CrossMark

An Overview of CHIME, the Canadian Hydrogen Intensity Mapping Experiment

The CHIME Collaboration,

Mandana Amiri[1], Kevin Bandura[2,3], Anja Boskovic[1], Tianyue Chen[4], Jean-François Cliche[5], Meiling Deng[1,6,7],
Nolan Denman[8], Matt Dobbs[5], Mateus Fandino[1,9], Simon Foreman[6,7], Mark Halpern[1], David Hanna[5],
Alex S. Hill[6,10], Gary Hinshaw[1], Carolin Höfer[1], Joseph Kania[11], Peter Klages[8], T. L. Landecker[6],
Joshua MacEachern[1], Kiyoshi Masui[4,12], Juan Mena-Parra[4], Nikola Milutinovic[1], Arash Mirhosseini[1],
Laura Newburgh[13], Rick Nitsche[1], Anna Ordog[6,10], Ue-Li Pen[7,14,15,16], Tristan Pinsonneault-Marotte[1], Ava Polzin[17],
Alex Reda[13], Andre Renard[15], J. Richard Shaw[1], Seth R. Siegel[5], Saurabh Singh[5,18], Rick Smegal[1],
Ian Tretyakov[19], Kwinten Van Gassen[1], Keith Vanderlinde[8,15], Haochen Wang[4,12], Donald V. Wiebe[1],
James S. Willis[15], and Dallas Wulf[5]

[1] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada
[2] Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA
[3] Center for Gravitational Waves and Cosmology, West Virginia University, Morgantown, WV, USA
[4] MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA; jdmena@mit.edu
[5] Department of Physics, McGill University, Montreal, QC, Canada
[6] Dominion Radio Astrophysical Observatory, Herzberg Astronomy & Astrophysics Research Centre, National Research Council Canada, Penticton, BC, Canada
sforeman@perimeterinstitute.ca
[7] Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada
[8] David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON, Canada
[9] Department of Physical Sciences, Thompson Rivers University, Kamloops, BC, Canada
[10] Department of Computer Science, Math, Physics, and Statistics, University of British Columbia-Okanagan, Kelowna, BC, Canada
[11] Department of Physics and Astronomy, West Virginia University, Morgantown, WV, USA
[12] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA
[13] Department of Physics, Yale University, New Haven, CT, USA
[14] Canadian Institute for Theoretical Astrophysics, Toronto, ON, Canada
[15] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Toronto, ON, Canada
[16] Institute of Astronomy and Astrophysics, Academia Sinica, Taipei, Taiwan
[17] Department of Astronomy, Yale University, New Haven, CT, USA
[18] Raman Research Institute, Sadashivanagar, Bengaluru, India
[19] Department of Physics, University of Toronto, Toronto, ON, Canada

Abstract

The Canadian Hydrogen Intensity Mapping Experiment (CHIME) is a drift scan radio telescope operating across the 400–800 MHz band. CHIME is located at the Dominion Radio Astrophysical Observatory near Penticton, BC, Canada. The instrument is designed to map neutral hydrogen over the redshift range 0.8–2.5 to constrain the expansion history of the universe. This goal drives the design features of the instrument. CHIME consists of four parallel cylindrical reflectors, oriented north–south, each 100 m × 20 m and outfitted with a 256-element dual-polarization linear feed array. CHIME observes a two-degree-wide stripe covering the entire meridian at any given moment, observing three-quarters of the sky every day owing to Earth's rotation. An FX correlator utilizes field-programmable gate arrays and graphics processing units to digitize and correlate the signals, with different correlation products generated for cosmological, fast radio burst, pulsar, very long baseline interferometry, and 21 cm absorber back ends. For the cosmology back end, the $N_{\rm feed}^2$ correlation matrix is formed for 1024 frequency channels across the band every 31 ms. A data receiver system applies calibration and flagging and, for our primary cosmological data product, stacks redundant baselines and integrates for 10 s. We present an overview of the instrument, its performance metrics based on the first 3 yr of science data, and we describe the current progress in characterizing CHIME's primary beam response. We also present maps of the sky derived from CHIME data; we are using versions of these maps for a cosmological stacking analysis, as well as for investigation of Galactic foregrounds.

Unified Astronomy Thesaurus concepts: Cosmology (343); Baryon acoustic oscillations (138); Radio telescopes (1360); Astronomical instrumentation (799); Dark energy (351); H I line emission (690); Interferometric correlation (807); Radio interferometers (1345); Telescopes (1689)

1. Introduction

The emergence of cosmic acceleration—the increasingly rapid expansion of the universe since redshift ∼1.5—has

signaled either that a gravitationally repulsive dark energy dominates the energy density of the universe today or that Einstein's general relativity does not correctly describe gravity on cosmological scales. The impact of this discovery on fundamental physics and astrophysics is revolutionary, and decoding the physics of cosmic acceleration requires new, higher-quality measurements of the expansion rate of the universe as a function of time.

Nature has provided a standard ruler with which to measure the expansion history of the universe: the baryon acoustic oscillation (BAO) scale (Seo & Eisenstein 2003, 2007). Acoustic waves propagated through the primordial plasma in the early universe for a fixed amount of time—379,000 yr— until the plasma cooled and became neutral gas, primarily hydrogen. The distance these waves traveled has been precisely measured in the cosmic microwave background (CMB) radiation (Hinshaw et al. 2013; Planck Collaboration et al. 2020). These waves imparted slight baryonic overdensities on the BAO scale that are imprinted in the large-scale distribution of matter in the universe. By measuring cosmic structure as a function of time (i.e., redshift), we can deduce the apparent size of the BAO scale as a function of cosmic epoch and hence the expansion history of the universe.

The signature of BAO was first detected in large-scale structure, at redshift $z \approx 0.35$ (Eisenstein et al. 2005) and $z \approx 0.2$ (Cole et al. 2005), using galaxies as tracers. More recently, measurements of the BAO scale at redshifts up to $z \sim 0.8$ have been made by observing the distribution of optically detected galaxies, using either spectroscopic (Percival et al. 2007; Beutler et al. 2011; Blake et al. 2011; Anderson et al. 2012; Padmanabhan et al. 2012; Ross et al. 2015; Alam et al. 2017, 2021) or photometric (Seo et al. 2012; DES Collaboration et al. 2019, 2022) catalogs, and at higher redshifts in Lyα systems (e.g., Busca et al. 2013; Slosar et al. 2013; du Mas des Bourboux et al. 2020) and quasars (Ata et al. 2018; Neveux et al. 2020). All of these efforts produce measurements of the distance–redshift relation that are consistent with the notion that the dark energy is a cosmological constant with an equation of state $p_{DE} = -\rho_{DE}$ ($w = -1$; Alam et al. 2021). However, improved precision in the distance–redshift relation is still possible owing to the fact that only a small fraction of the accessible large-scale structure has been mapped to date, especially at redshifts greater than 1. Several efforts are ongoing to map ever-larger volumes of large-scale structure to yield improved precision, particularly by the ground-based experiments the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016) and the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) and the space-based telescopes Roman (Akeson et al. 2019), Euclid (Amendola et al. 2018), and SPHEREx (Doré et al. 2014).

A complementary way to map the large-scale distribution of matter, called hydrogen intensity mapping, has been successfully demonstrated by several analyses (Pen et al. 2009; Chang et al. 2010; Masui et al. 2013; Switzer et al. 2013; Anderson et al. 2018; Wolz et al. 2022). The technique uses modest angular resolution observations of redshifted 21 cm emission from the hyperfine transition of neutral hydrogen to trace the distribution of hydrogen gas, and thus matter, in the universe. Hydrogen intensity mapping allows the apparent angular and radial BAO scale to be measured through cosmic history without the expensive and time-consuming step of resolving individual galaxies.

While the intensity mapping technique was first demonstrated using conventional radio telescopes, a dedicated instrument is needed to make a measurement of cosmic acceleration with the sensitivity required to test dark energy models. In order to reduce power spectrum uncertainties due to sample variance, we need to map cosmic hydrogen over nearly half the sky, which requires a telescope with a *much* higher mapping speed than previously existed.
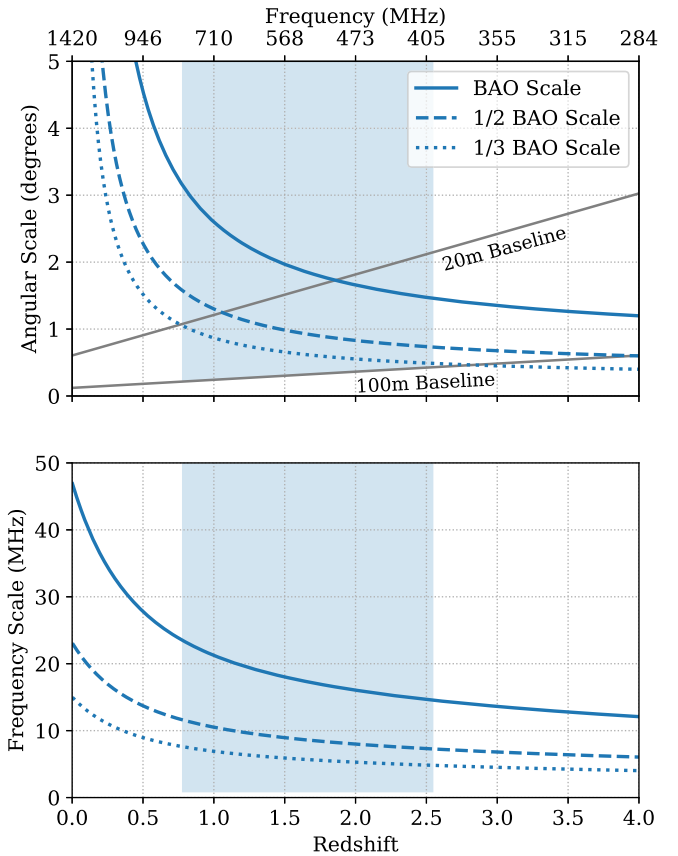

Figure 1. The BAO scale ($r_{BAO} \approx 150$ Mpc comoving) compared to CHIME's angular and frequency resolution. Top: the blue solid curve shows the angular scale associated with r_{BAO}, while the other line styles show the first few harmonics (corresponding to the peaks of successive BAO "wiggles" in $k_\perp$ Fourier space, located at multiples of $k_{BAO} \approx 2\pi/r_{BAO}$). The shaded region shows the range of angular scales accessible to CHIME as a function of frequency, for antenna baselines ranging from 0.3 to 100 m. The gray straight lines show the angular resolution associated with feed separations of 20 and 100 m. Bottom: the solid curve shows the frequency separation associated with the line-of-sight BAO diameter for 21 cm radiation as a function of redshift. The other line styles indicate the frequency resolution required to resolve the first two BAO harmonics in $k_\parallel$ Fourier space. CHIME's frequency resolution is 390 kHz. For all curves, we take $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. In both panels the shaded region denotes the frequency and redshift coverage of CHIME.

As described in this paper, the Canadian Hydrogen Intensity Mapping Experiment (CHIME) consists of an array of four 20 m × 100 m cylindrical telescopes, with no moving parts or cryogenic systems, which can observe the northern sky every day over the frequency range 400–800 MHz. As shown in Figure 1, CHIME's angular resolution of $\sim$40′ and frequency resolution of 390 kHz are well suited to measuring the BAO scale in 21 cm emission over the redshift range $0.8 \leqslant z \leqslant 2.5$. This range covers the important epoch in cosmic history when the expansion transitioned from decelerating to accelerating (Riess et al. 2004).

CHIME's large-scale structure map will constitute the largest survey of the universe ever undertaken. In addition to facilitating measurements of the BAO scale, CHIME data will constitute a rich data set for cross-correlating with other probes of large-scale structure. In a companion paper, we present a CHIME detection of cosmological 21 cm emission in cross-correlation with three separate tracers of large-scale structure

extracted from the Sloan Digital Sky Survey (CHIME Collaboration et al. 2022a).

The main challenge associated with 21 cm intensity mapping is the very bright synchrotron foreground emission from the Milky Way and from other nearby galaxies (e.g., Santos et al. 2005; Liu & Tegmark 2012). We are investigating several approaches to foreground filtering and subtraction, which rely in various ways on recognizing the difference between the smooth Galactic spectrum and the chaotic BAO spectrum along each line of sight (e.g., Shaw et al. 2015). Separately, we note here that CHIME provides a detailed and high signal-to-noise ratio data set for probing the interstellar medium.

CHIME will map the northern sky in polarization, and we will apply the Faraday synthesis technique (Brentjens & de Bruyn 2005) to obtain 3D information about magnetized interstellar structures in the Galaxy. This data set will be without precedent in the northern hemisphere and will form a component of the Global Magneto-Ionic Medium Survey (GMIMS). GMIMS is the first effort to measure the all-sky 3D structure of the Galactic magnetic field, using telescopes around the world to obtain maps with sensitivity to the range of Faraday depth structures we expect in the diffuse medium (Wolleben et al. 2019, 2021); the CHIME frequency range is a critical component of GMIMS.

CHIME has the same collecting area as the Green Bank telescope and also has a large fractional bandwidth and large instantaneous field of view. It scans the entire sky visible from southern Canada at daily cadence with submillisecond sampling. The data from CHIME are passed commensally to separate instruments that search for fast radio bursts (FRBs), monitor known pulsars visible from the site, and search at high spectral resolution for 21 cm line absorption systems. Additionally, CHIME supports very long baseline interferometry (VLBI; Cassanelli et al. 2022) observations with other telescopes.

In Section 2, we present an overview of the CHIME instrument, including its mechanical design, analog and digital systems, and low-level data processing. In Section 3, we describe recent progress in characterizing CHIME's primary beam response. Section 4 is devoted to various performance metrics based on the first 3 yr of science data, including sources of data loss, gain stability, thermal noise, excision of radio-frequency interference (RFI), and preliminary sky maps. We conclude in Section 5, discussing the outlook for future 21 cm measurements and showing an idealized forecast for the precision with which CHIME could measure the cosmic expansion history in the absence of foregrounds or systematics. (The details of this forecast are included in the Appendix.)

2. Instrument and Low-level Processing

CHIME is a transit radio telescope. It consists of linear arrays of feeds along the focus of each of four cylindrical parabolic reflectors. The optical system has no moving parts, and CHIME scans the sky as the Earth turns. A photograph of the telescope and surrounding site is shown in Figure 2.

In this section, we walk through the design of the instrument, showing how its main features have been designed coherently to meet the performance requirements established in Section 1. The signal flow is captured schematically in Figure 3, and we will follow this same path in our description: from reflectors that define the field of view, through feeds and analog electronics, to an FX correlator and the digital back end we call the data receiver.

As described in the introduction, the frequency coverage of CHIME is chosen to interrogate the epoch when dark energy first emerged in the dynamics of the universe. A wide observing bandwidth increases the total cosmic signal power and allows interrogation of a wide range in redshift. Limiting the frequency range to cover a factor of two eases the challenges in antenna design and allows digital sampling in the second Nyquist zone, which permits slower sampling and a substantial savings in the cost of electronics. CHIME takes advantage of the historic drop in the cost of low-noise amplifiers (LNAs) and digital electronics to fill the aperture of its cylindrical reflectors with radio feeds in one dimension. In this geometry, every feed scans the full north–south (N–S) meridian synchronously and simultaneously, and the instrument scans the full overhead sky every day with no moving parts, reducing systematic errors.

2.1. Site

CHIME is built at the Dominion Radio Astrophysical Observatory (DRAO), near Penticton, BC, Canada. DRAO is operated as a national facility for radio astronomy by the National Research Council Canada. Working at the DRAO has provided the CHIME team with very welcome connections to a community of experienced radio astronomers and engineers.

The site is in the White Lake Basin, within the traditional and unceded territory of the Syilx/Okanagan people. Prior to construction we walked the land with elders, and during initial excavation Okanagan Nation observers were present. The site offers flat land protected from RFI by federal, provincial, and local regulation and by surrounding mountains. The climate is semiarid, with low snowfall levels (relative to other places in Canada), important for a stationary telescope. The DRAO's John A. Galt Telescope, a 26 m steerable single-dish telescope with an equatorial mount, is located 230 m east of the center of CHIME, and 20 m north. We use the Galt Telescope for holographic beam mapping. The DRAO supports CHIME with roads, AC power, machine shop access, well-equipped electronics laboratories, office space, and staff accommodation.

The mountains around the observatory shield the site from RFI from nearby cities, but a significant portion of the CHIME frequency band is still contaminated by satellites, airplanes, wireless communication, and TV broadcasting bands. This includes LTE bands in the 730–755 MHz range, TV station bands between 480 and 580 MHz, and UHF repeaters around 450 MHz. These features are clearly visible in the spectrum shown in Figure 4. Besides cell-phone and TV station bands that are static in nature, there are many sources of intermittent RFI events such as direct transmission from satellites and airplanes, as well as scattering of distant ground-based sources. One such event is visible in Figure 4 from 460 to 600 MHz at around 165 s. These scattering events typically appear as 6 MHz wide bursts that last for a few seconds and are caused by the reflection of distant broadcast TV bands from meteor ionization trails or aircraft.

2.2. Mechanical and Optical Design

The design of CHIME is focused on enabling the measurement of BAO across the redshift range where dark energy begins

Figure 2. A photograph of CHIME looking north. The parabolic reflecting surface of each of the four cylinders is 20 m aperture, 5 m focal length, and 100 m long. The two people standing at the southeast corner (right, foreground) help to show the physical size. A total of 256 dual-polarization antennas are placed along the central 78 m of the focal line of each cylinder, beneath an 88 m × 0.65 m ground plane. The focal lines are covered by and suspended from the walkways visible along each cylinder axis. Signals are amplified at each feed and brought by low-loss coaxial cables to receiver huts located in commercial RF-shielded rooms within customized RF-protective shipping containers, one located between the first and second cylinders, another between the third and fourth. After band-defining amplification, analog-to-digital conversion, a time-to-frequency transform, and half of a "corner turn," signals are brought from the two receiver huts to additional RF-shielded rooms within the white shipping containers seen at the right, where the corner turn is completed and a spatial transform and other processing are performed. The gray and black structure at the far right is an ambient air heat exchanger associated with the water-cooling system for the X-engine in the adjacent RF-shielded rooms. Behind that, also gray, is a 0.5 MW power substation to power the instrument. CHIME is located at the Dominion Radio Astrophysical Observatory, which is protected by law and the adjacent hills from terrestrial radio interference. In the background one can see five dishes of the DRAO Synthesis telescope and a solar radio monitor.

to impact the dynamics of the universe. The spectral response, reflector geometry, and radio frequency (RF) feeds are designed together to form an instrument tuned to perform this measurement in a way that allows control and characterization of systematic errors. Total estimated cost was also a strong design driver.

Measuring BAO in the redshift range from 0.8 to 2.5 covers the region of interest for probing dark energy and fills in a redshift gap that is sparsely covered by optical measurements. At these wavelengths, sufficient angular resolution to resolve BAO features in the power spectrum of the sky is easily achieved by a 100 m baseline (see Figure 1).

An east–west (E–W) array of cylindrical, 100 m long reflectors each coupled to a linear feed array along its focus meets these needs. Such a system scans a N–S stripe of the sky interferometrically and observes most of the 3/4 of the celestial sphere visible from our site every day as the Earth turns. Given that each feed in this system requires a feed response of ± 1 rad along the cylinder axis, choosing a reflector shape to be an $f/0.25$ parabola allows the use of feeds with approximately symmetric angular response patterns. At this f-ratio, the focus is level with the edges of the reflector, protecting the feed array from terrestrial radiation.

The required E–W separation of feed arrays can be achieved by varying the number of cylinders and the aperture of each. Deploying four 20 m aperture reflectors was chosen as a reasonable compromise of costs of the reflectors and costs of the electronics to collect and process the signals while still providing massively redundant measurements of the most important (u, v) baselines. This redundancy simultaneously provides lower system noise and protection from minor variations of the response of individual elements of the instrument.

We describe the layout of the telescope in a 3D Cartesian system with $+Z$ pointing to the zenith, $+X$ to the east, and $+Y$ to the north. Thus, the linear feed arrays are oriented along the y-axis with X and Y polarization directions. When we describe the angular response of the telescope, we use the orthographic projected angles x and y defined in Section 3.2.

A steerable telescope can be turned to low elevation angles to shed snow, but this is not possible with the CHIME reflectors. Therefore, the reflector surface is formed with wire mesh to allow snow to fall though. Larger gaps in the mesh shed snow with more assurance but also allow thermal radiation from the ground to leak through to the focus, raising the system temperature. Heavy wire gauge lowers the RF
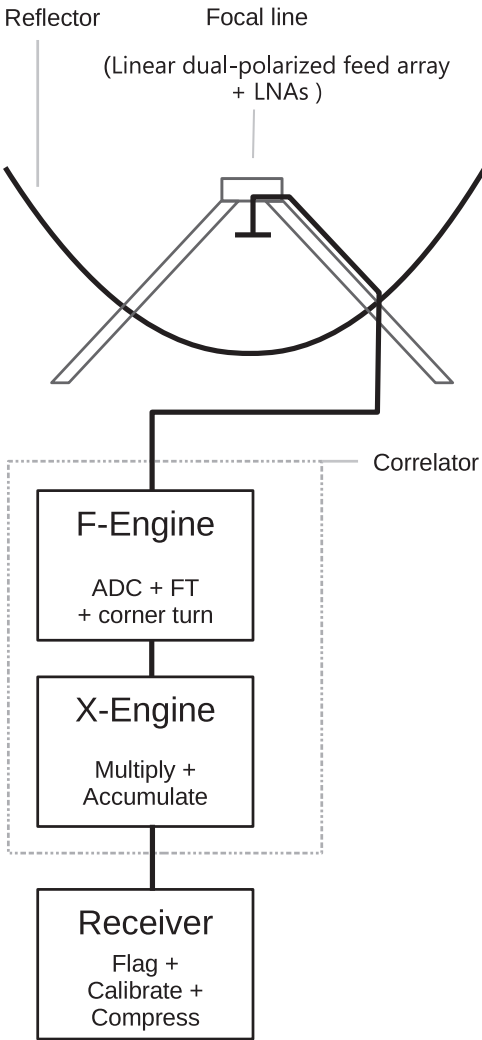
Figure 3. Schematic diagram of data flow through CHIME. Signals focused onto a linear array of broadband dual-polarization antennas are amplified (each polarization separately) using room-temperature receivers with a noise performance below 30 K that amplify and filter the signals to 400–800 MHz. The correlator is an FX design, where the F-engine digitizes and channelizes the signals from the 2048 analog receivers and also implements the majority of a corner-turn network that rearranges the channelized data for spatial correlation. The X-engine completes the corner turn and performs the cross-multiplications and averaging to compute the N_{feed}^2 spatial correlation matrix separately at each frequency. The X-engine also performs additional real-time data processing operations to beamform and to increase spectral resolution for the pulsar, FRB, and absorber back-end instruments.
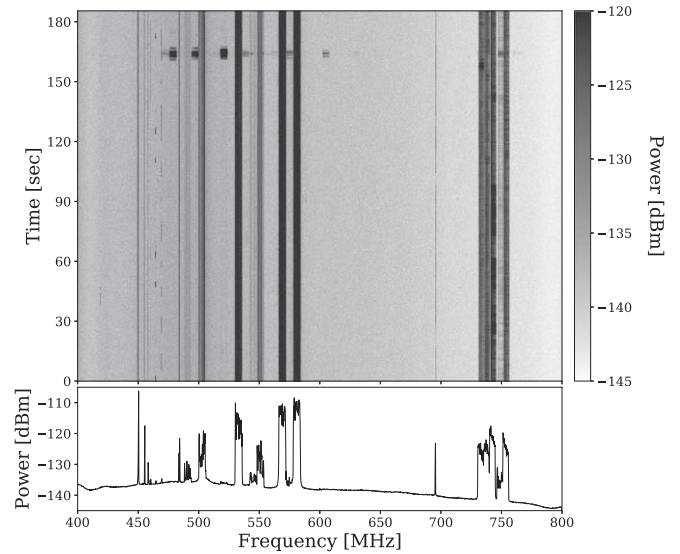

Figure 4. Example of the RFI measured by CHIME antennas. Top: dynamic spectrum obtained from the squared magnitude of the visibility between two E–W polarized antennas located near the southern end of two adjacent cylinders. The frequency resolution is 195 kHz, obtained by direct transform of the raw data rather than through the CHIME PFB pipeline, and the time resolution is 100 ms, averaged from the 1 ms cadence at which this data set was collected. The gray scale denotes power in units of dBm at a feed in the focal line. Data were collected on 2019 May 4 at 16:00 PDT. An intermittent RFI event is visible at 460–600 MHz at around 165 s. Bottom: the median value over time of the image shown in the top panel. Notable features include LTE bands at 730–755 MHz, several 6 MHz TV station bands between 480 and 580 MHz, and UHF repeaters around 450 MHz. The narrow line at 690 MHz is the local oscillator of a nearby synthesis telescope.

leakage. Using tools from Mumford (1961), we evaluated RF leakage across the CHIME band of commercially available sheets of heavy-duty mesh, settling on 19 mm spacing woven mesh made of 2.2 mm diameter galvanized steel. This material is easily available in large flat sheets. The leakage through these sheets adds from 1 to 2 K to the system temperature across the CHIME band.

The central 78 m of each focal line is instrumented with feeds and LNAs. The 100 m long reflectors intercept the beams of the end feeds out to a zenith angle of 65°. These end feeds do see more RFI and more thermal loading than typical feeds, and this is accounted for in our analysis pipeline (see Section 4.1).

The reflector structure itself was designed in collaboration with Empire Dynamic Systems, Coquitlam, BC, a civil engineering firm with substantial experience building astronomical facilities using standard steel fabrication techniques. Each 8 m long section of the reflector is formed from three panels. These are rolled steel beams connected by 8 m long purlins running parallel to the axis, assembled on site and lifted into place. The mesh reflector surface is bolted to the purlins once the structure of an entire cylinder is complete. The structure is supported on steel legs that stand on cement footings placed deep enough that the base is below the anticipated frost depth.

The surface accuracy, shown in Figure 5, corresponds to between $\lambda/50$ and $\lambda/100$ across the CHIME band. The surface errors are dominated by two terms: a consistent imperfect shape formed by the purlins welded to the curved steel frames and by almost 1 cm of sag of the mesh in each of the 1 m gaps between purlins. These perturbations are coherent for the full length of each cylinder in the N–S direction and were measured by tracking a retro-reflector across the full surface using a surveyor's total station.

The ground plane of the linear feed array, at the focus of the cylinder, is just wide enough that it can shield the narrowest building-code-compliant walkway placed above it. Removable panels of the walkway facilitate access to amplifiers and cables. Access stairs at the north end of every focal line are in line with the optic axis and the same width as the ground plane.

Observations of bright point sources acquired with CHIME exhibit an unexpected phase error that scales linearly with E–W baseline distance, frequency, and the sine of the source's zenith angle. This can be explained by a clockwise rotation (looking down from the sky) of the telescope structure by 0°.071 ± 0°.004 with respect to the true astronomical N–S direction. Alternatively, it can be explained by a linear offset in the N–S
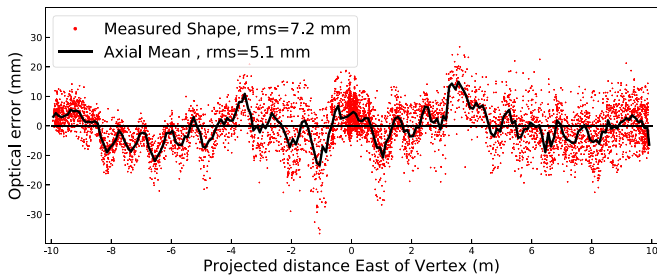
Figure 5. Measured surface error of CHIME cylinder A compared to a best-fit parabola plotted against cylinder *X*, the horizontal distance east of the vertex. Points on the surface are measured with a surveyor's total station tracking a retro-reflector on a small wheeled cart as it moves over the reflector surface. The survey accuracy is nominally 3 mm in 100 m, which has not been subtracted from the scatter seen here. The quantity plotted is half the optical delay error from the sky to the reflector to the focus, equivalent to simple surface error for a flat mirror. There are two main terms in the shape visible here. The mesh that forms the surface appears to sag approximately 1 cm in each of the 1 m gaps between the supporting purlins compared to the desired parabolic shape. Additionally, one can see that the rolled parabolic truss is formed of three segments that also depart from the desired shape by nearly 1 cm. The net surface deviation is 7.2 mm rms, or $\lambda/50$ at CHIME's shortest wavelength. These deviations are clearly coherent over the entire structure of a cylinder. Each of the four cylinders looks similar to this one example in all the key features.

positions of the feeds from one cylinder to the next of -2.73 ± 0.15 cm per cylinder (from west to east). The quoted values were measured by minimizing the phase of visibilities when beamformed to the location of 24 bright point sources ranging in decl. from 5° to 65°. We are currently unable to distinguish between these two explanations owing to confusion between this effect and the phase of the beam response as a function of hour angle. We assume an overall rotation of the telescope when constructing the baseline distances that are used in our analyses.

2.3. Analog System

The analog signal path consists of 256 dual-polarized cloverleaf antennas (Deng & Campbell-Wilson 2014; Deng 2020) in a linear feed array along the focus of each cylinder, with each linear polarization coupled to an LNA, coaxial cables, a band-defining filter and amplifier (FLA), and the input to an analog-to-digital converter (ADC). A single channel is shown in Figure 6. The system components have been designed together to optimize overall performance for interferometric measurement of the BAO. With 256 dual-polarized antennas per cylinder and four cylinders, there are 1024 antennas and 2048 analog signal chains.

Each cloverleaf antenna, together with its image antenna in the ground plane, has an effective focus nominally located at the ground plane, independent of frequency. The radiating board, whose current pattern is shown in Figure 7, is designed to have a smooth petal shape in order to be free of resonances and match to the CHIME LNA over the octave bandwidth from 400 to 800 MHz. Deng (2020) described this optimization. For each linear polarization, pairs of balanced signals from the four petals are combined via a tuned set of microstrip transmission lines (a balun) to form a single-ended signal at the input to the LNA on the base of the antenna. The petals are printed on the top and bottom surface thin (0."031) FR4 printed circuit board (PCB) material and liberally connected with vias, while the stem and base are printed on low-loss Arlon DiClad 880
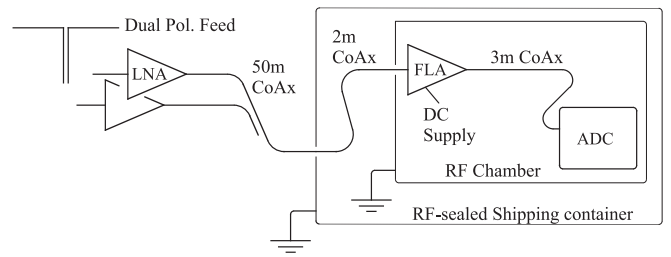

Figure 6. Block diagram of one channel of the analog front end. A signal from one port of a dual-polarization antenna is connected to an LNA and carried via a ~1 m cable and a 50 m low-loss coaxial cable into the double-shielded receiver hut. FLAs, mounted on the inside surface of the inner RF chamber wall, define the instrument passband, provide additional gain, and transmit the signal to the ADCs. A DC Power Supply in the RF chamber is used to power the FLAs, which, in turn, provide DC power to the LNAs over the coaxial cables using the built-in bias tees in the LNAs and FLAs. Any FLA/LNA chain can be turned on or off remotely. The antennas, amplifiers, and 50 m cables are all labeled with barcodes, which are scanned upon assembly, allowing their interconnections to be documented in a database. S-parameters have been measured for every individual component over the full CHIME band.
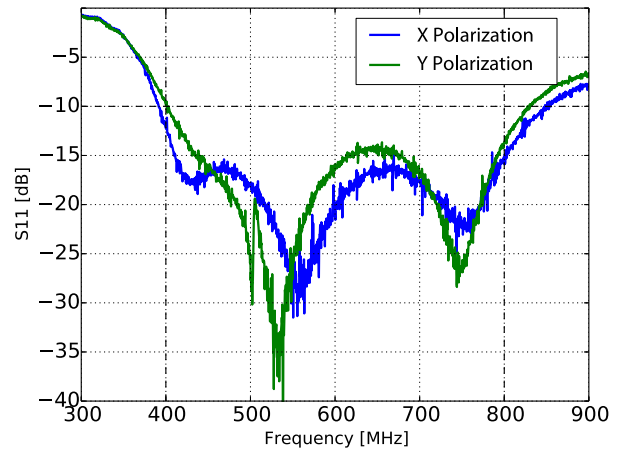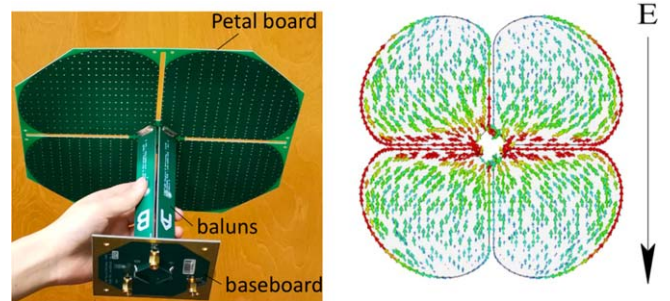

Figure 7. Top left: a photograph of a CHIME cloverleaf antenna element. Top right: the simulated current pattern on the petal-shaped radiating board of the cloverleaf antenna at 600 MHz. Feeds are constructed using commercial PCB materials and techniques, resulting in precise and economic antennas. Bottom: measured S11 of the two polarizations of the cloverleaf antenna. The design substantially exceeds the goal to have a return loss of more than 10 dB over the full CHIME band, illustrated by the horizontal dashed line.

($Dk = 2.2$) material using ordinary printed circuit techniques (Leung 2008).

Feeds are 305 mm apart along the focal line (the telescope *y*-axis) and communicate with one another with coupling coefficients that depend on separation, polarization, signal frequency, and angle of incidence. Coupling between feeds
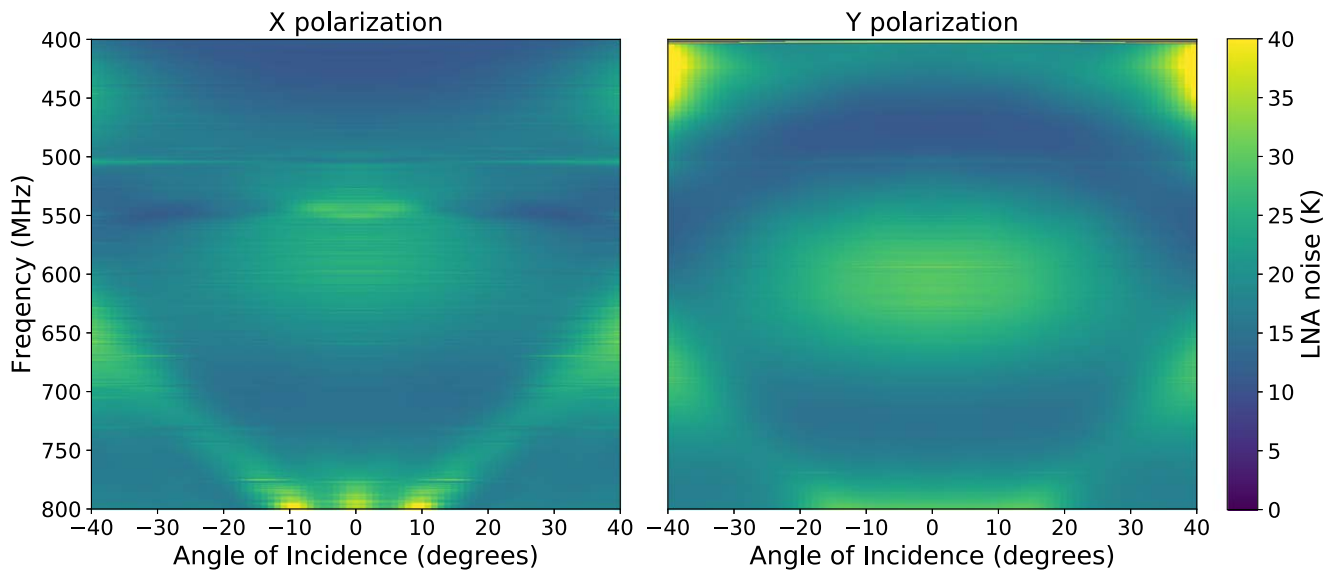
Figure 8. Modeled LNA noise temperature for the central element of a linear feed array. Using measured feed-to-feed coupling parameters (S_{21}, S_{31}, S_{41}, ...), the effective impedance for the central feed in a linear array has been calculated as a function of frequency and incident angle. The noise is calculated using this impedance and a high-fidelity model of our LNA performance. Because of the stronger coupling for *X* polarization, particularly in the vicinity of the feature near 550 MHz, 15 elements are used in the *X*-impedance model, and 13 for *Y*. The sharp feature at 500 MHz in both polarizations is a property of isolated CHIME antennas.
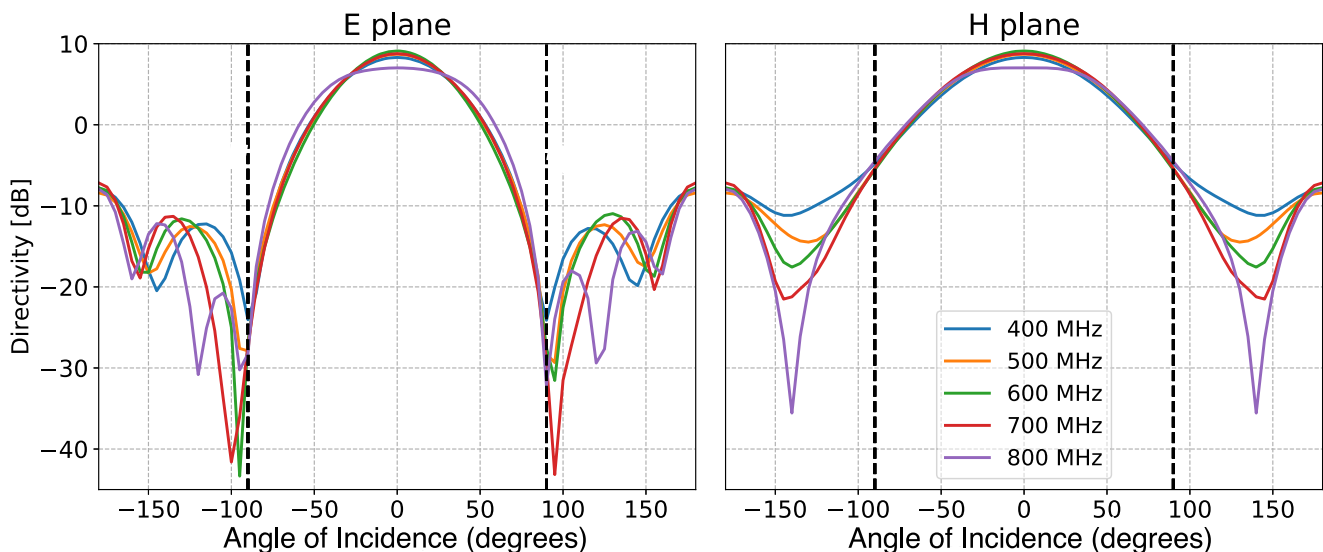

Figure 9. Sections of the modeled angular response of a CHIME feed in the E and H planes for several frequencies across the CHIME band. Although the dual-feed antenna is symmetric with respect to its *x*- and *y*-axes, each beam is slightly elliptical between its E and H planes, and therefore *Y*-polarized and *X*-polarized beams illuminate the reflector differently. The vertical dashed lines in the E and H panels are at $\pm 90°$, corresponding to the edges of the reflector for *X*- and *Y*-polarized radiation.

separated by as much as five times the basic interval is not negligible. The baluns are designed to produce an effective impedance of each element of the linear antenna array, including these coupling terms, which is noise-optimal for our LNA. Balun designs are therefore different for *X*- and *Y*-polarized elements because interfeed coupling is stronger for the *X* ($E\perp$ to separation) polarization than for *Y* ($E \parallel$ to separation). The calculated noise temperature for the central element of a linear array is shown in Figure 8 as a function of frequency and incident angle.

Figure 9 shows models of the angular response of an individual feed, modeled using CST Studio,[20] for several frequencies across the CHIME band. As desired for feeds facing an $f/0.25$ cylindrical reflector, the beam shape is broad and the beamwidth is largely independent of frequency over the CHIME band. Notice that the E-plane and H-plane beam widths are slightly different from each other. Therefore, the *X*-polarized and *Y*-polarized channels have slightly different illumination patterns on the reflector and slightly different far-field angular response patterns. The consequences of this variation will be discussed in Section 3.

The amplification and phase response of the remaining analog chain are plotted in Figure 10. The very sharp band edges at 400 and 800 MHz are designed to allow half-Nyquist sampling of the signal. The response is achieved with a custom bandpass filter built for CHIME by Mini-circuits,[21] model

---

[20] https://www.cst.com/
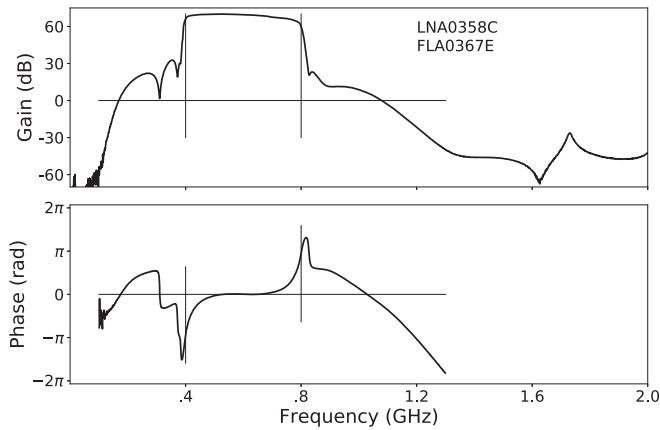
[21] https://www.minicircuits.com/

Figure 10. Top: gain of the analog chain from the LNA at the CHIME feed to the ADC input. The vertical dashed lines show the edges of the second Nyquist band for the CHIME ADC sampling cadence of 800 MHz, corresponding to the CHIME bandwidth of digital signals. The chief elements in this analog chain are an LNA with a peak gain of 42 dB and a gentle roll-off above $f = 1$ GHz, a filter amplifier with a peak gain of 38 dB and a well-defined passband provided by a custom filter from Mini-circuits (BPF-600-2+), 50 m of low-loss LMR-400-type coaxial cable, and 5 m of higher-loss cable located within the receiver huts. Bottom: the sum of measured phase shifts of all components of the analog chain plotted against frequency. A single delay term is subtracted to show a flat phase curve at the center of the band. Phase shifts associated with the very steep edges of the CHIME band-defining filters are evident.

BPF-600-2+, and installed following the first gain stage of the second-stage amplifiers (FLA). One sees in Figure 8 that the LNA noise across the CHIME band is roughly 20 K. The gains of the LNA and FLA are chosen so that all other noise contributions are minor. The FLA contributes 0.6 K at the very top end of the CHIME band. Cable losses and ADC input noise are less than this.

The nonlinear response coefficients for the CHIME analog chain are plotted in Figure 11, with all coefficients referred to the LNA input. By design, the system third-order intercept point (IP3) within the CHIME band is dominated by that of the ADC. The LNA and the first stage of the FLA are not protected by the bandpass filter, so in principle strong out-of-band RFI could produce in-band harmonics from nonlinear response of the front end. Extreme care has been taken with the nonlinearity of the front-end electronics to avoid this. RFI at the CHIME site does not reach the levels that would produce a nonlinear response in our electronics.

It is worthwhile to make a few remarks about the technical details of deploying 4000 amplifiers and a similar number of cables over a 100 m square. The LNA and FLA are built into folded steel boxes that are soldered shut. A small slab of RF absorber is glued inside the FLA boxes to suppress oscillations of the final stage to which earlier generations of our amplifier were prone. Aluminum segments of the focal line that we call cassettes, consisting of four antennas, eight LNAs, and associated 1 m long SMA-to-N-type cables, are assembled indoors and carried to the focal line, where they are mounted in place and bolted to each other. Thus, the interfeed spacing is set by digital machining. The 50 m low-loss N-type coaxial cables connecting the LNAs to the FLAs at the receiver hut are cut to be the same length to within 0.1%, and the optical delay of each cable has been measured separately. Excess cable length for the antennas nearest the hut is stored in cable trays running the length of each cylinder in a geometry we call an optical trombone. A full set of S-parameters is measured at the factory for each cable, and serial numbers for
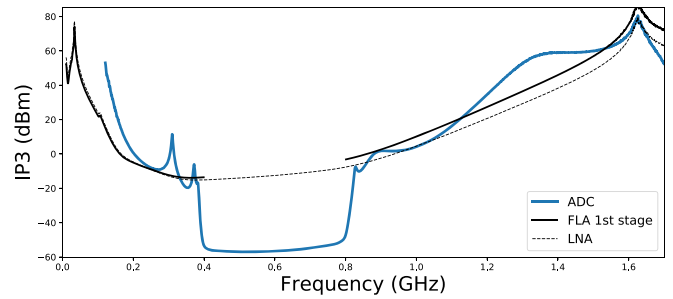

Figure 11. Analog chain linearity parameters, referred to the LNA input, are plotted against frequency. The nonlinearity parameter IP3 is 13 dBm at the input of the CHIME ADCs, where amplified RF power is highest. The output coefficients, OP3, for the FLA and LNA are measured to be 35 and 30 dBm, respectively, nearly independent of frequency. These coefficients are more useful referred to a common point, and so we have referred them all to the equivalent coefficients at the input of the LNA, taking account of gains, bandpasses, and cable losses in front of each element. We have nearly achieved our design goal that the system limit is set by the ADC at all frequencies. In normal operation RFI signals at CHIME do not reach these levels either in band or nearby out of band.

each are recorded on barcodes. This is the practice for all components of the analog chains. During system assembly, pairwise connectivity of all analog components is recorded using a hand scanner and an interactive script operating on a mobile device.

The FLAs sit within an RF-shielded room, with their input connectors protruding through a bulkhead in the wall. DC power is supplied to the LNA from the FLA over the coaxial cable. The amplifiers of each individual signal chain can be powered off by remote command if desired. The RF room provides 100 dB of attenuation and houses the ADC and F-engine. Once installed, physical access to any antenna or LNA is available by lifting the floorboards of a walkway along each focal line. This system is less waterproof than we wish, and in heavy rains water can get to the baseboards of the antennas, causing temporary unacceptable performance. The focal-line structure, consisting of an elevated enclosed dry volume mildly heated by the LNAs, is a nearly ideal bird habitat; consequently, we have found that it is very important that there are no holes as large as 2 cm diameter anywhere in the structure since these would would allow starlings to enter.

2.4. FX Correlator

CHIME employs an FX correlator in which the time-domain signal from each feed is transformed to form a frequency spectrum in a part called the F-engine. At each frequency, data from every feed are collected at a single designated computation node and a spatial transform is made of these signals to form visibilities. This spatial transform is performed in a part of the instrument called an X-engine. These two processes are described below. The F-engine consists of eight 16-card electronics crates housed in two separate RF-shielded rooms located in modified, cooled, 20-foot shipping containers between pairs of cylinders. These two containers are connected by optical fiber to the X-engine, which is housed in a pair of RF-shielded rooms enclosed in 40 foot shipping containers, adjacent to the telescope. The X-engine is built from 256 graphics processing unit (GPU) nodes and is water cooled.

2.4.1. F-Engine

The F-engine is implemented using the ICE (Bandura et al. 2016a) platform. ICE uses a field-programmable gate array
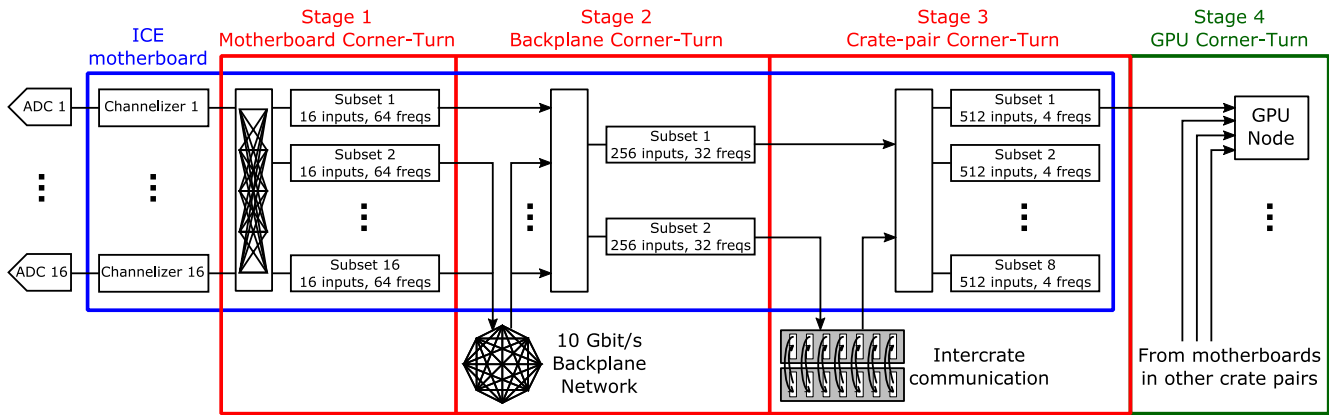
Figure 12. Data flow through the F-engine. A total of 128 ICE motherboards are required to process 2048 sky signals. These motherboards are installed in eight crates, with each crate handling the signals for one polarization from every antenna on one cylinder. Each motherboard digitizes 16 analog signals into 8 bits at 800 MSPS. The data stream from each digitized signal is fed to an FFT/PFB that splits the 400 MHz bandwidth into 1024 frequency channels. A four-stage corner-turn network rearranges the data to allow spatial cross-multiplication and integration at each frequency in the X-engine. In stage one, each motherboard creates 16 new data streams, each one having 64 frequency channels from each of the 16 input signals. In stage two, motherboards within a crate exchange data through a high-speed backplane network such that each board holds the data for 64 unique frequency channels from all of the 256 inputs processed by that crate. In stage three, each motherboard sends the data from half of its frequency channels to a sister motherboard in an adjacent crate. With this intercrate data exchange, each board within a crate pair contains the data for a subset of 32 unique frequency channels and 512 inputs. Stage four is completed within the X-engine GPU nodes. Each ICE motherboard reorders the data into eight subsets, each containing four frequency channels for 512 inputs. Each subset is sent to a different GPU node. Each of the 256 GPU nodes receives data from four different motherboards such that it ends up with the information from all the 1024 polarized antennas for four unique frequency channels.

(FPGA) and is a general-purpose astrophysics hardware and software framework that is customized to implement the data acquisition, frequency channelization, and corner-turning operations of the CHIME correlator.

A schematic diagram of the data flow through the F-engine is shown in Figure 12. The core of the system is built around ICE motherboards that handle signal processing and networking using Xilinx Kintex-7 FPGAs. Each motherboard supports two custom ADC daughter boards. FPGA firmware and software are customized for the CHIME application. Each ICE motherboard digitizes 16 analog signals into 8 bits at 800 million samples per second (MSPS). Thus, the 400–800 MHz sky signals are directly sampled in the second Nyquist zone.

The data stream from each digitized signal is fed to the FPGA, which implements a polyphase filter bank (PFB) efficiently using a fast Fourier transform (FFT; Parsons et al. 2008). Data are processed in frames of 2048 samples, separately for each stream. A PFB is more compact in frequency than a simple FFT would be, greatly aiding RFI excision by localizing any disturbance. At the cadence of individual data frames, the PFB applies a sinc-Hamming window to four consecutive data frames and outputs a single frame of 1024 complex values, one value per 390 kHz wide frequency channel, in 18 + 18 bit real and imaginary format. After the PFB, the data are rounded to 1024 4+4 bit complex values per frame. Adjustable scaling factors (complex gains) are applied to each frequency channel before this step in order to optimize the data compression (Mena-Parra et al. 2018).

After the frequency channelization, each ICE motherboard holds the data for 1024 frequency channels of signals from 16 analog inputs. However, in the X-engine for each frequency, data from every input must be presented to one processor in order to compute the cross-multiplications and averaging required to form the visibilities. A total of 6.6 Tbit s^{-1} of data needs to be rearranged and transmitted to the X-engine, an operation performed in a four-stage corner-turn network (Bandura et al. 2016b). The first stage is performed in each ICE motherboard, where the frequency-domain data from each input are split into 16 subsets, each containing one-sixteenth of the frequency channels from all 16 inputs.

Each group of 16 ICE motherboards is packaged in a crate, and all the boards within a crate are interconnected through a custom backplane that implements a passive high-speed full-mesh network. CHIME uses a total of eight crates, or 128 ICE motherboards. The second corner-turn stage is a data exchange between the boards in the crate, after which each board has all the data from 256 inputs for 64 of the frequency channels.

The third stage is a data exchange between pairs of ICE motherboards located in adjacent crates using high-speed serial links. After this third stage, the data from 512 inputs are split into 256 subsets distributed through the ICE motherboards of the two crates, and each subset contains four unique frequency channels. Each crate pair contains all the data for one-quarter of the CHIME array, both polarizations from one cylinder.

The fourth stage of the corner-turn network takes place inside the GPU nodes of the X-engine. Each ICE motherboard sends its data stream to eight different GPU nodes through two active 100 m multimode optical fiber QSFP+ to 4 × SFP+ cables. Each GPU node receives one frequency subset from one ICE motherboard in each crate pair and recombines the data to compute the correlation matrix for data from all 2048 inputs in four unique frequency channels.

The four F-engine crate pairs are housed in independent racks distributed between two separate RF-shielded rooms installed within 20-foot modified, RF-shielded shipping containers, known as receiver huts. Each receiver hut serves two cylinders and is placed between them at their midpoint. This arrangement minimizes the total length of coaxial cables running from the focal line of the cylinders to the receiver huts.

A GPS-disciplined, oven-controlled crystal oscillator provides the 10 MHz clock for the F-engine system. The GPS receiver also generates the IRIG-B timecode signal used to insert time stamps in the data. A copy of the clock and absolute time signals is sent to each of the F-engine crates. From there, the signals are distributed to each ICE motherboard and digitizer daughter board through a low-jitter distribution network. A broadband noise source system, which will be

described in Section 2.7, is used to monitor and correct for drift between copies of the clock provided to each digitizer daughter board.

The F- and X-engines communicate over 256 optical fibers. Each fiber cable contains four strands that connect one ICE motherboard to four different GPU nodes. These are carried within a waterproof cable tray that goes underneath the cylinders and above the huts. Also within the cable tray are the coaxial cables that distribute a clock and absolute time signals to the F-engine huts. The mapping of which RF frequencies are sent to which nodes in the X-engine is adjustable. This allows, for example, for sending the data from frequency channels heavily corrupted by RFI to nodes that are temporarily down for repair, preserving useful bandwidth.

2.4.2. X-engine

The CHIME X-Engine performs spatial correlations and other real-time signal processing operations, using 256 nodes, each with four GPU chips. Details of the nodes and support infrastructure can be found in Denman et al. (2020). These nodes run a soft real-time pipeline built using the KOTEKAN framework (Renard et al. 2021; A. Renard et al. 2022, in preparation), which handles the X-engine, RFI flagging, and multiple real-time beamforming operations. The processes performed by each node are shown in Figure 13.

Data arrive at each of the the nodes from the F-engine on four 10 Gbit s^{-1} fiber SFP+ links. Each link conveys data from 512 feeds from four frequency bins from each of the four F-engine crate pairs. Packet capture is handled in KOTEKAN using the DPDK[22] library to reduce UDP packet capture overhead normally associated with using Linux sockets. Once in the system, the packet data from each link is split into four different staging memory frames, one for each frequency, which completes the final corner turn. Following packet capture, there are four frames each with data from one frequency channel and from all 2048 feeds, for 49,152 time samples. These frames are transferred to the GPU chips, resulting in each GPU chip processing data for exactly one of the frequency channels.

Once the data frames are on the GPU, a number of operations are applied to the data using OPENCL and hand-optimized GPU kernels. The primary operation is the creation of the visibility matrix by the correlation kernel, using about 75% of the processing time. For each frequency channel, the complex data from each feed are multiplied by the complex conjugate of the corresponding signal from each other feed to create the visibility matrix. This is a Hermitian matrix, and only the upper triangle is directly computed.

This calculation dominates the computational cost in CHIME, and we worked hard to optimize it. Data are processed independently in blocks of 32 × 32 feeds, distributed across 64 collaborating computational instances ("work items" in a "work group"). These work items employ Cannon's algorithm (Cannon 1969), collectively loading eight sequential time steps for all 32 + 32 inputs under consideration, and sharing these over high-speed local interconnects. Unsigned 4 bit values can be packed into 32 bit registers, allowing efficient multiplication and in situ accumulation (Klages et al. 2015). Ultimately, six of the eight arithmetic operations required for a complex multiply accumulate (cMAC) operation are performed in a single GPU
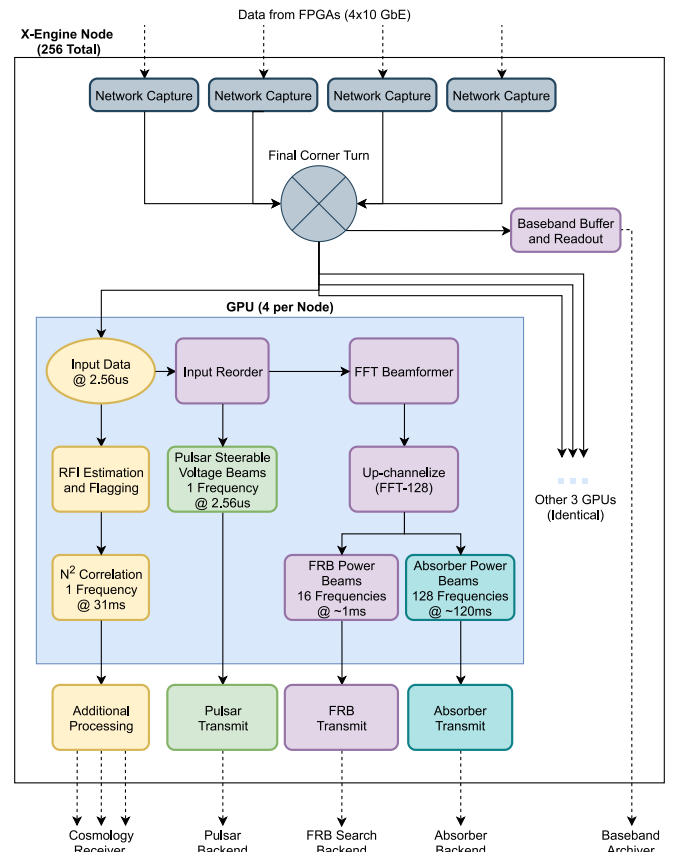
Figure 13. Processes performed by each X-engine node. Data arrive from the F-engine, the final corner turn is performed by the CPU in the X-engine, and signals from all 2048 feeds within one single-frequency channel are transferred to one of the four GPUs. On each GPU the spectral kurtosis is computed as an estimation of RFI, and contaminated samples are removed. After flagging the data for RFI at ~0.6 ms cadence, the data are correlated to produce an $N_{\rm feed}^2$ visibility matrix and summed over 31 ms. Each 31 ms correlation product is copied off the GPU and tested again, this time for long-duration RFI, which is either removed or processed further (see Figure 14). The data are also branched off to two distinct beamforming engines, a tracking voltage beamformer with 12 steerable beams and an FFT spatial beamformer that generates 1024 power beams at increased frequency resolution. Those power beams are further split into two combinations of frequency and temporal resolution. The tracking voltage beamformer is used primarily for the CHIME/Pulsar back end, and the FFT beamformer is used for both the CHIME/FRB search back end and a 21 cm narrowband absorber search back end. A buffer of the most recent 33 s of data from the F-engine is updated in RAM. When triggered by the FRB search engine, the raw voltage data in this buffer, corresponding to one event, are transmitted to an archive.

instruction. The remaining two are paired with another cMAC, for a total of three instructions per pair of cMAC computations. These intermediate products are accumulated in active registers, with top bits periodically peeled off and accumulated to high-speed local memory to prevent overflow. Products are summed in time over 12,288 input time samples, before being unpacked and read out, to produce visibility products with a temporal resolution of roughly 31 ms. To maximize throughput, this kernel was directly implemented in AMD's assembly level Instruction Set Architecture (ISA), and the resulting high performance both left space for additional processing kernels (e.g., beamforming, RFI) and allowed for a substantial reduction in observatory power envelope via low-power operation of the GPUs.

To excise RFI-contaminated data prior to the correlation operations, a spectral kurtosis value is computed over all inputs
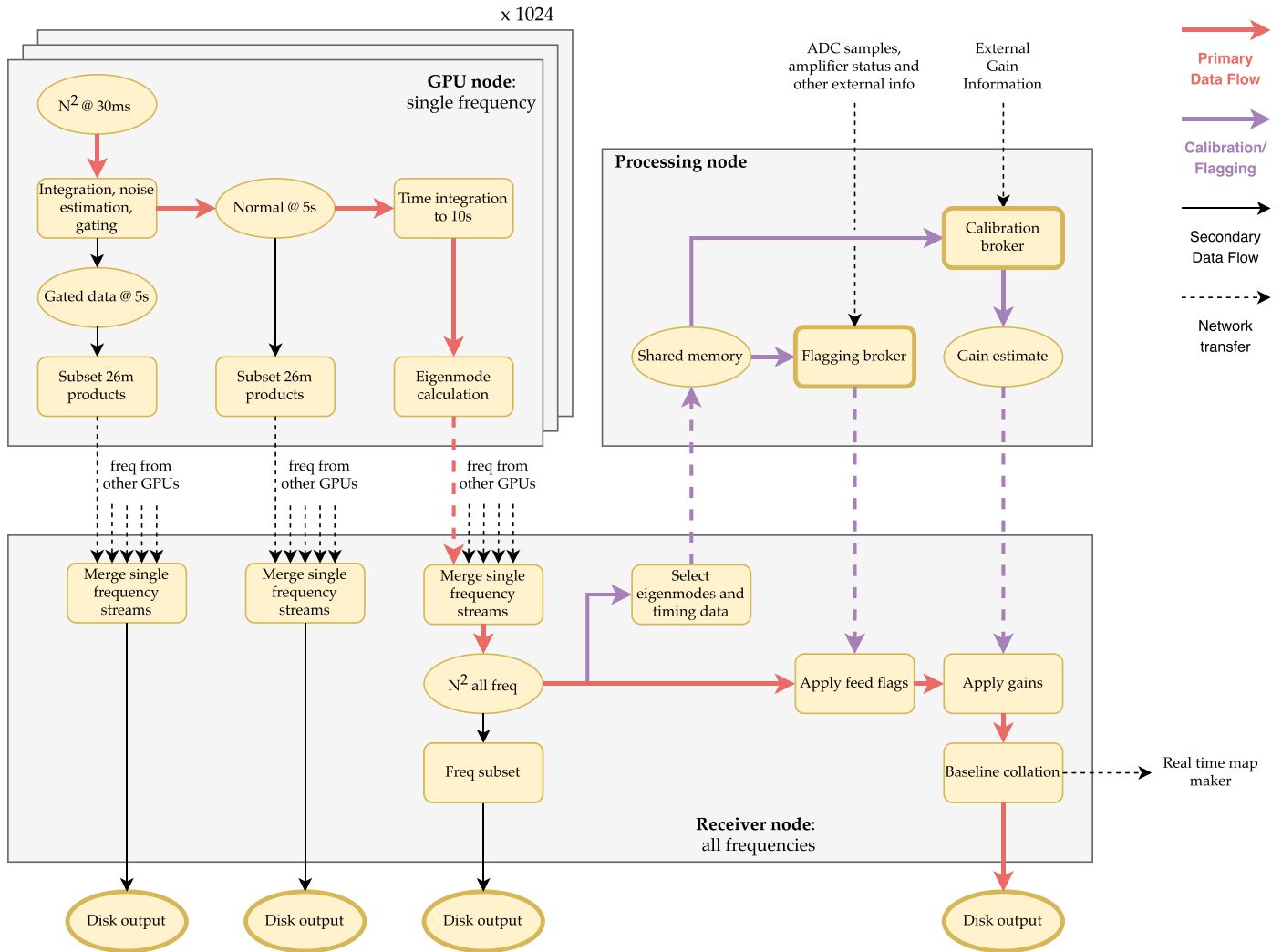
Figure 14. A diagram of the data flow through the CHIME postcorrelation receiver system. The receiver system processes a full N_{feed}^2 correlation matrix for each of the 1024 frequency channels every ~31 ms from the X-engine (see Section 2.4.2). Initially this stream is processed within the CPUs of each GPU node to accumulate the data up to a 10 s cadence, estimate its noise, and, if desired, perform gating for pulsar observations. Cross-correlations with the Galt 26 m Telescope used for bright source and pulsar holography can be extracted at 5 s cadence to prevent fringe smearing. For use in calibration, we solve for the highest four eigenvalues and vectors of each N_{feed}^2 frame. The data for all of this are sent over the network to a single receiver node for further processing, including flagging, calibration, and baseline stacking before being written to disk. Flags for bad correlator inputs are derived by a broker process running on a separate node that assimilates various sources of data-quality information into a mask for each correlator input. Similarly, gain solutions for calibration are derived by a broker that uses eigenvector and noise source timing data from the correlation products, as well as environmental data to produce gains that are applied in real time to the N_{feed}^2 data.

and 256 successive time samples (total ~ 0.66 ms; Taylor et al. 2018). Each 0.66 ms block of data with a kurtosis value deviating from the expected value by a configurable threshold is given 0 weight. The amount of data that are excised or otherwise lost (e.g., to lost network packets) is accounted for in the metadata and normalized later in the pipeline. These kurtosis values are extracted from the GPU and used in a second-stage RFI test that can drop entire 31 ms samples after they leave the GPU based on the statistics of the 48×0.66 ms spectral kurtosis samples within. This second stage is designed to excise RFI events that are lower in power but longer in duration than those found in the first stage. This second-stage excision is turned off during solar transit.

The 31 ms visibility frames that are not excised are processed in the CPU associated with each node and transmitted to a receiver system running another configuration of KOTEKAN that does further processing. See Figure 14 and Section 2.5 for more detail.

In addition to the correlation, RFI estimation, and flagging, the GPUs perform two kinds of beamforming operations. The first type is a tracking voltage beamformer, which takes R.A. and decl. coordinates and generates a set of dynamic phases that are applied to input voltage data and summed over all feeds to generate a single coherent beam used to observe celestial sources while in the CHIME field of view. Currently CHIME forms 12 of these beams simultaneously. The data from these formed beams are scaled to $4 + 4$ bit complex data at full 2.56 μs time resolution and transmitted over the 1 Gigabit Ethernet (GbE) links on the nodes. The data streams from 10 of these beams are sent to 10 CHIME/Pulsar processing nodes. The remaining two beams are used for other operations such as VLBI and calibration.

The second type of beamforming operation is an FFT-based spatial imaging beamformer (Ng et al. 2017) that generates 1024 power beams in fixed terrestrial coordinates for use in the FRB engine and the high-resolution absorber search. A spatial

FFT is performed for the data from each cylinder to generate 512 beams for each polarization. Of these, 256 are selected to achieve roughly achromatic pointing. A four-way transform is computed across all these beams in rows between cylinders. This combination produces 1024 beams at each frequency, and for each of these 128 successive temporal samples are Fourier-transformed to extract higher frequency resolution. For the high frequency resolution absorber search, the data are squared at full 128 subfrequency spectral resolution (~3 kHz) and integrated to ~120 ms time resolution. After leaving the GPU, these high-resolution data are integrated again to 10 s and stored on a back end running a special configuration of KOTEKAN, to enable a search for 21 cm narrow-line absorbers. For the FRB search engine, the data are squared, summed over polarizations, and summed over 16 frequency bins and 384 time samples to produce 1024 power beams with 16 subfrequency bins (~24 kHz) per original CHIME channel at ~1 ms time resolution. This tuning of sampling time and frequency resolution is made to match the data to the conflicting goals in the FRB engine of resolving short pulses and performing dedispersion in a discretely sampled spectrum. These data are sent from each GPU to the FRB search back end in custom UDP packets over a 1 GbE link to be searched in real time for FRBs.

2.5. Real-time Processing

The ensemble of the 1024 GPUs generate $N_{\rm feed}^2$ correlation products at a 31 ms cadence for each of 1024 frequencies. This amounts to a raw data rate of ~4.6 Tbit s^{-1}. It is not feasible to write out and store such a fire hose of data. The receiver system is tasked with aggregating and processing the data stream in preparation for archiving. In the process it produces ancillary data products that are tapped for system and data-quality monitoring. Figure 14 provides a schematic representation of the receiver system. The various stages are distributed across multiple computers (aka nodes). The first of them occur on the GPU nodes themselves (executed on the CPU) before being transmitted over the network to the single receiver node, where the remainder of the pipeline occurs. Another computer, the processing node, hosts parallel processing tasks that are not time-critical for subsets of data. Notably this includes deriving the calibration solutions that are fed back into the main receiver node pipeline. The final data products are sent over the network to an archive node. Aside from a few exceptions, all of these stages are built on the KOTEKAN framework.

Accumulation and gating. In order to reduce the data rate, the first stage following the GPU co-adds RFI-cleaned 31 ms frames for 5 s. A later stage co-adds samples further to the final 10 s cadence, but optionally the subset of the data composed of correlation products with the Galt 26 m Telescope are kept at the finer time resolution to avoid smearing due to the faster fringing of the ~230 m baseline between Galt and CHIME. This is the last chance for any operations on the fast-cadence data. The variance over the 31 ms samples is calculated to estimate the noise level in the accumulated frame and passed along with it. Gated accumulation is also supported, where samples are weighted and binned into on and off gates and the difference of the two is returned at the end of the integration window. Gating can be initiated, or its parameters updated on the fly without interrupting data acquisition. Currently, gating is used for simultaneous observations with the Galt telescope of slow ($P > 300$ ms) pulsars for beam holography (Section 3.2).

Eigendecomposition. The four leading eigenvalues/vectors of the $N_{\rm feed}^2$ visibility matrix are estimated for every time sample and passed on down the pipeline. It is necessary to perform this step in the X-engine in order to distribute the computational load over the 256 CPUs located there. The eigenvectors represent the response of every individual array element to the dominant modes on the sky at that moment, making them a valuable tool for real-time calibration. Importantly, it is not possible to perform this decomposition after the redundant baseline collation step, and the full $N_{\rm feed}^2$ visibility matrix is only stored for a small number of frequencies, so these eigendata are important for off-line analysis as well. Since noise coupling between nearby feeds is significant and will outweigh the sky modes, the diagonal values of up to 30 feed separations are excised from the matrix prior to the decomposition. To avoid biasing the result, an iterative scheme is employed to progressively complete the masked region.

Calibration broker. A daily complex gain calibration for every sky signal is derived from the transit of a bright astronomical point source. The calibration broker is a service running on the processing node that produces gain solutions by fitting the eigenvector data immediately following the transit of a chosen point source. The eigenvectors are continuously provided to the broker via a shared memory ring buffer, and the broker can access a time stream spanning the transit by reading the buffer file approximately 20 minutes after transit. During transit, the bright source is the dominant contribution to the sky signal, and the visibility matrix can be approximated as an outer product of the input gain vector (a rank-1 approximation), identified as the leading eigenvector. A complication is that the 2048 sky signals include two polarizations, so there are in fact two near-orthogonal components to the matrix. There is no guarantee that these two vectors neatly divide the inputs by polarization as is required to interpret the eigenvectors as gain solutions. An additional orthogonalization with respect to the 2D space of polarizations must be performed by the broker to isolate them. The intrinsic flux density of the source across the band is corrected for using the measurements of Perley & Butler (2017). Frequencies affected by RFI are flagged by comparing the ratio of the eigenvalue on- and off-source, and those with anomalous gain amplitudes are also flagged. Gains for the flagged frequencies are recovered by interpolating between the gain solutions for adjacent good frequencies. The four brightest sources are processed in this way at every transit, but only one is used for calibration. The choice of which source is used changes throughout the year to avoid calibrators near the Sun, and any differences in the primary beam patterns are corrected using the average ratio of past gains from the source to past gains from Cygnus A (Cyg A). The calibration procedure therefore normalizes the primary beam pattern at each frequency to unity on meridian at the decl. of Cyg A. See Section 2.7 for additional corrections applied later in the pipeline.

Flagging broker. The role of the flagging broker is to perform real-time identification of correlator inputs that should be excluded from further analysis. It runs on the processing node and provides regular updates to the relevant stages of the receiver pipeline. It uses a variety of data products and housekeeping metrics to repeatedly evaluate 10 different tests, with each test designed to identify malfunctioning or otherwise anomalous correlator inputs. Below we list its data sources and

briefly summarize the corresponding tests. Note that there can be multiple tests derived from a single data source.

1. *Layout database:* Reject inputs that are not currently connected to an antenna or that have been flagged manually by a user.
2. *Power server:* Reject inputs whose amplifiers are not currently powered.
3. *ADC data:* Reject inputs whose raw ADC data have an outlier rms, histogram, or spectrum.
4. *RFI broker:* Reject inputs determined to have highly non-Gaussian statistics based on a monitoring stage internal to the X-engine.
5. *Calibration broker:* Reject inputs for which the complex gain calibration failed, or whose gain amplitudes exhibit large, broadband changes relative to its median over the past 30 days.
6. *Autocorrelation data:* Reject inputs that have outlier noise or whose autocorrelation shows large, broadband changes relative to past values.

If a correlator input fails any one of the tests, all baselines formed from that input will be given zero weight when averaging over redundant baselines.

Gain/flag application and redundant baseline collation. The N_{feed}^2 and 26 m streams from the GPU nodes are merged into all-frequency streams as they arrive at the receiver node. The 26 m streams undergo no further processing, as do a subset of four frequencies from the N_{feed}^2 visibility calculation that are output at this stage to preserve some of the full array information. Keeping N_{feed}^2 terms for all frequencies, amounting to a data rate of over 200 TB day^{-1}, is not feasible because of storage constraints. A lossy compression is effected by averaging redundant baselines within each cylinder pair together. Baselines are not combined between the six cylinder pairs to maintain the possibility of correcting for any nonredundancy between the cylinders or between the signal paths that are routed to separate receiver huts. The daily rate of data archived is thus reduced to ∼1 TB. Prior to collating visibilities along redundant baselines, the gain calibration and flags generated by their respective brokers are applied to the data. This compression method is lossy owing to any nonredundancy that might arise from support structures, edge effects, and imperfections in the reflectors, or nonuniformity in the feed responses, as well as imperfect calibration.

Real-time map. A subset of 64 frequencies is tapped from the main pipeline following the baseline collation stage and transmitted to the processing node, where a separate pipeline beamforms the visibilities to generate a real-time data stream we call a ring map. The ring map is a representation of the data as a time stream of formed beams, visualizing the sky as it drifts through the field of view of the cylindrical reflectors (see Section 4.5 for details). The maps for those frequencies are buffered over a period of 24 hr and can be displayed using a data monitoring web viewer. They are useful for assessing recent data quality at a glance, and we study them every day.

Output data sets. The branching points in the pipeline lead to three main data products. The stack data set is output by the baseline collation stage and contains the total of CHIME's sensitivity, with all nonflagged baselines contributing over the entire band. The N2 data set holds complete uncompressed visibility matrices for four frequencies. It is useful for instrument characterization and understanding the effects of

baseline collation. The gated and ungated 26m data sets contain only the cross-products with inputs from the Galt telescope, and at a 5 s cadence, twice that of the other data sets. These are produced only during simultaneous observations of point sources for beam holography (see Section 3.2).

Compression and archiving. The final module of the real-time pipeline is an archiving service that packages the data into a structured archive format, applies another stage of compression, and registers files with the archive database. It takes advantage of the relatively slow rate of change of the measured sky gradually drifting through the field of view by ordering the data with time as the fastest varying index and compressing the redundant information between nearby time samples. All the data are truncated at a specified fraction of the measured noise level to excise the high variability in the (noise-dominated) least significant bits and thus further improve the effectiveness of the compression. The BITSHUFFLE algorithm (Masui et al. 2015) compresses these data on a bitwise basis, resulting in a typical size reduction of ∼2–5 times for stack files. Data are stored on site for up to 6 months and indefinitely at archives located at Compute Canada centers. Archive files are tracked in an SQL database, and all file operations are mediated by a software daemon that validates the integrity of the data and ensures storage redundancy.

2.6. System Monitoring

CHIME is a complex instrument with 2048 analog signal chains processed by nearly 400 separate computers spread over six physical locations on site. To keep the experiment running 24 hr and 7 days a week, it is important to identify and rectify inevitable failures in a timely manner. In this section, we explain how the CHIME operations are monitored in almost real time to assess instrumental and experimental health.

Instrument health monitoring. The instrumental health can be monitored by verifying that various hardware and software subsystems are running, data are written to disk, equipment huts are thermally stable, and there is no failure that is an emergency and needs immediate attention, e.g., coolant leak or fire. An array of auxiliary sensors are deployed across CHIME to probe various environmental parameters. These include temperature sensors across one cylindrical reflector; ambient temperature, humidity, smoke, and leak sensors in equipment huts; and a weather station with wind and rain-accumulation sensors. Data from all these sensors are streamed in real time into a central database. In addition, metrics are collected from various hardware and software components, including but not limited to power supplies, operating systems, and network statistics from switches. Almost every software and firmware component also generates its own set of internal health metrics.

CHIME uses PROMETHEUS[23] and GRAFANA[24] for managing and monitoring the housekeeping data in real time. PROMETHEUS is an open-source monitoring system and time-series database. The data collected by PROMETHEUS can be displayed through web-based dashboards in the GRAFANA environment. PROMETHEUS allows defining rules for alert conditions and expressing them as a PROMETHEUS query that can invoke an alert to an external service. Alerts are handled by ALERTMANAGER, which sends out notifications through SLACK and email to targeted team members when thresholds set on various metrics are violated.

---

[23] HTTPS://PROMETHEUS.IO/
[24] HTTPS://GRAFANA.COM/GRAFANA

This combination of PROMETHEUS and GRAFANA environment provides the ability to monitor the operation remotely. As there is only one telescope operator on site during working hours, and no one otherwise, the CHIME team provides nearly 24 hr remote monitoring of the operation by taking on shifts on a rotating basis after regular work hours. The person on duty responds to alerts in situ only if they are critical and causing interruption in the data acquisition. As an example, temperature control in equipment huts is quite sophisticated. As both X-engine and F-engine hardware are cooled by liquid coolant, the greatest attention is paid to detecting any potential leaks in the plumbing. If leaks are detected, valves automatically cut the supply of coolant into huts to minimize any potential damage to the system. Similarly, if smoke or flood sensors are persistently tripped, the power is automatically shut to receiver huts. This way the system automatically reacts to catastrophic events, ensuring the safety of subsystems.

A subset of housekeeping data stored in PROMETHEUS is exported and written to an HDF5 file on a daily basis. These files are then archived to be used during off-line data analysis.

Experimental health and data-integrity monitoring. Considering the amount of data that CHIME generates, it is challenging to check the data quality and integrity in real time. The focus of this operation is to highlight only those data-quality issues that can be addressed and improved by acting swiftly and adjusting certain configurable hardware or software parameters. The time frame for these assessments can be seconds (e.g., rms of sky signal), minutes (e.g., spectra waterfall, correlation triangle), or a day (e.g., calibration quality, downward trend in noise integration). Data quality and integrity are monitored though a mix of manual checks and a set of automated quick data analyses on a daily basis by the remote operator(s).

DIAS[25] is a software framework for data-integrity analysis and generation of daily plots. It runs as a service that schedules the execution of data analyzers. This framework replaces slow on-demand script execution with an automated pre-generation of a set of data products, which are not archived and are only available for a few months. A lightweight package for generating web-based plots, THEREMIN, is developed in house and used to view these data products.

Using DIAS and THEREMIN, we are able to monitor the quality of the data itself in near real time. This includes estimates of the RFI environment, the full array sensitivity derived from subintegration variances, bright source spectra, and real-time sky maps derived directly from the saved CHIME data products. This allows the CHIME team to get rapid feedback on the end-to-end performance of the instrument and to make timely adjustments if needed.

2.7. Off-line Processing

Postprocessing of the CHIME data is done via a Python-based, YAML-configurable, off-line pipeline. The basic infrastructure is available in CAPUT (Shaw et al. 2020b), and most of the non-CHIME-specific functionality is available in DRACO (Shaw et al. 2020c). CHIME-specific parts of the pipeline are found in CH_PIPELINE (Shaw et al. 2020a). The pipeline structure is flexible, being used not only for the main data product pipeline but also for a variety of functions such as instrument simulations, holography and cross-correlation analysis, foreground removal, and power spectrum estimation.

The main data pipeline for CHIME runs on Compute Canada's Cedar[26] cluster, where one of our science data archives is located. The data are processed in units of local sidereal days (LSD).[27] The first step of the pipeline is to locate and load all files pertaining to a particular LSD into memory. A number of calibration and transformation operations are performed in the order presented below.

A timing correction is applied to each file to account for differences in timing between the two receiver huts.[28] The final step of redundant baseline stacking is performed, in which redundant baselines corresponding to different pairs of cylinders are stacked together (this step is delayed to this point to allow for the timing calibration to occur). At this point, an off-line stage of RFI masking is applied to the data, complementary to the real-time RFI excision that takes place in the receiver pipeline (see Section 2.5). This stage derives a figure of merit for sensitivity estimates based on the radiometer equation applied to cross-polarization data. This figure of merit is fed to a sum-threshold algorithm (Offringa et al. 2010) in frequency-time space that outputs a single mask for all baseline stacks. This stage also includes a specific search for intermittent RFI with the 6 MHz wide bands, characteristic of TV stations.

To allow for later stacking of multiple sidereal days, the data are resampled to go from the original time-of-day basis to R.A. This regridding is done via an inverse Lanczos interpolation that takes the data from the native resolution of around 10 s to approximately $5'$ in R.A. The regridded data corresponding to a full sidereal day are combined into a sidereal stream, the final data product that is written to disk for analysis and long-term archiving. A few additional products are saved alongside each sidereal stream visibility data. These include ring maps (see Section 4.5), delay power spectra, and bright point-source spectra, as well as the sensitivity figure of merit and the RFI mask derived from them.

An independent second-stage pipeline exists to combine many sidereal streams into higher-sensitivity full sidereal day products called *sidereal stacks*. Initially, all sidereal streams in a specified time range are selected. These are specified to be times of mostly uninterrupted observation in which the telescope was operating in a stable mode. For instance, we require all the data that go in a sidereal stack to have been calibrated on the same source (see Section 2.5 for calibration details).

Before stacking multiple days, an extra step of cleaning is applied to each sidereal stream to remove all daytime data, as well as any times flagged as potentially corrupted by a range of environmental indicators (rain, excessive site RFI, bad calibration due to instrument restart, etc.). The data are combined into aggressively cleaned, Sun-free, sidereal stacks, which are the main science-ready data products of the CHIME data pipeline. Corrections for thermally induced phase shifts as described in Section 4.2 can be applied at this point.

---

[25] https://github.com/chime-experiment/dias

[26] https://docs.computecanada.ca/wiki/Cedar

[27] CHIME uses sidereal days referenced to a 2013 November 15 start.

[28] The timing differences and the corrections applied are described in detail in Section 4.2.2.

3. Beams

The biggest challenge for detecting extragalactic 21 cm emission is filtering out the much brighter foreground emission, dominated by diffuse Galactic emission and extragalactic radio sources (Liu & Tegmark 2011). To do so, it is crucial to have precise knowledge of the instrumental beam response. Estimates by Shaw et al. (2015) indicate that this response must be characterized to roughly a part in 10^4 in power units, and this has motivated the pursuit of a number of parallel strategies for beam measurement and modeling, as well as efforts to quantify the required precision in more detail. In this section, we first describe how CHIME's instrument design determines the general features of the beam response, and then we present the current status of our ongoing work to characterize this response.

3.1. General Features of the CHIME Beams

We define the "base" beam to be the illumination on the sky (amplitude, phase, and polarization) that results when a single feed broadcasts with all other feeds along the focal line shorted (Deng & Campbell-Wilson 2014). Although CHIME never operates as a transmitter, this is a useful construct for understanding the beam properties. In the absence of multipath effects, discussed below and in Section 3.3, this base beam produces a nearly elliptical illumination of the sky: ~120° long in the unfocused N–S direction, along the cylinder axis, and a few degrees wide with frequency-dependent diffraction side-lobes in the E–W direction, perpendicular to CHIME's cylinder axis.

Multipath and other coupling effects alter this simple description by as much as 50% at some frequencies. The physical origin of the multipath interference is radiation interacting with the focal-line assembly, which consists of the linear feed array and a common ground plane. In this environment, a signal broadcast by a feed will reflect off the cylinder, and a large fraction of that signal will go directly to the sky, but a small portion strikes the focal plane assembly, where some is absorbed by a neighboring feed and the rest is reflected and/or reradiated by the assembly, eventually reaching the sky. The details of the latter interaction are complex and are still actively being characterized. Nonetheless, the "primary" beam is the illumination on the sky one gets when these effects are accounted for. The "synthesized" beam is the illumination produced by coherently combining the signal from multiple feeds, each with their own (nearly identical) primary beams. In this section we focus on characterizing CHIME's primary beam.

Since multipath propagation is occurring within a 5 m cavity (CHIME's focal length), new interference fringes arise roughly every 30 MHz in frequency, as seen below. In the remaining sections we present the data sets used to calibrate CHIME's primary beam and discuss approaches to modeling the full response, informed by these data.

3.2. Data Sets for Beam Calibration

Ideally, the CHIME primary beam calibration would be based on direct measurements of the telescope's response to a bright (relative to the sky confusion), polarized point source along every direction in the far field, at every frequency. However, a sufficiently complete population of such sources is not available; instead, we make use of several direct measurements, each of which provides beam information in a different regime. Importantly, these regimes often overlap, which allows for multiple cross-checks on the results. Thus far, the most useful information has been obtained from three data sets: holography of bright point sources, which allows beam amplitude and phase measurements for each feed along a limited number of 1D tracks through the beam; transits of bright point sources, which trace the feed-averaged beam response on meridian; and transits of the Sun, which provide similar information to holography (without the phase information) but with near-continuous sampling over a specific range of decl.

When plotting 2D beam measurements over a large angular extent, we use an orthographic projection with its origin at zenith. This projection has the advantage of not distorting the apparent beamwidth at different elevations. Moreover, the projected coordinates x and y in the tangent plane remain parallel to east and north, respectively. For the unit vector pointing to hour angle ha and decl. δ, the corresponding angular coordinates are given by

$$x = -\cos\delta\sin\mathrm{ha} \tag{1}$$

and

$$y = \cos\ell\sin\delta - \sin\ell\cos\delta\cos\mathrm{ha}, \tag{2}$$

where ℓ is the latitude of the observer ($+49°.3$ for CHIME).

3.2.1. Holography

Holography is an established technique for making accurate measurements of the amplitude and phase of antenna beams at RFs (e.g., Bennett et al. 1976; Scott & Ryle 1977; Baars 2007). We use this technique by tracking a celestial source with a nearby moving telescope while the source transits through the stationary CHIME beam. The correlation between the signals from each stationary feed and the tracking reference telescope traces the response of CHIME along the path of the source. For CHIME holography, the John A. Galt 26 m Telescope, located 230 m east of CHIME, is used as the tracking system. For these observations, a 400–800 MHz dual-polarization modified CHIME receiver is mounted on the Galt Telescope (Berger et al. 2016). The resulting cross-correlations yield CHIME's co-polar and cross-polar far-field beam response (amplitude and phase) per feed, per frequency, along a track in hour angle at the decl. of each observed source.

The data collected to date comprise 1888 tracks of 24 celestial sources since holographic observations began in 2017 October, typically spanning ±40° or more in hour angle and −21° to +65° in decl. (−70° to +16° in zenith angle). The data are fringe-stopped (phase-shifted to account for Earth rotation) and binned to a celestial grid, with the resulting average and variance per bin stored on disk. Data from successive observations of a given source can be combined, reducing measurement noise. A sample holographic measurement of Cyg A is presented in Figure 15, which shows the amplitude and phase of the co- and cross-polar beams in each CHIME cylinder.

For each frequency and co-polar correlation product in the holography data, we fit the sum of a Gaussian profile and a constant offset to the amplitude response as a function of hour angle. The resulting centroid and Gaussian FWHM parameters are shown in Figures 16 and 17, respectively, for all feeds and frequencies.
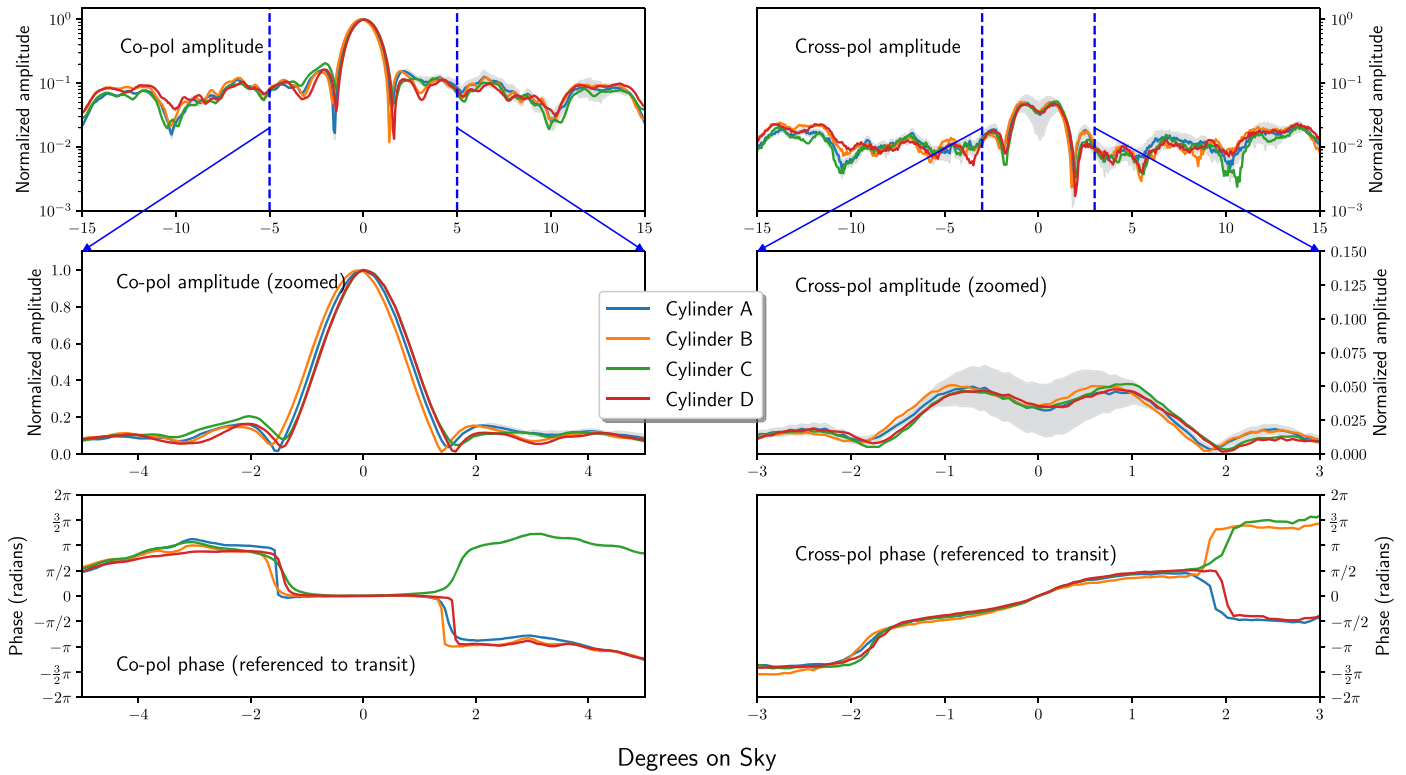
Figure 15. The CHIME X-polarized beam response at 717 MHz from the holographic measurements of a Cyg A transit, taken on 2018 September 28. (See Section 3.2 for a discussion of CHIME's holographic measurement methodology.) Each panel shows the median response taken over all feeds within a cylinder, normalized such that the co-polar response is $1 + 0j$ at transit. Top left: the median co-polar amplitude of all feeds per cylinder, normalized by the Gaussian-fit peak height, over the full extent of the observation, converted to degrees on sky, HA $\cdot \cos(\delta_{\rm CygA})$. Top right: the median cross-polar amplitude from the product of a CHIME feed with the opposite polarization on the Galt receiver. The data have been scaled by the same factor as applied to the co-polar response, so the curves give an indication of the level of cross-polarization in the beam. Middle left: same as the top left panel, but zoomed in to a smaller hour angle range and plotted on a linear scale. Middle right: same as the top right panel, but zoomed in to a smaller hour angle range and plotted on a linear scale. Bottom left: the median co-polar phase as a function of scaled hour angle, taken over all feeds in a cylinder (the median was evaluated for the real and imaginary parts separately before evaluating the phase). Bottom right: same as the bottom left panel, but for the cross-polar phase. The phase difference between cylinders, after accounting for phase wrap, is only large near the first zero crossing of the field. The gray bands in the amplitude plots indicate the standard deviation over all the Cyg A holography tracks of cylinder A's median feed response. (Cylinder A is representative.)
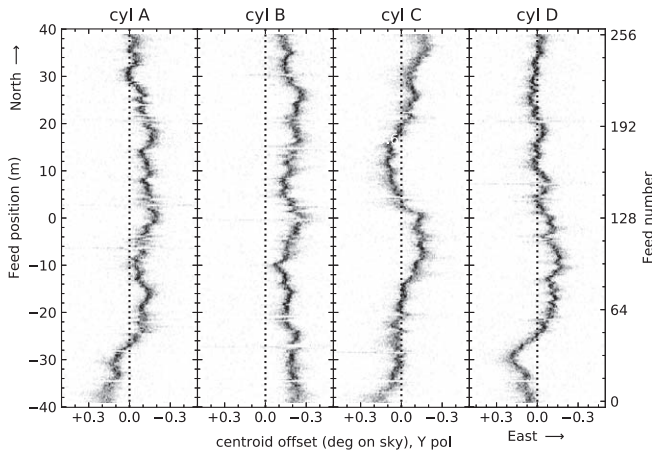

Figure 16. Per-feed measurements of the CHIME E–W beam centroid obtained from a Gaussian fit to the holographic measurements of a Cyg A track. For each cylinder A–D (left to right panels), the best-fit centroid is shown as a function of feed position along the cylinder. Multiple points per feed show results for each nonflagged frequency that was processed for that feed. The spread with frequency arises from a small but statistically significant oscillation in the centroid with a periodicity of 30 MHz, indicating a small E–W asymmetry in the signal multipath. The dominant effect, however, is the position-dependent variation that arises from imperfections in the cylinder surface and primarily from a few millimeters of E–W position offsets of feeds on the focal line. The Y polarization is shown; the X polarization shows a similar trend with a slightly larger frequency variation.

The centroid parameter shows a small but significant dependence on focal-line position that is correlated for nearby feeds (Figure 16). This suggests that the centroid offsets are due to physical displacements of the focal lines and/or cylinder structures from their design positions. Note that, given the 5 m focal length of CHIME, a 0°.2 centroid offset requires an effective position offset of 1.7 cm between the E–W feed position and the symmetry plane of the cylinder. In cylinders A, C, and D, the median centroid offset (taken over feed number) is close to zero, whereas in cylinder B all feeds are offset to the east (i.e., toward negative hour angle), implying that the focal line as a whole is offset by ~1 cm from the symmetry plane of cylinder B's parabolic figure. Multiple points for each feed on a given cylinder in Figure 16 show measurements for that feed at different frequencies, and the spread of these points represents a small frequency dependence in the E–W centroid. This variation has a periodicity of ~30 MHz, which arises from an E–W asymmetry in CHIME's signal multipath. Multipath effects are discussed in Section 3.3.

Figure 17 shows the FWHM parameter as a function of frequency for both polarizations, with multiple points per frequency representing measurements for all the nonflagged feeds for that frequency. As expected given the dipole illumination pattern of the feed, the FWHM is roughly twice as large at 400 MHz as at 800 MHz and ~20% higher in the X polarization than in the Y polarization. Multipath effects cause
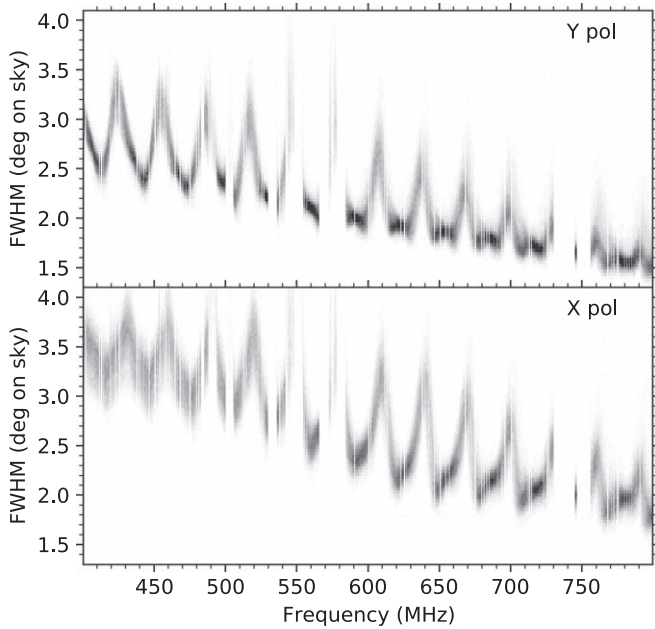
Figure 17. Measurements of the CHIME X–Z plane (E–W) beam FWHM obtained from a Gaussian fit to the holographic measurements of a Cyg A track, plotted as a function of frequency. Multiple points per frequency show results for each nonflagged feed that was processed for that frequency. The top and bottom panels show results for the Y and X polarizations, respectively. The dominant variation in the FWHM arises from signal multipath that introduces a 30 MHz periodicity in the beam response. Characterizing this multipath is the dominant ongoing effort in the CHIME beam calibration program.

the ~30 MHz ripple in the FWHM for both polarizations. There is a larger spread in the FWHM measurements for the X polarization, especially at low frequencies. This difference between polarizations remains after including flags for feeds near structural elements like support struts, so the exact cause of the larger spread in X polarization, as well as its impact on the cosmology data analysis, remains under investigation.

3.2.2. Celestial Sources near Transit

There are 37 bright point sources in CHIME's decl. range with flux greater than 10 Jy at 600 MHz, which is significantly above our estimated confusion noise of ~0.1 Jy. These sources span zenith angles of 37° north of zenith to 38°.9 south of zenith. We measure the spectra of these sources at transit by phasing the CHIME array to the decl. of the source and recording the observed spectrum as a separate data set. Given our Cyg A calibration strategy, the ratio of the observed spectrum to its spectrum reported in the literature gives the ratio of CHIME's on-meridian beam response at the zenith angle of the source to its on-meridian response at the zenith angle of Cyg A. Examples of these data are shown in Figure 18, along with a preliminary fit to a "coupling model" described in Section 3.3.2.

This technique can be extended to a much larger number of fainter sources if we restrict attention to intercylinder baselines that have a large E–W baseline component and therefore lower confusion noise from diffuse synchrotron emission. For the cross-correlation of CHIME data with large-scale structure traced by the extended Baryon Oscillation Spectroscopic Survey (eBOSS; CHIME Collaboration et al. 2022a), we used this technique to produce a model of CHIME's main lobe response from the north to south horizon. A detailed description of the

procedure and the model is given there, so we provide only a brief summary here.

Intercylinder baselines with a large E–W component are largely insensitive to diffuse sky signals, such as Galactic synchrotron emission. Thus, one can approximate the emission measured by these baselines as solely composed of radio point sources (ignoring the subdominant cosmological signal). We construct a model of this sky using catalogs of source spectra measured by the VLA Low-frequency Sky Survey (VLSS; Cohen et al. 2007), the Westerbork Northern Sky Survey (WENSS; Rengelink et al. 1997), the NRAO VLA Sky Survey (NVSS; Condon et al. 1998), and the Green Bank survey (GB6; Gregory et al. 1996). This sky model is put into a simulation pipeline that produces mock (noise-free) visibilities that have no CHIME beam convolution applied. Then, as described in Appendix A of CHIME Collaboration et al. (2022a), we form beams on the sky using both the simulated and measured visibilities and regress the two data sets to infer the primary beam response in the data. The resulting beams are filtered to remove small-scale features that likely originate from flux errors in the catalog.

At present, the model is only derived for hour angles less than roughly 2°, but in principle it can be extended to cover the dominant E–W sidelobes. Figure 19 shows the beam response obtained from this method for the Y polarization at 600 MHz. Our interpretation of the main features of this beam is given in Sections 3.1 and 3.3.

3.2.3. Solar Response

The Sun provides a complementary data set to astrophysical point sources for beam mapping. Every 6 months, the Sun moves between ±23°.5 decl., providing quasi-continuous spatial sampling over this decl. range. Additionally, the brightness of the Sun (>100 kJy) permits unconfused hour angle coverage comparable to the holographic measurements. The flux of the Sun varies with time, but this can be calibrated at every decl. that has a sufficiently bright astrophysical source. Variability between such calibrations limits the accuracy of these data, as does the finite angular size of the Sun, but even this qualitative information is invaluable for guiding beam modeling efforts. Data collected in the fall of 2019 are shown in Figure 20. A more detailed description of CHIME's solar data processing is presented in CHIME Collaboration et al. (2022b).

3.3. Beam Modeling

Ultimately, we seek to use the data sets described above to construct a single comprehensive beam model. The biggest challenge in this endeavor is accurately accounting for the multipath and coupling effects that modulate the simple elliptical base beam. In the following, a few complementary approaches to this problem are described: a data-driven approach, where we attempt to extrapolate the data sets described above to the 2π sr above the horizon, and a semianalytic approach, where we model the coupling between separate feeds with a physically motivated parameterization. Note that this work is ongoing, and further details are deferred to forthcoming papers.

The models described below are intended to describe a typical feed's beam response. The response of individual feeds will deviate from this owing, for example, to perturbations in
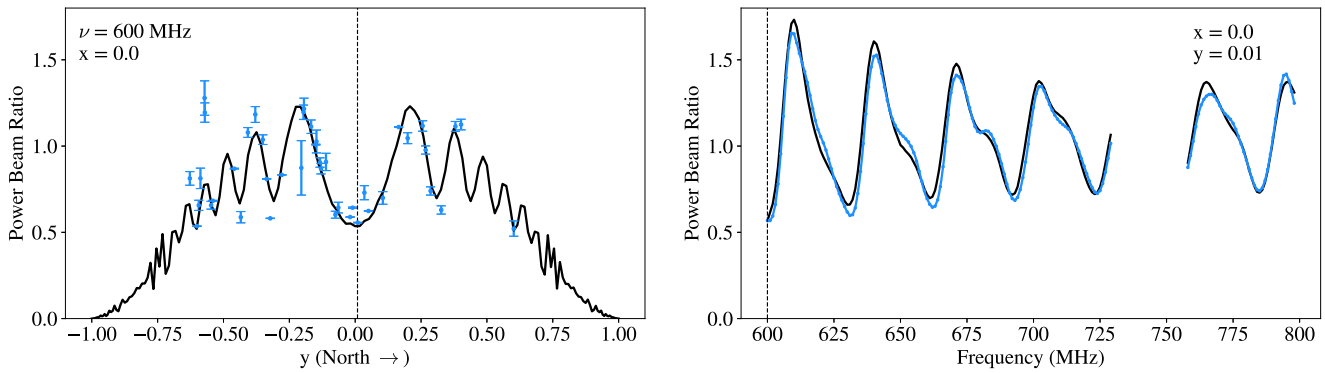
Figure 18. The on-meridian power beam response as determined from 37 bright celestial sources (blue points and curve) and from a coupling model (black curves, Section 3.3.2). Left: the response as a function of orthographic y at 600 MHz. Right: the response as a function of frequency for a source, 3C 147, within $0°5$ of zenith. The vertical dashed lines show the zenith angle and frequency used for constructing 1D beams in the right and left panels, respectively. The blue points and curve are obtained from the beamformed response to 37 point sources at transit, divided by their expected flux from the literature. The uncertainties are dominated by uncertainties in the literature flux densities and are highly correlated across the band. The current best-fit coupling model (Section 3.3.2), fit to all 37 sources in the range 600–800 MHz, is shown in black. This relatively simple model clearly captures the main features of CHIME's on-meridian response.

the cylindrical reflector shape (e.g., Figure 5) and/or to feed position and orientation offsets (e.g., Figure 16). Additionally, the presence of structural elements in the vicinity of some feeds, e.g., support struts, can scatter radiation and alter the beam response of those feeds (Landecker et al. 1991). Given that CHIME measures numerous redundant visibilities (i.e., correlation products with the same baseline), feed-to-feed variations will average down in the stacked data. The extent to which these variations must be accounted for when filtering foregrounds remains to be quantified.

3.3.1. Data-driven Extrapolation

We exploit the fact that CHIME's beam response is nearly separable in orthographic (x, y) angular coordinates, and we use singular value decomposition (SVD) of the solar data to derive a set of beam modes that can be continued to regions not covered by the solar data. The extrapolations can be guided by additional data, e.g., the holography data (Section 3.2.1) and/or the celestial source data (Section 3.2.2), and/or by theory, e.g., the coupling model (Section 3.3.2). We have been developing a few approaches to this extrapolation problem, which we outline below. However, we have yet to settle on a single approach, so we defer the details to a forthcoming paper.

In one approach we form a set of basis functions at a target frequency, derived from the solar data in a small frequency range centered on the target frequency. We use the coupling model to extrapolate these functions to 2π sr and fit them to a combination of the holography and celestial source data described above. The viability of this model rests on the fact that $\sim99\%$ of the variance in the solar data can be described by a linear combination of three modes that are separable in (x, y) coordinates. However, our ability to accurately extrapolate these modes to the rest of the sky relies on a model that has known limitations. Further, our ability to assess the quality of the model is limited by the available holography and source data, which have limited sky coverage. Figure 21 shows a current estimate of the 2π sr beam response at 678 MHz.

In a second approach, we exploit the fact that the sidelobe signal in the solar data, as a function of orthographic x—once rescaled by frequency, i.e., $x' \equiv x \cdot (\nu/600 \text{ MHz})$—is well described by a linear combination of three functions of x' over the entire range of (y, ν) measured by the solar data. We fit these three modes to the near-meridian celestial source data

depicted in Figure 19, at each y and ν separately. The result is a 2π sr model that is visually similar to Figure 21. A detailed description and comparison are deferred to a forthcoming paper.

3.3.2. Coupling Model

This approach is a phenomenological one inspired by physical optics: we form a parameterized model of the base beam and of multipath effects and fit those parameters to the data described in Section 3.2. In its simplest form—called the coupling model—the multipath is attributed entirely to cross talk between pairs of feeds along the focal line. In the time domain, we may express this as a superposition of base beam profiles, delayed by specific amounts in time,

$$\tilde{A}_i(\boldsymbol{n}, t) = A_i(\boldsymbol{n}, t) + \sum_j \alpha_{ij} A_j(\boldsymbol{n}, t + \tau_{ij}), \qquad (3)$$

where $A_i(\boldsymbol{n}, t)$ is the electric field produced by feed i (thought of here as a transmitter) in the absence of neighboring feeds; $\boldsymbol{n}$ is a directional unit vector; $A_j(\boldsymbol{n}, t + \tau_{ij})$ is the electric field produced by neighboring feed j, delayed by a time τ_{ij}; and α_{ij} is a coupling coefficient that describes the strength of the coupling. In the frequency domain, the time delay transforms to a phase factor. In the model's simplest form, we assume that all feeds produce the same pattern, $A(\boldsymbol{n})$, and that there are two coupling paths between any pair of feeds: a "direct" path via signals propagating parallel to the ground plane with delay $\tau_{ij} = |\Delta y_{ij}|/c$, where Δy_{ij} is the N–S separation between feeds i and j, and a "1-bounce" path via signals reflecting once off the cylinder as they travel from feed i to feed j, with a delay set by analogous geometric arguments. The model is parameterized in terms of coupling coefficients for different coupling paths and their associated fall in strength as a function of feed separation. An example of this model, fit to the source transit data and evaluated on meridian, is presented in Figure 18. Typical coupling strengths between adjacent feeds are found to be $\sim15\%$ and $\sim3\%$ for the direct-path and 1-bounce-path cases, respectively. The coupling strength as a function of antenna separation falls differently for the two cases and is estimated to be $\sim 1/|\Delta y_{ij}|^2$ and $\sim 1/|\Delta y_{ij}|^{1/2}$ for the direct and 1-bounce paths, respectively. Multibounce paths couple at less than 1%. Further details about the parameterization and
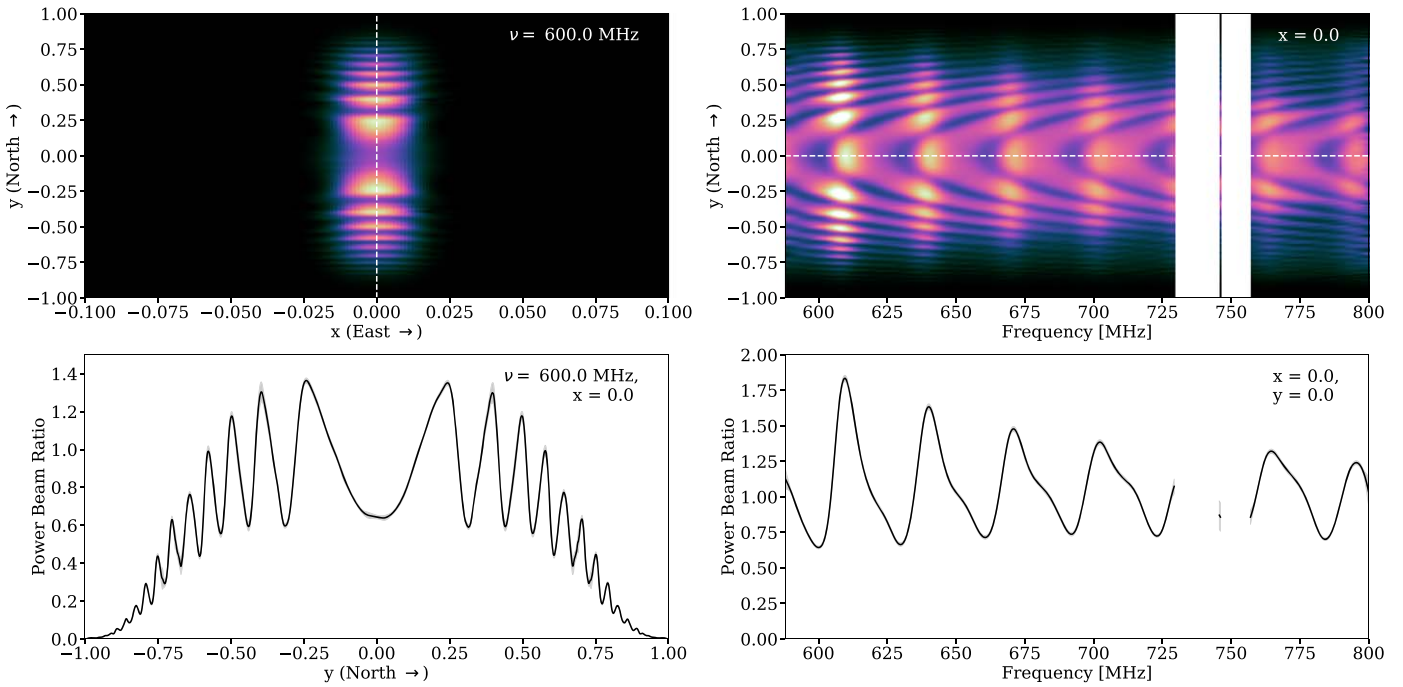
Figure 19. Model for the near-meridian primary beam of the *Y*-polarized array, obtained by fitting the visibilities measured with long E–W baselines to a model for the radio emission from extragalactic point sources. Top left: beam model at 600 MHz as a function of orthographic angular coordinates *x* and *y*. Top right: beam model on meridian as a function of frequency and *y*. The bottom panels show 1D slices through the beam at the location indicated by the white dashed line in the panels above. The gray contours in the bottom panel provide an estimate of the uncertainty (68% confidence interval). The color scale in the top panel spans the range shown on the *y*-axis in the bottom panel. The beam model is in power units in all cases and has been normalized to 1.0 on meridian at the decl. of Cyg A (40°.734) at each frequency in order to match how the data are normalized by the calibration procedure. The *X* polarization response exhibits the same general features but is slightly wider in both the *x*- and *y*-directions and also has a lower response at zenith because the dipole is oriented perpendicular to the axis of the cylinder.
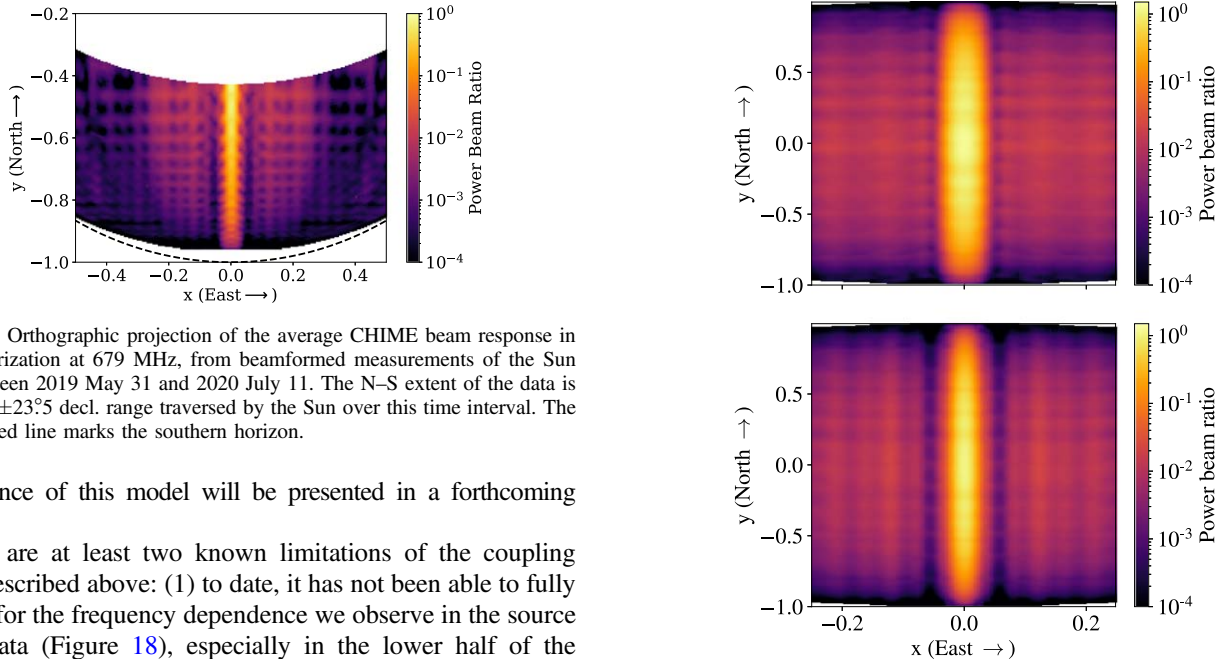

Figure 20. Orthographic projection of the average CHIME beam response in the *Y* polarization at 679 MHz, from beamformed measurements of the Sun taken between 2019 May 31 and 2020 July 11. The N–S extent of the data is set by the ±23°.5 decl. range traversed by the Sun over this time interval. The black dashed line marks the southern horizon.

performance of this model will be presented in a forthcoming paper.

There are at least two known limitations of the coupling model described above: (1) to date, it has not been able to fully account for the frequency dependence we observe in the source transit data (Figure 18), especially in the lower half of the frequency band, and (2) it predicts a N–S response modulation that is independent of E–W direction on the sky, which is inconsistent with the solar data (Figure 20). There are at least two possible explanations for this: (1) the coupled feeds, *j*, (re) radiate a different base beam, $A_j(\boldsymbol{n})$, than does the source feed, *i*, and/or (2) in addition to coupled feeds reradiating the source signal, there is also a reflected signal that bounces directly off the ground plane and back to the cylinder before reaching the

Figure 21. Orthographic projection of the modeled CHIME beam response in *X* (top panel) and *Y* polarization (bottom panel) feed, generated at 678 MHz using the data-driven model described in Section 3.3. It is modeled using basis functions derived from solar data measurements, which are fit to independent measurements of the beam.

sky. This reflected signal could have a slightly different delay parameter than the 1-bounce coupled signal, and it is expected to have a different E–W profile than the coupled signal.

We are in the process of developing a richer model that incorporates these effects, parameterized by the electric field distribution in the cylinder aperture, as informed by the commercial software packages CST and GRASP.[29] From preliminary studies, it appears that the aperture field can be parameterized relatively compactly and that the resulting model is qualitatively successful at fitting the features seen in the solar data. Specifically, with ∼20 parameters to describe the aperture field, single-frequency fits to the solar data in Figure 20 produce a model with residual errors of $\sim10^{-3}$ in the solar data, but which can be evaluated over the full sky. Future work will involve using the holography and celestial source data in the model fits, so a detailed discussion of this effort will be deferred to a forthcoming paper. Note that the coupling model described above is a special case of this more general multipath model.

3.4. Beam Model Usage

In this section we summarize how various beam models developed for CHIME have been used in scientific analyses to date.

1. The celestial source model depicted in Figure 19 was developed for the stacking analysis presented in CHIME Collaboration et al. (2022a).
2. The detection of an exceptionally bright radio burst from a Galactic magnetar (CHIME/FRB Collaboration et al. 2020) occurred when the object was 22° off of CHIME's meridian. Characterization of this rare event requires knowledge of the instrumental beam well off axis. We use the solar data (CHIME Collaboration et al. 2022b) and Taurus A (Tau A) holography data to measure CHIME's beam response there, enabling a measurement of the burst flux/fluence.
3. The first CHIME FRB catalog (CHIME/Pulsar Collaboration et al. 2021) gives an estimate of flux/fluence of each FRB. A beam model that gives the beam solid angle as a function of forward gain and frequency is required to model the statistical distribution of their brightness. An early version of the 2π sr model depicted in Figure 21 is used for this work. This enables a measurement of the FRB sky rate, one of the main results from the paper.
4. CHIME/FRB is able to perform polarimetry on some events (Mckinven et al. 2021). While polarized beam models are not yet used in these measurements, CHIME's beam data have informed which systematic effects need to be included in the polarization fits as nuisance parameters. The most important of these is the differential response of the X- and Y-polarized beams near their half-power points, seen clearly in all CHIME beam measurements.
5. The FRB team is building *outrigger* cylindrical telescopes to provide a steady stream of subarcsecond localizations of FRBs. Data from CHIME holography show a lack of significant beam phase variation within a few degrees of meridian (Figure 15). This result is crucial input to the design of the CHIME outriggers, meaning that the optical design of the outriggers could differ somewhat from CHIME's design and not require beam phase recalibration.

---

[29] https://www.ticra.com/software/grasp/

Table 1
Primary Sources of Data Loss

Axis	Source of Data Loss	Fraction Lost
Time	Downtime	17%
	Daytime	50%
	Rain/snowmelt	20%
	Phase miscalibration due to correlator restart	6%
	Total	**69%**
Frequency	Nonoperational GPU nodes	10%
	RFI	42%
	Total	**48%**
Feed	Nonoperational or malfunctioning feed	3%
	Feed at ends of cylinder	6.25%
	Total	**9%**

Note. Since each source of data loss is largely independent of all other sources, the total fraction of data lost is given by $f_{\text{total}} = 1 - \Pi_i(1 - f_i)$, where i runs over all the sources that are listed.

4. Performance

In this section, we evaluate the performance of the instrument using data acquired over the first two years of operation and a number of dedicated measurements. This performance evaluation includes an examination of the main sources of data loss, an assessment of the stability of the complex receiver gains, a characterization of the system temperature, an investigation into the effectiveness of the real-time RFI excision algorithm, and finally, a presentation of maps of the radio sky created from the CHIME stack data set.

4.1. Data Loss

CHIME has been operating continuously since its first-light ceremony on 2017 September 7. The first year of operations was dedicated to commissioning the instrument, developing the real-time pipeline, and developing the calibration and flagging strategies. Acquisition of data for the cosmological analysis began on 2018 October 7. Since then, the daily data acquisition rate has averaged ≈ 1 TB day^{-1}. Of this daily total, approximately 600 GB is the stack data set containing the primary science data. The remainder is calibration, beam holography, housekeeping, and other engineering data sets.

Table 1 summarizes the main sources of data loss between 2018 October 7 and 2020 October 7. During this 2 yr period, the instrument was down for a total of 127 days (17%). The majority of this time (102 days) was due to planned hardware maintenance and software upgrades, which occurred approximately five times per year. A further 25 days were unintended interruptions due to power failures, cooling failures, and other accidental outages.

The radio signal from the Sun dominates over the signal from the rest of the sky, even when the Sun is in the far sidelobes. The signal from the Sun can be modeled and subtracted to a large extent; however, feed-to-feed variations in the gain or beam and inaccuracies in the model for the extended emission yield residuals that are significant compared to the noise and signal from the rest of the sky. As a result, data acquired when the Sun is above the horizon are currently excluded from the cosmological analysis.

Precipitation at the telescope site causes deterioration of the detected signal as a result of water pooling around the focal-line electronics. This signal deterioration is broadband and characterized by a reduction in gain, an increase in noise, and, occasionally, gain oscillations with periods ranging from seconds to minutes. Accumulation of dry snow does not cause analog signal deterioration, but snowmelt, which is more difficult to detect using weather data alone, does produce the same signal deterioration as rain. Signals from the 2048 feeds are monitored for broadband, differential increases in their autocorrelations. This signature is used to identify and flag wet feeds before collating redundant baselines. After each rain or snowmelt, roughly 4%–12% (interquartile range (IQR)) of the inputs are flagged for 3–21 hr (IQR), effectively until they dry. It is not yet clear whether data acquired during these wet periods can be used in the science analysis. Excluding them results in a 20% reduction in observing time, preferentially occurring in months when nights are longest and therefore when we have the most useful data. Steps are being taken to improve focal-line waterproofing.

The synchronization procedure implemented by the FPGAs does not guarantee that the phase of a common signal measured by two inputs on different ADC chips will remain constant through an FPGA restart. This change in phase after FPGA resynchronization is observed in the noise source data, and the size of the phase change can be large compared to the requirements on instrument stability outlined in Section 4.2.2. As a result, we mask the interval between each FPGA restart and the following point-source calibration. This results in an approximately 6% reduction in observing time.

Table 1 also lists the average fraction of the 400–800 MHz band that is masked due to RFI (as detailed in Section 4.4) and lost due to nonoperational GPU nodes. Note that in 2020 June the correlator software was upgraded to allow for much greater flexibility in the mapping between frequency channels and GPU nodes. This gave us the capability to send frequency channels already contaminated by persistent RFI to the set of GPU nodes that are off-line at any given time, which recovers a large fraction of the 10% of the band that was previously lost due to nonoperational GPU nodes.

Finally, Table 1 provides estimates of the fraction of the 2048 correlator inputs that are masked prior to collating redundant baselines. The flagging broker masks approximately 3% of inputs because they fail one or more of the tests described in Section 2.5. In addition, in 2019 December we began applying a static mask that consists of the eight feeds at the edge of each cylinder because it was determined that these feeds exhibit a highly nonredundant beam pattern.

4.2. Stability

The stability of the instrument is assessed using the complex gains measured by the *calibration broker* (described in more detail in Section 2.5). The broker computes and stores gains using data from four bright source transits every day: Cassiopeia A, Cygnus A, Taurus A, and Virgo A, henceforth referred to as Cas A, Cyg A, Tau A, and Vir A, respectively. Figure 22 shows an example of N_{feed}^2 visibility data acquired during a Cyg A transit after applying the complex gains derived from the transit. On any given day, one source is chosen as the primary calibrator (typically the brightest source to transit at night), but all of the transits are analyzed off-line to assess stability. To help assess and maintain phase stability, a
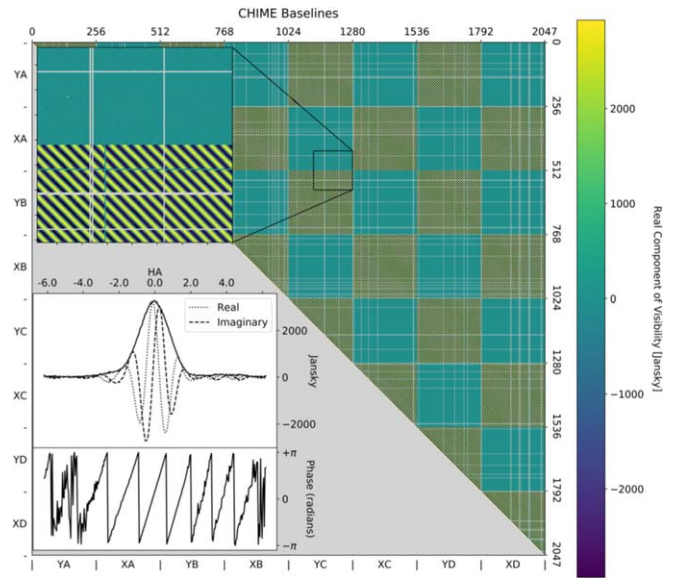

Figure 22. Real component of the calibrated visibilities at 758.203 MHz during the transit of Cyg A. Inset is the visibility for a single 22 m pure E–W baseline as a function of the hour angle of Cyg A, with the amplitude (solid), real (dotted), and imaginary (dashed) components shown in the top panel and the phase shown in the bottom panel.

broadband noise source system is also employed, as described below.

We use end-to-end simulations to determine our stability requirements. Simulation of a CHIME-sized telescope is challenging owing to computer resource limitations; therefore, we have performed simulations of a scaled-down instrument (with roughly one-quarter of CHIME's collecting area) to investigate these requirements, examining the anticipated accuracy of the 21 cm power spectrum measured after the application of the Karhunen–Loève foreground filter described in Shaw et al. (2015). This work found the requirement for fractional variations in the complex gain to be less than 1%, which translates into phase errors smaller than 0.007 rad and amplitude errors smaller than 0.7%. However, these requirements are derived from a simulation whose gain variations are constant across the band and uncorrelated from input to input. Furthermore, it is unclear how these requirements scale with the size of the telescope, and neither of these conditions holds in the observed gain variations presented below; thus, these requirements serve as a rough guide only. More realistic simulations designed to better reflect some of the observed complex gain variations have since been performed, and the resulting requirements are noted below where applicable.

4.2.1. Amplitude

Figure 23 shows the fractional gain amplitude variations (standard deviation) for all correlator inputs as derived from the calibration broker gains over a full year from 2018 June to 2019 July. These data include 259 gain solutions (94 from Cas A, 89 from Cyg A, and 76 from Tau A), which have been scrubbed of RFI-contaminated transits and anomalous gains mostly related to wetness of the instrument during rainy periods.

The blue curve indicates the intrinsic gain variations after outliers are removed but prior to applying any calibration corrections. It shows a pronounced slope with frequency that ranges from 1% at 400 MHz to 1.8% (standard deviation) at
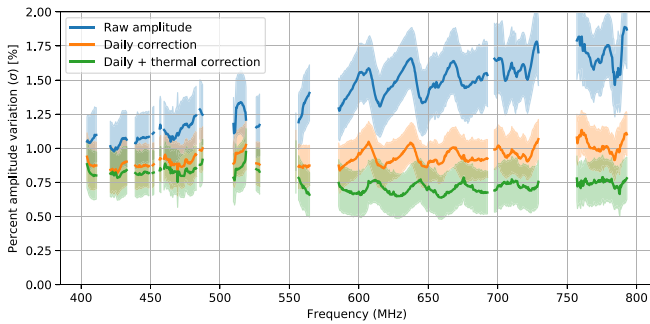
Figure 23. Gain amplitude stability. The blue band shows the standard deviation of the fractional gain amplitude variations as determined from 259 transits of Cas A, Cyg A, and Tau A from 2018 June to 2019 July. The rms stability is evaluated input by input; the central curve shows the mean stability across inputs, while the band indicates the 1σ spread across inputs. The orange band shows the corresponding information with the gains corrected once per day using a previous transit. The green band shows the result of applying an additional correction to the orange data based on a linear regression to the ambient temperature change since the previous transit. Gaps in the data correspond to known RFI-dominated bands.

800 MHz. A substantial portion of this variation is due to the thermal susceptibility of the instrument.

The orange curve shows the residual gain amplitude variations for the same data, but after applying a daily correction similar to that which is applied to the archived visibility data. This gives an indication of the gain variations present in the stored data prior to applying any subsequent corrections (see below). To produce this curve, we take the difference between each transit's gain and a solution from the previous 48 hr (if available) and compute the standard deviation of the difference. This procedure brings the fluctuations down to a nearly flat 0.9%–1% level.

The green curve shows the residuals after correcting the orange data using the measured thermal susceptibilities and the ambient temperature change since the previous transit. This brings the variations down to $\sim 0.7\%$ (standard deviation). The thermal correction flattens the residuals considerably, a consequence of the fact that the temperature susceptibility of the system gain rises with frequency from 0.06% K^{-1} at 400 MHz to 0.2% K^{-1} at 800 MHz. This measured stability achieves the preliminary requirement described above, but with no margin.

By construction, the data tracked by the orange curve remove gain variations slower than ~ 1 day due to *any* source, while the data tracked by the green curve remove variations correlated with ambient temperature on all timescales. We find that thermal regression applied to raw data (blue curve) and the daily corrected data (orange curve) produced similar residuals. This suggests that most of the variation on timescales longer than a day is thermal in origin and that variations on shorter timescales are not well correlated with ambient temperature.

The analysis discussed above is carried out input by input, assuming nothing about how correlated the gain variations are across inputs. An SVD analysis of the raw gain variations over input and time reveals a single dominant mode followed by a closely packed mode spectrum. The dominant mode accounts for about 60% of the data variance at the lower end of the band and grows to over 80% of the variance halfway to the high end of the band. The singular vector of the dominant mode is highly correlated with the ambient temperature, implying that an ambient-temperature-based correction largely accounts for the common-mode portion of the variance. Thus, the residual

variability after thermal regression (the green band in Figure 23) gives a good estimate of the non-common-mode variations in the system. These residual gain variations show some degree of correlation across inputs. Making use of this to further improve the correction is under study.

The frequency structure of the gain stability depends on the decl. of the source used to derive the gains. This appears to be due to a time and/or thermal dependence of the primary beam response of the instrument. This would be expected if feed-to-feed cross talk depended on time and/or temperature, which, in turn, could result from thermal expansion and contraction of the CHIME structure. (See Sections 3 and 4.2.2 for further discussion of these effects.) Efforts to model this dependence are ongoing. The results shown in Figure 23 are computed from the gains derived from the three brightest sources, so the frequency structure shown there is a weighted average of the response to these three sources.

4.2.2. Phase

Figure 24 summarizes the phase stability of CHIME as inferred from the response of each correlator input to the two brightest calibration sources (Cyg A and Cas A). The measured phase variations are highly correlated across frequencies and, to first order, can be described by delay-type variations of the form

$$\delta\phi_{ij}(t, \nu) = 2\pi\nu \; \delta\tau_{ij}(t), \tag{4}$$

where $\delta\phi_{ij}$ is the change in the relative phase between inputs i and j at time t for RF ν due to a change in the relative delay $\delta\tau_{ij}$. If we perfectly corrected all delay-type variations, the phase stability of the instrument would improve from the red curve to the black curve in Figure 24. The dominant sources of delay variations are relative drifts between copies of the 10 MHz clock that defines the sampling rate of the ADCs, expansion and contraction of the telescope with ambient temperature, and changes in the electrical length of the 50 m coaxial cables with temperature. We describe these three sources of delay variation in turn and outline the methods used to partially correct for them, to stabilize the phase. After applying the corrections, the resulting stability is given by the blue curves in Figure 24.

The dominant source of phase instability on timescales $\lesssim 20$ minutes is relative drifts between the eight copies of the 10 MHz clock that are separately distributed to each of the eight FPGA crates. Each clock defines the sampling rate of the 256 ADCs within a crate. The drifts are measured and corrected using a single broadband noise source that is distributed to one input on each of the eight FPGA crates through a passive system of coaxial cables and power splitters. The correlator computes the covariance of the noise source inputs over a 10 s integration for each of the 1024 native resolution frequency channels. The largest eigenvector of this covariance matrix is used to estimate the response of the eight inputs to the signal from the noise source. The phase of the response is referenced to the time of the last point-source calibration to remove static ripples caused by reflections in the distribution network. Then, for each 10 s integration, the phase as a function of frequency is fit to Equation (4) to extract the delay as a function of time, $\delta\tau_{ij}(t)$, for each FPGA crate i relative to a reference j. This is used as a proxy for the drift in the clock copy provided to that crate relative to the reference ADC input on the reference crate.
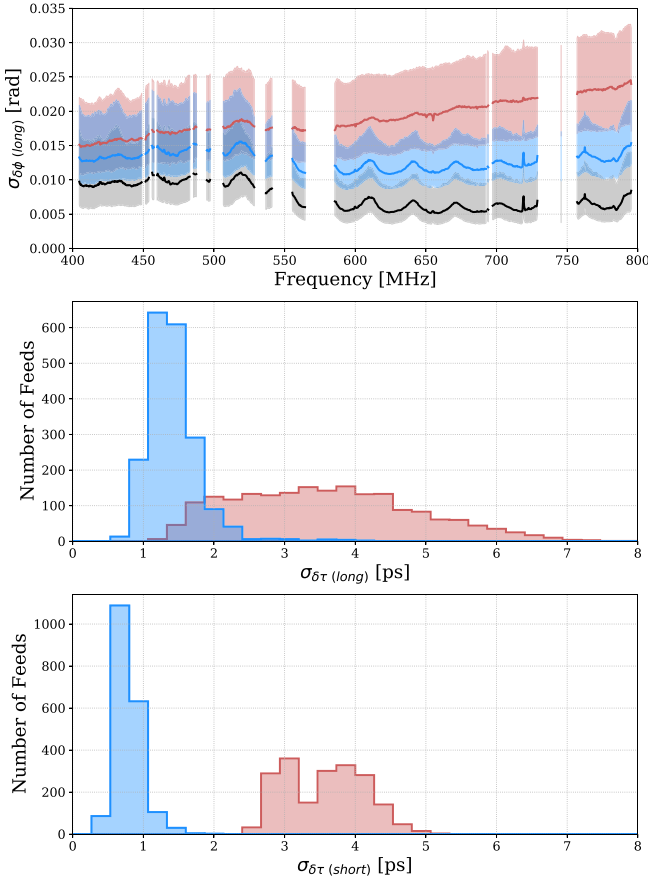
Figure 24. Gain phase stability. Top: the standard deviation over 74 days of CHIME's phase response to Cas A at transit after applying a daily calibration derived from Cyg A. Lines denote the median and bands denote the central 68% over the 2048 CHIME feeds. Red indicates the raw phase variations. Blue indicates the residual phase variations after correcting for delay variations caused by drift between copies of the 10 MHz clock, thermal expansion of the focal line, and thermal susceptibility of analog receiver chain as tracked by the ambient temperature (see the text for a discussion of each of these effects). Black indicates the residual phase variations after removing all delay-type variations by fitting and subtracting a model for the phase variations that scales linearly with frequency from each transit. The phases are referenced to the average phase over feeds of a given polarization. Middle: the standard deviation of the delay variations, shown as a histogram over feeds. The red histogram indicates the raw delay variations, while the blue histogram shows the residual delay variations after applying the three corrections listed above. Bottom: same as the middle panel, but showing delay variations on short timescales ($\lesssim$20 minutes), obtained by examining a window around the transit of Cyg A or Cas A when these sources are in the primary beam. The black curve in the top panel, which corresponds to the perfect removal of all delay-type variations, is by definition equal to zero for all feeds in the bottom two panels.

Examining the relative delay variations between the four crates within a single receiver hut, we find that the variations exhibit a sawtooth pattern with an 8 minute (east receiver hut) or 6 minute (west receiver hut) periodicity that mimics the temperature variations in that hut. This periodicity tracks the cooling cycle of the chiller system in each hut. This produces a relative delay variation of 1–2 ps (standard deviation) between crates in the same hut. Since the temperatures of the two huts cycle at different periods, the relative delay variations between crates in different huts are significantly larger: approximately 6–8 ps (standard deviation).

A suite of simulations is used to estimate the bias in the 21 cm power spectrum due to realistic clock drifts. We find that

the clock drifts must have a standard deviation of $\lesssim$1 ps to ensure negligible bias in the power spectrum. The bottom panel of Figure 24 shows the improvement in the short-timescale delay noise that is achieved by regressing the delay variations obtained from point-source observations against the delay measured by the broadband noise source. The residual delay variations have standard deviation < 1.5 ps and are thus close to meeting our requirements.

Thermal expansion and contraction of the focal line introduce a temperature dependence to the N–S baseline distance that manifests as delay variations on timescales $\gtrsim$20 minutes. We can model this with the following expression:

$$\delta\tau_{ij}^{\text{focal}-\text{line}}(t, \delta, \text{ha})$$
$$= [\cos\ell\sin\delta - \sin\ell\cos\delta\cosh a] \times \frac{\Delta y_{ij}}{c} \epsilon \, \delta T(t), \qquad (5)$$

where ℓ is the latitude of the telescope, δ is the decl. of the source, ha is the hour angle of the source, Δy_{ij} is the nominal N–S baseline separation, c is the speed of light, ϵ is the linear thermal expansion coefficient of the focal line, and $\delta T(t)$ is the difference between the ambient temperature and the nominal temperature. Fitting the delay variations obtained from the point-source transits to Equation (5) yields a thermal expansion coefficient of $\epsilon = 21 \times 10^{-6}$ K^{-1} for the focal line. This is approximately equal to the coefficient for aluminum and roughly twice that of steel. The focal-line structure itself is made of steel, while the cassettes that hold groups of four antennas to the focal line are made of aluminum and are bolted to each of their neighbors. The interplay of these components as the temperature changes is still under study, but our model fits the sky data well, so we adopt the best-fit ϵ as a description of the instrument. The resulting delay error is the same for all redundant baselines, so the correction for this effect can be done off-line, after collating these baselines. However, the correction depends on sky position, so it needs to be implemented at the mapmaking stage. This work is currently under development.

After controlling for drift in the clocks and thermal expansion of the focal line, the residual delay variations exhibit a correlation with ambient temperature. Based on thermal chamber measurements of the components of the signal path, changes in the electrical length of the 50 m coaxial cables are expected to be the dominant source of thermally induced delay variations. These changes in electrical length are the result of changes in the physical length of the cable from expansion of the center conductor and changes in the dielectric constant due to a reduction in the dielectric density from expansion of the outer conductor. To first order, the observed delay variations can be modeled as

$$\delta\tau_{ij}^{\text{coax}}(t) = (\alpha_i - \alpha_j) \bar{T}(t) + \bar{\alpha}[T_i(t) - T_j(t)], \qquad (6)$$

where α is the thermal susceptibility of the coaxial cable, T is the temperature of the coaxial cable, subscripts i and j refer to specific inputs, and a bar indicates the average over all inputs. The first term is due to differences in the thermal susceptibility between cables, while the second term is due to differences in the effective temperature of the cables.

In order to gauge the relative importance of the two terms in Equation (6), we have installed three "cable monitors" that consist of two 50 m coaxial cables that are routed to the focal

line and then back along the same path, with one end connected to the noise source described above and the other end connected to the correlator. There is one cable monitor routed to each of cylinders A, B, and C. The cable monitor data are processed in the same manner as the noise source data described above. The resulting delays are divided by 2 to account for the fact that the length of coaxial cable in the cable monitors is twice that of the CHIME on-sky inputs. The measured delays are regressed against the ambient temperature in order to measure the thermal susceptibility of the three cable monitors. The average thermal susceptibility over cable monitors is $\overline{\alpha} = 2.93$ ps K^{-1}. The standard deviation over cable monitors is $\sigma_\alpha = 0.04$ ps K^{-1}, which will result in relative delay variations with a standard deviation of $\sim$0.25 ps given the temperature variations on a typical night. Residual delay variations that are not explained by differences in thermal susceptibility are attributed to differences in the effective temperature of the cables. These residuals have a standard deviation of $\sim$1.0 ps, which implies effective temperature variations with a standard deviation of $\sim$0.3 K given the value of $\overline{\alpha}$ quoted above.

We characterize the difference in thermal susceptibility between CHIME correlator inputs by regressing the change in delay between point-source transits against the change in ambient temperature. The standard deviation of the thermal susceptibility over inputs is 0.3 ps K^{-1}, which is much larger than we would expect from the scatter in the value of α measured for the three cable monitors. If we randomly draw thermal susceptibilities for three inputs from the sample of 2048, the probability that they are all within 3% like the cable monitors is <2%. This indicates that the analog receiver chain likely has some other source of susceptibility to the ambient temperature beyond the coaxial cables that is highly dependent on input. Nevertheless, this thermal susceptibility is well characterized using the point-source observations; we estimate that our uncertainty on the thermal susceptibility is ~ 0.05 ps K^{-1} using bootstrap resampling methods.

Figure 24 shows in blue the residual delay variations after correcting for clock drift, expansion and contraction of the focal line, and thermal susceptibility of the analog receiver chain. We find a standard deviation of < 1.5 ps on <20 minute timescales and $1-2$ ps on 3 hr timescales. The cable monitor data suggest that differences in the temperature of the coaxial cables are a significant contributor (~ 1 ps) to the residual delay variation on long timescales. We are actively investigating new techniques to measure and correct for the differences in coaxial cable temperature.

We characterize the phase stability of the instrument on longer timescales by examining changes in the phase between night-time transits of other pairs of bright point sources observed between 2019 February and 2020 March. The transit times of the four brightest point sources are spaced apart such that their various differences probe timescales ranging from 0 to 24 hr, with a roughly 3 hr sampling. The worst performance occurs on 18 hr timescales, where the postcorrection delay variations have an rms of 2.0 ± 0.6 ps (mean $\pm$ standard deviation over feeds). This is a small degradation in the 1.6 ± 0.6 ps rms delay stability observed on 3 hr timescales and shown in Figure 24. If we expand the analysis to also include daytime transits, then we find a significant degradation in the delay stability, with the worst performance occurring on 10 hr timescales, where the rms is 3.4 ± 1.1 ps. This is a secondary

reason to exclude the daytime data from the cosmology analysis, with the primary reason being contamination from solar radio emission.

4.3. Noise

Measuring the system temperature of the CHIME receivers using observations of the radio sky alone is challenging because it requires knowledge of the effective area of the antenna beam pattern. Instead, we perform an in situ measurement of the system temperature referred to the LNA input of four CHIME receivers (two polarizations on each of two antennas) by temporarily disconnecting the LNA from the antenna under test and connecting it to well-matched cold, ambient temperature and hot loads at 80, $\sim$300, and 373 K. We observe each regulated load for approximately 10 minutes, reconnect the LNA to the antenna, and resume normal observations.

The autocorrelations recorded by the CHIME correlator during the measurement are corrected for bias due to quantization to 4 bit real + 4 bit imaginary, which is insignificant for sky measurements but is a significant correction for the hot and ambient temperature measurements. The autocorrelations are converted to units of Jy using the gains obtained from the visibility matrix at the transit of Cyg A occurring approximately 6 hr before the measurements and regressed against the load temperature. The slope of the regression is used to estimate the Jy K^{-1} factor that converts between flux density on the sky and temperature at the input to the LNA. The intercept divided by the slope is used to estimate the *receiver temperature*, by which we mean the noise temperature of the LNA, FLA, cables, and ADC, referred to the LNA input. The Jy K^{-1} calibration factor is applied to the autocorrelations collected the night following the measurement to estimate the system temperature. The resulting *system temperature* measurements, referred to the LNA input for the two polarizations of one of the antennas, are shown in Figure 25. The results for both channels of the other antenna are consistent with these at the 5% level.

The receiver temperature increases from approximately 20 K at 400 MHz to 25 K at 800 MHz. This is in good agreement with measurements of the LNA temperature made in the laboratory and described in Section 2.3, indicating that the LNA dominates the receiver noise, as expected from the design. The system temperature when a dim part of the radio sky is transiting overhead is approximately 50 K but shows significant spectral structure that can be broadly separated into 150 MHz and 30 MHz ripples. The approximately 30 K difference between the system temperature and the receiver temperature includes contributions from the radio sky, loss in the antenna balun, ground spillover, transmission through the mesh, noise coupled from neighboring feeds, and antenna impedance mismatch in order of most significant to least significant contribution.

The radiometer equation can be used to estimate the noise given the system temperature presented above and the number of baselines, integration time, and bandwidth. In what follows, the variance of the data on different timescales is estimated directly and compared to our expectation based on the radiometer equation. On short timescales, the variance of each visibility is estimated by differencing even and odd time samples at 31 ms cadence (see Section 2.5). The radio sky does not change appreciably on these timescales and thus drops out of the difference. This "fast-cadence noise estimate" shows
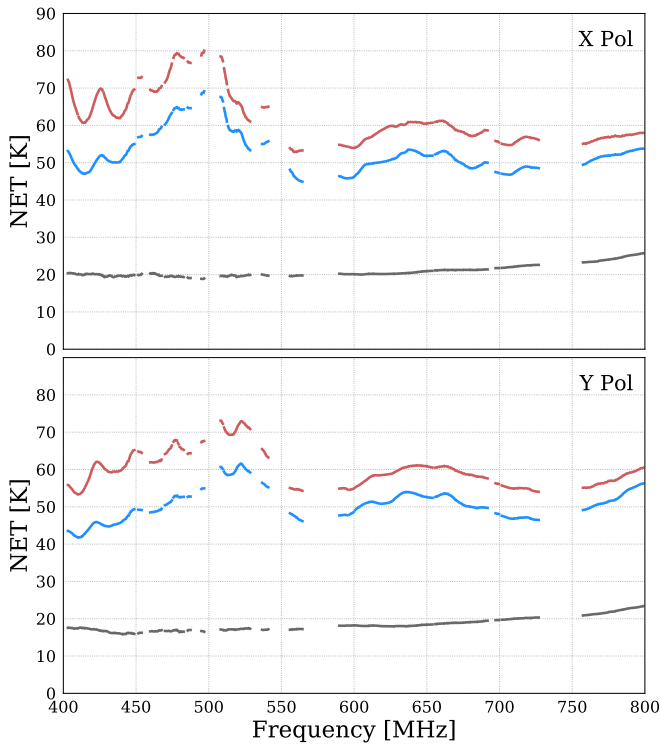
Figure 25. Noise equivalent temperature (NET) as a function of frequency for the two polarizations of a single CHIME antenna. The antenna is located 11 m south of the center of cylinder B. The receiver temperature (see text) is shown in black. The system temperature when a dim part of the radio sky is transiting through the primary beam is shown in blue (corresponding to the median value between LST of 14:00 and 14:20). The system temperature when a bright part of the radio sky is transiting through the primary beam is shown in red (corresponding to the median value between LST of 20:20 and 20:40). All quantities are obtained from autocorrelation data collected on 2019 May 30 and calibrated to units of kelvin using hot, cold, and ambient temperature loads as described in the text.

good agreement with our expectation based on the radiometer equation after excluding events that are localized in time and frequency in a manner characteristic of RFI. On longer timescales, the variance can be estimated by differencing visibilities acquired at the same local sidereal time (LST) on different sidereal days. In general, these day-to-day variations are consistent with the fast-cadence noise estimate and integrate down with the number of redundant baselines that are stacked. There are a few exceptions. The residual complex gain instabilities described in Section 4.2 dominate the day-to-day variations when the four brightest point sources are in the primary beam. In addition, visibilities measured by the shortest intracylinder baselines, specifically those with a N–S distance less than 10 m, are dominated by variations in sky brightness due to residual complex gain instabilities and also variations in the noise coupled between the feeds that form the baseline.

Similar results are obtained with beamformed data, differencing "ring maps" of the sky (see Section 4.5) produced on different sidereal days. For maps constructed with intercylinder baselines only (thus excluding the shortest intracylinder baselines mentioned above), the day-to-day variation over most of the sky is consistent with the fast-cadence noise estimate after accounting for the number of baselines that are used to produce the maps. The exception are pixels brighter than a few Jy beam^{-1}, for which the noise is dominated by residual complex gain instabilities. The noise can be further reduced by stacking maps produced on multiple days. In an

analysis of 38 daily, intercylinder ring maps spanning an interval of 73 days, the noise was observed to integrate down with the number of stacked days.

4.4. RFI

The real-time RFI excision algorithm described in Section 2.4.2 was deployed in 2019 October. To evaluate its performance, the Gaussianity of the autocorrelations is compared before and after applying the RFI excision. The Gaussianity test value (GT) for signal i is defined as

$$\mathrm{GT}_i = \sqrt{\frac{(1-f)\Delta t\,\Delta\nu}{2\,(N-2)}\sum_{j=1}^{N-1}\left(\frac{V_{ii}(t_{j+1}) - V_{ii}(t_j)}{V_{ii}(t_{j+1}) + V_{ii}(t_j)}\right)^2} - 1, \quad (7)$$

where $\Delta\nu = 390$ kHz is the channel bandwidth; V_{ii} is the autocorrelation evaluated at N times t_j; $\Delta t (\sim 10$ s) is the integration time, with $(1-f)\Delta t$ remaining on average after high-speed excision; and f is the real-time excision fraction. For a perfect Gaussian distribution the test will return ~ 0, and a large deviation from 0 indicates non-Gaussianity of the data. The results of the test for a single input are shown in Figure 26. Gaussianity of the data improves at all frequencies after applying the RFI excision, particularly in the 600–700 MHz band, where excising less than 1% of the samples significantly improves the quality of the data. The algorithm excises 15% of the data on average.

The off-line RFI excision algorithm described in Section 2.7 masks frequencies and times where the measured subintegration variance averaged over all cross-polar baselines deviates significantly from our expectation for radiometer noise. Over 188 nights in 2019 the average fraction of the band that was masked was 42%, with little night-to-night variation. During this interval, the real-time RFI excision was turned off. Figure 27 shows as a solid line the cumulative distribution of frequency bins as a function of fraction of time masked over this interval. About 29% of frequency bins are always masked, corresponding to the persistent sources of RFI discussed in Section 2.1. Figure 27 also shows as a dashed line this same quantity for 27 nights in mid-2020 when the real-time RFI excision was turned on. The fraction of the band that is always masked increased to 35% because of a degradation in the RFI environment at DRAO, primarily due to (i) the appearance in early 2020 of the downlink for Rogers 600 MHz band, which introduced persistent RFI in a part of the spectrum that was previously clean (617–627 MHz), and (ii) the transition from partial to complete occupation of the LTE band at 782–788 MHz. For the cleanest half of the CHIME band the fraction of time that was masked decreased from almost 15% to less than 5% with use of real-time RFI excision.

Since the real-time excision operates on the 0.66 ms and 31 ms frames, it is able to mask transient RFI events while discarding a much smaller fraction of the data than the off-line algorithm that operates on the 10 s data frames does. At present, the average fraction of the band that is masked is roughly the same with either method, 42%, but the fraction of the passband that is more than 95% free of RFI is much higher with rapid excision.
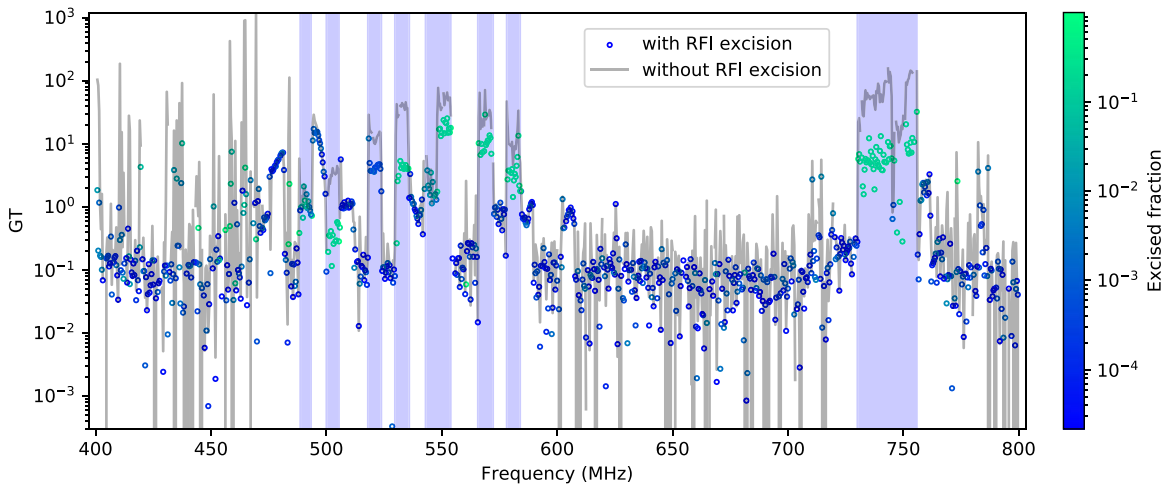
Figure 26. Result of the GT, Equation (7) for a single input on 2019 October 11 from 00:00 to 01:30 PDT before and after kurtosis-based RFI excision. The color of circles shows the average excised fraction over these 1.5 hr for each frequency channel. LTE and TV station bands are shown in purple. The Gaussianity of the data has improved in many frequency channels by excising less than 1% of the samples, i.e., their GT value is getting closer to zero after RFI excision. Notice that almost all the data are automatically excised within the TV and LTE bands. While this heavy excision improves the GT values for what remains, frequency channels in those bands still fail and are excised.
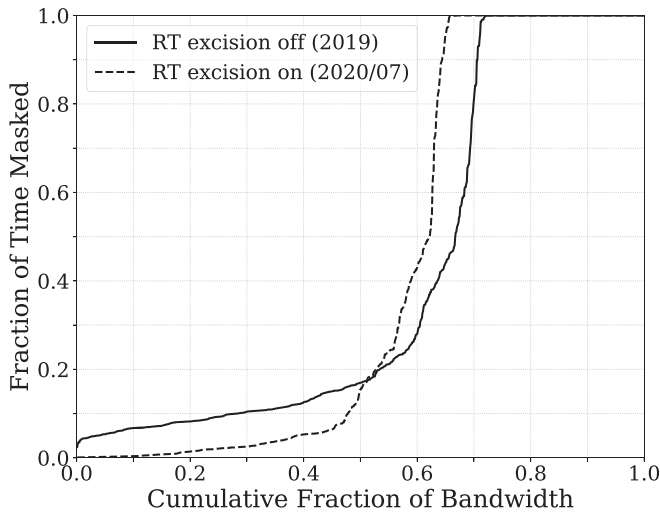
Figure 27. The cumulative fraction of the 400–800 MHz band that is masked less than a particular fraction of time by the off-line (second-stage) RFI excision algorithm. The solid line indicates that all frequencies were masked at least 5% of the time over 188 nights between 2018 December 22 and 2019 September 30 when the real-time, kurtosis-based RFI excision was turned off. The dashed line indicates that nearly 40% of the CHIME band was masked less than 5% of time over 27 nights between 2020 June 26 and 2020 July 21 when the real-time RFI excision was turned on. The difference between the vertical asymptotes (30% always masked in 2019, 40% in 2020) is due to new radio transmitters nearby.

4.5. Sky Maps

We generate maps of the sky for data-quality assessment, for instrument characterization, and as the starting point for Galactic science with CHIME; all have short-term and long-term goals. Our basic product is the "ring map." We generate 1D images along the meridian by Fourier-transforming one image for every 10 s time sample, and we assemble these into an all-sky image. These maps employ visibilities directly. The process is described in detail in CHIME Collaboration et al. (2022a). The cosmological stacking analysis is based on ring maps with intracylinder baselines excluded in order to filter out diffuse Galactic emission and reduce the impact of noise cross talk.

4.5.1. Single-day Maps

We show a ring map produced from *YY* visibilities using a single sidereal day of data in Figure 28. The map is shown in the time-sin(za) coordinate system, where za is the zenith angle, and with corresponding R.A. and decl. labels on the top and right. We show 24 sidereal hours of data, with time increasing from left to right: R.A. also increases from left to right, opposite to the astronomical convention for sky images.

The ring map of Figure 28 highlights a number of features of the Galaxy, our observing strategy, and instrumental features and artifacts. The large features of the radio sky dominate the map. The Galactic plane stretches across the sky, and the North Polar Spur rises from it. The Galactic plane and the North Polar Spur appear once at their true declinations and again at the top of the image, the result of aliasing. The spacing of feeds along the focal line is 30 cm, more than half a wavelength for frequencies higher than 500 MHz; the Fourier transform therefore produces an aliased response across much of the band. The bright sources—the Sun, Cas A, Cyg A, Tau A, and Vir A—also have aliased versions; all except the Sun are unresolved by the CHIME beam and can be treated as point sources. Cas A is circumpolar, and a lower transit image and its alias are also seen. All bright sources are seen both at transit and in the sidelobes for several hours on either side of transit. Away from transit these sources appear to be at higher decl., producing the characteristic "smile" features on the ring map. The point sources show a bright peak at the source R.A. and fainter peaks before and after transit, produced by the grating lobes; all the smile features have a dotted appearance. The shape of the smile is geometric and therefore frequency independent, but the positions of the grating lobe peaks along the smile are frequency dependent. Each time slice is an interferometric image lacking zero-spacing information and therefore must average to zero. Consequently, the transit of each of the bright sources produces a vertical dark stripe of negative values across the map. Similar negative regions are
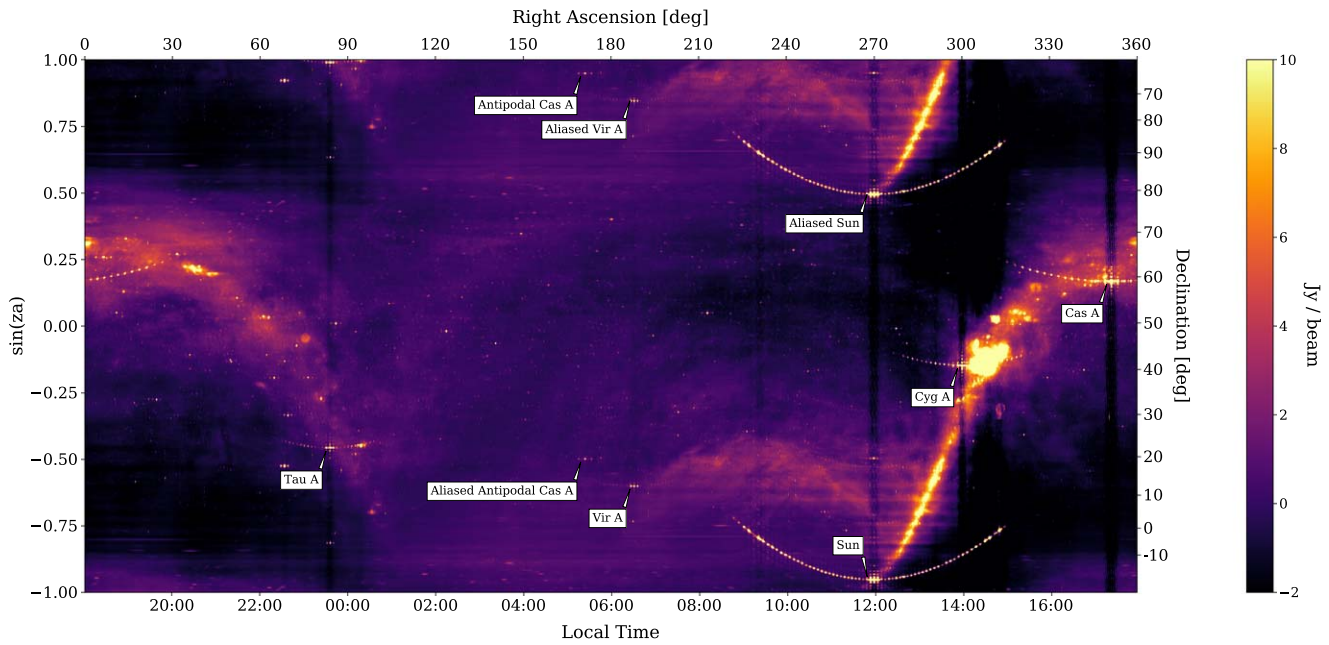
Figure 28. Map of the northern radio sky at 679 MHz constructed from data collected by CHIME over a single sidereal day (2018 December 21/22), obtained by beamforming all *YY* visibilities (excluding autocorrelations) for each 10 s integration to a grid of 2048 declinations along the meridian, spanning from horizon to horizon and equally spaced in sin(za). The map has been minimally processed, and no attempt has been made to deconvolve the transfer function of the instrument. The Sun and the four brightest point sources and their aliases are identified. The map is shown with time (Pacific Standard Time, UTC−8) increasing from left to right to illustrate the CHIME observing strategy; therefore, R.A. increases from left to right, opposite the astronomical convention. The image is plotted in the native time–sin(za) coordinates; decl. and R.A. (or, equivalently, because all observations are at hour angle zero, LST) are labeled on the right and top.
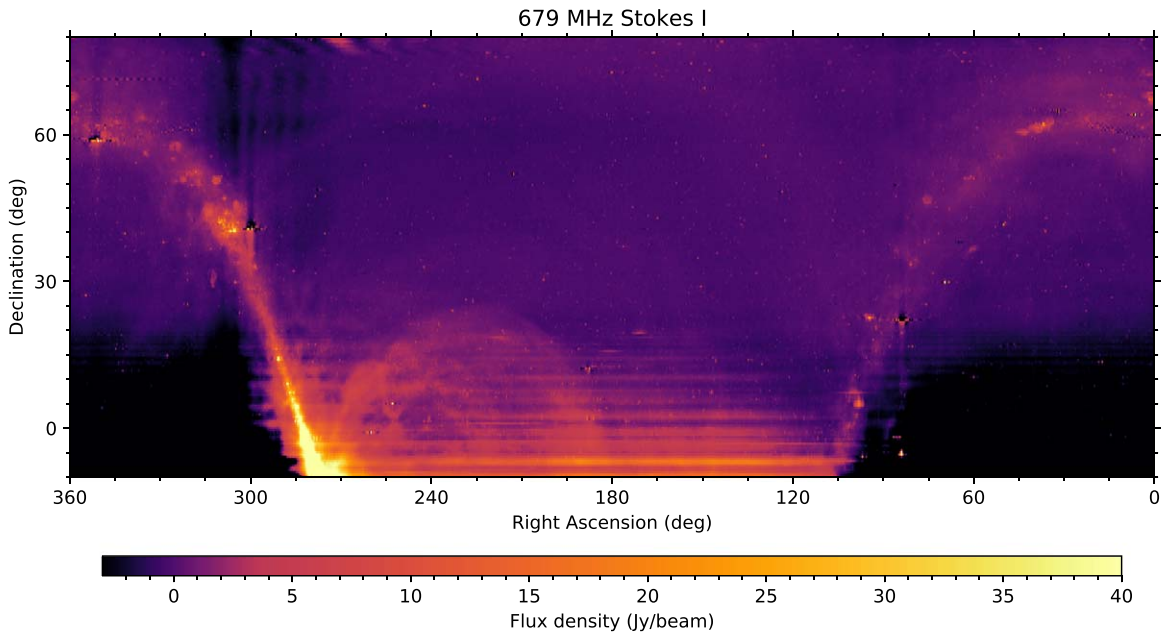
Figure 29. Map of the northern radio sky constructed from data collected by CHIME over 52 nights and stacked. This is a deconvolved Stokes $I = (XX + YY)/2$ ring map obtained from all *XX* and *YY* visibilities using the stacked data, plotted in celestial coordinates in a plate carrée projection. The image shows most of the northern sky, oriented in the conventional way for astronomical images with R.A. increasing to the left (unlike Figure 28).

evident at the right ascensions of particularly bright Galactic emission.

Cross talk between adjacent feeds (see Figure 18) produces ripples in zenith angle that are evident as horizontal stripes in an uncorrected ring map. We have reduced the striping by subtracting the median at each decl. from the image. This process is quite effective at $|za| \lesssim 25°$, but some striping is still evident at larger zenith angles.

4.5.2. Stacked Maps

In Figure 29 we show a stacked map, formed from data from nearly 2 months of observations. This is also a ring map, but the data are combined as visibilities before the formation of the map. The stack uses nighttime data from 52 periods of 24 sidereal hours (we call them "days" for brevity), divided into contiguous sets of days from different periods in the year chosen to provide complete coverage of the sidereal day (see
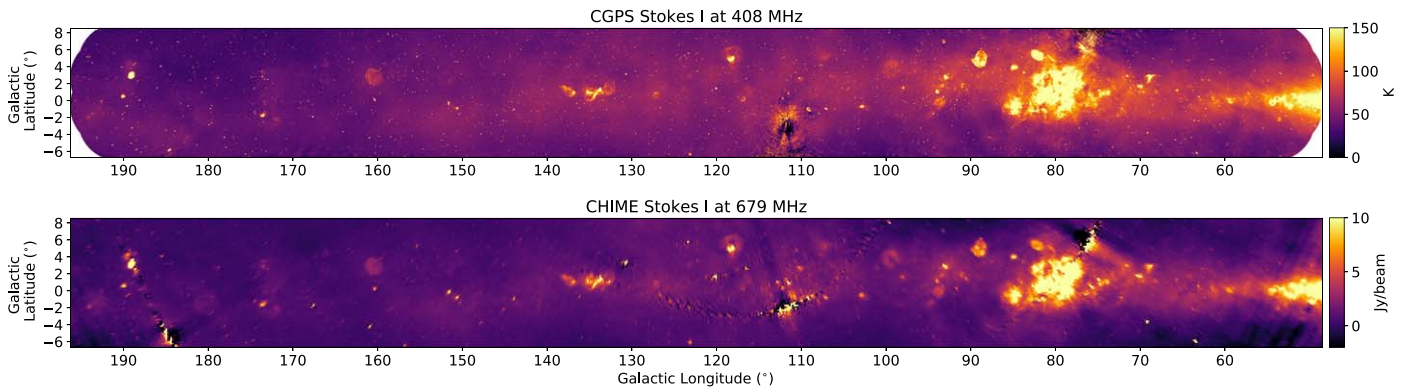
Figure 30. Stokes *I* maps of the Galactic plane from the CGPS at 408 MHz (top panel; Tung et al. 2017) and CHIME at 679 MHz (bottom panel; same data as in Figure 29) in Galactic coordinates.

Section 2.7 for an overview of the daily processing pipeline). The stacking proceeds in two steps: first, days within a contiguous set are averaged together, and second, all these averages are combined.

In the first step, averaging over contiguous blocks, data deemed bad are masked (arising from the presence of the Sun or Moon, RFI, or data-quality flags), and any day with less than 70% coverage after masking is discarded. Bias due to cross talk is estimated by calculating the median visibility at each zenith angle in a 1 hr region of R.A. where the sky signal is at low intensity. This value is subtracted from each individual day of data before stacking. Ideally, the same R.A. range would be used for all averages, but this is clearly impossible because the part of the sky transiting at night is changing with time of year. We compromised by choosing two R.A. ranges for the estimate of the cross-talk contribution. To ensure consistency, we use a set where both these ranges transit at night; we derive an additive correction from that average and apply it to all averages. Daily calibration is based on either Cas A or Cyg A. To account for different beam responses at the locations of these two sources, we derive a multiplicative amplitude correction at every frequency and apply it to all the averages prior to stacking. To remove the most prominent "smile" artifacts for display purposes, we subtracted Cas A, Cyg A, and Tau A in visibility space.

The deconvolved ring map at each frequency and decl. is approximately given by the 1D convolution of the sky with the E–W profile of the primary beam at the corresponding frequency and decl., as described in detail in CHIME Collaboration et al. (2022a). By this method, we attain an estimate of the true sky at each decl. by deconvolving the beam profile from each row of the ring map.

In the 52 day map of Figure 29 we see all the features that are evident in the 1 day map of Figure 28, illustrating the fact that CHIME achieves a high signal-to-noise ratio even in 1 day. Both the single-day and stacked maps are confusion limited; a major benefit of stacked maps for Galactic science is the full sky coverage even with the elimination of daytime data. Within the envelope of the diffuse emission along the Galactic plane we can identify many well-known supernova remnants and H II regions; these are evident in more detail in Figure 30. The combination of the visibility-space subtraction of the brightest three point sources and the deconvolution removes the grating lobe copies of all point sources and the saturation of the image at the R.A. of the brightest sources.

In Figure 30, we show the map from Figure 29 zoomed in on the Galactic plane and compared to a 408 MHz Stokes *I* map of the Galactic plane from the Canadian Galactic Plane Survey (CGPS; Tung et al. 2017). The CGPS 408 MHz data, obtained with the DRAO Synthesis Telescope, have an angular resolution of $\approx 3'$ and cover the area $52° \leqslant l \leqslant 193°$, $-6°.5 \leqslant b \leqslant 8°.5$. Short spacings for the CGPS map are incorporated from the Haslam et al. (1982) single-antenna data. There is good overall agreement between the CHIME and CGPS maps in the Galactic plane. Discrete objects such as supernova remnants and more extended objects such as the W3/4/5 H II region and the Cygnus X complex of H II regions and stellar clusters are distinctly visible in the CHIME data and are well matched with the CGPS data in terms of structure and relative brightness. Although the CHIME data lack zero-spacings, and thus sensitivity to the largest-scale structures, much of the diffuse emission visible in the CGPS map is also clearly discernible in the CHIME map. This is especially true of the bright extended emission at the low-longitude end of the CGPS coverage. The bright radio sources, Cyg A and Cas A, produce artifacts in both the CHIME and CGPS maps, although these are more easily mitigated in the CGPS through mosaicking of fields with a sufficiently dense sampling of pointings in those regions. While CHIME does not match the high angular resolution of the CGPS, its spectral coverage far exceeds that of the CGPS,[30] allowing for more in-depth exploration of frequency-dependent phenomena in the Galaxy over a larger spatial extent.

Sky maps like these will be the main data product for science involving noncosmological foregrounds. We will have all-sky images at hundreds of frequencies across an octave obtained with the same telescope, allowing analyses of spectral indices of point sources, extended objects, and diffuse emission. The Galactic signal is dominated by synchrotron emission, linearly polarized at its source, and Faraday rotated by the intervening magneto-ionic medium along virtually every line of sight. A major scientific goal is to apply Faraday synthesis (Brentjens & de Bruyn 2005) to the polarization data. We will derive Stokes *Q* and *U* maps, which will provide a valuable data set for Faraday synthesis across the whole sky; the wavelength-squared range and resolution of the CHIME data provide the Faraday depth resolution to isolate discrete magnetic features, with Faraday depth resolution $\delta\phi \approx 3.8/\Delta(\lambda^2) \approx 9$ rad m^{-2},

---

[30] The CGPS has a bandwidth of 3.5 MHz at 408 MHz and a bandwidth of 35 MHz at 1420 MHz.

while retaining sensitivity to extended Faraday depth features, with $\phi_{\text{max-scale}} \approx \pi\lambda_{\min}^{-2} \approx 22$ rad m^{-2}, in the Galaxy (Schnitzeler et al. 2009). Therefore, it will be possible to distinguish between extended structures and multiple narrow features in Faraday depth space. Exploration of this parameter space is only beginning (Dickey et al. 2019; Thomson et al. 2019). The 400–800 MHz polarization maps with $\approx 40'$ angular resolution from CHIME will form a component of the GMIMS survey, which includes a southern sky data set obtained with the CSIRO Parkes Telescope (Wolleben et al. 2019) and a 1280–1750 MHz northern sky data set observed with the Galt Telescope (Wolleben et al. 2021). If we are able to combine data across the 400–1800 MHz range, we will achieve $\delta\phi \approx$ 7 rad m^{-2} and $\phi_{\text{max-scale}} \approx 110$ rad m^{-2}, providing sensitivity to an unprecedented range of Faraday depth scales.

CHIME is an interferometer: it has coverage of the (u, v) plane down to 30 cm baselines, but not to zero baseline because autocorrelations of the signal from each feed are excluded from the analysis. To provide information on Galactic structure at the largest angular scales, a companion polarization survey will be made with a 15 m radio telescope at DRAO, covering 350–1050 MHz. These data, calibrated to an absolute scale of brightness temperature, will also provide the calibration of CHIME polarization data.

In addition, by observing the entire sky every day, we are sensitive to slow transients. We are cataloging daily fluxes of 2723 point sources, primarily quasars, to characterize variability.

5. Conclusions and Outlook

We have built and are operating an extremely high mapping speed instrument designed to measure the 3D distribution of neutral hydrogen over the full northern hemisphere and the redshift range $0.8 \leqslant z \leqslant 2.5$ with enough accuracy to provide useful constraints of the expansion history of the universe.

The instrument has been collecting data for cosmological analysis since late 2018. First results measuring the distribution of neutral hydrogen in 3D correlation with redshift catalogs of quasars and galaxies, using data from 2019, are presented in a companion paper (CHIME Collaboration et al. 2022a). CHIME is also monitoring the variability of 2723 sources with daily cadence and has produced confusion-limited maps of polarized Galactic emission across the 400–800 MHz band.

To quantify the cosmological constraining power of CHIME under ideal conditions, in Figure 31 we show an updated forecast for the statistical precision of CHIME in measuring the cosmic expansion history using the BAO feature in the 21 cm power spectrum. Compared to previous forecasts in the literature, these results use a more accurate version of CHIME's feed layout (Section 3.2.1), updated models for the mean 21 cm brightness temperature and linear H I bias, and an empirically derived estimate of CHIME's total system temperature, based on measurements presented in Section 4.3. We describe the methodology of these forecasts, which mostly follows Bull et al. (2015), in the Appendix. In particular, Table 2 lists the CHIME instrumental and survey characteristics used in these forecasts. Note that these forecasts assume perfect foreground subtraction and the absence of systematic errors. (Persistent RFI bands are indicated in the figure.)

We also show the expected precision of a combined galaxy and quasar sample from DESI (DESI Collaboration et al. 2016), computed within the same forecasting formalism; Lyα
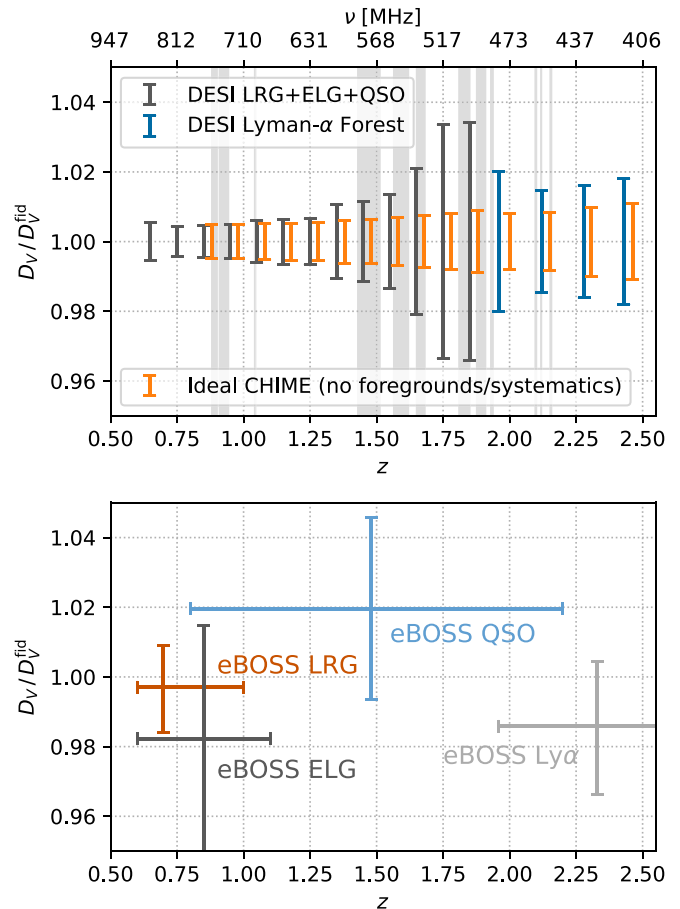

Figure 31. Top panel: projected constraints on the cosmic expansion history, parameterized using the spherically averaged distance measure D_V as a function of redshift, shown relative to a fiducial ΛCDM cosmology. For CHIME, the forecast error bars (orange) were calculated for 1 yr of integration time using the Fisher matrix approach of Bull et al. (2015), assuming perfect foreground subtraction and no systematics. Each error bar is statistically independent. We also show projections for the DESI clustering measurements (black), computed using the same formalism and based on combined constraints from the three clustering samples that overlap with CHIME's redshift coverage, and DESI Lyα forest measurements (blue), which we take from DESI Collaboration et al. (2016). (See the Appendix for the details of these forecasts.) Shaded gray bands denote regions inaccessible to CHIME owing to persistent sources of RFI. Bottom panel: expansion history measurements from the final eBOSS, taken from the compilation in Zhao et al. (2022). Comparison to the CHIME forecasts in the top panel indicates that the intrinsic *statistical* precision of CHIME is highly competitive with that of existing and near-future expansion history measurements. The challenge is to understand systematic effects well enough that statistical errors dominate.

forest measurements expected from DESI (which we do not recompute but take from DESI Collaboration et al. 2016); and state-of-the-art measurements by eBOSS (Alam et al. 2021; specific measurements taken from du Mas des Bourboux et al. 2020; Bautista et al. 2021; de Mattia et al. 2021; Hou et al. 2021, and summarized in Zhao et al. 2022). Figure 31 shows that CHIME's intrinsic statistical precision is competitive with DESI and that CHIME on its own is in principle capable of percent-level BAO measurements over most of its band.

Efforts in the coming years will be focused on realizing this potential, but we emphasize that we will need to overcome several challenges to do so. Foreground subtraction remains the primary obstacle to producing measurements that exploit CHIME's statistical power, and it is the main focus of our analysis effort. The path to seeing BAO through a haze of

Table 2
The Instrumental and Observing Parameters of CHIME Used in Our Forecasts

Parameter	Value
Sky coverage, S_{sky} (deg^2)	31,000
Redshift range, $[z_{min}, z_{max}]$	[0.8, 2.5]
Channel width (kHz)	390
Number of redshift bins, n_{bin}	15
System temperature,[a] T_{sys} (K)	55
Integration time, t_{tot} (yr)	1
Number of antennas per cylinder, N_{ant}	256
Number of polarizations per antenna, n_{pol}	2
Number of cylinders, N_{cyl}	4
Cylinder width, w_{cyl} (m)	20
Cylinder spacing (edge-to-edge) (m)	2
Physical cylinder length (m)	100
Illuminated cylinder length,[b] l_{cyl} (m)	78
Antenna spacing, d_{ant} (m)	0.3048
Minimum baseline, b_{min} (m)	0.3048
Maximum baseline, b_{max} (m)	102

Notes.
[a] Note that the quoted system temperature includes both instrumental and sky contributions and is based on the noise measurements in Section 4.3.
[b] In our chosen forecasting formalism (Bull et al. 2015), the length of the cylinder that is instrumented with feeds is relevant; for CHIME, this length is $l_{cyl} = 256 \times d_{ant} \approx 78$ m.

Galactic emission many orders of magnitude brighter is to filter out the spectrally smooth Galactic components and keep the spectrally chaotic BAO signal. Any systematic error that produces a rough or poorly understood spectral response mixes the Galactic and BAO signals. Thus, great care in the design has been taken to build a stable instrument with smooth, well-characterized response. Very precise measurement of the angular response of CHIME will be necessary to perform component separation at the level required to characterize the BAO because poorly understood frequency dependence of the angular response would lead to frequency variation of the Galactic contribution along an inferred line of sight. CHIME Collaboration et al. (2022a) describe a set of beam measurements and analysis methods that have allowed an initial detection of the 21 cm signal, and work is underway to improve on these methods. Other areas requiring further attention include mitigation of noise cross talk between nearby feeds, RFI mitigation in the lower half of the CHIME frequency band, and development of analysis methods that are robust to residual uncertainties in gain calibration and beam knowledge.

Overcoming these challenges has the potential to unlock a rich array of science targets accessible to 21 cm intensity mapping. Beyond BAO, there is potential to constrain the linear growth rate of structures as a way to test general relativity (Obuljen et al. 2018; Castorina & White 2019; Chen et al. 2019), constrain models of cosmic inflation through signatures in the primordial power spectrum of fluctuations (Xu et al. 2016; Beutler et al. 2019) or non-Gaussian statistics in large-scale structure (Xu et al. 2015; Karagiannis et al. 2020), and probe the nature of dark matter (Carucci et al. 2015; Bauer et al. 2021). In addition, "tidal reconstruction" techniques, which reconstruct large-scale (foreground-obscured) modes from the correlations they induce between smaller-scale modes (Zhu et al. 2018; Modi et al. 2019; Darwish et al. 2021), can greatly expand the opportunities for cross-correlations with

surveys of the CMB or photometric galaxy redshifts. Additionally, lower-frequency observations of the 21 cm line are well suited to probing the era of reionization (Furlanetto et al. 2019b) or, more ambitiously, the cosmic "dark ages" up to $z \sim \mathcal{O}(100)$ (Furlanetto et al. 2019a).

The instrument described here also acts as the front end for several other systems, providing calibrated data to an FRB detector (CHIME/FRB Collaboration et al 2018), a 10-beam system that monitors all pulsars visible from Canada with up to daily cadence (CHIME/Pulsar Collaboration et al. 2021), a system to search for cold clouds acting as 21 cm absorption-line systems, and a VLBI station (Cassanelli et al. 2022). Among the accomplishments these new instruments have made is the discovery of half a dozen Galactic pulsars; detection of an exceptionally bright radio burst from a Galactic magnetar (CHIME/FRB Collaboration et al. 2020), pointing to possible similarities of magnetars and FRBs; and publication of the first substantial catalog of FRBs (CHIME/Pulsar Collaboration et al. 2021). This broad range of additional scientific impact comes directly from achieving the sensitivity, large fractional bandwidth, and enormous field of view that hydrogen intensity mapping requires.

We have shown that CHIME is capable of generating a multitude of scientific results and have demonstrated that one can build a very powerful instrument for a comparatively small cost when a clear scientific goal drives the design. We expect a steady flow of further results in the years to come.

[31] https://www.westgrid.ca
[32] https://www.computecanada.ca
[33] https://www.cmc.ca

Software: NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), HDF5 (https://www.hdfgroup.org/HDF5/), h5py (Collette et al. 2021), Matplotlib (Hunter 2007), scikit-rf (Arsenovic et al. 2022), Skyfield (Rhodes 2019), caput (Shaw et al. 2020b), ch_pipeline (Shaw et al. 2020a), draco (Shaw et al. 2020c), kotekan (Renard et al. 2021), CST (https://www.cst.com/), GRASP (https://www.ticra.com/software/grasp/), Prometheus (https://prometheus.io/), and Grafana (https://grafana.com/grafana).

Appendix A
Details of BAO Forecast

A.1. Fisher Matrix Formalism

We project the constraints on BAOs from CHIME, in comparison with DESI, which observes in the optical and has overlapping sky and redshift coverage with CHIME. We mainly follow the Fisher matrix method of Bull et al. (2015), using a modified version of their publicly available forecast code.[34]

The Fisher matrix can be written as (Seo & Eisenstein 2007)

$$F_{ij} = \int \frac{d^3k}{(2\pi)^3} V_{\text{eff}} \frac{\partial C^S(\mathbf{k}, z)}{\partial \theta_i} \frac{\partial C^S(\mathbf{k}, z)}{\partial \theta_j}, \quad (A1)$$

where the effective volume V_{eff} is given by

$$V_{\text{eff}} = f_{\text{sky}} \int_{z_{\min}}^{z_{\max}} \frac{dV}{dz} dz \left[\frac{C^S(\mathbf{k}, z)}{C^S(\mathbf{k}, z) + C^N(\mathbf{k}, z)} \right]^2. \quad (A2)$$

In the above expression, f_{sky} is the fractional sky coverage; $\int_{z_{\min}}^{z_{\max}} \frac{dV}{dz} dz$ is the physical volume of the survey; C^S and C^N are the signal and noise covariance, respectively; and θ is the set of cosmological parameters to be constrained. In our case, θ is the following set of parameters:

$$\theta = \{D_A(z), H(z), A(z), [b\sigma_8](z), [f\sigma_8](z), \sigma_{\text{NL}}\}. \quad (A3)$$

The redshift-dependent parameters are constrained within redshift bins with width $\Delta z = 0.1$ for $z \leqslant 1.8$ and $\Delta z = 0.16$ for $z > 1.8$, in order to match the binning of the DESI forecasts in DESI Collaboration et al. (2016). The Hubble rate $H(z)$ and angular diameter distance $D_A(z)$ are transformed into the volume distance D_V and Alcock−Pacynski term F through

$$D_V(z) = \left[(1 + z)^2 D_A(z)^2 \frac{cz}{H(z)} \right]^{\frac{1}{3}},$$
$$F(z) = (1 + z) D_A(z) \frac{H(z)}{c}, \quad (A4)$$

and we forecast the fractional error bars on measurements of D_V in each redshift bin. The amplitude $A(z)$ is defined by decomposing the matter power spectrum P_m into a smooth template P_{smooth} and oscillatory BAO factor f_{BAO},

$$P_m(k, z) = [1 + A(z) f_{\text{BAO}}(k)] P_{\text{smooth}}(k, z), \quad (A5)$$

implemented using the method from Bull et al. (2015). The linear bias b, linear growth rate f, and fluctuation amplitude σ_8 have their usual meanings, while σ_{NL} is the redshift-space

damping scale defined in the next section. In each redshift bin, the D_V forecasts marginalize over the other parameters (F, A, $b\sigma_8$, $f\sigma_8$, and σ_{NL}) with no priors.

A.2. Signal Models

For CHIME, we take the H I signal covariance to be

$$C^S(\mathbf{k}, z) = T_b(z)^2 (b_{\text{H I}}(z) + f(z)\mu^2)^2 e^{-k^2\mu^2\sigma_{\text{NL}}^2} P_m(k, z), \quad (A6)$$

where $T_b(z)$ is the H I brightness temperature and $b_{\text{H I}}(z)$ is the linear bias of H I. For $T_b(z)$, we use the expression from Hall et al. (2013), with the fitting formula for the mean H I density $\Omega_{\text{H I}}(z)$ from Crighton et al. (2015), and for $b_{\text{H I}}(z)$, we use the model from Cosmic Visions 21 cm Collaboration et al. (2018), which smoothly interpolates between measurements from the IllustrisTNG simulation at $z < 2$ (Villaescusa-Navarro et al. 2018) and the analytical approximation from Castorina & Villaescusa-Navarro (2017) at $z > 2$. The large-scale effect of redshift-space distortions is accounted for in the $f(z)\mu^2$ term in Equation (A6), where $f(z)$ is the linear growth rate and μ is the angle of the wavevector to the line of sight. At smaller scales, the exponential factor roughly accounts for the "Finger-of-God" effect that suppresses the observed clustering power beyond the cutoff scale σ_{NL}. The linear matter power spectrum $P_m(k)$ is calculated using the Code for Anisotropies in the Microwave Background (Lewis et al. 2000).

For DESI, we use

$$C^S(\mathbf{k}, z) = (b_g(z) + f(z, k)\mu^2)^2 e^{-k^2\mu^2\sigma_{\text{NL}}^2} P(k) \quad (A7)$$

for the galaxy signal covariance, where $b_g(z)$ is the linear galaxy bias and the other components are the same as in Equation (A6). For these forecasts, we combine the luminous red galaxy (LRG), emission-line galaxy (ELG), and quasar (QSO) samples, by summing their expected number densities in each redshift bin, as given for the DESI baseline survey in Section 2.4.2 of DESI Collaboration et al. (2016), and using a number-density-weighted mean of the corresponding linear bias factors, also taken from DESI Collaboration et al. (2016). (We do not consider the bright galaxy sample because its redshift range does not overlap with that of CHIME.)

For these forecasts, we adopt fiducial cosmological parameters from the Planck CMB-only best-fit ΛCDM model (Planck Collaboration et al. 2020),

$$h = 0.6732, \ \Omega_m = 0.3158, \ \Omega_K = 0, \ \Omega_b = 0.0494,$$
$$w = -1, \ n_s = 0.966, \ \sigma_8 = 0.812, \ N_{\text{eff}} = 3.046. \quad (A8)$$

Following Bull et al. (2015), we choose the nonlinear dispersion scale to be $\sigma_{\text{NL}} = 7$ Mpc, corresponding to power being significantly damped at $k \gtrsim 0.14$ Mpc^{-1}. This value is higher than recent values from the literature, for both H I and DESI-like galaxies (e.g., CHIME Collaboration et al. 2022a), but is justified here because it limits the sensitivity of our forecasts to nonlinear scales where the assumptions in Equations (A6) and (A7) break down. In addition, we make use of the BAO information only, instead of the full shape of the H I or galaxy power spectrum (e.g., Sailer et al. 2021). While a full-shape analysis would provide increased constraining power, it it also more likely to be affected by foregrounds

---
[34] https://github.com/philbull/RadioFisher

and systematics, so we aim to be conservative in that respect by restricting to BAO only.

A.3. Noise Models

We mainly follow Bull et al. (2015) in approximating the noise covariance for CHIME as

$$C^N(\mathbf{k}, z) = \frac{T_{\text{sys}}(z)^2}{\nu_{21} n_{\text{pol}} t_{\text{tot}}} \frac{\lambda^4 S_{\text{sky}}}{A_e^2 S_{\text{FOV}}} \frac{1}{n(\mathbf{u})}, \tag{A9}$$

where ν_{21} is the H I line emission rest frequency, n_{pol} is the number of polarizations per antenna, and $\lambda(z)$ is the observing wavelength corresponding to emission from redshift z. For the system temperature $T_{\text{sys}}(z)$, we use a constant 55 K, based on the observations in Section 4.3; note that this includes both instrumental and sky contributions, which are usually modeled separately in forecasts. We take the total integration time t_{tot} to be 1 yr. The sky coverage of CHIME, S_{sky}, is 31,000 deg^2, corresponding to $f_{\text{sky}} \approx 0.75$. We approximate the instantaneous field of view of a single cylinder as $S_{\text{FOV}} \approx 90° \times \lambda/w_{\text{cyl}}$ (Newburgh et al. 2014), where w_{cyl} is the cylinder width. The effective collecting area per antenna is denoted by A_e, and in this formalism, it takes the following form for a cylinder telescope:

$$A_e = \eta \frac{l_{\text{cyl}}}{N_{\text{ant}}} w_{\text{cyl}}, \tag{A10}$$

where η is the aperture efficiency, assumed to be 0.7 in our case (following Bull et al. 2015); l_{cyl} is the length along the cylinder axis that is instrumented with feeds; and N_{ant} is the number of antennas per cylinder. $n(\mathbf{u})$ is the (u, v)-plane baseline number density of CHIME, calculated using the code from Bull et al. (2015) with details given in their Appendix C, accounting for the fact that adjacent CHIME cylinders are separated by 2 m (implying that the shortest nonzero E–W baseline is 22 m). We are only interested in the ideal statistical constraining power of CHIME, so we assume perfect foreground cleaning and no systematics, so that the noise covariance includes only the instrumental thermal noise. In addition, we assume Gaussian beams with equal response across the sky and neglect the intrinsic H I shot noise, which is far smaller than the thermal noise (Villaescusa-Navarro et al. 2018). Table 2 summarizes the instrumental parameters used in our forecast for CHIME.

The noise covariance for a galaxy survey is dominated by the shot noise owing to the limited number of galaxies detected in the observed region at a particular redshift. We assume that this noise covariance is Poissonian for DESI and is thus

$$C^N(z) = \frac{1}{n(z)}, \tag{A11}$$

where $n(z)$ is the comoving galaxy number density at redshift z. We convert the quoted values for $dN/(dz\, d\text{deg}^2)$ from Section 2.4.2 of DESI Collaboration et al. (2016) into $n(z)$ values within each redshift bin and sum over the LRG, ELG, and QSO samples. We adopt the sky coverage of the DESI baseline survey at 14,000 deg^2.

ORCID iDs

Mandana Amiri https://orcid.org/0000-0001-6523-9029
Kevin Bandura https://orcid.org/0000-0003-3772-2798
Tianyue Chen https://orcid.org/0000-0003-0173-6274
Jean-François Cliche https://orcid.org/0000-0001-6509-8430
Meiling Deng https://orcid.org/0000-0001-8123-7322
Nolan Denman https://orcid.org/0000-0003-2381-9804
Matt Dobbs https://orcid.org/0000-0001-7166-6422
Mateus Fandino https://orcid.org/0000-0002-6899-1176
Simon Foreman https://orcid.org/0000-0002-0190-2271
Mark Halpern https://orcid.org/0000-0002-1760-0868
David Hanna https://orcid.org/0000-0002-8513-5603
Alex S. Hill https://orcid.org/0000-0001-7301-5666
Gary Hinshaw https://orcid.org/0000-0002-4241-8320
Carolin Höfer https://orcid.org/0000-0003-4887-8114
Joseph Kania https://orcid.org/0000-0002-3354-3859
Peter Klages https://orcid.org/0000-0003-3405-7470
T. L. Landecker https://orcid.org/0000-0003-1455-2546
Joshua MacEachern https://orcid.org/0000-0001-8064-6116
Kiyoshi Masui https://orcid.org/0000-0002-4279-6946
Juan Mena-Parra https://orcid.org/0000-0002-0772-9326
Nikola Milutinovic https://orcid.org/0000-0001-8292-0051
Arash Mirhosseini https://orcid.org/0000-0002-2626-5985
Laura Newburgh https://orcid.org/0000-0002-7333-5552
Anna Ordog https://orcid.org/0000-0002-2465-8937
Ue-Li Pen https://orcid.org/0000-0003-2155-9578
Tristan Pinsonneault-Marotte https://orcid.org/0000-0002-9516-3245
Ava Polzin https://orcid.org/0000-0002-5283-933X
Alex Reda https://orcid.org/0000-0001-6967-7253
Andre Renard https://orcid.org/0000-0003-3463-7918
J. Richard Shaw https://orcid.org/0000-0002-4543-4588
Seth R. Siegel https://orcid.org/0000-0003-2631-6217
Saurabh Singh https://orcid.org/0000-0001-7755-902X
Rick Smegal https://orcid.org/0000-0002-6873-2094
Ian Tretyakov https://orcid.org/0000-0002-7436-2325
Keith Vanderlinde https://orcid.org/0000-0003-4535-9378
Haochen Wang https://orcid.org/0000-0002-1491-3738
Donald V. Wiebe https://orcid.org/0000-0002-6669-3159
Dallas Wulf https://orcid.org/0000-0001-7314-9496

References

Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:1902.05569
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alam, S., Aubert, M., Avila, S., et al. 2021, PhRvD, 103, 083533
Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, LRR, 21, 2
Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, MNRAS, 476, 3382
Anderson, L., Aubourg, E., Bailey, S., et al. 2012, MNRAS, 427, 3435
Arsenovic, A., Hillairet, J., Anderson, J., et al. 2022, IMMag, 23, 98
Ata, M., Baumgarten, F., Bautista, J., et al. 2018, MNRAS, 473, 4773
Baars, J. W. 2007, The Paraboloidal Reflector Antenna in Radio Astronomy and Communication (Berlin: Springer)
Bandura, K., Bender, A. N., Cliche, J. F., et al. 2016a, JAI, 5, 1641005
Bandura, K., Cliche, J. F., Dobbs, M. A., et al. 2016b, JAI, 5, 1641004
Bauer, J. B., Marsh, D. J. E., Hložek, R., Padmanabhan, H., & Laguë, A. 2021, MNRAS, 500, 3162
Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2021, MNRAS, 500, 736
Bennett, J. C., Anderson, A. P., McInnes, P. A., & Whitaker, A. J. T. 1976, ITAP, 24, 295
Berger, P., Newburgh, L. B., Amiri, M., et al. 2016, Proc. SPIE, 9906, 99060D
Beutler, F., Biagetti, M., Green, D., Slosar, A., & Wallisch, B. 2019, PhRvR, 1, 033209
Beutler, F., Blake, C., Colless, M., et al. 2011, MNRAS, 416, 3017
Blake, C., Kazin, E. A., Beutler, F., et al. 2011, MNRAS, 418, 1707

Brentjens, M. A., & de Bruyn, A. G. 2005, A&A, 441, 1217
Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, ApJ, 803, 21
Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96
Cannon, L. E. 1969, PhD thesis, Montana State Univ.
Carucci, I. P., Villaescusa-Navarro, F., Viel, M., & Lapi, A. 2015, JCAP, 2015, 047
Cassanelli, T., Leung, C., Rahman, M., et al. 2022, AJ, 163, 65
Castorina, E., & Villaescusa-Navarro, F. 2017, MNRAS, 471, 1788
Castorina, E., & White, M. 2019, JCAP, 2019, 025
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Natur, 466, 463
Chen, S.-F., Castorina, E., White, M., & Slosar, A. 2019, JCAP, 2019, 023
CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022a, arXiv:2202.01242
CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022b, ApJ, 932, 100
CHIME/FRB Collaboration, Amiri, M., Bandura, K., et al. 2018, ApJ, 863, 48
CHIME/FRB Collaboration, Andersen, B. C., Bandura, K. M., et al. 2020, Natur, 587, 54
CHIME/Pulsar Collaboration, Amiri, M., Bandura, K. M., et al. 2021, ApJS, 255, 5
Cohen, A. S., Lane, W. M., Cotton, W. D., et al. 2007, AJ, 134, 1245
Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505
Collette, A., Kluyver, T., Caswell, T. A., et al. 2021, h5py/h5py: 3.5.0, v3.5.0, Zenodo, doi:10.5281/zenodo.5585380
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Cosmic Visions 21 cm Collaboration, Ansari, R., Arena, E. J., et al. 2018, arXiv:1810.09572
Crighton, N. H. M., Murphy, M. T., Prochaska, J. X., et al. 2015, MNRAS, 452, 217
Darwish, O., Foreman, S., Abidi, M. M., et al. 2021, PhRvD, 104, 123520
de Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2021, MNRAS, 501, 5616
Deng, M. 2020, PhD thesis, Univ. of British Columbia
Deng, M., & Campbell-Wilson, D. 2014, in 2014 16th Int. Symp. on Antenna Technology and Applied Electromagnetics (Piscataway, NJ: IEEE)
Denman, N., Renard, A., Vanderlinde, K., et al. 2020, JAI, 09, 2050014
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
DES Collaboration, Abbott, T. M. C., Abdalla, F. B., et al. 2019, MNRAS, 483, 4866
DES Collaboration, Abbott, T. M. C., Aguena, M., et al. 2022, PhRvD, 105, 043512
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Dickey, J. M., Landecker, T. L., Thomson, A. J. M., et al. 2019, ApJ, 871, 106
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, ApJ, 901, 153
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Furlanetto, S., Bowman, J. D., Mirocha, J., et al. 2019a, arXiv:1903.06212
Furlanetto, S., Carilli, C. L., Mirocha, J., et al. 2019b, arXiv:1903.06204
Gregory, P. C., Scott, W. K., Douglas, K., & Condon, J. J. 1996, ApJS, 103, 427
Hall, A., Bonvin, C., & Challinor, A. 2013, PhRvD, 87, 064026
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hou, J., Sánchez, A. G., Ross, A. J., et al. 2021, MNRAS, 500, 1201
Hunter, J. D. 2007, CSE, 9, 90
Karagiannis, D., Slosar, A., & Liguori, M. 2020, JCAP, 2020, 052
Klages, P., Bandura, K., Denman, N., et al. 2015, in 2015 IEEE 26th Int. Conf. on Application-specific Systems, Architectures and Processors (ASAP) (Piscataway, NJ: IEEE), 164
Landecker, T. L., Anderson, M. D., Routledge, D., Smegal, R. J., & Trikha, P. 1991, RaSc, 26, 363
Leung, M. 2008, PhD thesis, Univ. of Sydney
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Liu, A., & Tegmark, M. 2011, PhRvD, 83, 103006
Liu, A., & Tegmark, M. 2012, MNRAS, 419, 3491
Masui, K., Amiri, M., Connor, L., et al. 2015, A&C, 12, 181
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20
Mckinven, R., Michilli, D., Masui, K., et al. 2021, ApJ, 920, 138
Mena-Parra, J., Bandura, K., Dobbs, M. A., Shaw, J. R., & Siegel, S. 2018, JAI, 7, 1850008
Modi, C., White, M., Slosar, A., & Castorina, E. 2019, JCAP, 2019, 023
Mumford, W. W. 1961, Proc. IRE, 49, 427
Neveux, R., Burtin, E., de Mattia, A., et al. 2020, MNRAS, 499, 210
Newburgh, L. B., Addison, G. E., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91454V
Ng, C., Vanderlinde, K., Paradise, A., et al. 2017, in 2017 32nd General Assembly and Scientific Symp. of the Int. Union of Radio Science (URSI GASS) (Piscataway, NJ: IEEE)
Obuljen, A., Castorina, E., Villaescusa-Navarro, F., & Viel, M. 2018, JCAP, 2018, 004
Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155
Padmanabhan, N., Xu, X., Eisenstein, D. J., et al. 2012, MNRAS, 427, 2132
Parsons, A., Backer, D., Siemion, A., et al. 2008, PASP, 120, 1207
Pen, U.-L., Staveley-Smith, L., Peterson, J. B., & Chang, T.-C. 2009, MNRAS, 394, L6
Percival, W. J., Cole, S., Eisenstein, D. J., et al. 2007, MNRAS, 381, 1053
Perley, R. A., & Butler, B. J. 2017, ApJS, 230, 7
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Renard, A., Shaw, R., Ng, C., et al. 2021, Kotekan: A framework for high-performance radiometric data pipelines, v2021.11, Zenodo, doi:10.5281/zenodo.5842660
Rengelink, R. B., Tang, Y., de Bruyn, A. G., et al. 1997, A&AS, 124, 259
Rhodes, B. 2019, Skyfield: High precision research-grade positions for planets and Earth satellites generator, Astrophysics Source Code Library, ascl:1907.024
Riess, A. G., Strolger, L.-G., Tonry, J., et al. 2004, ApJ, 607, 665
Ross, A. J., Samushia, L., Howlett, C., et al. 2015, MNRAS, 449, 835
Sailer, N., Castorina, E., Ferraro, S., & White, M. 2021, JCAP, 2021, 049
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
Schnitzeler, D. H. F. M., Katgert, P., & de Bruyn, A. G. 2009, A&A, 494, 611
Scott, P. F., & Ryle, M. 1977, MNRAS, 178, 539
Seo, H.-J., & Eisenstein, D. J. 2003, ApJ, 598, 720
Seo, H.-J., & Eisenstein, D. J. 2007, ApJ, 665, 14
Seo, H.-J., Ho, S., White, M., et al. 2012, ApJ, 761, 13
Shaw, J. R., Masui, K., Hincks, A. D., et al. 2020a, chime-experiment/ch_pipeline: 20.10.0, v20.10.0, Zenodo, doi: 10.5281/zenodo.5846379
Shaw, J. R., Masui, K., Nitsche, R., et al. 2020b, radiocosmology/caput: 20.10.0, v20.10.0, Zenodo, doi: 10.5281/zenodo.5846375
Shaw, R., Nitsche, R., Siegel, S. R., et al. 2020c, radiocosmology/draco: 20.10.0, v20.10.0, Zenodo, doi: 10.5281/zenodo.5828373
Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, PhRvD, 91, 083514
Slosar, A., Iršič, V., Kirkby, D., et al. 2013, JCAP, 2013, 026
Switzer, E. R., Masui, K. W., Bandura, K., et al. 2013, MNRAS, 434, L46
Taylor, J., Denman, N., Bandura, K., et al. 2018, JAI, 8, 1940004
Thomson, A. J. M., Landecker, T. L., Dickey, J. M., et al. 2019, MNRAS, 487, 4751
Tung, A. K., Kothes, R., Landecker, T. L., et al. 2017, AJ, 154, 156
Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ApJ, 866, 135
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wolleben, M., Landecker, T. L., Carretti, E., et al. 2019, AJ, 158, 44
Wolleben, M., Landecker, T. L., Douglas, K. A., et al. 2021, AJ, 162, 35
Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, MNRAS, 510, 3495
Xu, Y., Hamann, J., & Chen, X. 2016, PhRvD, 94, 123518
Xu, Y., Wang, X., & Chen, X. 2015, ApJ, 798, 40
Zhao, C., Variu, A., He, M., et al. 2022, MNRAS, 511, 5492
Zhu, H.-M., Pen, U.-L., Yu, Y., & Chen, X. 2018, PhRvD, 98, 043511