



The CHIME Fast Radio Burst Project: System Overview

The CHIME/FRB Collaboration

M. Amiri¹, K. Bandura^{2,3}, P. Berger^{4,5}, M. Bhardwaj^{6,7}, M. M. Boyce^{6,7}, P. J. Boyle^{6,7}, C. Brar^{6,7}, M. Burhanpurkar⁸, P. Chawla^{6,7}, J. Chowdhury^{6,7,9}, J.-F. Cliche^{6,7}, M. D. Cranmer^{6,7}, D. Cubranic¹, M. Deng¹, N. Denman^{10,11}, M. Dobbs^{6,7}, M. Fandino¹, E. Fonseca^{6,7}, B. M. Gaensler¹¹, U. Giri¹², A. J. Gilbert^{6,7}, D. C. Good¹, S. Guliani¹, M. Halpern¹, G. Hinshaw¹, C. Höfer¹, A. Josephy^{6,7}, V. M. Kaspi^{6,7}, T. L. Landecker¹³, D. Lang^{10,11,12}, H. Liao⁶, K. W. Masui¹, J. Mena-Parra^{6,7}, A. Naidu^{6,7}, L. B. Newburgh¹⁴, C. Ng¹¹, C. Patel⁶, U.-L. Pen^{4,11,12}, T. Pinsonneault-Marotte¹, Z. Pleunis^{6,7}, M. Rafiei Ravandi¹², S. M. Ransom¹⁵, A. Renard¹¹, P. Scholz¹³, K. Sigurdson¹, S. R. Siegel^{6,7}, K. M. Smith¹², I. H. Stairs¹, S. P. Tendulkar^{6,7}, K. Vanderlinde^{10,11}, and D. V. Wiebe¹

¹ Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada

² Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA

³ CSEE, West Virginia University, Morgantown, WV 26505, USA

⁴ Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada

⁵ Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H4, Canada

⁶ Department of Physics, McGill University, 3600 University Street, Montréal, QC H3A 2T8, Canada; vkaspi@physics.mcgill.ca

⁷ McGill Space Institute, McGill University, 3550 University Street, Montréal, QC H3A 2A7, Canada

⁸ Harvard University, Cambridge MA, 02138, USA

⁹ Ming Hsieh Department of Electrical Engineering, University of Southern California, 3740 McClintock Avenue, EEB 102, Los Angeles, CA 90089-2560, USA

¹⁰ Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

¹¹ Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

¹² Perimeter Institute for Theoretical Physics, 31 Caroline Street N, Waterloo, ON N2L 2Y5, Canada

¹³ Dominion Radio Astrophysical Observatory, Herzberg Astronomy & Astrophysics Research Centre, National Research Council of Canada, P.O. Box 248, Penticton, V2A 6J9, Canada

¹⁴ Department of Physics, Yale University, New Haven, CT 06520, USA

¹⁵ NRAO, 520 Edgemont Road, Charlottesville, VA 22903, USA

Received 2018 March 16; revised 2018 June 10; accepted 2018 July 3; published 2018 August 9

Abstract

The Canadian Hydrogen Intensity Mapping Experiment (CHIME) is a novel transit radio telescope operating across the 400–800 MHz band. CHIME is composed of four 20 m × 100 m semicylindrical paraboloid reflectors, each of which has 256 dual-polarization feeds suspended along its axis, giving it a $\gtrsim 200$ deg² field of view. This, combined with wide bandwidth, high sensitivity, and a powerful correlator, makes CHIME an excellent instrument for the detection of fast radio bursts (FRBs). The CHIME Fast Radio Burst Project (CHIME/FRB) will search beam-formed, high time and frequency resolution data in real time for FRBs in the CHIME field of view. Here we describe the CHIME/FRB back end, including the real-time FRB search and detection software pipeline, as well as the planned offline analyses. We estimate a CHIME/FRB detection rate of 2–42 FRBs sky^{−1} day^{−1} normalizing to the rate estimated at 1.4 GHz by Vander Wiel et al. Likely science outcomes of CHIME/FRB are also discussed. CHIME/FRB is currently operational in a commissioning phase, with science operations expected to commence in the latter half of 2018.

Key words: instrumentation: interferometers – methods: observational – radio continuum: general – techniques: interferometric – telescopes

1. Fast Radio Bursts

The fast radio burst (FRB) phenomenon was first recognized when Lorimer et al. (2007) discovered a bright (30 Jy) radio burst of duration <5 ms well off the southern Galactic plane, in a search for single pulses using the Parkes radio telescope. The burst had a dispersion measure (DM) of 375 pc cm^{−3}, far in excess of the maximum expected along this line of sight from the Galaxy (25 pc cm^{−3}) according to the NE2001 model for the Galactic electron density distribution (Cordes & Lazio 2001). The large excess DM suggests an extragalactic origin, as first noted by Lorimer et al. (2007). Subsequently, Thornton et al. (2013) reported four additional FRBs, also detected using the Parkes telescope, each with a DM greatly in excess of the total line-of-sight Galactic electron column density. Following this, FRBs became widely accepted as a newly recognized astrophysical phenomenon. The identification of particularly unusual FRB-like “peryton” events with radio interference

local to the Parkes telescope (Petroff et al. 2015c) ultimately solidified the astrophysical nature of the remaining, non-peryton sources. The discovery of an FRB using the Arecibo Observatory¹⁶ (Spitler et al. 2014) laid to rest concerns regarding FRBs being a telescope-specific phenomenon. Since then, one FRB has been discovered using the Green Bank Telescope (Masui et al. 2015), and an additional two dozen have been found using Parkes, the UTMOST telescope (e.g., Caleb et al. 2017), and ASKAP (Bannister et al. 2017). This brings the total number of formally reported FRBs¹⁷ to approximately 30. Of these, most were discovered at a radio frequency of 1.4 GHz, except the GBT and UTMOST events, which were detected near 800 MHz.

Although the number of detected FRBs today is fewer than three dozen, the inferred sky rate at 1.4 GHz is estimated to be

¹⁶ www.naic.edu

¹⁷ See the online FRB catalog at www.frbcat.org.

between several hundred and a few thousand per sky per day (e.g., Thornton et al. 2013; Rane et al. 2016; Lawrence et al. 2017) at existing fluence thresholds (~ 1 Jy ms). The discrepancy between the high inferred rate and the relatively small number of detections is largely due to the small fields of view (FOVs) of existing telescopes. There are indications that the overall rate is substantially lower at frequencies below ~ 400 MHz (Karastergiou et al. 2015; Rowlinson et al. 2016; Chawla et al. 2017). However, in the 400–800 MHz range, there exist only weak constraints on the FRB rate (Amiri et al. 2017).

Assuming standard estimates for the intergalactic free electron density (e.g., Inoue 2004), as well as reasonable estimates for the DM contribution from host galaxies, the events appear to be from cosmological distances (redshifts $z \sim 0.1$ – 2). The estimated volumetric FRB rate does not match with those measured for other known transient events such as gamma-ray bursts or supernovae (Kulkarni et al. 2014). The sources of FRBs are thus currently unknown, but models typically include compact objects, due to the short timescales involved, with some invoking cataclysmic events (see Rane & Lorimer 2017, for a recent review). Being thus far discovered primarily using diffraction-limited single-dish radio telescopes at decimeter wavelengths, or interferometers with limited baselines, FRB positional uncertainties are typically many arcminutes, prohibiting identifications of multiwavelength counterparts or host galaxies.

Thus far, only the Arecibo-discovered FRB 121102 has been observed to repeat (Scholz et al. 2016; Spitler et al. 2016). This discovery demonstrates that at least this particular source does not have its origin in a cataclysmic event. Many dozen repeat bursts, all at a consistent DM, have now been detected and show strong evidence for clustering in time (Scholz et al. 2016; Chatterjee et al. 2017; Opperman & Pen 2017; Michilli et al. 2018). Detection of multiple bursts interferometrically enables a subarcsecond localization (Chatterjee et al. 2017; Marcote et al. 2017) in a dwarf galaxy at redshift $z \approx 0.2$ (Tendulkar et al. 2017). This provides strong support for the interpretation that the large excess DMs of other FRBs imply cosmological distances and suggests that FRBs may be useful new probes of galaxy halos, large-scale structure, the intergalactic medium, and cosmological parameters (e.g., McQuinn 2014; Masui & Sigurdson 2015; Fialkov & Loeb 2016; Ravi et al. 2016; Yang & Zhang 2016; Prochaska & Neeleman 2018; Shull & Danforth 2018; Yoshiura & Takahashi 2018).

Whether all FRBs repeat is currently unknown. Apart from FRB 121102, observations of several FRB positions for many hours (e.g., Petroff et al. 2015b) have not revealed repeated bursts from the same source. In one case, a second event, FRB 140514, was observed from the same arcminute-area beam on the sky as FRB 110220, but at a very different DM. Petroff et al. conclude that they are likely unrelated (but see Maoz et al. 2015, for an alternative discussion). Thus, whether the repeating nature of FRB 121102 is unique is unknown, as are many basic characteristics describing the population, such as their true sky distribution, the luminosity, DM, or scattering time distributions, and whether any of these properties are correlated.

Here we describe a project underway to enable the Canadian Hydrogen Intensity Mapping Experiment (CHIME)¹⁸—originally

designed to map baryon acoustic oscillation (BAO) features in redshifted neutral hydrogen gas as a measure of dark energy (Newburgh et al. 2014)—to be, in parallel, a sensitive FRB detector: the CHIME Fast Radio Burst Project (hereafter CHIME/FRB). CHIME’s large ($\gtrsim 200$ deg²) FOV, large collecting area, wide radio bandwidth, and powerful correlator, which provides 1024 independent beams within the telescope’s primary beam, will enable this FRB back end to detect multiple FRBs per day, given current estimates of the FRB rate. The instrument should also detect many Galactic rotating radio transients (RRATs; McLaughlin et al. 2006). Moreover, its sensitivity across the wide 400–800 MHz band is ideal for studying the frequency dependence of the FRB sky rate. In Section 2 we provide a brief description of the CHIME telescope structure and analog system. Section 3 describes the powerful CHIME correlator, whose original design for cosmology was upgraded significantly to accommodate intensity beam-forming for the FRB searches. Section 4 describes the CHIME/FRB real-time search, detection, and analysis procedure. Planned analysis features are described in Section 5, and the commissioning status of CHIME/FRB is described in Section 6. Section 7 provides an up-to-date estimate of the expected CHIME/FRB event rate, together with a discussion of the science to be probed by CHIME/FRB data.

In addition to the cosmology experiment and CHIME/FRB, the CHIME telescope will also perform daily timing observations of known radio pulsars and RRATs using a dedicated, independent back end: “CHIME/Pulsar.” Information about the latter effort is described by Ng & CHIME/Pulsar Collaboration (2017).

2. The CHIME Telescope Structure, Feeds, and Analog-signal Path

The CHIME telescope is located on the grounds of the Dominion Radio Astrophysical Observatory (DRAO) near Penticton, British Columbia. The choice of operating frequency, collecting area, and angular resolution for the CHIME telescope was driven by the original motivation for the project: hydrogen intensity mapping of the entire northern hemisphere to probe the accelerating expansion of the universe over the redshift range where dark energy began to exert its influence, $z = 0.8$ – 2.5 . Since the BAO signal is weak, and large sky coverage is needed to overcome sample variance, exceptionally fast mapping speed is required, driving the instrument to a design with many hundreds of feeds to achieve the mapping goal in a reasonable amount of time. As 100 m class telescopes are expensive, and no positions on the sky are favored, a transit telescope with no moving parts is the preferred option. Table 1 provides a summary of the telescope’s key properties. Detailed telescope performance metrics will be provided in a future publication.

The CHIME telescope (see Figure 1) consists of four 20 m wide and 100 m long cylindrical paraboloidal reflectors. The intercylinder gaps are 2 m. The cylinders are stationary, aligned north–south. The focal length was chosen to be 5 m ($f/D = 0.25$) to place the feeds in the plane of the aperture to minimize cross-talk and ground pickup. With no moving parts, the telescope structure could be built at low cost. The structure is steel, built of conventional sections and components, to the dimensional accuracy common for commercial buildings. The reflecting surface is galvanized steel mesh with 16 mm openings, a compromise between allowing snow to fall through

¹⁸ www.chime-experiment.ca

Table 1Key Properties of the CHIME Telescope Relevant to the CHIME/FRB Project^a

Parameter	Value
Collecting area	8000 m ²
Longitude	119°37'25"25 West
Latitude	49°19'14"52 North
Frequency range	400–800 MHz
Polarization	orthogonal linear
E–W FOV	2°5–1°3
N–S FOV	~110°
Focal ratio, f/D	0.25
Receiver noise temperature	50 K
Number of beams	1024
Beam width (FWHM)	40'–20'
FRB search time resolution	0.983 ms
FRB search frequency resolution	24.4 kHz
Source transit duration	Equator: 10–5 minutes 45°: 14–7 minutes North celestial pole: 24 hr

Note.

^a Where two numbers appear, they refer to the low- and high-frequency edges of the band, respectively.

and minimizing ground noise leaking through to the feeds. The measured surface roughness is ~ 9 mm ($< 2\%$ of the observing wavelength), adequate for operation in the CHIME band.

A schematic diagram of the CHIME telescope signal path is shown in Figure 2 and further explained in the following sections. A total of 256 dual-polarization feeds, spaced by 30 cm, are placed along 80 m of the focal line of each of the four cylinders, giving a total of 2048 signal paths. Digital signal processing of these signals generates the multiple beams that make this telescope a powerful instrument for FRB research (see Section 3.2). The feed design (Deng et al. 2014) achieves nearly equal beamwidths in both polarizations and excellent matching over the octave operating band of the receiver. Mutual coupling between adjacent and nearby feeds was considered in the design of the baluns and matching networks. The entire feed is constructed from printed-circuit materials, appropriate for mass production.

The focal line is deliberately kept simple, housing only feeds and low-noise amplifiers. Two low-noise amplifiers accept the signals from each feed, and those signals are carried through coaxial cables of equal length (50 m) to receiver huts (shipping containers that have been modified with radio frequency shielding and liquid cooling). One receiver hut serves two cylinders and is placed between them. Within each of the two receiver huts, signals from 1024 inputs are further amplified in a stage that includes a bandpass filter and are digitized and split into 1024 frequency channels by the custom F-Engine electronics (see Section 3.1). Temperatures inside the receiver huts are controlled with a liquid-cooled system, but no attempt is made to control the temperature of focal-line components.

3. Upgraded CHIME Correlator

The CHIME correlator was originally designed to handle the 2048 inputs from the CHIME antennas to map the CHIME-visible

sky in redshifted 21 cm emission. The CHIME correlator, of hybrid FX design, uses custom FPGA boards to digitize and channelize the data (the “F-Engine”), while a GPU cluster provides the spatial correlation (the “X-Engine”). The CHIME correlator was designed to be capable of recording visibilities across 400 MHz of bandwidth divided into 1024 frequency channels at ~ 20 s cadence, sufficient for BAO mapping and radio frequency interference (RFI) excision.

FRB detection, however, demands significantly higher time and frequency resolution, as well as spatially localized sky beams, albeit with less stringent calibration requirements. First, typical FRBs have durations of at most a few milliseconds, so FRB surveys require at least comparable time resolution. Additionally, FRBs are dispersed by free electrons along the line of sight; without correction by dedispersion, their signals are rendered undetectable at the CHIME operating frequencies. Dedispersion of channelized intensity data, and hence detection, demands high-frequency resolution but leaves a residual dispersive smearing within each frequency channel. Minimizing this intrachannel smearing requires very narrow frequency channels, especially at CHIME’s low operating frequencies. Figure 3 shows the predicted intrachannel dispersive smearing time versus DM for a variety of FRB surveys. The CHIME/FRB project has opted for 1 ms cadence and 16,384 (hereafter 16k) frequency channels (see Table 1) in order to minimize dispersion smearing in the most relevant part of phase space.

The baseline CHIME correlator required upgrades to provide the independent high-cadence, high frequency resolution, and spatially discrete data streams necessary for FRB detection. The raw data output rate from the CHIME F-Engine is 6.5 Tb s^{-1} , making it difficult to duplicate or distribute beyond the X-Engine. The additional processing must therefore take place inside this system. The next section is a description of the correlator system, with emphasis on the modifications required for FRB detection.

3.1. Digitizer, F-Engine, and Corner-turn

The custom electronics that perform the digitization, frequency channelization, and “corner-turn” (see below) are housed in two 20 ft steel shipping containers outfitted with radio-frequency-shielded enclosures that provide > 100 dB of shielding of the focal line from the high-speed electronics within. These enclosures (labeled the East and West receiver huts) are located midway between the first and second cylinders and the third and fourth cylinders, halfway along their lengths, to minimize and equalize the coaxial cable lengths that connect to the feed lines.

The F-Engine system consists of 128 “ICE” motherboards (Bandura et al. 2016a) housed in eight rack-mounted crates and interconnected with custom high-speed, full-mesh backplanes. The system is described in detail by Bandura et al. (2016b). Briefly, 16 amplified and filtered analog sky signals are digitized in the second Nyquist zone on daughter cards attached to each motherboard at a data rate of 800 MHz with 8-bit accuracy. This information is transmitted to a Field Programmable Gate Array (FPGA) located on each motherboard. The total data rate digitized by the F-Engine is 13.1 Tb s^{-1} for the 2048 time streams. The 400 MHz bandwidth of each time stream is channelized into 1024 frequency bins, each 390 kHz wide, using a polyphase filter bank. A programmable gain and phase offset are applied to each frequency channel, and the data



Figure 1. Photograph of the CHIME telescope on 2016 September 15, looking northwest. The shipping containers housing the X-Engine and CHIME/FRB back end can be seen adjacent to the rightmost cylinder. The receiver huts containing the F-Engine are beneath the reflectors and cannot be seen here. The DRAO Synthesis Telescope (Kotthes et al. 2010) can be seen in the background. See Table 1 for detailed properties of CHIME.

are rounded to $4 + 4$ bit complex numbers, providing a total data rate of 6.5 Tb s^{-1} .

The “corner-turn” modules in the F-Engine reorganize the channelized data from all the motherboards in a crate in order to concentrate the data for a subset of frequencies into a single FPGA. A second corner-turn module reorganizes the data between a pair of crates located in the same enclosure before offloading the data to the X-Engine GPU nodes (see Section 3.2) through 1024 optical high-speed transceivers using standard 10 Gb s^{-1} ethernet protocol. Data from all eight ICE motherboard crates are assembled in each GPU node, forming the last stage of the corner-turn. At this point, data for four frequency bins originating from all 2048 digitizers are assembled in a single GPU node, ready to be spatially correlated.

3.2. X-Engine

The X-Engine consists of 256 processing nodes, each receiving 25.6 Gb s^{-1} of frequency-channelized data on $4 \times 10 \text{ GbE}$ ports. An Intel Xeon E5-2620v3 CPU handles data transfer, using the Intel Data Plane Development Kit (DPDK¹⁹) to reliably achieve high throughput. Each node has 64 GB of DDR4 RAM, sufficient to buffer resampled data for up to 20 s; an upgrade to at least 96 GB (for a 31 s buffer) is planned for the near future (see Section 5.3). The nodes occupy custom 4U rack-mount chassis and have no persistent local storage, but instead, a local set of file servers boot the nodes over the network.

Each node processes four input frequency channels of $\sim 390 \text{ kHz}$ bandwidth each. These are processed by two dual-chip AMD FirePro S9300x2 GPUs, with each chip operating independently to perform all spatial and in-band processing on a single frequency channel. The processed data, including the resampled stream for FRB searching, are exported to the back ends over a pair of GbE links.

The X-Engine is entirely liquid cooled, using direct-to-chip cooling on the CPUs and GPUs and coupling the coolant to the ambient outside air with a $3 \times 120 \text{ mm}$ radiator in the front of each node, without any active chilling. Each rack operates a sealed and independent coolant loop, which is coupled to externally circulating coolant through a CoolIT CHx40 liquid-handling unit. The external coolant exhausts heat to the outside air through a high-capacity dry cooler. The X-Engine is housed in two 40 ft shipping containers directly adjacent to the cylinders (see Figure 1). To limit self-generated interference, a custom-built Faraday cage in each container provides $>100 \text{ dB}$ of shielding from 1 MHz to 10 GHz.

3.3. L0: Beam-forming and Frequency Channelization

In order to take full advantage of the large instantaneous FOV of CHIME while maintaining the full sensitivity, we have developed a hybrid beam-forming pipeline (hereafter referred to as the “Level-0” or L0 process; see Figure 4) to be employed in the X-Engine correlator. Details of this pipeline are described in Ng et al. (2017). In summary, we synthesize 256 formed beams via a fast Fourier transform (FFT) algorithm along the north–south direction. The formalism of FFT beam-forming in the context of a radio interferometer can be found in, e.g., Tegmark & Zaldarriaga (2009) or Masui et al. (2017). We zero-fill the FFT by a factor of two and resample the result to improve spatial alignment of the 256 formed beams across the 400 MHz observing bandwidth. The final N–S beams are tiled across a runtime-defined range of angles (nominally $\sim 110^\circ$), evenly spaced in $\sin \theta$ space, where θ is the zenith angle. In the east–west direction, we form four beams via exact phasing. These give a total of 1024 discrete formed static beams that are closely spaced—with the exact spacing a tunable parameter—and tile the entire primary beam continuously. For beams spaced evenly in $\sin \theta$ from $\theta = -60^\circ$ to $\theta = +60^\circ$, beam centers are separated by 0.4° , to be compared with the beam FWHM of 0.5° at 400 MHz and 0.25° at 800 MHz. Thus, there is significant beam overlap at the lower

¹⁹ dpdk.org

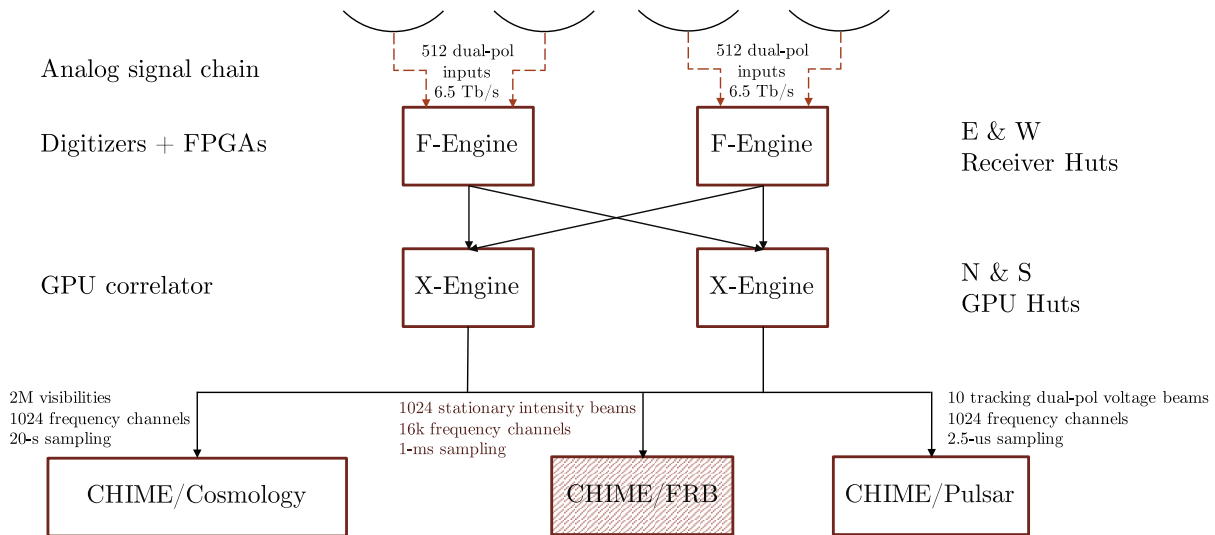


Figure 2. Schematic of the CHIME telescope signal path. The four cylinders (black arcs), the correlator (F- and X-Engines), and the back-end science instruments are shown. The dashed orange segments depict analog-signal-carrying coaxial cables from the 256 feeds on each cylinder to the F-Engines in the corresponding East or West receiver huts beneath the cylinders. The black segments depict digital data carried through copper and fiber cables. Networking devices are not shown. The X-Engine is housed in two shipping containers (labeled North and South) adjacent to the cylinders. The HI intensity mapmaking (CHIME/Cosmology) and CHIME/Pulsar back ends are housed in a shielded room in the DRAO building, and the CHIME/FRB back end (hatched red) is in a third shipping container adjacent to the cylinders (see Figure 1). Note that the total input data rate into the F-Engine is 13 Tb s^{-1} . The data rate into the CHIME/FRB back end is 142 Gb s^{-1} .

frequencies. (For reference, for $\theta = -90^\circ$ to $\theta = +90^\circ$, which is not obviously desirable in spite of the larger FOV because of primary beam fall-off and occultation by nearby mountains, the beam separation would be 0.46° .) Simulations show that this planned beam spacing results in detections in 6–10 beams for bright (several janskys or more) events that are near the center of a beam, and typically just one for faint events. For bright sources in between beams, as many as 12 beams may detect it.

To increase spectral resolution, 128 successive $2.56 \mu\text{s}$ voltage samples are collected and Fourier-transformed. The square of the magnitude of this spectrum is downsampled in frequency by a factor of 8. Three successive downsampled transforms are averaged together, and the two orthogonal polarizations are summed, producing a Stokes I data stream with a cadence of 0.983 ms and a spectral resolution of 24.4 kHz . The output data thus consist of 1024 total intensity beams, with 16k frequency channels at $\sim 1 \text{ ms}$ cadence. These data are scaled, offset, and packed into 8-bit integers for transmission. In principle, it is possible to perform coherent dedispersion to a selected DM on phase data in the X-Engine; however, this feature has yet to be implemented.

4. CHIME/FRB Instrument and Software Pipeline

The CHIME/FRB instrument is the system built to search for FRBs in real time, after receiving the 16k frequency channels at 1 ms cadence for the 1024 CHIME telescope beams from the upgraded X-Engine correlator. The pipeline is split into four further stages or levels, named L1 through L4. Figure 4 schematically represents the different components of the pipeline and the flow of data. Each correlator node calculates the intensities for four frequency channels for all beams, and each L1 node runs FRB searches on the full frequency data for eight independent beams. Thus, the data from the L0 nodes to the L1 nodes are cross-distributed across the clusters via the networking described in Section 4.2. L1 performs per-beam RFI rejection and dedispersion using a highly optimized tree algorithm (K. M. Smith et al. 2018, in

preparation), identifying candidate events in the DM/time plane. L1 processing is performed on each of the 1024 formed beams by a dedicated cluster of 128 compute nodes, and candidate events are consolidated at L2. L2 groups events seen simultaneously in different beams at the same DM and improves localization based on the strength of the signal in multiple beams. L3 classifies the detection and selects among different actions—including alerting the community within a few seconds of the detection—based on source properties. L4 performs the selected actions and hosts a database that stores astrophysical events, including individual pulses from radio pulsars and RRATs, for further offline analysis. The L0 and L1 stages include buffers so that baseband (i.e., voltage) and intensity data, respectively, can be retrieved upon request and further analyzed offline when an event is detected.

Below we provide details of the CHIME/FRB instrument and of each key processing step. The CHIME/FRB hardware is described in Section 4.1, our network in Section 4.2, and L1 through L4 in Sections 4.3 through 4.6. Retrieval of intensity data is described in Section 4.7, and our RFI mitigation strategy is given in Section 4.8.

4.1. CHIME/FRB Instrument

Similar to the X-Engine, the CHIME/FRB instrument consists of a total of 132 compute nodes housed in a customized RF-shielded 40 ft shipping container sitting adjacent to the telescope cylinders (Figure 1). Each node contains two 10-core Intel E5-2630v4 CPUs and 128 GB of RAM in a 4U height (17.8 cm) case. L1 processing, namely, dedispersion (Section 4.3), is the most computationally intensive and runs on 128 of the nodes (see Figure 4). Each L1 node processes eight beams using 16 cores. The remaining four cores in each node are utilized for assembling the incoming packets, processing the outgoing headers, and intensity data retrieval (see Section 4.7). L2 and L3 share a node, and L4 runs on a node of its own. All nodes are network-booted for flexibility to allow dynamic mapping of nodes to

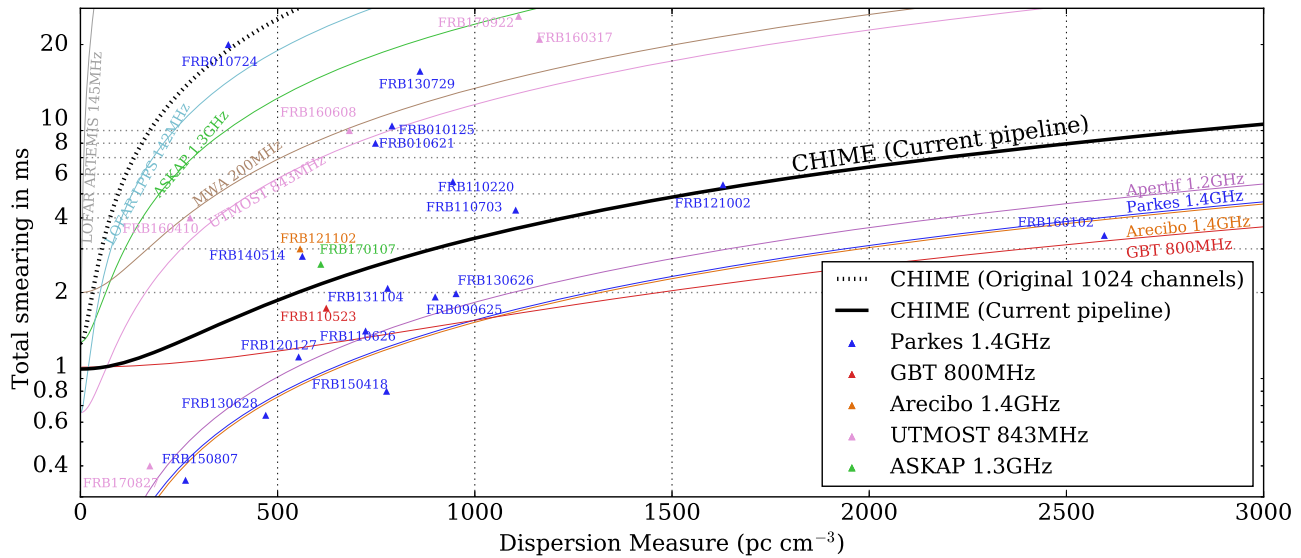


Figure 3. Intrachannel dispersive smearing timescale as a function of DM for a variety of FRB searches, including CHIME/FRB. The black dotted line would be the CHIME/FRB smearing if the native CHIME frequency resolution (1024 channels) were used. Measured widths at 1.4 GHz of detected FRBs (color-coded by detection survey) are indicated. Observed FRBs having widths far below the black dotted line would be artificially broadened to the black dotted line at the native resolution, rendering them difficult if not impossible to detect. To mitigate this problem, rechannelization to 16k frequency channels is performed (see Section 3.3). The smearing for the current pipeline having 16k frequency channels is shown by the solid black line. FRBs with widths that lie above the CHIME/FRB line would be detected by the system with no flux degradation. FRBs with widths below the line may be detected, but they would suffer decreased significance. The colored lines are the intrachannel smearing of other FRB surveys, as labeled.

specific functions. To alleviate single-point failures, the L2/L3 and L4 nodes each have a backup node that will automatically take over processing in the event of a node failure. For on-site data storage, a Storinator S45 Turbo unit (manufactured by 45 Drives Inc.) is fitted with 45 10 TB Western Digital Gold drives. The X-Engine and FRB systems are monitored and controlled in part using the Intelligent Platform Management Interface (IPMI),²⁰ using the Prometheus²¹ and Grafana²² software packages.

The FRB instrument is liquid cooled, using the same direct-to-chip cooling as for the X-Engine (Section 3.2). However, as this system does not have GPUs, there is far less heat loading per node (~ 200 W) compared to the X-Engine (~ 1000 W). A smaller 240×120 mm radiator couples the cooling fluid to forced air circulating in each node. One CoolIT CHx40 liquid-handling unit is used to service two full racks that together contain a total of 20 nodes.

4.2. CHIME/FRB Network

The X-Engine simultaneously transmits data to CHIME/Cosmology data-handling nodes, the FRB search back end, and CHIME/Pulsar. For the FRB search back end, each X-Engine node processes 64 24.4 kHz frequency channels for all 1024 beams, while each L1 node processes all 16k frequencies for eight beams. Thus, each L0 node must send data to each L1 node, which is achieved by employing three layers of network switches, as shown in Figure 5. Groups of 20 L0 nodes are connected, via two 1 Gb s^{-1} links, to a CISCO Catalyst switch (model no. 3650-48TQ-L), which routes the data to a central data switch (CISCO Nexus 3132Q-X) via $4 \times 10 \text{ GbE}$ outputs.

The data are then routed through a second set of Catalyst switches to groups of 10 L1 nodes on four 1 Gb s^{-1} links. In total, this network consists of 21 Catalyst (14 for L0 and 7 for L1) and two Nexus switches. One of the 1 Gb s^{-1} links on L0 is dedicated to transmitting data to the FRB and Pulsar back ends, at a rate of 0.55 and 0.25 Gb s^{-1} , respectively, with the transmission of each data packet timed to avoid data loss of the UDP packets. Each L1 node receives data at a rate of 1.1 Gb s^{-1} for a total data rate for the FRB system of 142 Gb s^{-1} . The CHIME/Cosmology data and FRB baseband data (when requested—see Section 5.3), are transmitted on the second L0 1 Gb s^{-1} link to archive machines. The network traffic is continuously monitored for missing or corrupted data packets so that these effects can be included in the sensitivity calculations for the FRB search (Section 5.2).

4.3. L1: Dedispersion and Candidate Event Identification

The most computationally expensive part of the CHIME/FRB search is the “dedispersion transform,” which converts intensity data from time and frequency into time and DM, allowing efficient detection of dispersed impulse signals (see Figure 4). In our implementation of the dedispersion transform, there are five key FRB parameters: DM and arrival time (each of which ranges over a large number of trial values), spectral index, scattering time, and intrinsic width (which range over a few trial values). Thus, the output of the dedispersion transform is a 5D array of signal-to-noise ratios (S/Ns), to which we apply a tunable threshold, currently set at 10σ given our current experience with fainter pulsar signals (see Section 6), to identify candidate events for processing by subsequent stages of the pipeline.

The computational challenge of the CHIME/FRB search is immense: the input data rate is 1.5 PB day^{-1} , and the dedispersion transform computes 10^{11} S/N values per second (total for all beams). This is orders of magnitude larger than

²⁰ See, e.g., www.intel.com/content/www/us/en/servers/ipmi/ipmi-technical-resources.html.

²¹ prometheus.io

²² grafana.com

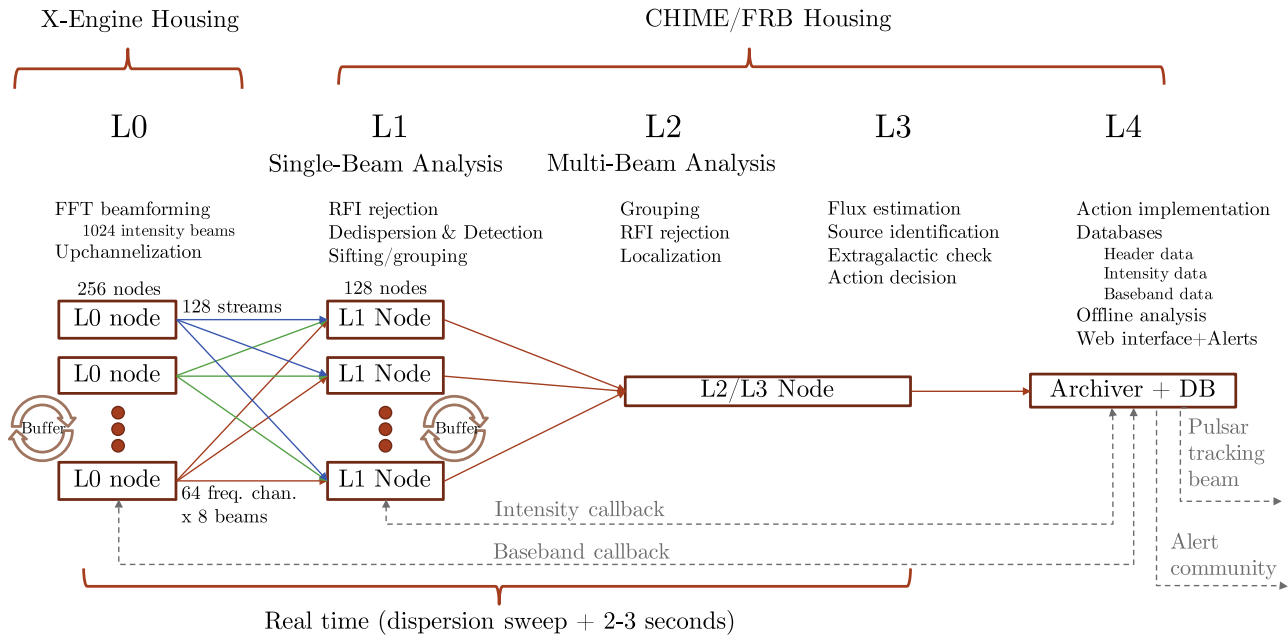


Figure 4. Schematic of the CHIME/FRB software pipeline. L0 performs the beam-forming and up-channelization described in Sections 3.2 and 3.3 with a limited bandwidth in each node. L1 performs RFI rejection and dedispersion described in Section 4.3 using the full bandwidth for a limited number of beams. Portions labeled L2, L3, and L4, described in Section 4.4 through Section 4.6, combine information from multiple beams, classify the detections, provide alerts, and store events. The L1 buffer, available in each node, is for intensity data callback (see Section 4.7), and the L0 buffer is for the baseband callback (see Section 5.3).

any FRB search to date, and comparable to the future search planned for the Square Kilometer Array.²³ For example, the data rate of the 36-beam ASKAP “fly’s-eye” search (Bannister et al. 2017) is 1700 times smaller than for CHIME/FRB.

To meet these challenges, we have developed a new FRB search code “bonsai,” which we will release publicly in the near future. This code will be described in detail in a forthcoming paper (K. M. Smith et al. 2018, in preparation). Here we briefly summarize its main attributes.

There are two general approaches to dedispersion: direct algorithms (see, e.g., Manchester & Taylor 1977) and tree algorithms (Taylor 1974; Zackay & Ofek 2017). The computational cost of tree dedispersion is $\mathcal{O}(TF \log F)$, versus $\mathcal{O}(TF^2)$ for direct dedispersion, where T and F are the number of time and frequency samples, respectively. Thus, tree dedispersion is parametrically faster but suffers from two problems in practice. First, tree dedispersion tends to perform badly owing to memory bandwidth bottlenecks. In a straightforward implementation, a large memory region (larger than the Level 3 cache available in the CPU of our nodes) must be processed in each of $\log_2(F)$ iterations of the algorithm. The cost of these memory transactions ends up being much larger than the computational cost. Second, in some implementations, tree dedispersion may make approximations to the ν^{-2} dispersion delay, which results in loss of S/N.

Our implementation of tree dedispersion, *bonsai*, includes two new features that solve these problems. To address memory bandwidth bottlenecks, we have implemented a version of tree dedispersion that is “blocked,” in the sense that the data are processed in a small number of passes, in blocks having size tuned to that of the CPU cache. This minimizes the number of times the data go through the CPU Level 3 cache, resulting in a factor of ~ 30 speed-up on multicore machines. An interesting property of the blocked tree

algorithm is that it is also incremental: given sufficiently large S/N, the search triggers within a few seconds, even for an FRB whose dispersion delay in the CHIME band is much larger. This low latency is critical for the triggered baseband recording system (Section 5.3), since the X-engine can currently only buffer data for 20 s, and this window must include both the dedispersion delay of the FRB and the latency of the search.

Second, we have made tree dedispersion close to statistically optimal through appropriate choices of internal weightings. For CHIME/FRB, we choose parameters that are a compromise between computational cost and statistical optimality. We search to a maximum DM of $13,000 \text{ pc cm}^{-3}$ and a maximum pulse width of 100 ms, with two trial spectral indices, -3 and $+3$, to provide added sensitivity to bursts with nonuniform spectra. Though DM $13,000 \text{ pc cm}^{-3}$ is well beyond the maximum DM yet observed, it requires little extra computational time to reach this value over, for example, $3,000 \text{ pc cm}^{-3}$, near the maximum yet seen (Bhandari et al. 2018). After initial RFI removal (see Section 4.8), the search consists of multiple dedispersion trees with different levels of time downsampling, to cover different parts of the (DM, pulse width) parameter space. The search is $>80\%$ optimal over most of this parameter space, occupying $0.75 \text{ cores beam}^{-1}$, and a memory footprint of 7 GB beam^{-1} . This more than suffices to meet the initial science goals of the CHIME/FRB search.

The *bonsai* code produces S/N estimates for a grid of DMs and times for a discrete set of spectral indices, scattering times, and intrinsic width parameters. As a detail, the output array of S/N estimates is “coarse-grained”: we divide the grid of trial arrival times and DMs into coarse cells and take the maximum of all S/N values over each cell. We set a (tunable) threshold at 10σ and isolate local maxima within a neighborhood of size $\Delta \text{DM} \approx 10 \text{ pc cm}^{-3}$ and $\Delta t \approx 0.25 \text{ s}$. Once we have identified the peak S/N parameters, we generate a lightweight description of the coarse-grained candidate event, the “L1 Header.” This includes the DM, time, and coordinates

²³ www.skatelescope.org

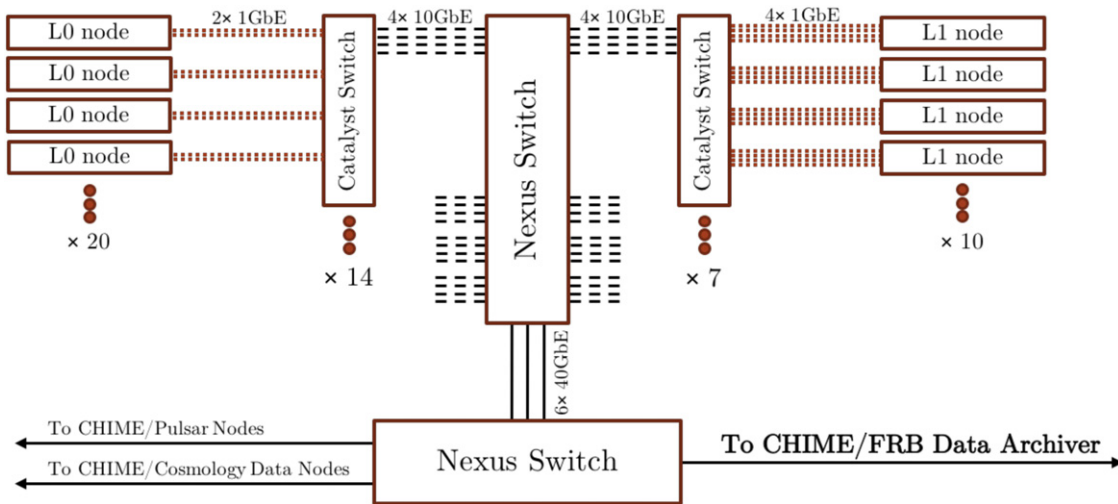


Figure 5. Schematic of the CHIME/FRB data network. Each L0 node is connected via two 1 GbE connections (dotted orange lines) to a Catalyst switch, which in turn is connected to the backbone Nexus switch via four 10 GbE links (dashed black lines). Similarly, each L1 node is connected via four 1 GbE connections to the Catalyst switches, which connect to the Nexus switch. The two backbone Nexus switches are connected via six 40 GbE links (solid black lines). CHIME/Pulsar computation nodes, CHIME/Cosmology data storage nodes, and CHIME/FRB data archivers are mounted on the network via the Nexus switches.

of the trigger; an RFI rating (see Section 4.8); and S/N values for slices along each parameter dimension (holding the remaining parameters fixed at their peak values).

To validate the accuracy and sensitivity of our candidate identification, we have performed injection analyses and processed CHIME Pathfinder (Bandura et al. 2014) acquisitions containing pulsations from PSR B0329+54 and PSR B0531+21. Compared to a pipeline based on RRATTrap (Karako-Argaman et al. 2015) and PRESTO²⁴ single-pulse search (e.g., Chawla et al. 2017), we find in initial analyses that the CHIME/FRB L1 pipeline has a lower rate of RFI false positives and is more effective at finding faint pulses (see Section 6).

4.4. L2: Grouping of Multibeam Events

Per-beam event detection reports from the L1 *bonsai* dedispersion analysis (see Section 4.3) are streamed to a single endpoint where they are buffered. These reports consist of L1 headers or null results. Once all beams have been reported for a given data block, the existence of multibeam events is investigated by grouping candidates in time, DM, and sky position. The cluster detection is performed with a simplified implementation of the DBSCAN²⁵ algorithm (Ester et al. 1996), which has $\mathcal{O}(n \log n)$ complexity, where n is the number of events above our threshold and is capable of handling large event rates without issue. For any event in a group, there exists another whose differences in time, DM, and sky position are all below dimension-specific thresholds. For position, the threshold ensures that grouped events are spatially connected through beam adjacency. The DM and time thresholds reflect the uncertainties built into the coarse-grained event properties (see Section 4.3). After cluster detection, the event is classified as either RFI (see Section 4.8) or astrophysical, and the L2 header is formed, containing one or more L1 headers from the grouping. RFI events are sent straight to the L4 database

(Section 4.6), while astrophysical events are passed to L3, after position refinement and event flux estimation, described next.

Position refinement is performed by comparing the detected S/Ns from each beam to what is expected from a frequency-dependent beam model. For multibeam events, the detected S/Ns are compared to those in a precomputed look-up table containing relative S/Ns for a grid of sky positions and spectral indexes to form a refined estimate of these parameters. χ^2 minimization is then used to refine the initial guess and estimate the uncertainty region. For single-beam events, a precomputed mapping between event S/N and the location and uncertainty region is used. The mapping is constructed from the distribution of simulated events drawn from the beam model that produce nondetections in adjacent beams. For very bright ($S/N \sim 1000$) events having spectra well described by a power law and that hit randomly with respect to beam centers, simulations of our real-time pipeline show that positions can in principle be determined to ~ 0.3 , and spectral indexes to within ± 0.5 .

With a refined position and estimate of the on-axis S/N, radio fluxes of events are then estimated in real time. The estimate is based on the radiometer equation for single pulses (see, e.g., Cordes & McLaughlin 2003). Besides instrument and pipeline-specific parameters, we use the S/N and pulse width as estimated by *bonsai* as input. The S/N is corrected for the beam sensitivity at the location of the event, and the receiver bandwidth is corrected for the masked fraction from RFI excision. The sky temperature at the location of the event is interpolated from the reprocessed Haslam 408 MHz sky temperature map (Remazeilles et al. 2015) and scaled to our survey central frequency (600 MHz) using a power-law index of -2.6 . The intrinsic pulse width is estimated by correcting for the sampling time, intrachannel smearing, and scattering, in quadrature. Intergalactic scattering times are estimated as in Section 3.11 of Yao et al. (2017), where we estimate the Galactic scattering contribution using the estimated Galactic DM contribution and the empirical DM–scattering relation from Krishnakumar et al. (2015), ignoring potential scattering in the host galaxy of the source.

²⁴ github.com/scottransom/presto

²⁵ Density-based spatial clustering of applications with noise.

4.5. L3: Identification of Extragalactic Events and Action Determination

Astrophysical events can be of Galactic or extragalactic origin and can come from either a known or an unknown source. The processing-related actions triggered for an event depend on its identification, and, as detailed below, the pipeline makes this decision in L3 based on the determined DM and localized position. By automatically recognizing events that come from known sources, we can provide alerts (see Section 5.1) only for FRBs, and not for Galactic objects. However, all astrophysical L1 events are stored, and additional data products for non-FRB events can also be stored for selected interesting sources.

We maintain a database of known sources that initially consists of all known pulsars, RRATs, and FRBs and their relevant parameters recorded in the ATNF pulsar catalog²⁶ (Manchester et al. 2005), the “RRATalog,”²⁷ and the FRB catalog²⁸ (Petroff et al. 2016). As time progresses, discoveries by CHIME/FRB and other FRB searches will be added to the database. For each astrophysical event, we compare the sky position and DM, weighted by the measurement error, to the same parameters of the sources in the database. The likelihood ratio (i.e., the Bayes factor) of association with neighboring known sources is calculated. A threshold for association is determined based on the Receiver Operating Characteristic (ROC) curve (the true-positive vs. the false-positive rate) for a set of simulated events. Currently, only the sky location and DM of an event are being compared with those of the known sources, but the framework allows for future additions (such as a comparison of the periodicity or the flux).

For events that cannot be associated with a known source, the predicted maximum Galactic DM along the line of sight is calculated using models described by Cordes & Lazio (2001) and Yao et al. (2017) of the Galactic free electron density. The difference in predicted maximum Galactic DM between the two models is taken as a systematic uncertainty; this is then added in quadrature with the L1-estimated DM uncertainty to yield a full uncertainty, σ . We impose the condition that a σ less than a minimum value (currently set to 20%) is set to this minimum, to avoid underestimating the true uncertainty when the two models agree coincidentally. In the real-time pipeline, a source is deemed extragalactic if its measured DM exceeds the maximum Galactic DM predicted by *both* models by at least 5σ . Sources with a measured DM only 2σ – 5σ in excess of the predicted maximum Galactic DM fall in an ambiguous class of sources. Though otherwise not treated differently, this class is invoked to highlight sources that may be Galactic, warranting a multiwavelength study to identify possible Galactic foregrounds such as H II regions. Alternatively, they may be extragalactic, but relatively nearby, and hence potentially good targets for multiwavelength counterpart identification.

Finally, a set of predefined rules determine which actions to trigger for each detection. Possible actions are to ignore an event, store the event header in the events database (the default action), call back for buffered intensity data from L1 (see Section 4.7), dump buffered baseband data from L0 (see Section 5.3), ask the CHIME/Pulsar observing system to track the source position in search mode, or send out an alert

(Section 5.1). The rules can be set for individual sources or for groups of sources, with all rules being stored in a configuration database that is accessible through a web interface. Some examples of action rules are the following: “if the flux of a binary millisecond pulsar pulse is 100 times higher than its nominal flux, store the event header and trigger an intensity callback,” and “if the source is unknown and the event extragalactic, store the event header, trigger an intensity callback and a baseband dump, and alert the community.”

4.6. L4: Action Implementation and Event Archiving

L4 first implements the actions selected in L3 for each event, such as requests to L1 for intensity data callbacks, requests for monitoring with CHIME/Pulsar, or sending out real-time alerts to the community for FRB detections. L4 is also host to the CHIME/FRB archive, which saves the header information and its associated analysis products in a relational database for every event sent past the L1 stage regardless of whether it is classified as astrophysical or RFI in the L2 stage. Each event header that arrives at the L4 stage of the pipeline is assigned a unique event number to identify the event and track all data products associated with it.

The size and quantity of data products stored for each event depend on the actions and priorities determined at the previous stages of the pipeline. Due to the current limited bandwidth out of the observatory, a local archiver with 450 TB of disk space has been installed. The data will be regularly transferred to a long-term offsite staging facility. In order to manage file transfers between different locations, we employ a custom file management software *alpenhorn*²⁹ (Hincks et al. 2015). *Alpenhorn* ensures reliable data transfer between multiple originating, intermediate, and archival storage sites. Among its features are keeping at least two copies of the data at different sites at all times, verifying data integrity through checksums, and handling data transfer among sites. The FRB pipeline database is periodically updated with the current status of the data files tracked by *alpenhorn*, so that a user accessing the live website has access to their location.

The database is integrated into a web display that allows for searching and filtering of events, displaying the best data measurements, interacting with data plots, tracing the location, and requesting downloads of raw data. The web display also incorporates user accounts, allowing users to rank, classify, and comment on candidate events. Users can also update action criteria and known source properties.

4.7. Intensity Data Callback

Since the data rate of CHIME/FRB is very high, we cannot save to disk all the intensity data taken by the telescope. Instead, each L1 node stores a buffer of recent data that can be retrieved if an interesting event is detected by L3 (see Section 4.5). We save data from the beams in which the event is detected independently, as well as the immediately adjacent beams.

CHIME/FRB searches up to a DM of $13,000 \text{ pc cm}^{-3}$. At this maximal DM, a pulse can take up to $\sim 250 \text{ s}$ to sweep through the CHIME frequency band. The L1 nodes do not have enough memory to store this much intensity data at full resolution, so we implemented a telescoping ring buffer: older

²⁶ www.atnf.csiro.au/people/pulsar/psrcat/

²⁷ astro.phys.wvu.edu/rratalog/

²⁸ www.frbcat.org/

²⁹ github.com/radiocosmology/alpenhorn

data are progressively downsampled and stored with coarser time resolution to optimize memory usage. Initially, we keep 60 s of the most recent data at full resolution, 120 s binned by a factor of two, and 240 s binned by a factor of four.

As the L1 nodes do not have local storage, called-back intensity data are written to a network-shared archiver (see Section 4.6), where they are available for additional offline analysis and visualization. Apart from generating the dynamic spectra (“waterfall plots”) for visualization of single-pulse events, we refine the event parameters (DM, arrival time, pulse width) determined by the L1 pipeline. We perform an improved localization using the intensity data from neighboring beams (considering detections as well as nondetections) and obtain more accurate values for spectral index and flux.

4.8. RFI Excision

The CHIME/FRB software pipeline mitigates RFI using different criteria at different stages in order to minimize the misclassification of terrestrial signals as astrophysical transients.

In L1, we remove RFI signals from the intensity data by iteratively detrending and subtracting a time series obtained by assuming zero DM (Eatough et al. 2009) from the intensity data, while clipping amplitude and standard-deviation outliers. We have developed fast C++/assembly kernels that form a sequence of “transforms” at a high level. Each transform operates on arrays of intensity data with a parallel array of weights (where masked intensities are represented as zeros). The output of each transform is the input to the next. We have utilized low-level ring buffering and transparent resizing and resampling of the data, since the transforms need not use the same block size or downsampling.

Our current default RFI removal scheme consists of a sequence of nine clipping transforms (five with 3σ threshold and four with 5σ threshold), which are iterated six times. The sequence is repeated twice; after each iteration, the intensity is detrended by two polynomial (degree 4) and spline (equivalent degree 12) transforms, one along the time axis and one along frequency, respectively. We have found empirically that this transform chain removes RFI in data we have captured so far (e.g., from the CHIME Pathfinder, the Galt 26 m telescope, or with a CHIME incoherent beam—see Section 6) to a sufficient level that simulated FRBs near detection threshold dominate the output of the dedispersion transform. Our real-time detrending and RFI excision pipeline is currently in a proof-of-concept stage, and we are still experimenting with the implementation.

After candidate events have been identified (see Section 4.3), and before they are forwarded to L2 (see Section 4.4), a second round of RFI excision is performed by subjecting the S/N behavior surrounding significant detections to a machine-learning classifier—in this case, a support vector machine (Cortes & Vapnik 1995). Currently, we include two features with the goal of capturing both local and global S/N behaviors. For the former, we use the S/N fall-off at nearby but suboptimal DMs. For the latter, we consider the distribution of other above-threshold locations in the DM–time plane, noting that RFI tends to manifest as vertical streaks of events covering a large range of DMs. As our chosen machine-learning method requires supervision, we have developed tools to allow for efficient event labeling; training the classifier and

refining feature definitions will be an ongoing process throughout commissioning.

The candidate events detected in single beams are transmitted from all L1 nodes to the L2 node. At this stage we group events detected in spatially neighboring beams and within a DM and time threshold. We further classify these events as astrophysical or RFI based on the distribution of S/Ns in neighboring beams and the shape of the grouped beams. Astrophysical events are in the far field of the telescope ($\gtrsim 25$ –50 km for CHIME’s operating frequency band) and are expected to have a sharply focused intensity distribution pattern, whereas RFI events are typically in the near field and likely detected in the horizon-directed sidelobes, causing a significantly broader intensity distribution pattern.

In L2, we use a machine-learning classifier, implemented using `scikit-learn`³⁰ (Pedregosa et al. 2011), to calculate the probability of an event being astrophysical or RFI. Ultimately, multiple different classifiers can be loaded, and their calculated probabilities are assumed to be independent by the framework and multiplied together. Currently, the classifier is based on a stochastic gradient descent classifier with a linear combination of the following features: (1) the pattern of beams in a 3×3 grid in which the event was detected above threshold with respect to the beam with the highest S/N, (2) the ratio of the second-highest S/N to the highest S/N in the group, (3) the ratio of the average S/N of the detected beams to the highest S/N in the group, (4) the total number of beams where the event was detected, and (5) the highest L1 grade associated with the group from the L1 classifier. For events detected in single beams, no extra information is available, so the grade from the L1 classifier is the only value used for classification. In the future, we may implement features based on (1) the total L1 event activity (number of candidates detected in the same time window) at the time of event detection, (2) ratio of the FFT-formed beam intensity to that of an incoherent beam formed from all CHIME feeds, and (3) location of the beams on the sky (as RFI seems more likely to be detected at lower elevation beams).

As with L1, we are developing event labeling and generation of training sets for the classifier. At the L2 stage, all L1 candidates that are sent have their `bonsai`-determined information (S/N, DM, width, beam coordinates, etc.) stored in the database irrespective of their RFI/astrophysical classification. This allows us to reclassify and retrain the classifiers as required in the future.

5. Planned CHIME/FRB Features

Here we describe features of the CHIME/FRB analysis pipeline that either are currently being developed or are still in a design phase.

5.1. Real-time Alert System

Rapid reporting of FRB discoveries is important for coordination of detections with FRB searches at other radio bands, for finding repeat bursts from FRBs, and for discovering or ruling out possible electromagnetic or gravitational wave counterparts. Specifically for CHIME’s FOV, facilities such as the Owens Valley Long Wavelength Array (23–88 MHz; G. Hallinan et al. 2018, in preparation) or the DSA-10

³⁰ www.scikit-learn.org

(1.4 GHz; V. Ravi et al. 2018, in preparation) could provide constraints on the spectral indices, spectral ranges, absorption, and scattering properties of FRBs. The non-Poissonian clustering of FRB 121102 (Scholz et al. 2016) makes the detection of repeat bursts more likely after the initial detection. If this characteristic holds for other FRBs as well, the monitoring of CHIME/FRB-discovered sources with tracking telescopes would be an important test of repetition, as well as an opportunity for interferometric localization (e.g., with the Jansky Very Large Array³¹ in New Mexico as in Chatterjee et al. 2017).

The CHIME/FRB real-time alert system will follow the VOEvent implementation suggested by Petroff et al. (2017) to be compatible with the other FRB search efforts around the world. After a period of human confirmation at the beginning of the CHIME/FRB program, we plan to announce every detected FRB along with the significance of the detection and other relevant characteristics in real time. The latency of the CHIME/FRB real-time pipeline from the arrival of the lowest frequency signal is $\sim 2\text{--}3$ s. FRBs for which the intensity data are stored and reanalyzed (see Section 4.7), or for which there has been a baseband dump and subsequent improved parameter determination (see Section 5.3), will have updates to their reported characteristics announced as soon as possible.

5.2. Calibration, Sensitivity, and Completeness Measurement

CHIME/FRB is a complex instrument with multiple components that must be regularly calibrated and monitored to reach its science goals. Here we describe the three planned stages of the process: phase calibration of CHIME, monitoring of the pipeline performance to track CHIME/FRB's exposure to the sky, and measurement of the sensitivity and completeness of the CHIME/FRB pipeline.

In order to form phased beams, each RF input signal must be phase-calibrated regularly. While ultimately milliradian phase calibration will be available for CHIME in order to meet the stringent requirements of the cosmology experiment, this calibration accuracy will not be available at the commencement of FRB searching. For the CHIME/FRB system, we will determine the complex gains in a near real-time system using transits of bright effective point sources like Cassiopeia A and Cygnus A (Trotter et al. 2017), approximately once per day. We will observe such a calibration source with a full set of N ($N - 1$)/2 visibilities, where N is the number of feeds, using the CHIME cosmology X-Engine at a full complement of radio frequencies. We will then compensate for the delay expected from the physical locations of the antennas and the calibrator sky location. Any observed residual delay in the response to the calibrator will be instrumental, which we record and compensate for in the FRB beam-former in L0. The observations of the calibrator sources will also be used to calibrate the CHIME bandpass.

Most scientific goals of CHIME/FRB—measurements of the FRB rate and of property distributions—require a thorough understanding of CHIME/FRB's detection sensitivity as a function of time and sky location (i.e., the exposure of the survey) and the completeness of its survey as a function of burst fluence, burst width, DM, etc. We take two approaches that allow us to cross-check our results. To determine the exposure and the sensitivity as a function of time, we have to

monitor metrics from the data flow and compute a sensitivity as a function of time and beam location using the radiometer equation (McLaughlin & Cordes 2003). However, the multiple complex stages of RFI filtering and excision make it challenging to understand and study the completeness, especially in the presence of rapidly changing RFI that is correlated across phased beams. Hence, we also plan to create a real-time multibeam FRB injection and detection system to estimate the survey completeness in real time.

To account for the total sensitivity and exposure of CHIME/FRB, we have to keep track of the sensitivity of each coherent beam and the number of beams that are being searched for FRBs at a given time. At the L0 stage, CHIME/FRB will monitor which feeds from the telescope are usable (informing beam shape and sensitivity) and which L0 nodes are active (informing the bandwidth and usable frequency channels). At the L1 stage, we monitor the UDP packet loss from L0 nodes to the L1 nodes (again informing usable bandwidth and frequencies), the fraction of data masked by the first-stage RFI removal algorithm, and the number of L1 nodes that are actively searching for FRBs, informing the sensitivity of each beam on the sky. At the L2/L3 stage, the number and identity of the reporting L1 nodes and the delay in TCP packets are tracked.

Most of these metrics will be reported on a few-minute timescale and stored in a time-tagged database. A database scraping script will then collect the metrics and calculate the sensitivity and exposure of CHIME/FRB to the sky.

In order to test the response of our RFI excision and filtering, we plan to inject FRB signals of known properties, accounting for the spatial and spectral responses of the coherent beams, into a small number of the beams and verify that we recover them as expected. Figure 6 shows a schematic of the injection pipeline that will run in parallel to the main CHIME/FRB search pipeline. A grid of 4×4 coherent beams will be chosen randomly among the 1024 beams to sample different north-south locations in the primary beam. An injection server will generate a set of FRB parameters (location, flux, DM, width, scattering, and spectral index) from a prior distribution. The effective signal strength in each beam and frequency band will then be computed based on the coherent beam model. A simulator code will inject the signal into the intensity data from the corresponding beams. The rest of the data analysis will proceed identically to the CHIME/FRB search pipeline. The injected parameters and the recovered parameters will be stored in a temporary database where the detection completeness of the pipeline as a function of FRB parameters can be analyzed. A detailed paper reporting the measured sensitivity of the CHIME/FRB search pipeline will be forthcoming once commissioning (Section 6) is complete.

5.3. Planned Triggered Baseband Recording System

We plan to have triggered recordings of baseband data containing FRBs. This system will provide CHIME/FRB with sensitivity to polarization, greatly improved ability to localize sources, and spectral and time resolution constrained only by the uncertainty principle $\Delta f \Delta t \geq 1$, and it will open possibilities to perform very longbaseline interferometry (VLBI) of FRBs with other telescopes. This baseband system is currently under development and will be commissioned once CHIME/FRB science operations are underway. Here we provide our current plans for this system.

³¹ www.vla.nrao.edu/

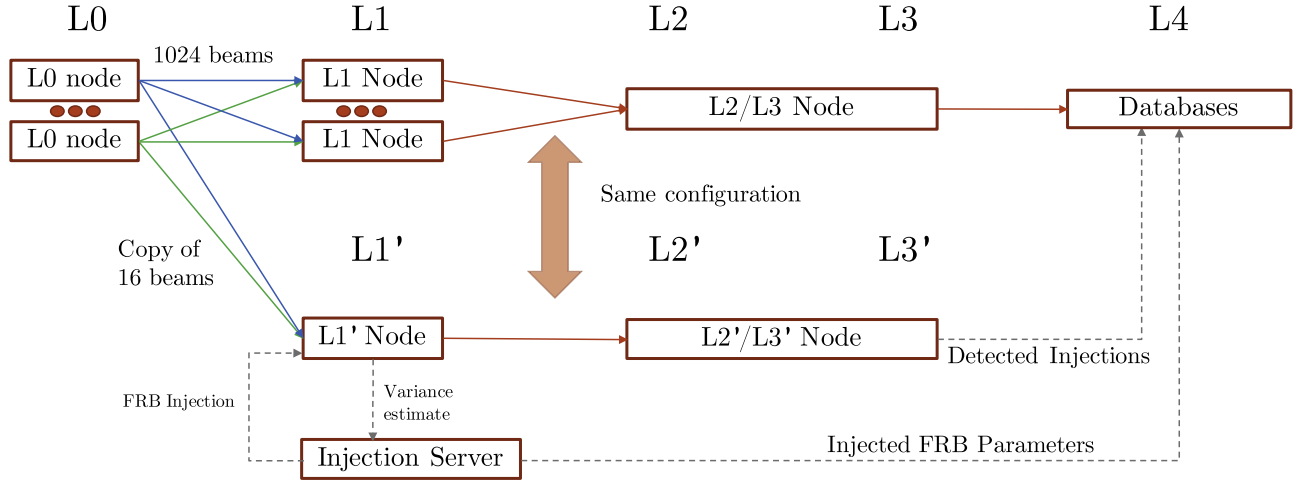


Figure 6. Schematic of the planned CHIME/FRB injection pipeline. A fraction of the CHIME/FRB beams (typically a 4×4 grid) is copied to a separate L1 node where FRB signals with a large range of fluence, DM, width, and scattering (convolved with modeled beam responses) are injected into the beams. The processing of the injected data will be identical to that in the main CHIME/FRB pipeline, and the detections and nondetections of the injected FRB signals will be tracked in a database.

Baseband data from all 2048 correlator inputs that have been spectrally channelized by the F-Engine will be buffered in the X-Engine prior to beam-forming. These data will be written to disk on an archiver system upon receipt of a trigger from the FRB search pipeline (see Section 4.5). Ultimately, this system will allot approximately 90 GB of random access memory in each X-Engine node, providing a 31 s ring buffer—long enough for the few seconds of latency in the FRB search pipeline, as well as the dispersion delay across the band for DMs up to $\sim 1500 \text{ pc cm}^{-3}$. To economize network resources and disk space, only the 100 ms of data surrounding the burst will normally be read out and stored to disk, for each frequency channel. These data will be analyzed offline in the initial implementations of the system. The buffer system will be designed so that upon receipt of a trigger only a small block of memory containing the desired 100 ms of data will be frozen, with the rest of the memory continuing to participate in the ring buffer. As such, the system will have no down time during readout of baseband data, only a slight shortening of the buffer.

Among the most compelling applications for the baseband system is improved localization precision of FRB sources. Rather than sources being localized to a formed beam width of $\sim \lambda/D$ (where λ is the observing wavelength and D is the telescope extent, notably different in the north/south versus east/west direction), the baseband system permits interferometric localizations with precision (Masui et al. 2017)

$$\Delta\theta = \frac{\sqrt{6}}{2\pi} \frac{\lambda}{D} \frac{1}{S/N} \text{ rad.} \quad (1)$$

Here S/N is the S/N of the burst as detected in the baseband data, and the geometric factor of $\sqrt{6}$ accounts for CHIME’s baseline distribution. This will amount to arcminute uncertainties for most bursts.

The baseband system, through the use of coherent dedispersion (Hankins & Rickett 1975), will also permit the identification of structure in the FRB signals at time resolution well under the 1 ms provided to the CHIME/FRB search engine. Any such structure will constrain FRB emission physics. Similarly, the baseband data will allow the detailed investigation of FRB scattering and scintillation timescales. These

properties can be used to constrain the astrophysical environments in which FRBs occur (e.g., Masui et al. 2015).

Finally, baseband data will contain full Stokes information, allowing us to determine the degree to which each FRB is polarized (Masui et al. 2015; Petroff et al. 2015a; Ravi et al. 2016; Michilli et al. 2018). In cases where an FRB has a linearly polarized component, the dependence of the polarization angle on observing frequency can then be used to calculate the Faraday rotation measure (RM). The RM corresponds to the integral of the electron-density-weighted line-of-sight magnetic field between the observer and the FRB and thus is a powerful probe of an FRB’s immediate environment, host galaxy, and the intervening intergalactic medium (Petroff et al. 2015c; Akahori et al. 2016; Ravi et al. 2016; Michilli et al. 2018).

6. CHIME/FRB Commissioning

The CHIME/FRB pipeline development has followed a phased approach. First releases of software were developed on a two-node L1 system at McGill University in early 2017. A 10-node system was constructed at University of British Columbia (UBC) to allow for testing before this system was moved to the CHIME site in late 2017. The deployment of the 10-node system has allowed initial integration with the CHIME X-Engine and L0 system, as well as testing of the backbone network and pipeline configurations. Initial testing of the beam-forming and phase calibration is underway with the CHIME telescope. Observations with an incoherent beam (summing the intensities of all feeds without consideration of their phases) for limited times have been done. Single pulses from multiple different radio pulsars have been detected at approximately the expected signal strength, noting the greatly reduced sensitivity compared with that expected for coherently formed beams. The left panel of Figure 7 shows a synopsis of observations from 2018 January 28 showing the transits of PSR B0329+54 and the Crab pulsar (see figure caption for details), and the right panel shows a dynamic spectrum/waterfall plot for one bright pulse from PSR B0329+54. The detection of single pulses from several radio pulsars with the 10-node system has allowed an initial commissioning with 64 beams on the sky, as well as verification of basic system operations, including time-tagging, event clustering and identification, and archiving. The full

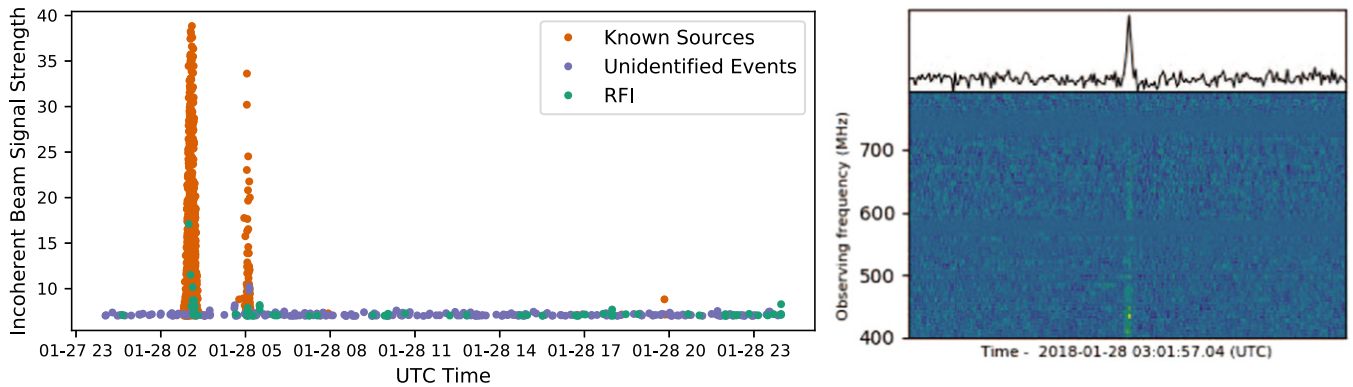


Figure 7. CHIME/FRB incoherent beam commissioning and test observations of transits of PSR B0329+54 and the Crab pulsar (orange circles), obtained on 2018 January 28 and 29. Left: approximate incoherent beam signal strength (greatly reduced compared to the expected coherent beam signal strength once available) vs. time over 24 hr, showing one transit of both pulsars (PSR B0329+54 on the left and the Crab pulsar to its right). The x -axis label denotes UTC date and time. Right: waterfall plot of one bright single pulse from PSR B0329+54 observed with the CHIME/FRB 10-node system on 2018 January 28.

system node-level components are being integrated at McGill University in early 2018 prior to deployment at the CHIME site in spring 2018. Science operations are expected to commence in the latter half of 2018.

7. Predicted CHIME FRB Event Rate and Envisioned FRB Science

FRBs have been detected mostly at 1.4 GHz (e.g., Thornton et al. 2013; Spitler et al. 2014), with some near 800 MHz (Masui et al. 2015; Caleb et al. 2017); however, thus far none have been seen below 400 MHz in spite of significant exposures with large telescopes at those frequencies (e.g., Karastergiou et al. 2015; Rowlinson et al. 2016; Chawla et al. 2017). The reasons for the apparent low-frequency dearth are not yet understood, which makes CHIME’s band particularly important for making progress on FRBs. Amiri et al. (2017) reported on an FRB search in the 400–800 MHz band, but this was using the CHIME Pathfinder telescope in an incoherent mode, which would have been able to detect only extremely bright FRBs. CHIME/FRB will have several orders of magnitude more sensitivity than that study.

There are three main reasons why the FRB rate in the CHIME band may be significantly lower than at higher radio frequencies. The first is that multipath scattering of radio waves by inhomogeneities in interstellar, or in principle intergalactic, plasma causes pulses to be broadened significantly. This is strongly dependent on radio frequency ν , with broadening time varying as $\nu^{-4.4}$ (e.g., Rickett 1977). It is possible that some FRBs are broadened beyond detectability in the CHIME band, particularly at the lower end. However, many observed 1.4 GHz FRB widths are below 1 ms (see Figure 3) and still are limited by intrachannel dispersion smearing, suggesting that scatter broadening may not be a hindrance to detection in the CHIME band, especially given that we plan to search to widths of up to 100 ms (see Section 4.3). If scattering is significant in the CHIME range, CHIME/FRB will see clear evolution of the pulse morphology over the band, which spans a factor of $2^{4.4} \simeq 21$ in broadening time. The detection of significant scattering beyond that expected from the Milky Way alone in a large fraction of FRBs would argue strongly that the population is found in disks of spiral galaxies, or special regions of high plasma turbulence such as near supermassive black holes, or in supernova remnants.

Another possible reason for a reduced FRB rate in the CHIME band is free–free absorption. This effect is important in dense, ionized plasma where some models, notably those involving young neutron stars in supernova remnants (e.g., Piro 2016), have proposed that FRBs are located. The free–free optical depth varies as $\tau_{ff} \propto EM \nu^{-2.1}$, where EM is the emission measure $\int n_e^2 dl$, with n_e the electron density. Since the absorbed radio flux has intensity that varies as $e^{-\tau_{ff}}$, even a relatively small τ_{ff} can lead to fading toward the lower end of the spectrum, and hence a steep inverted spectrum (see Kulkarni et al. 2014, for a discussion). This would manifest itself as a rapid fading in the CHIME band, with width otherwise constant, in contrast to pure scattering, for which fluence remains constant but pulse width increases. However, there are no hints of absorption in the vast majority of FRB bursts at 1.4 GHz. Though FRB 121102 has shown steep inverted spectra, these have been established as being intrinsic to the source, whose spectrum is highly variable, with steeply rising spectra observed as commonly as falling ones (Scholz et al. 2016; Spitler et al. 2016). The detection of significant free–free absorption in many FRB spectra would indicate that the population inhabits regions of high plasma densities, strongly suggesting a very young population.

Finally, on average, FRB spectra may be intrinsically fainter at lower radio frequencies. Without a detailed source emission model, the likelihood of this cannot be determined. This possibility can be distinguished from the free–free absorption case by CHIME/FRB observations, since the latter should follow strictly the predicted free–free frequency dependence and may also show a correlation with DM after correcting for the Milky Way and intergalactic contributions. There is no reason to think that an intrinsic spectral change should have the same frequency dependence as free–free absorption, or be correlated with source environment.

Chawla et al. (2017) have done simulations of FRB populations subject to various assumptions regarding intrinsic spectral index, free–free absorption, and scattering measure distributions. For example, they have predicted CHIME/FRB event rates of order 10 day^{-1} for a Euclidean flux distribution and a scattering measure distribution like that observed in the Milky Way, with a limiting intrinsic spectral index constrained by the absence of detections at 350 MHz in the GBNCC survey (see Chawla et al. 2017, for details). The uncertainty is an order of magnitude in the 700–800 MHz band and approximately two

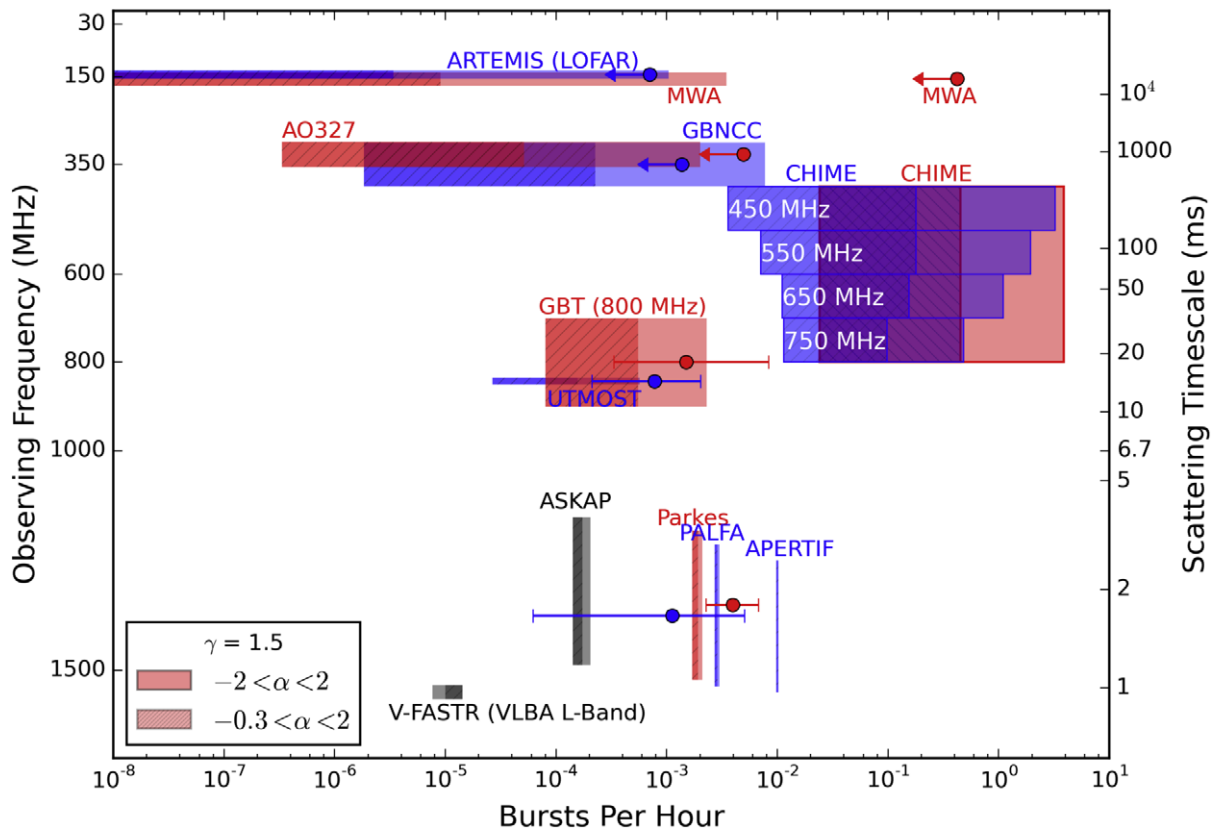


Figure 8. Estimated CHIME/FRB event detection rate, displayed alongside those of other FRB detection efforts, with radio frequency on the vertical axis. This represents an update from a similar plot presented by Chawla et al. (2017) but based on the revised all-sky rate estimated by Lawrence et al. (2017). The red CHIME box is for the full CHIME bandwidth, whereas the blue rectangles show estimates in four 100 MHz sub-bands. A Euclidean flux distribution ($\gamma = 1.5$) is assumed. The hatched boxes denote rate estimates assuming a spectral index range, $-0.3 < \alpha < 2$, that was estimated in Chawla et al. (2017). Different colors are used to distinguish different experiments in the same frequency range.

orders of magnitude in the 400–500 MHz band. The predicted CHIME/FRB event rates from Chawla et al. (2017) are broadly consistent with those based on a simpler analysis by Connor et al. (2016), who considered the 700–900 MHz band only.

However, the predictions presented in Chawla et al. (2017) are derived from the Crawford et al. (2016) 1.4 GHz FRB rate, which is a factor of several higher than the more recent estimate by Lawrence et al. (2017) of 587^{+337}_{-315} bursts $\text{sky}^{-1} \text{day}^{-1}$ above a peak flux density of 1 Jy. Based on this revised rate and assuming a Euclidean flux distribution, our revised analysis predicts that CHIME/FRB will detect 0.6–11 bursts per day, as shown in Figure 8. This prediction is obtained using simulations following the same methodology as in Chawla et al. (2017), except that the simulated bursts have an intrinsic width of 3 ms with scattering timescales drawn from a distribution similar to the Galactic scattering time distribution at 1 GHz and scaled to the appropriate frequency. A significant reduction in the detection rate (0.3–2 bursts per day) is expected if FRBs are detectable only in the upper part of the CHIME band (700–800 MHz). Although these updated CHIME/FRB rate predictions are significantly lower than previous estimates, it is important to note that our revised rates have been calculated using an all-sky event rate. Actual detection rates could be as much as a factor of four higher since 75% of the CHIME-visible sky corresponds to high Galactic latitudes ($|b| > 15^\circ$), where the observed rate is estimated to be an order of magnitude greater than the low-latitude rate:

2866^{+4462}_{-1745} bursts $\text{sky}^{-1} \text{day}^{-1}$ for a threshold of 1 Jy (Vander Wiel et al. 2016). This implies a CHIME/FRB rate of $2\text{--}42 \text{ day}^{-1}$.

If CHIME/FRB detects many FRBs per day, thousands of FRBs over the currently envisioned ~ 3 yr lifetime of the project (currently limited only by available operations funding) will be detected. This will allow detailed sky distributions, DM distributions, and a $\log N / \log S$ distribution with the benefits of a uniform survey with a single well-determined sensitivity function. This will also enable the study of potential correlations between parameters such as DM and flux, DM and scattering measure, or width and flux. The FRB DM distribution, for example, has been suggested to be informative of the location of the so-called missing cosmic baryons (McQuinn 2014) and could provide a new probe of the three-dimensional clustering of matter in the universe (Masui & Sigurdson 2015). Yang & Zhang (2016) suggest that, given a large number of FRBs with redshift determinations, one can infer cosmological parameters even without knowing the host galaxy DM contribution. The detection of many FRBs at very high redshifts ($z > 6$) could allow a determination of the size distribution of ionized bubbles in the intergalactic medium near the epoch of reionization (Yoshiura & Takahashi 2018). High-redshift detections by CHIME/FRB seem possible given that the smearing time for DMs near $10,000 \text{ pc cm}^{-3}$ (e.g., Fialkov & Loeb 2016) is well under the maximum width searched by our pipeline, and IGM scattering times may be comparable to or lower than the smearing time (see, e.g., Bhandari

et al. 2018). Sky distributions of FRBs will also allow us to understand Galactic propagation effects such as plasma lensing (Cordes et al. 2017) and separate those from intrinsic FRB characteristics. Additionally, we can determine robust upper limits on short radio transient emission in the CHIME FOV in the absence of a detection by our pipeline. This will be useful for setting upper limits on radio emission from counterparts from sources that trigger, e.g., gravitational wave detectors or other electromagnetic transient detectors.

On the other hand, CHIME/FRB’s angular resolution alone is insufficient for host galaxy identification, a key ingredient in making progress in this field. However, there are several potential paths to improved localization and/or redshift determination. Because CHIME sees the full sky above declination $\sim -20^\circ$ each day, CHIME/FRB will be sensitive to repeat bursts, enabling identification of repeating FRBs like FRB 121102. These can be followed up by long baseline interferometers for localization and subsequent host galaxy identification as was done for FRB 121102 (Chatterjee et al. 2017; Tendulkar et al. 2017). Note that statistics of repeaters will be well studied thanks to CHIME/FRB’s regular daily exposure of fixed duration with stable sensitivity. Another avenue for localization will be through VLBI using the CHIME/FRB baseband recording, in conjunction with another telescope operating in the CHIME band and watching the same portion of sky. “Outrigger” stations are currently under study.

One way CHIME/FRB could potentially determine a host galaxy redshift independently is through H I absorption of a bright FRB’s spectrum, from the host galaxy’s hydrogen gas. For an FRB in the redshift range 0.8–2.5, assuming a flux of 30 Jy, the absorption is expected to have $S/N \gtrsim 15$ approximately 10% of the time, assuming that FRBs are in gas-rich spiral galaxies viewed randomly (Margalit & Loeb 2016). Note that FRBs with fluxes of up to 120 Jy (albeit at 1.4 GHz) have been reported (e.g., Ravi et al. 2016). The above S/N estimate assumes the planned 16k frequency channels across the CHIME bandwidth, but with baseband data we could improve the resolution further, rendering detection easier. Moreover, for repeating FRBs, we can stack spectra to improve the S/N of a possible absorption feature. However, a redshift detection via H I absorption requires that not all FRBs be in dwarf galaxies, as is the case for FRB 121102 (Tendulkar et al. 2017), for which the effect would likely be undetectable.

8. Conclusions

The CHIME/FRB project has the potential to yield major progress in our understanding of the FRB mystery. This is remarkable for a custom-designed telescope whose basic properties were chosen explicitly for independent science. The telescope’s utility for simultaneous pulsar timing via CHIME/Pulsar is further evidence of the scientific value of a “software” radio telescope given today’s massive computing capabilities. Computing power and storage are likely to grow given massive consumer demand, suggesting that CHIME is easily scalable, which, in the FRB context, argues that far deeper surveys may one day be possible. Whether efforts in this direction are well motivated will be determined by FRBs’ utility as cosmic probes, something CHIME/FRB should be able to resolve. Crucial to this endeavor is the ability to follow up FRBs to identify host galaxies and redshifts, something CHIME/FRB cannot do alone because of its relatively poor localization capabilities. However, interferometric follow-up, especially of repeating FRBs, which


CHIME/FRB is well placed to discover, is a promising avenue for localization (as has been accomplished for FRB 121102) and is being pursued by our team. Another attractive option, especially for nonrepeaters, is the correlation of CHIME/FRB baseband data with those obtained from an array of CHIME/FRB “outriggers” located across North America. Such an effort is currently under consideration. Another future avenue for enhancement of CHIME/FRB’s baseline capabilities is also under investigation: the use of machine-learning algorithms at various places in the CHIME/FRB pipeline, namely, for improved RFI mitigation and improved real-time as well as offline event identification. This may prove to be yet another manifestation of the power of data-rich software telescopes in that they permit great creativity and flexibility in data handling and hence in scientific output.

We are very grateful for the warm reception and skillful help we have received from the staff of the Dominion Radio Astrophysical Observatory, which is operated by the National Research Council Canada. The CHIME/FRB Project is funded by a grant from the Canada Foundation for Innovation 2015 Innovation Fund (Project 33213), as well as by the Provinces of British Columbia and Quebec. Additional support was provided by the Canadian Institute for Advanced Research (CIFAR) Gravity & Extreme Universe Program, McGill University and the McGill Space Institute, University of British Columbia, and University of Toronto Dunlap Institute. The Dunlap Institute is funded by an endowment established by the David Dunlap family and the University of Toronto. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. J.C. was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award. P.C. is supported by a Mitacs Globalink Graduate Fellowship. M.D. is supported by an NSERC Discovery Grant, CIFAR, and Fonds de Recherche du Québec—Nature et technologies Centre de Recherche en Astrophysique du Québec (FRQNT/CRAQ). B.M.G. is supported by an NSERC Discovery Grant and the Canada Research Chairs program. M.H. and G.H. are supported by CIFAR. V.M.K. is supported by a Lorne Trottier Chair in Astrophysics & Cosmology, a Canada Research Chair, CIFAR, an NSERC Discovery Grant and Herzberg Award, and FRQNT/CRAQ. U.-L.P. is supported by CIFAR. K.W.M. is supported by the Canadian Institute for Theoretical Astrophysics National Fellows program. S.M.R. is supported by CIFAR and the NSF Physics Frontiers Center award 1430284. I.H.S. is supported by CIFAR and an NSERC Discovery Grant. P.S. is supported by a DRAO Covington Fellowship from the National Research Council Canada.

Software: CHIME/FRB pipeline, alpenhorn (Hincks et al. 2015), bonsai (K. M. Smith et al. 2018, in preparation), Grafana (grafana.com), Prometheus (prometheus.io), RRAT-Trap (Karako-Argaman et al. 2015), PRESTO (github.com/scottransom/presto).

ORCID iDs

P. Chawla  <https://orcid.org/0000-0002-3426-7606>

B. M. Gaensler  <https://orcid.org/0000-0002-3382-9558>

V. M. Kaspi  <https://orcid.org/0000-0001-9345-0307>
 S. M. Ransom  <https://orcid.org/0000-0001-5799-9714>

References

- Akshori, T., Ryu, D., & Gaensler, B. M. 2016, *ApJ*, **824**, 105
- Amiri, M., Bandura, K., Berger, P., et al. 2017, *ApJ*, **844**, 161
- Bandura, K., Addison, G. E., Amiri, M., et al. 2014, *Proc. SPIE*, **9145**, 914522
- 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
- Bannister, K. W., Shannon, R. M., Macquart, J.-P., et al. 2017, *ApJL*, **841**, L12
- Bhandari, S., Keane, E. F., Barr, E. D., et al. 2018, *MNRAS*, **475**, 1427
- Caleb, M., Flynn, C., Bailes, M., et al. 2017, *MNRAS*, **468**, 3746
- Chatterjee, S., Law, C. J., Wharton, R. S., et al. 2017, *Natur*, **541**, 58
- Chawla, P., Kaspi, V. M., Josephy, A., et al. 2017, *ApJ*, **844**, 140
- Connor, L., Lin, H.-H., Masui, K., et al. 2016, *MNRAS*, **460**, 1054
- Cordes, J. M., & Lazio, T. J. W. 2001, *ApJ*, **549**, 997
- Cordes, J. M., & McLaughlin, M. A. 2003, *ApJ*, **596**, 1142
- Cordes, J. M., Wasserman, I., Hessels, J. W. T., et al. 2017, *ApJ*, **842**, 35
- Cortes, C., & Vapnik, V. 1995, *Machine Learning*, **20**, 273
- Crawford, F., Rane, A., Tran, L., et al. 2016, *MNRAS*, **460**, 3370
- Deng, M., Campbell-Wilson, D. & CHIME Collaboration, T. 2014, *arXiv:1708.08521*
- Eatough, R. P., Keane, E. F., & Lyne, A. G. 2009, *MNRAS*, **395**, 410
- Ester, M., Kriegel, H.-P., Sander, J., et al. 1996, A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise in *Proc. of the Second Int. Conf. on Knowledge Discovery and Data Mining (Portland, OR: AAAI Press)*, 226
- Fialkov, A., & Loeb, A. 2016, *JCAP*, **5**, 004
- Hankins, T. H., & Rickett, B. J. 1975, *Methods in Computational Physics Volume 14—Radio Astronomy (New York: Academic)*, 55
- Hincks, A. D., Shaw, J. R. & CHIME Collaboration 2015, in *ASP Conf. Ser.* **495**, *Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV)*, ed. A. R. Taylor & E. Rosolowsky (San Francisco, CA: ASP), 523
- Inoue, S. 2004, *MNRAS*, **348**, 999
- Karako-Argaman, C., Kaspi, V. M., Lynch, R. S., et al. 2015, *ApJ*, **809**, 67
- Karastergiou, A., Chennamangalam, J., Armour, W., et al. 2015, *MNRAS*, **452**, 1254
- Kothes, R., Landecker, T. L., & Gray, A. D. 2010, in *ASP Conf. Ser.* **438**, *The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey*, ed. R. Kothes, T. L. Landecker, & A. G. Willis (San Francisco, CA: ASP), 415
- Krishnakumar, M. A., Mitra, D., Naidu, A., Joshi, B. C., & Manoharan, P. K. 2015, *ApJ*, **804**, 23
- Kulkarni, S. R., Ofek, E. O., Neill, J. D., Zheng, Z., & Juric, M. 2014, *ApJ*, **797**, 70
- Lawrence, E., Vander Wiel, S., Law, C., Burke Spolaor, S., & Bower, G. C. 2017, *AJ*, **154**, 117
- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, *Sci*, **318**, 777
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, *AJ*, **129**, 1993
- Manchester, R. N., & Taylor, J. H. 1977, *Pulsars (San Francisco: Freeman)*
- Maoz, D., Loeb, A., Shvartzvald, Y., et al. 2015, *MNRAS*, **454**, 2183
- Marcote, B., Paragi, Z., Hessels, J. W. T., et al. 2017, *ApJL*, **834**, L8
- Margalit, B., & Loeb, A. 2016, *MNRAS*, **460**, L25
- Masui, K., Lin, H.-H., Sievers, J., et al. 2015, *Natur*, **528**, 523
- Masui, K. W., Shaw, J. R., Ng, C., et al. 2017, *arXiv:1710.08591*
- Masui, K. W., & Sigurdson, K. 2015, *PhRvL*, **115**, 121301
- McLaughlin, M. A., & Cordes, J. M. 2003, *ApJ*, **596**, 982
- McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., et al. 2006, *Natur*, **439**, 817
- McQuinn, M. 2014, *ApJL*, **780**, L33
- Michilli, D., Seymour, A., Hessels, J. W. T., et al. 2018, *Natur*, **533**, 132
- Newburgh, L. B., Addison, G. E., Amiri, M., et al. 2014, *Proc. SPIE*, **9145**, 91454V
- Ng, C. & CHIME/Pulsar Collaboration 2017, in *Proc. IAU Symp.* **337**, *Pulsar Astrophysics—The Next 50 Years (Cambridge: Cambridge Univ. Press)*, in press (*arXiv:1711.02104*)
- Ng, C., Vanderlinde, K., Paradise, A., et al. 2017, *arXiv:1702.04728*
- Opperman, N., & Pen, U.-L. 2017, *arXiv:1705.04881*
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Petroff, E., Bailes, M., Barr, E. D., et al. 2015a, *MNRAS*, **447**, 246
- Petroff, E., Barr, E. D., Jameson, A., et al. 2016, *PASA*, **33**, e045
- Petroff, E., Houben, L., Bannister, K., et al. 2017, *arXiv:1710.08155*
- Petroff, E., Johnston, S., Keane, E. F., et al. 2015b, *MNRAS*, **454**, 457
- Petroff, E., Keane, E. F., Barr, E. D., et al. 2015c, *MNRAS*, **451**, 3933
- Piro, A. L. 2016, *ApJL*, **824**, L32
- Prochaska, J. X., & Neelaman, M. 2018, *MNRAS*, **474**, 318
- Rane, A., & Lorimer, D. 2017, *JApA*, **38**, 55
- Rane, A., Lorimer, D. R., Bates, S. D., et al. 2016, *MNRAS*, **455**, 2207
- Ravi, V., Shannon, R. M., Bailes, M., et al. 2016, *Sci*, **354**, 1249
- Remazeilles, M., Dickinson, C., Banday, A. J., Bigot-Sazy, M.-A., & Ghosh, T. 2015, *MNRAS*, **451**, 4311
- Rickett, B. J. 1977, *ARA&A*, **15**, 479
- Rowlinson, A., Bell, M. E., Murphy, T., et al. 2016, *MNRAS*, **458**, 3506
- Scholz, P., Spitler, L. G., Hessels, J. W. T., et al. 2016, *ApJ*, **833**, 177
- Shull, J. M., & Danforth, C. W. 2018, *ApJL*, **852**, L11
- Spitler, L. G., Cordes, J. M., Hessels, J. W. T., et al. 2014, *ApJ*, **790**, 101
- Spitler, L. G., Scholz, P., Hessels, J. W. T., et al. 2016, *Natur*, **531**, 202
- Taylor, J. H. 1974, *AAPS*, **15**, 367
- Tegmark, M., & Zaldarriaga, M. 2009, *PhRvD*, **79**, 083530
- Tendulkar, S. P., Bassa, C. G., Cordes, J. M., et al. 2017, *ApJL*, **834**, L7
- Thornton, D., Stappers, B., Bailes, M., et al. 2013, *Sci*, **341**, 53
- Trotter, A. S., Reichart, D. E., Egger, R. E., et al. 2017, *MNRAS*, **469**, 1299
- Vander Wiel, S., Burke-Spolaor, S., Lawrence, E., Law, C. J., & Bower, G. C. 2016, *arXiv:1612.00896*
- Yang, Y.-P., & Zhang, B. 2016, *ApJL*, **830**, L31
- Yao, J. M., Manchester, R. N., & Wang, N. 2017, *ApJ*, **835**, 29
- Yoshiura, S., & Takahashi, K. 2018, *MNRAS*, **473**, 1570
- Zackay, B., & Ofek, E. O. 2017, *ApJ*, **835**, 11