga https://orcid.org/0000-0001-8159-8208
R. Gualtieri https://orcid.org/0000-0003-4245-2315
J. E. Gudmundsson https://orcid.org/0000-0003-1760-0355
M. Halpern https://orcid.org/0000-0002-1760-0868
W. C. Jones https://orcid.org/0000-0002-3636-1241
J. S.-Y. Leung https://orcid.org/0000-0001-7116-3710
S. Li https://orcid.org/0000-0002-8896-911X
L. Moncelsi https://orcid.org/0000-0002-4242-3015
J. M. Nagy https://orcid.org/0000-0002-2036-7008
I. L. Padilla https://orcid.org/0000-0002-0024-2662
A. S. Rahlin https://orcid.org/0000-0003-3953-1776
T. M. Ruud https://orcid.org/0000-0002-7615-1900
J. D. Soler https://orcid.org/0000-0002-0294-4465
I. K. Wehus https://orcid.org/0000-0003-3821-7275

References

Abazajian, K. N., et al. 2016, arXiv:1610.02743
Ade, P. A. R., Aikin, R. W., Amiri, M., et al. 2015, ApJ, 812, 176
Ade, P. A. R., Pisano, G., Tucker, C., & Weaver, S. 2006, Proc. SPIE, 6275, 62750U
Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. 1983, PhRvD, 28, 679
Battistelli, E. S., Amiri, M., Burger, B., et al. 2008, JLTP, 151, 908
Benton, S. J., Ade, P. A., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91450V
Bergman, A. S., Ade, P. A. R., Akers, S., et al. 2018, JLTP, 193, 1075
Bicep2 Collaboration 2015, ApJ, 814, 110
BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., et al. 2014a, PhRvL, 112, 241101
BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., et al. 2014b, ApJ, 792, 62
BICEP2/Keck Array Collaboration, Ade, P. A. R., Ahmed, Z., et al. 2018, PhRvL, 121, 221301
Bond, J. R., Jaffe, A. H., & Knox, L. 2000, ApJ, 533, 19
Brandt, W. N., Lawrence, C. R., Readhead, A. C. S., Pakianathan, J. N., & Fiola, T. M. 1994, ApJ, 424, 1
Bryan, S. A., Ade, P. A. R., Amiri, M., et al. 2010a, Proc. SPIE, 7741, 77412B
Bryan, S. A., Ade, P. A. R., Amiri, M., et al. 2016, RSciI, 87, 014501
Bryan, S. A., Montroy, T. E., & Ruhl, J. E. 2010b, ApOpt, 49, 6313
Cardoso, J.-F., Le Jeune, M., Delabrouille, J., Betoule, M., & Patanchon, G. 2008, ISTSP, 2, 735
Choi, S. K., & Page, L. A. 2015, JCAP, 2015, 020
Chon, G., Challinor, A., Prunet, S., Hivon, E., & Szapudi, I. 2004, MNRAS, 350, 914
Cook, W. G., Glushchenko, I. A., Ijjas, A., Pretorius, F., & Steinhardt, P. J. 2020, PhLB, 808, 135690
de Korte, P. A. J., Beyer, J., Deiker, S., et al. 2003, RScI, 74, 3807
Delabrouille, J., & Cardoso, J. F. 2007, arXiv:astro-ph/0702198
Delabrouille, J., Cardoso, J. F., & Patanchon, G. 2003, MNRAS, 346, 1089
Delouis, J.-M., Pagano, L., Mottet, S., Puget, J.-L., & Vibert, L. 2019, A&A, 629, A38
Duivenvoorden, A. J., Gudmundsson, J. E., & Rahlin, A. S. 2019, MNRAS, 486, 5448
Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306
Efstathiou, G., & Gratton, S. 2019, OJAp, 4, 8
Feldman, G. J., & Cousins, R. D. 1998, PhRvD, 57, 3873
Filippini, J. P., Ade, P. A. R., Amiri, M., et al. 2010, Proc. SPIE, 7741, 77411N
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fraisse, A. A., Ade, P. A. R., Amiri, M., et al. 2013, JCAP, 2013, 047
Gambrel, A. E. 2018, PhD thesis, Princeton University
Gambrel, A. E., Rahlin, A. S., Song, X., et al. 2021, ApJ, 922, 132
Gandilo, N. N., Ade, P. A. R., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91452U
Gerbino, M., Lattanzi, M., Migliaccio, M., et al. 2020, FrP, 8, 15
Gorski, K. M., Hivon, E., Banday, A., et al. 2005, ApJ, 622, 759
Gualtieri, R., Filippini, J. P., Ade, P. A. R., et al. 2018, JLTP, 193, 1112
Gudmundsson, J. E., Ade, P. A. R., Amiri, M., et al. 2015, Cryo, 72, 65
Guth, A. H., & Pi, S. Y. 1982, PhRvL, 49, 1110
Hawking, S. W. 1982, PhLB, 115, 295
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Ijjas, A., & Steinhardt, P. J. 2018, CQGra, 35, 135004
Ijjas, A., & Steinhardt, P. J. 2019, PhLB, 795, 666
Jones, W. C., Ade, P. A. R., Bock, J. J., et al. 2006, ApJ, 647, 823
Jones, W. C., Montroy, T. E., Crill, B. P., et al. 2007, A&A, 470, 771
Kamionkowski, M., & Jaffe, A. H. 2001, IJMPA, 16, 116
Kamionkowski, M., & Kovetz, E. D. 2016, ARA&A, 54, 227
Katayama, N., & Komatsu, E. 2011, ApJ, 737, 78
Loken, C., Gruner, D., Groer, L., et al. 2010, JPhCS, 256, 012026
Lueker, M., Reichardt, C., Schaffer, K., et al. 2010, ApJ, 719, 1045
Mather, J. C., Fixsen, D. J., Shafer, R. A., Mosier, C., & Wilkinson, D. T. 1999, ApJ, 512, 511
Mukhanov, V. F., & Chibisov, G. V. 1982, ZhETF, 83, 475
Nagy, J., Ade, P., Amiri, M., et al. 2017, ApJ, 844, 151
Nagy, J. M. 2017, PhD thesis, Case Western Reserve University
Netterfield, C. B., Ade, P. A. R., Bock, J. J., et al. 2002, ApJ, 571, 604
Osherson, B., Filippini, J. P., Fu, J., et al. 2020, JLTP, 199, 1127
Padilla, I. L., Eimer, J. R., Li, Y., et al. 2020, ApJ, 889, 105
Page, L., Hinshaw, G., Komatsu, E., et al. 2007, ApJS, 170, 335
Peebles, P. 2012, ARA&A, 50, 1
Planck Collaboration 2016a, A&A, 594, A10
Planck Collaboration 2016b, A&A, 594, A1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A9
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016a, A&A, 594, A26
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, A&A, 594, A13
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A1
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020b, A&A, 641, A6
Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020c, A&A, 641, A10
Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020d, A&A, 641, A11
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020e, A&A, 641, A3
Ponce, M., van Zon, R., Northrup, S., et al. 2019, in Proc. Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning), PEARC '19 (New York, NY: Association for Computing Machinery)
Rahlin, A. S. 2016, PhD thesis, Princeton University
Rahlin, A. S., Ade, P. A. R., Amiri, M., et al. 2014, Proc. SPIE, 9153, 915313
Rocha, G., Contaldi, C. R., Bond, J. R., & Górski, K. M. 2011, MNRAS, 414, 823
Runyan, M. C., Ade, P. A. R., Amiri, M., et al. 2010, Proc. SPIE, 7741, 77411O
Seljak, U., & Zaldarriaga, M. 1997, PhRvL, 78, 2054
Shandera, S., Adshead, P., Amin, M., et al. 2019, BAAS, 51, 338
Shariff, J. A., Ade, P. A. R., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91450U
Shaw, E. C., Ade, P. A. R., Akers, S., et al. 2020, Proc. SPIE, 11453, 114532F

Soler, J. D., Ade, P. A. R., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91450T
Starobinsky, A. A. 1982, PhLB, 117, 175
Stiehl, G. M., Cho, H. M., Hilton, G. C., et al. 2011, ITAS, 21, 298
Tanabashi, M., Hagiwara, K., Hikasa, K., et al. 2018, PhRvD, 98, 030001
Tegmark, M. 1997, ApJL, 480, L87
Tegmark, M., & de Oliveira-Costa, A. 2001, PhRvD, 64, 063001
Ticra Tools 2020, User Manual, 19.1.1, https://www.ticra.com/
Tristram, M. 2006, in Proc. CMB and Physics of the Early Universe, Ischia, Italy, 20–22 April 2006, 63
Tristram, M., Banday, A. J., Górski, K. M., et al. 2021, A&A, 647, A128
Tristram, M., Macías-Pérez, J. F., Renault, C., & Santos, D. 2005, MNRAS, 358, 833
Young, E. Y. 2018, Studying the Cosmic Microwave Background with SPIDER's First Flight, PhD thesis, Princeton University