

Observations of Repeating Fast Radio Burst Sources with the CHIME/Pulsar Instrument

Andrew Sikora, Department of Physics
McGill University, Montreal

July, 2021

A thesis submitted to McGill University in partial fulfillment
of the requirements of the degree of

Master of Science

©July, 2021

Abstract

Since the beginning of its pre-commissioning phase in July 2018, the Fast Radio Burst team at the Canadian Hydrogen Intensity Mapping Experiment (CHIME/FRB collaboration) has detected over 500 fast radio bursts (FRBs) [9]. FRBs are highly energetic bursts of radio emission originating beyond the Milky Way. They are primarily characterized by their millisecond durations, high luminosities, and extragalactic dispersion measures (DMs). While most FRBs appear to be one-off events, some have been observed to repeat, leading to theories of two separate FRB progenitor populations. CHIME has no moving parts and thus cannot track sources across the sky. One way of overcoming this has been tackled by the CHIME/Pulsar team. CHIME/Pulsar receives 10 digitally formed tracking beams within CHIME’s primary beam. This allows for short term tracking of sources while they are in the primary beam. In this thesis, I present the development and results of a convolutional neural network (CNN) to search hundreds of hours of CHIME/Pulsar data taken of FRB repeaters. Following an introduction to FRBs, CHIME/FRB and CHIME/Pulsar, I will outline neural networks in general, the architecture of this CNN, the process taken to train it, and finally the results of its analysis. I will also describe upgrades

made to a scheduling algorithm for CHIME/Pulsar. Since CHIME/Pulsar can only track 10 locations at a time, and there are often more than 10 interesting sources in the main beam, a basic scheduling algorithm that eventually needed significant upgrades was developed. Here I will describe why such upgrades were needed, the process taken to implement them, tests of the upgraded schedule, and its current status.

Abrégé

Depuis le commencement des opérations durant la phase de pré-démarrage en juillet 2018, le projet portant sur les sursauts radio rapides (FRBs, de l'anglais *fast radio bursts*) de l'Expérience canadienne de cartographie de l'intensité de l'hydrogène (CHIME/FRB, pour *Canadian Hydrogen Intensity Mapping Experiment Fast Radio Burst Project*) a découvert plus de 500 sursauts [9]. Les FRBs sont de puissants flashes de lumière radio, provenant de sources lointaines à l'extérieur de la Voie Lactée, et sont principalement caractérisés par leur courte durée de quelques millisecondes, leur luminosité élevée, et leur mesure de dispersion extragalactique. Bien que, dans la majorité des cas, un seul sursaut ponctuel ait été observé par progéniteur, certaines sources émettent des FRBs de façon répétitive. Ceci donne lieu à des théories proposant deux populations distinctes de progéniteurs de FRBs. Ne comportant aucune pièce mobile, CHIME est un instrument complètement stationnaire et ne peut donc pas être orienté mécaniquement pour suivre un objet astronomique se déplaçant dans le champ focal du télescope. Pour surmonter cette limitation, l'équipe du projet CHIME/Pulsar a développé une solution consistant à former numériquement 10 faisceaux auxiliaires de pointage contenus à l'intérieur du faisceau lumineux principal, permettant

ainsi le suivi de sources durant la courte période où celles-ci se trouvent dans le faisceau primaire. Dans ce mémoire, je présente le développement et les résultats d'un réseau neuronal convolutif (CNN, de l'anglais *convolutional neural network*) conçu pour analyser et identifier des FRBs répétitifs dans un large ensemble de données collectées par le système CHIME/Pulsar. Suite à l'introduction des FRBs et des systèmes CHIME/FRB et CHIME/Pulsar, je présente dans ce mémoire les notions de base relatives aux réseaux neuronaux, l'architecture et la formation de mon CNN, et finalement, les résultats de l'analyse des données. Je décris également les mises à niveau effectuées sur un algorithme d'ordonnancement pour CHIME/Pulsar. Puisque celui-ci ne peut qu'observer et suivre 10 positions à la fois et qu'il y a très souvent plus de 10 sources d'intérêts dans le faisceau principal, un algorithme d'ordonnancement a dû être conçu pour CHIME/pulsar. Toutefois, cet algorithme nécessitait d'importantes améliorations. Dans ce mémoire, j'explique les raisons justifiant les mises à niveau ainsi que les techniques utilisées pour développer, tester et implémenter ces mises à niveau de l'algorithme.

Acknowledgements

First I would like to thank the McGill Department of Physics and the McGill Space Institute for providing me with an excellent work space filled with dedicated and passionate researchers. I also thank those at the Natural Sciences and Engineering Research Council of Canada for their generous support in the form of the Canada Graduate Scholarships-Masters grant. I thank the many people of the CHIME/FRB and CHIME/Pulsar collaborations for their endless patients and willingness to educate. Specifically, I thank Arun Naidu for his willingness to teach the basics of not only machine learning but also radio astronomy in general to myself who had little experience with both, and Chitrang Patel for his understanding, technical guidance, and contagious enthusiasm. I thank my incredible supervisor, Vicky Kaspi, for her dedication to myself and all her students, especially during the COVID-19 pandemic. Her guidance, positivity, and motivation have been instrumental in the development and completion of this thesis. I thank my family for being supportive and loving as I followed my passions. I also thank my fellow graduate students, specifically Alice Curtin, Hannah Fronenberg, Zarif Kader, Matheus Pessoa, Sabrina Berger, Tim Halllet, Tristan Menard and Thomas Mosher, for making this experience full of

friendship and positive memories during a negative time in the world's history.

Finally, I would like to thank the many doctors, nurses, and healthcare professionals in Canada and around the world who have dedicated the last 1.5 years of their lives to stopping the spread and protecting the public from COVID-19. It is thanks to them that people such as myself have been able to continue to work at relatively full capacity while remaining safe from the virus.

Table of Contents

Abstract	i
Abrégé	iii
Acknowledgements	v
List of Figures	xviii
List of Tables	xix
1 Introduction to Fast Radio Bursts	1
1.1 Radio Astronomy	1
1.2 Pulsars	2
1.3 The Lorimer Burst	3
1.4 Recent Advancements	5
1.5 FRB Properties	6
1.6 Repeating FRBs	10
1.7 Outline of this thesis	14
2 CHIME/FRB Overview	16
2.1 Introduction to CHIME	16
2.2 Why CHIME for an FRB search?	17
2.3 Physical Description	19

2.4	FRB Search Pipeline	19
2.4.1	Correlation	19
2.4.2	CHIME/FRB Detection Pipeline	21
3	CHIME/Pulsar Overview	27
3.1	Introduction to CHIME/Pulsar	27
3.2	Hardware and Backend	28
3.3	Software	30
3.3.1	Data Types	30
3.3.2	Tracking Beam Scheduler	31
3.4	Scientific Intent	32
4	Machine Learning Search for FRB Repeaters	34
4.1	Neural Networks	34
4.1.1	The Basics	34
4.1.2	Learning	37
4.1.3	Convolutional Neural Networks	40
4.1.4	A CNN for FRB Repeater Searching	44
4.1.5	Results	52
4.1.6	Conclusion	60
5	An Upgraded Scheduling Algorithm for CHIME/Pulsar	62
5.1	Motivation	62
5.1.1	Outline of Logic	65
5.1.2	Testing	69
5.1.3	Current Status	72

6 Conclusion	76
6.0.1 Next steps: Scheduler	79
6.0.2 Next steps: CNN	80

List of Figures

1.1	Radio intensity as a function of Frequency vs Time for the Lorimer Burst (FRB 010724). This burst marks the first detection of a Fast Radio Burst. The characteristic dispersion sweep is clearly visible. A dedispersed pulse profile can be seen in the inset. Taken from [26]	4
1.2	From [9]. a: DM _{excess} distribution of all FRBs in [9]. DMs of $\sim 500 \text{ pc cm}^{-3}$ are common but can extend to $> 1500 \text{ pc cm}^{-3}$. b: Distribution of temporal pulse width of FRBs in [9]. Pulses of $\sim 1 \text{ ms}$ in length are common, but can broaden to multiple seconds. c: Bandwidths of non-repeater and first detected repeater events. Repeaters tend to be wider temporally and narrower in frequency, suggesting different underlying sample populations. In all plots, blue regions denote bursts from non-repeating FRBs while the orange plots the first burst detected from each repeater.	9

1.3 Skymap of all CHIME discovered FRBs published in [9]. Triangles correspond to the 18 CHIME/FRB discovered repeaters. Circles designate the 474 newly published FRBs from [9]. There appears to be no favoured positions for FRBs, and they are evenly distributed on the sky [9] (Josephy et al 2021, in preparation)	10
2.1 a: Image of CHIME [1]. The four parabolic reflectors are clearly visible with the shipping containers containing the X portion of the CHIME FX-correlator and FRB compute nodes to the right (white boxes). b: Satellite view of DRAO and CHIME ^a . The CHIME telescope and main DRAO facility are marked by the blue and red pins, respectively. The Pathfinder telescope can be seen in the bottom right. The nodes containing L1, L2 and L3 segments of the FRB search pipeline are housed in the shipping containers next to the telescope. An underground network link connects to the L4, FRB analysis, and archiving nodes in the DRAO facility.	18
2.2 Conceptual block diagram of the FRB search pipeline, from [1]. Processes use a numbering scheme of LN, where N is process number, 0 being the most upstream. The main features of each process are listed.	22

4.1	Conceptual examples of a single neuron neural network (a) and a two neuron, single hidden layer NN (b). Both of these are examples of linear neural networks. Each neuron (circle) houses an activation function that is trained to help predict home or property value based on a set of input data (house size, property size, etc.)	36
4.2	Conceptual diagram of the input data to a convolutional neutral network. Instead of the linear input data seen in Figure 4.1a, the CNN input layer is a 2-dimensional grid of inputs, usually the pixels of an image.	41
4.3	The connection of a 5x5 kernel of input neurons to the first neuron in the first hidden layer. This kernel will then be stepped a configurable number of neurons to the right, and will connect the new 5x5 input neurons to the second neuron in the first hidden layer. This process will continue until the entire input layer has been scanned.	42
4.4	Conceptual image showing that the input layer can be scanned multiple times to produce multiple feature maps in the first hidden layer. Here the 2D grid of circles seen in Figure 4.2 is replaced with a square. These three feature maps in the first hidden layer are the 2D analog to the 2 neurons in the first hidden layer of Figure 4.1b. Each feature map corresponds to the scanning of the input layer with a fixed LRF. Thus, they search the input image for specific key features.	43

4.5 Similar to Figure 4.3, this figure shows the maximum pooling of a convolutional layer. Pooling is usually done in small kernel sizes which take either the maximum or average of the neurons in the kernel.	44
4.6 The rectified linear unit function (ReLU) define as $f(x) = \max(0, x)$. This function is often used in CNNs as the activation function due to it maintaining a high gradient when x is large. Another common activation function is $f(x) = \tanh(x)$, but this suffers from low gradients at large value of x . Maintaining a high gradient ensures faster training of the CNN model.	46
4.7 Examples of training data passed to the CNN to learn bursts (a) and RFI (b). RFI training data was derived from Gaussian noise and sifted through by eye to ensure any images with burst-like structures were removed. Burst training data were augmented in time, DM, and had low levels of Gaussian noise added. The time and DM augmentation can be clearly seen in a by the non-centring of the burst in time and the clear sub-optimal dedispersion (non-vertical burst). The entire training set totalled to 7664 images with equal parts bursts and non-bursts.	51

4.8	The 4 bursts found by the CNN for the repeater R2. SNRs for these bursts are 6.6, 7.5, 9.4, and 7.0 for a , b , c , and d respectively. They were detected on Feb 14 2019, Feb 21 2019, April 1 2019, and Sept 3 2019, respectively. These add to the 21 R2 bursts detected by CHIME/FRB.	54
4.9	Waterfall plots of the three R3 bursts found by the CNN that were not found by CHIME/FRB. These bursts come from observations on Nov 14, 2019 and Jan 4, 2020. Note that b and c occur on the same day. These burst of SNRs of 10.9, 8.8 and 13.7, respectively. The title space of each plot contains the modified Julian date (MJD) of the observation, the time corresponding to the start of this image relative to the beginning of the observation, the time downsampling used (in ms), and the score given by the CNN. Higher scores relate to a higher confidence by the classifier that the image contains an FRB. . .	56
4.10	Frequency vs time waterfall plot of the R5 burst found by the CNN that was not detected by CHIME/FRB. This burst has an SNR of 6.12 and comes from an observation on July 24, 2019. As in Figure 4.9, the title space the modified Julian date (MJD) of the observation, the time corresponding to the start of this image relative to the beginning of the observation, the time downsampling used (in ms), and the score given by the CNN.	57

4.11	The 4 bursts not detected by CHIME/FRB that were found by the CNN for R6. This set increases the existing 9 burst set by 44%. These additional bursts were detected on June 22 2019, Dec 5 2019, Jan 13 2020, and Feb 10 2020, and have SNRs of 25.0, 5.35, 8.0 , and 7.2 respectively. The titles spaces are similar to Figures 4.9 and 4.10.	59
4.12	Visualization of the dates when R6 was detected by CHIME/FRB (blue) and CHIME/Pulsar using the new CNN (orange). Note that the last observation of R6 with CHIME/Pulsar occurred on MJD 58925 (March 17, 2020).	60
5.1	Snapshot of the sources in the primary beam at ~11:00am, June 3, 2021. Yellow circles are locations of previously detected repeating and non-repeating FRBs, green circles are for known pulsars, and purple circles mark steady sources. The grey shaded region marks the CHIME primary beam. It is clear that there are often more than 10 interesting sources in the primary beam. Thus a scheduling algorithm for the 10 CHIME/Pulsar beams needed to be developed.	63
5.2	Diagram outlining the logic of the upgraded CHIME/Pulsar scheduling algorithm.	66

- 5.3 A visualization of \sim 24 hours of schedule. Each block corresponds to a source's transit. The y-axis is the beam number (0-9) and the x-axis is time marked in hours from now. The green coloured blocks have their observation status set to COMPLETE, OBSERVING observations are in yellow, and the faded multi-colour blocks are labelled SCHEDULED. The change in colour of the faded sources is simply due to their colours being randomized. 71
- 5.4 Histogram of the number of observations of 8 sources near the Galactic centre after 15 days of simulated schedule. Each bin corresponds to one of the 4 original priority values. Two sources for each level were selected. It can clearly be seen that higher priority sources are observed more often than lower priority ones. Note that even though *initially* priority 128 sources are 8 times as likely to be observed as priority 16 sources, as the priorities decay the distribution in the number of observations smooths out. 73
- 5.5 Before (a) and after (b) the addition of a source to the real-time source list to test ToO functionality of the scheduling algorithm. The realtime source is marked in cyan. 74

List of Tables

2.1	Key properties of the CHIME telescope. From [1]. Where two values are given, they are for the 400 MHz and 800 MHz edges of the band, respectively.	20
4.1	Summary table of the repeaters searched by the CNN, total number of observation hours for each repeater, current number of CHIME/FRB detections as of June 18 2021, and the number of CNN detections missed by CHIME/FRB.	58
5.1	A list of all the keys in a source table entry, with descriptions for each key.	69
5.2	A list of all the keys in an observation table entry, with descriptions for each key.	70

Chapter 1

Introduction to Fast Radio Bursts

1.1 Radio Astronomy

In 1931, Karl Jansky discovered unexpected static of extraterrestrial origin when studying radio emissions from thunderstorms for Bell Laboratories. The majority of this emission seemed to come from the disk of the Milky Way and the constellation Sagittarius. This work birthed the extensive field of radio astronomy [27].

Radio astronomy now plays a crucial role in any multi-wavelength astrophysical study. With emission mechanisms in all branches of astrophysics, from galaxy structure, to compact objects, to cosmology, radio astronomy has given astronomers access to some of the inner workings of the Universe's most interesting phenomena.

1.2 Pulsars

One of the largest fields in radio astronomy is the study of radio pulsars. First discovered by Jocelyn Bell-Burnell and Antony Hewish at Cambridge in 1967, pulsars are the result of a rotating neutron star having its magnetic field axis misaligned with its rotational axis [27]. The radio emission comes from the open magnetic field lines at the poles of the neutron star. Due to the misalignment of the magnetic and rotational axes, one observes a characteristic "lighthouse" effect. This effect is a periodic peak in radio intensity when the neutron star's magnetic pole crosses the line of sight. Known for hosting extreme magnetic environments, most pulsars have magnetic field strengths of 10^8 - 10^{13} G, while a sub-group known as magnetars can have field strengths up to 10^{15} G [21]. For the 10^8 - 10^{13} G group, known as rotationally-powered-pulsars (RPPs), typical rotational periods are between 1ms and 8s. Those with a period < 20 ms are usually referred to as millisecond-pulsars (MSPs). MSPs are the result of a neutron star in a binary system "recycling" material from its companion via accretion. This torques the neutron star and causes a large increase in spin [6]. The fastest MSP to date is PSR J1748-2446ad with a rotational period of ~ 1.397 ms [19]. Conversely, Tan et al. (2018) discovered the slowest rotating pulsar, PSR J0250+5854, with a period of 23.5s [44]. Pulsar astronomy is unique in that neutron stars are some of the only astronomical sources that can be observed across the entire electromagnetic spectrum. The extreme environments of these objects have made them a fruitful laboratory for extreme physics.

Pulsars are traditionally searched for with large radio telescopes. Due to their characteristic periodicity, pulsars are identified by stacking (folding) many hours of data over a set of possible periodicities. If any test period yields a significant signal, the pulsar is identified. This process is made possible by extraordinary fast-Fourier transforms (FFTs) and fast-folding algorithms (FFAs).

A less conventional technique is to search for single pulses, as opposed to a series of pulses at a certain period. It is this process that led to the discovery of RRATs (rapidly rotating radio transients) [34]. Although these objects have some underlying period, their sporadic emission timing warranted the use of single pulse searches.

Over 2000 pulsars have been found to date [30], but with an estimated $> 10^9$ neutron stars in the Milky Way, many more pulsar discoveries are expected. The use of single pulse searches to find the first RRATs encouraged widespread use of such algorithms on archived radio telescope data.

1.3 The Lorimer Burst

When performing a single pulse search for pulsars and RRATs in the Small Magellanic Cloud (SMC) on archival 2001 data from the Parkes radio telescope, Duncan Lorimer observed a pulse so bright it saturated the primary detection beam of the receiver. This pulse was determined to have a dispersion measure (DM) of 375 pc cm^{-3} and a peak flux density $> 30 \text{ Jy}$ [26]. Dubbed the "Lorimer burst" (Fig. 1.1), this burst was remarkable for its incredible brightness and implied distance. Such a DM (see FRB Properties,

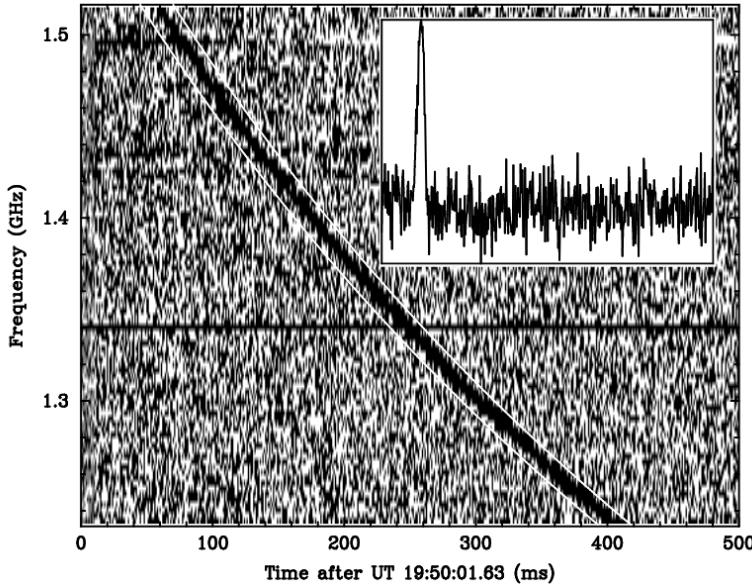


Figure 1.1: Radio intensity as a function of Frequency vs Time for the Lorimer Burst (FRB 010724). This burst marks the first detection of a Fast Radio Burst. The characteristic dispersion sweep is clearly visible. A dedispersed pulse profile can be seen in the inset. Taken from [26]

section 1.5) was estimated to be ~ 8 times greater than the maximum Galactic contribution (from free electrons in the Milky Way). Although close to the SMC, it was argued to be unassociated with it. This put the burst at an upper limit distance of 1 Gpc. With the now more conventional name of FRB 010724, the Lorimer burst is accepted as the first observed Fast Radio Burst (FRB).

1.4 Recent Advancements

For many years, FRB 010724 stood alone in the FRB population. The discovery of a similar pulse by Keane et al. in 2011 [22], and four high-DM pulses by Thornton et al. in 2013 [48], gave credence to such bursts being of astrophysical origin. It was by Thornton et al. that the term Fast Radio Burst (FRB) was first used to describe this phenomena.

In 2012, FRB 121102 was discovered by the late Arecibo telescope in Puerto Rico [41]. Followup efforts in May 2015 detected the first repeat bursts from FRB 121102, marking it as the first repeating FRB [42]. This quickly made describing FRB 121102 by a cataclysmic event difficult, as the source must be able to remain active for many years.

Since these initial discoveries, over 500 FRBs have been published. The vast majority of these bursts have been observed by the Canadian Hydrogen Intensity Mapping Experiment (CHIME). A brief description of this telescope is provided in Chapter 2, and a more thorough description can be found in [1]. Originally designed to map 21-cm emission in the Universe, the 400-800MHz telescope is an excellent FRB observatory. Coined CHIME/FRB, the FRB branch of the CHIME experiment is aimed at real time detection of FRBs at declinations >-20 deg.

With the discovery of repeat bursts from FRB 121102 and others (see Sec. 1.6) cataclysmic models were quickly called into question, and models allowing for repeated emission took centre stage.

1.5 FRB Properties

FRBs share much of the same properties as pulsars, but are of extragalactic distances and significantly higher energy. Of all the properties of FRBs that are to be mentioned in this subsection, one of, if not *the*, most important property for searching for and identifying FRBs is their high dispersion measure. The amount of dispersion in a sample of frequency vs time between radio frequencies ν_{high} and ν_{low} is defined as the time delay

$$\Delta t = \frac{e^2}{2\pi m_e c} (\nu_{low}^{-2} - \nu_{high}^{-2}) \text{DM} \approx 4.15 (\nu_{low}^{-2} - \nu_{high}^{-2}) \text{DM ms}, \quad (1.1)$$

where m_e is the mass of the electron, c is the speed of light, and the dispersion measure (DM), usually expressed in units of pc cm⁻³, is given as

$$\text{DM} = \int_0^d n_e(l) dx, \quad (1.2)$$

where n_e is the electron number density, l is the path length, and d is the distance to the FRB. This delay causes a sweeping pattern to the bursts with the lower frequency components arriving latest. This pattern can be clearly seen in the Lorimer burst in Figure 1.1. The delay occurs when the pulse travels through an ionized medium. In the case of pulsars, the majority of this plasma comes from the interstellar medium (ISM) and the local region of the pulsar. In the case of FRBs, due to their extragalactic origin, the dispersion measure of a burst can be separated into components

$$\text{DM} = \text{DM}_{Ion} + \text{DM}_{IPM} + \text{DM}_{ISM} + \text{DM}_{IGM} + \left(\frac{\text{DM}_{Host} + \text{DM}_{Local}}{1+z} \right). \quad (1.3)$$

Here, DM_{Ion}_o , DM_{IPM} , DM_{ISM} , and DM_{IGM} are the dispersion contribution from the Earth's ionosphere, the interplanetary medium, the interstellar medium, and the intergalactic medium, respectively. DM_{Host} and DM_{Local} are the DM contributions from the FRB's host galaxy and its local region, respectively. It is common to combine DM_{Ion}_o , DM_{IPM} , DM_{ISM} into one DM_{MW} component which measures the DM contribution from the Milky Way. Estimations of the Milky Way's DM contribution come from the YMW16 and NE2001 electron distribution models [50] [11]. These DM_{MW} estimates can be subtracted from the observed DM to get a DM-excess (DM_{ex}). If we consider only the DM contribution from the IGM, assume that all baryons are homogeneously distributed with ionization fraction $x(z)$, and assume Planck cosmological parameters, following Deng and Zhang (2014) [13] (as done in [39]), the mean IGM contribution is approximately

$$\text{DM}_{\text{IGM}} \simeq z \text{ } 1000 \text{cm}^{-3} \text{pc} \quad (1.4)$$

Following [39], for $z < 1$ we can assume the relation $d_L \simeq 2z(z + 2.4)$ Gpc. Thus, an upper limit for the luminosity distance to the FRB is

$$d_L < \left(\frac{\text{DM}}{500 \text{cm}^{-3} \text{pc}} \right) \left[\left(\frac{\text{DM}}{1000 \text{cm}^{-3} \text{pc}} \right) + 2.4 \right] \text{Gpc.} \quad (1.5)$$

The process of dedispersing to search for peaks in time series data is tackled differently for different telescopes. The process taken by CHIME/FRB is described in Chapter 2. Once the time series data are dedispersed, searches for pulses of high signal to noise ratio (SNR) are performed. With careful

calibration, the relative intensity of these peaks can be converted into a peak flux density, S_{peak} . Each FRB has some observed temporal width, W , which can be used to get the total energy, or fluence, $F = S_{peak}W = \int_{pulse} S(t) dt$.

The peak flux density can be combined with the luminosity distance calculated in Equation 1.5 to get the isotropic equivalent source luminosity,

$$L = \frac{4\pi d_L^2 S_\nu \Delta\nu}{1+z}. \quad (1.6)$$

This assumes the simple case of a flat spectrum source (constant S_ν).

Another key characteristic of FRBs, emphasized in Section 1.6, is the temporal broadening of bursts as a result of scattering. This arises from multipath propagation where parts of the signal that travel along a slightly longer path arrive later. This is usually seen as an FRB pulse with a one-sided exponential decay [39]. The decay time tends to follow the relation,

$$\tau \propto \nu^{-4}. \quad (1.7)$$

The distribution of these characteristics for the FRB population can be seen in Figure 1.2. It can be seen here that the population as a whole tends to prefer lower DMs and thus nearer positions, although this is likely due to lower detectability at greater distances and instrumental biases. Most burst durations are between 1 and 7 ms, but there is a notable spread in this distribution. Also, due to instrumental biases, wider bursts tend to be harder to detect. Differences in pulse width distributions between repeating and non-repeating FRBs is discussed in Section 1.6.

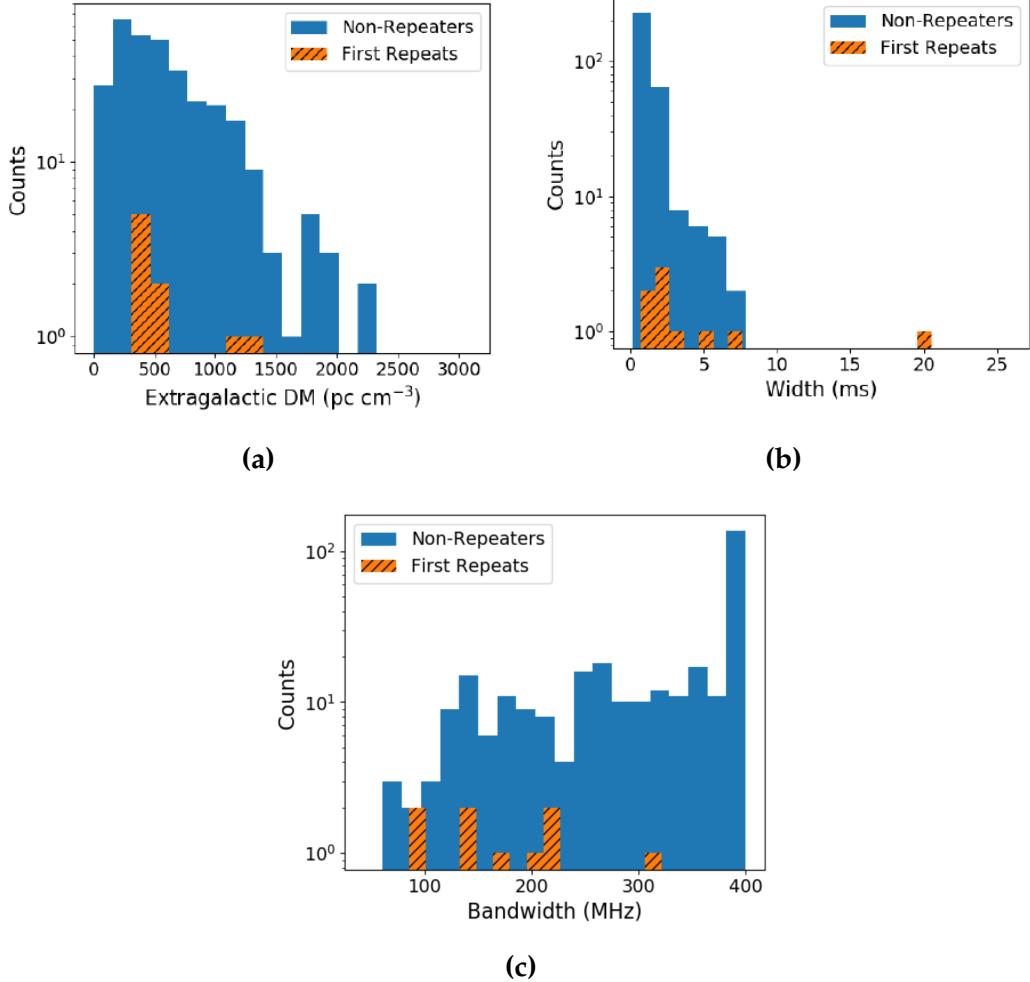


Figure 1.2: From [9]. **a:** $\text{DM}_{\text{excess}}$ distribution of all FRBs in [9]. DMs of $\sim 500 \text{ pc cm}^{-3}$ are common but can extend to $> 1500 \text{ pc cm}^{-3}$. **b:** Distribution of temporal pulse width of FRBs in [9]. Pulses of $\sim 1 \text{ ms}$ in length are common, but can broaden to multiple seconds. **c:** Bandwidths of non-repeater and first detected repeater events. Repeaters tend to be wider temporally and narrower in frequency, suggesting different underlying sample populations. In all plots, blue regions denote bursts from non-repeating FRBs while the orange plots the first burst detected from each repeater.

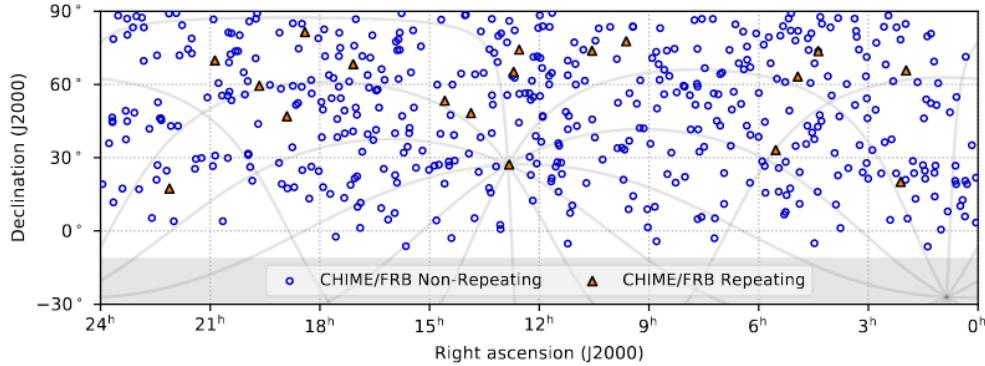


Figure 1.3: Skymap of all CHIME discovered FRBs published in [9]. Triangles correspond to the 18 CHIME/FRB discovered repeaters. Circles designate the 474 newly published FRBs from [9]. There appears to be no favoured positions for FRBs, and they are evenly distributed on the sky [9] (Josephy et al 2021, in preparation)

FRBs tend to be evenly distributed on the sky, with no regions favoured over any other. The sky distribution of FRBs published to date can be seen in Figure 1.3. The current all sky FRB rate is estimated to be $818 \pm 64 \text{ sky}^{-1} \text{ day}^{-1}$ for bursts with a fluences above 5 Jy ms, scattering times less than 10 ms at 600 MHz, and DMs above 100 pc cm $^{-3}$.

1.6 Repeating FRBs

Naturally, it was of common interest to perform follow-up observations of FRBs to determine if these new phenomena were in fact one-off events or if they repeated. After deep observations of the Lorimer burst (FRBs 010724), Keane burst (FRB 010621), and other Parkes discovered FRBs showed no such repetitions, it seemed that the last hope for detecting repetition was

FRB 121102. In May of 2015, a followup study of FRB 121102 detected 10 new bursts from the source, 6 of which occurred within a single 10-min observation [42]. This posed harsh restrictions on FRB theories. Namely, this meant that any repeating FRBs could not be explained by a cataclysmic event, and the idea of multiple FRB populations (repeaters and non-repeaters) began to grow.

With the discovery of repeat bursts from FRB 121102, researchers kept a keen eye out for repeat bursts from other FRBs. With the new CHIME/FRB instrument coming online, with its \sim 200 deg² field of view and 400-MHz bandwidth, repeater searches have become a more fruitful.

To date there have been 20 published FRB repeaters, 18 of which have been discovered by CHIME/FRB [8,16]. Although this number is quite low, population studies have been done on these 20 repeaters to determine if they are from the same physical population as non-repeaters.

Initially, the DM distribution of repeaters and non-repeaters was shown to have no statistically significant differences [8,16]. This conclusion has recently been supported by [9] with the analysis of \sim 500 FRBs. This suggests that repeater and non-repeater spatial distributions and associated local environments do not differ strongly.

Two aspects where repeaters and non-repeaters differ are their intrinsic temporal widths and bandwidths. Figure 1.2b (from [9]) shows the distribution in the widths of FRBs for repeating and non-repeating populations in the CHIME band (400 - 800 MHz). Although the repeater population is fully enveloped by the non-repeaters, Kolmogorov-Smirnov and Anderson-Darling tests give probabilities of $p_{AD}=1.5\times10^{-4}$ and $p_{KS}=2.2\times10^{-4}$ respec-

tively of the two populations being from the same underlying distribution. Similar tests on the bandwidth distributions seen in Figure 1.2c give $p_{AD}=1.3\times10^{-4}$ and $p_{KS}=2.3\times10^{-4}$. These differences suggest that repeaters and non-repeaters do not have the same origin [9]. This result agrees with that from [16]. The cause for such a difference could be that repeaters come from a different emission mechanism than non-repeaters, or there could be differences in the circumburst media between the types. The former of these is taken as the more likely hypothesis [16], (Pleunis et al 2021, in preparation).

It must be stressed that all of the above population analyses done to the repeater population used only the 20 repeaters published thus far. Although there appears to be statistically significant results in some regimes, the results from using such a small number of repeaters could change significantly as more are discovered. Thus, it is crucial that searches for FRB repeaters are performed, even if only to increase the sample size of repeat bursts to aid population analyses.

Repeaters rule out or severely constrain cataclysmic models. Recently proposed models include highly magnetized neutron stars that interact with their environments [3, 28, 35, 36], massive black holes [51] or internal magnetar instabilities affecting the magnetosphere [29]. This wide range of possible progenitors is due to the relatively limited data on the FRB repeater population. Thus, discoveries of new repeating FRBs as well as the continued discovery of new bursts from established repeaters are crucial for constraining the available models.

To date, repeating FRB followup has been quite successful. Of the 20 published bursts previously mentioned, two have a distinct periodicity. The first periodic FRB, FRB 180916, was discovered to have a period of 16.35 ± 0.15 days with a 0.6-day active phase window [7]. This period and phase window have been confirmed through multiple subsequent CHIME/FRB observations of FRB 180916. Thanks to its repeating behaviour, using very long baseline interferometric (VLBI) techniques, the European VLBI Network (EVN) localized FRB 180916 to a star-forming region in a massive star-forming galaxy at redshift $z=0.0337 \pm 0.0002$ [32]. Recently, Hubble Space Telescope observations show it to be just outside the star-forming region [47].

This periodicity discovery gave credence to the idea of all repeating FRBs having some level of periodicity, similar to pulsars and RRATs. It was not long until the next periodic FRB was discovered, FRB 121101, the first repeater. FRB 121102 was found to have a period of 157 days with an 88-day active phase window [40], much larger than FRB 180916, but with a lower significance. Similar to FRB 180916, FRB 121102 was localized to a star-forming region, but in a dwarf host galaxy at redshift $z=0.193$ [46]. Due to the discrepancies in the periods and host galaxies of these FRBs, it is clear that more periodicity searches and localizations are needed to search for common characteristics in the population.

In order to perform and constrain periodicity searches, many bursts from a given repeater are needed. One method is to use a machine learning approach to search for bursts in data already searched using other methods such as PRESTO’s single-pulse-search algorithm. This ML method was

used by the Breakthrough Listen team on archived FRB 121102 data. They discovered 72 additional pulses to the 21 already discovered in the dataset [52]. This discovery validated the machine learning approach, and showed that there could be a large number of bursts in archived data that conventional search algorithms have missed.

1.7 Outline of this thesis

Following this introduction, this thesis is separated into 5 additional chapters. In Chapter 2 I will describe the CHIME/FRB instrument, its role in FRB search efforts, and some of its notable accomplishments. In Chapter 3 I will describe CHIME/FRB’s sister instrument, CHIME/Pulsar. In brief, CHIME/Pulsar receives baseband data in 10 independent directions within the CHIME primary beam. This allows for the tracking of 10 sky locations at any given time. Initially devised to search and study pulsars, CHIME/Pulsar is also able to perform FRB-repeater followup.

In Chapter 4 I describe the main work of this thesis; the development of a convolutional neural network (CNN) to search archived CHIME/Pulsar data of FRB repeaters. I will explain CNNs in general, and outline the details of the one used in this search. I will go through the extensive training process, give examples of training data, provide performance statistics, and present the findings from its search.

Chapter 5 will discuss work done to improve a scheduling algorithm used by CHIME/Pulsar to decide which sources will be observed by its 10 beams at any given time. Such an algorithm is needed as there are often

more than 10 desirable sources in the primary beam at any given time. Thus, a robust scheduling algorithm is needed to ensure sources are observed in an organized manner. A similar algorithm has been in existence since the start of CHIME/Pulsar operations, but upgrades were needed, so a new algorithm was written.

The 6th and final chapter will review the current state of these projects and outline their potential uses in the future. It will also summarize and conclude the thesis.

Chapter 2

CHIME/FRB Overview

2.1 Introduction to CHIME

The Canadian Hydrogen Intensity Mapping Experiment (CHIME) is a radio telescope located at the Dominion Radio Astrophysical Observatory near Penticton, British Columbia, Canada. This telescope was originally designed to map baryon acoustic oscillation (BAO) features in neutral hydrogen gas at high redshifts, the purpose of which is to study dark energy. Although this is still underway at CHIME, two additional projects have recently come online to take advantage of some of CHIME's features. The first of which, and the topic of this chapter, is the FRB search project at CHIME, named CHIME/FRB. CHIME/FRB utilizes a state of the art FRB search backend to scan the entire northern sky with daily cadence for FRBs. The second project, CHIME/Pulsar, performs similar but more targeted observations of pulsars and RRATs with daily cadence. A brief outline of CHIME/Pulsar can be found in Chapter 3. The CHIME/FRB collabora-

tion has published a detailed overview paper (see [1]). This chapter will cover the key aspects of the FRB search pipeline at CHIME and the reader is encouraged to read the full overview paper if interested in the finer details.

2.2 Why CHIME for an FRB search?

Although many upgrades were needed for a CHIME based FRB search to happen, some of the core characteristics of the telescope are what drew researchers to propose such a project. Chief of these characteristics is CHIME’s exceptionally large $\gtrsim 200 \text{ deg}^2$ field of view. Considering the telescopes location, this allows CHIME to scan the entire northern sky down to a declination of $\sim 20 \text{ deg}$. every day. With such a large FoV, and no region on the sky being favoured for CHIME’s BAO operations, the structure was built as a transit telescope with no moving parts. Thus, each point in CHIME’s FoV is observed daily as it crosses the telescope’s zenith.

Another reason for CHIME’s excellent FRB detection potential is its large bandwidth and powerful correlator. CHIME operates in the 400-800 MHz band, a very wide range compared to other telescopes used to search for FRBs. This wide range allows for exceptional sensitivity to broadband signals, as most FRBs are. This frequency range is also ideal for the search of Galactic Rotating Radio Transients (RRATs) and the study of pulsars, handled by CHIME/Pulsar. CHIME’s correlator provides 1024 independent beams within the telescopes primary beam. Even so, due to its initial cosmology focus, this correlator required significant upgrades to accommodate the intensity beam-forming needed in FRB searches.



(a)



(b)

Figure 2.1: **a:** Image of CHIME [1]. The four parabolic reflectors are clearly visible with the shipping containers containing the X portion of the CHIME FX-correlator and FRB compute nodes to the right (white boxes). **b:** Satellite view of DRAO and CHIME^a. The CHIME telescope and main DRAO facility are marked by the blue and red pins, respectively. The Pathfinder telescope can be seen in the bottom right. The nodes containing L1, L2 and L3 segments of the FRB search pipeline are housed in the shipping containers next to the telescope. An underground network link connects to the L4, FRB analysis, and archiving nodes in the DRAO facility.

^a=Google Maps, 2021: www.google.ca/maps

Together, its FoV, daily observation cadence, and powerful correlator gave CHIME the potential to detect multiple FRBs per day.

2.3 Physical Description

Table 2.1 contains some of the key characteristics of the CHIME telescope. Images of CHIME can be seen in Fig 2.1. The CHIME telescope consists of four parabolically shaped dishes oriented N-S. Each cylinder is 20-m wide and 100-m long with an inter-cylinder gap of 2-m and a focal length of 5-m. As mentioned, these cylinders are stationary and scan the sky as the Earth rotates. The reflecting surface was chosen to be galvanized steel with 16-mm openings to allow for snow to fall though while also minimizing ground noise. Along the focal line of each cylinder lie 256 dual-polarization feeds, for a total of 2048 signal paths (256 feeds \times 4 cylinders \times 2 polarizations).

2.4 FRB Search Pipeline

2.4.1 Correlation

The signal from the feeds is carried via coaxial cable to 2 special receiver huts (one every 2 cylinders) modified for radio frequency shielding. These shipping containers contain a custom F-engine built by the McGill Cosmology team lead by Prof. Matt Dobbs. This F-engine amplifies and bandpass corrects the signal before digitization. This occurs every $2.56 \mu\text{s}$.

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°- 1.2°
N-S FoV	~ 110
Focal ratio	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 min 45°: 14-7 min North Celestial Pole: 24 hr

Table 2.1: Key properties of the CHIME telescope. From [1]. Where two values are given, they are for the 400 MHz and 800 MHz edges of the band, respectively.

The output from these two F-engine huts is sent for spatial correlation by the GPU arm of the CHIME correlator: the X-engine. The X-engine resides in two commercial shipping containers located next to the telescope (Fig 2.1). Here, 256 processing nodes, powered by two dual-chip AMD Fire-Pro S9300x2 GPUs, each receive four input frequency channels of ~390 kHz bandwidth. These powerful machines first upchannelize the data to 16384 frequency channels then spatially correlate and create the 1024 formed beams

within the CHIME main beam. For each dish, the 256 north-south beams are formed to be evenly spaced in $\sin\theta$ (θ =zenith angle) and the 4 east-west beams are formed via exact phasing. Each X-engine node receives data from the F-engine at a rate of ~ 25.6 Gb/s. To avoid extremely high data rates into the later parts of the FRB search pipeline, the data is downsampled in time to ~ 1 ms. Thus, the upgraded CHIME correlator outputs 1024 total intensity beams with 16k frequency channels at 1-ms cadence.

2.4.2 CHIME/FRB Detection Pipeline

In this section I will outline the CHIME/FRB detection pipeline. As mentioned, for a more detailed review please see [1]. Generally, the pipeline is separated into 4 components. Each of these components will be reviewed in the following subsections. The components are labelled L1-L4. These stages follow the digitization, channelization, and correlation of the data by the F/X-engines. Each L-stage is responsible for a key process in the FRB search. In brief, L1 handles the computationally intensive dedispersion and event detection, L2 groups detections of multi-beam events, L3 determines the nature of the event (extragalactic or otherwise) and decides which action to take, and L4 implements the desired action and handles data archiving. A diagram explaining the process can be seen in figure 2.2. These processes happen on a total of 132 compute nodes housed in yet another RF-shielded shipping container adjacent to the telescope.

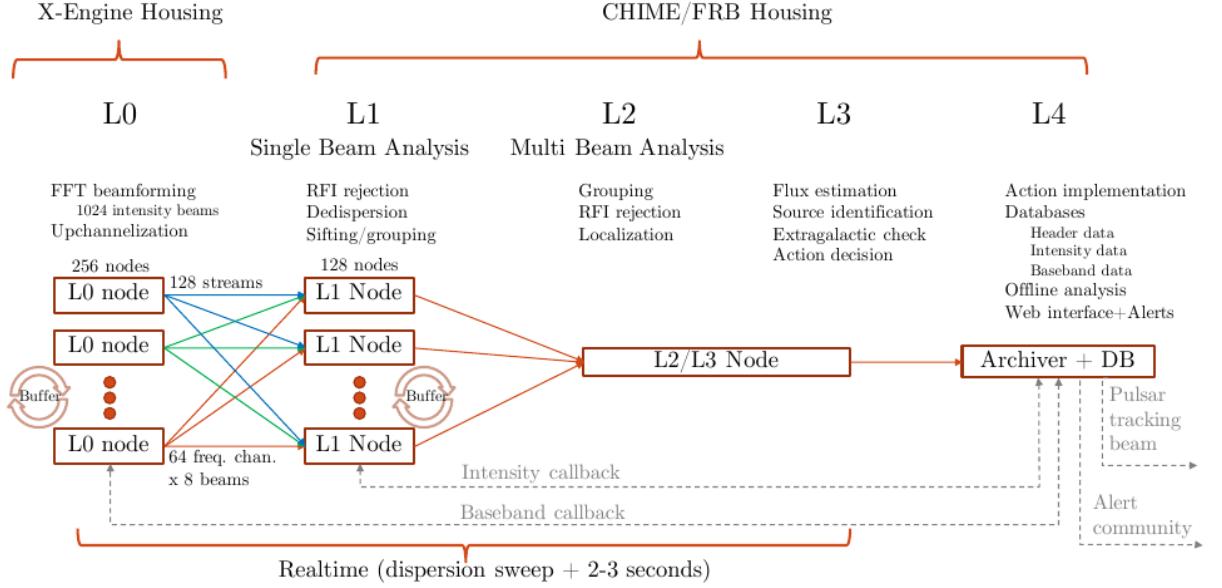


Figure 2.2: Conceptual block diagram of the FRB search pipeline, from [1]. Processes use a numbering scheme of LN, where N is process number, 0 being the most upstream. The main features of each process are listed.

L1: Dispersion and Event Identification

The L1 process uses 128 of the 132 compute nodes, with each node processing 8 beams. The fundamental purpose of this process is to convert the time and frequency data into time and DM which allows for the detection of dispersed signals. 5 parameters are considered in the transform, DM, arrival time, spectral index, scattering time, and intrinsic width. The computational cost largely comes from the large range of trial DMs, from 0-13000 pc cm⁻³. This cost is mitigated by the newly developed bonsai FRB dedispersion/search algorithm.

bonsai uses a tree dedispersion technique as opposed to direct dedispersion. This is due to the decreased $\mathcal{O}(TF \log F)$ computational cost as

opposed to $\mathcal{O}(TF^2)$ for directed dedispersion, where F and T are the number of frequency and time channels respectively. For a comprehensive review of tree dedispersion techniques see [45] and [2]. Different amounts of time downsampling are invoked to allow for dedispersion up to 13000 pc cm⁻³ and pulse widths of up to 100 ms. After RFI removal, pulses with signal-to-noise ratio (SNR) > 9 are passed to L2 (note that this is the current SNR threshold and has been changed many times throughout CHIME/FRB operations).

L2: Multi-Beam Event Grouping

Of the 132 compute nodes, L2 shares a node with L3. They are often referred collectively as L2/L3. In L2, candidates are grouped based on arrival time, DM, and sky position. This clustering is performed by the DBSCAN algorithm [15]. Classification of events as either RFI or astrophysical is then handled by a machine learning classifier. The classifier is trained on the SNR vs DM behaviour of the candidate. RFI tends to cause above threshold peaks at a very wide range of DMs. The distribution of SNRs in neighbouring beams is also considered. FRBs tend to have a focused distribution while RFI can be dispersed more evenly across many beams, including non-adjacent beams. In all, the ML classifier is based on a stochastic gradient descent classifier with a linear combination of the pattern of beams in a 3x3 grid around the highest SNR beam, the ratios of the second highest SNR and the average SNR to the highest SNR in the group, the total number of beams above threshold, and the grade given to the event by L1. This algorithm is implemented using `scikit-learn`. Events classified as RFI are

sent straight to L4 while astrophysical events get a position refinement and flux estimation before being sent to L3.

L3: Action Determination

The purpose of L3 is to identify if the event is of Galactic or extragalactic origin, as well as determine if it came from an already known source. Known source identification is handled by referencing lists of known pulsars (ATNF catalog, [31]), known RRATS (RRATatalog¹), and known FRBs (FRBCat, [38]). CHIME/FRB also references its own internal database to identify repeat bursts from its unpublished FRBs. Sky position and DM are specifically compared, weighted by their measurement errors, to the sources in the catalogs.

It is then determined if the source is of Galactic or extragalactic origin. Source DM is compared to the estimated maximum galactic DM calculated along the line of sight using the NE2001 [11], and YMW16 [50], models individually. If the source DM exceeds the maximum galactic DM from both of these models by more than 5σ the source is classified as extragalactic. If the source DM exceeds these models by $2-5\sigma$ the source is classified as ambiguous. If the source DM does not exceed the predicted max DM by 2σ the source is classified as Galactic.

An action is then taken depending on the characteristics of the pulse (classification, DM, SNR, etc). These actions include ignoring the event, storing the meta-data of the event in the database, calling back buffered L1 intensity data, dumping buffered baseband data from L0, sending an

¹<http://astro.phys.wvu.edu/rratalog/>

alert to the community, or requesting that CHIME/Pulsar track the source position (see Chapter 3 for source tracking with CHIME/Pulsar).

L4: Taking Actions and Event Archiving

The L4 process runs on 1 of the 132 previously mentioned compute nodes. This process receives and implements the actions chosen in L3. This process also hosts the CHIME/FRB archive. This is where all meta data and analysis products for each event are stored, for every event passing L1, even those classified as RFI in L2.

Every potential FRB candidate is then sent for human verification. Events are examined by a member of the collaboration to confirm or override the automatic classification. This person also has the opportunity to manually adjust some of the burst plotting metrics such as DM, downsampling factor, and zap noise channels. This is particularly useful for confirming faint bursts ($\text{SNR} \lesssim 10$) as slight dedispersion errors or noisy channels can wash out the burst. Figure 2.3 shows an example of a clear, high SNR burst, a burst which required significant channel masking, and a clear RFI event.

Each event requires two individuals to separately classify it before it is confirmed. Due to storage limitations, events classified by two humans as RFI are removed from the database. It is from this database that data are pulled for scientific study.

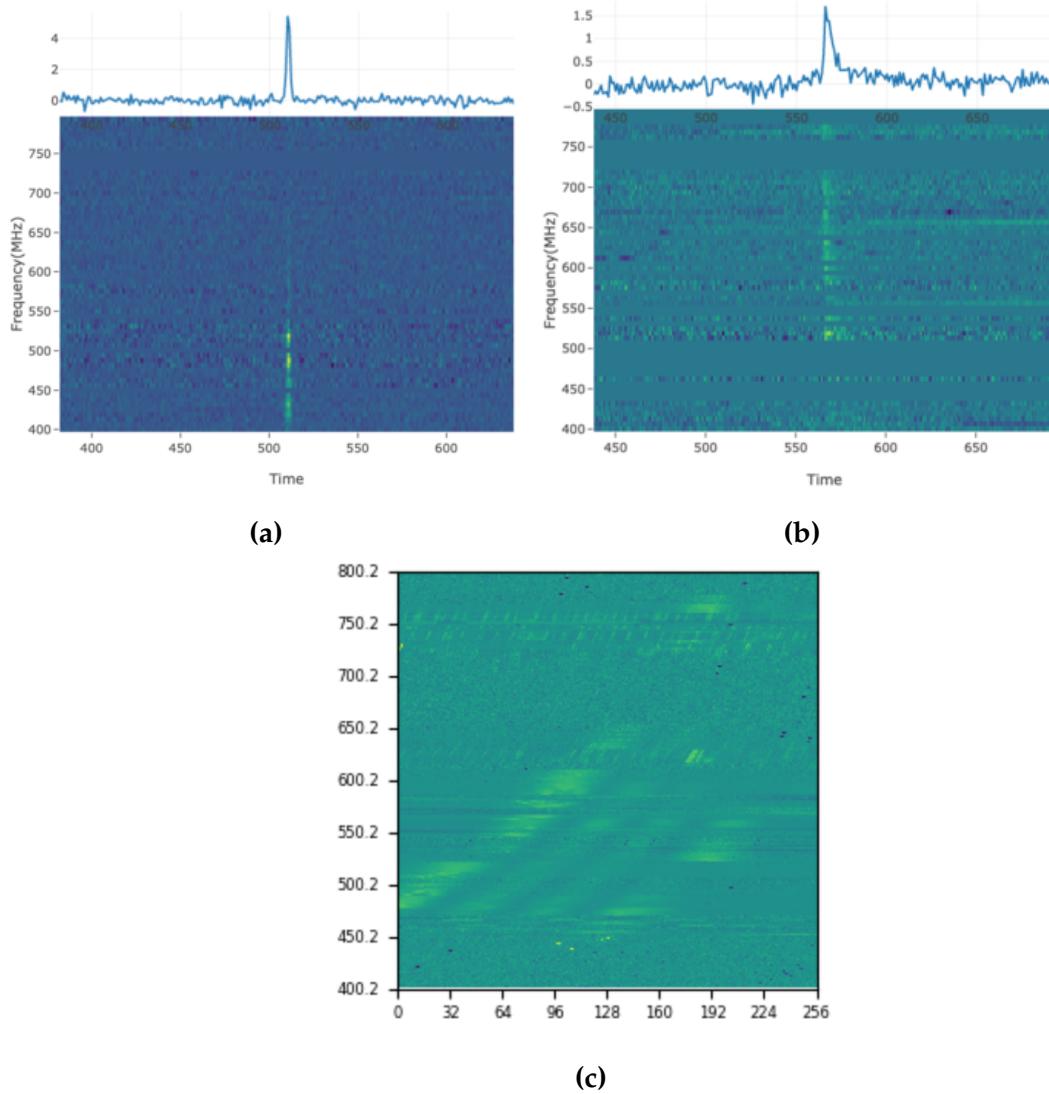


Figure 2.3: **a:** A FRB detected by the CHIME/FRB search pipeline on May 19, 2019. This burst has a particularly high SNR of 39.7 and a DM of 1170.8. **b:** A FRB detected on June 6, 2019 with a DM of 277.3. It is clear by the blank horizontal streaks that this event required significant channel removal in order to suppress RFI. **c:** An example of clear RFI. X and Y axes are the same as the latter two plots (frequency vs time in ms). The sweeps in intensity data correspond to the DM applied to it, meaning this “burst” corresponds to a DM of 0, common with RFI. Events such as this are common when an airplane transits the CHIME beam.

Chapter 3

CHIME/Pulsar Overview

3.1 Introduction to CHIME/Pulsar

A sister project to CHIME/FRB, CHIME/Pulsar denotes a separate backend system built for the CHIME telescope. The main purpose of this project is to use CHIME’s daily observing cadence of all sources having declinations greater than -20 degrees to study pulsars in the Milky Way. In practice, CHIME/Pulsar can observe up to 900 sources every sidereal day and observes all sources in the Northern Hemisphere every \sim 2 weeks. The only project similar to this in scale is the newly developed MeerKAT pulsar timing program in the Southern Hemisphere¹.

This chapter will closely follow the detailed CHIME/Pulsar system outline in [10]. I will first discuss some of the hardware comprising the CHIME/Pulsar instrument. This section is intentionally kept brief as CHIME/Pulsar shares much of the same hardware as CHIME/FRB, outlined in Chapter 2. I

¹<https://science.ska.ac.za/meerkat>

will then provide an overview of the CHIME/Pulsar backend, placing emphasis on CHIME/Pulsar’s abilities to observe FRB repeaters. This Chapter will conclude with an outline of the main scientific topics CHIME/Pulsar hopes to explore.

3.2 Hardware and Backend

Most of the hardware comprising CHIME/Pulsar is the same as those for CHIME and CHIME/FRB. Where CHIME/Pulsar differs is after the data stream is split and sent to the various CHIME projects. For CHIME/Pulsar, the data are sent to 10 independent compute nodes equipped with a Supermicro motherboard, and Intel Xeon CPU, and a solid-state drive. The data are buffered on 128 GB of RAM (per node) and then sent for coherent dedispersion using a fast-Fourier-transform (FFT) based algorithm on an NVIDIA Titan X GPU. The data are then written to a custom archiver consisting of redundant arrays of independent disks (RAIDs) at an average rate of 67 Mbps, for short term storage. Eventually, the data are transferred via hard disk shipments to CHIME/Pulsar institutions and Compute Canada² for offline analysis and long term storage. Alternatively, the option to bypass the coherent dedispersion and skip to saving the data to the archiver can also be chosen. A schematic diagram of how the CHIME telescope data stream connects to the CHIME/Pulsar backend can be seen in Figure 3.1.

²www.computecanada.ca

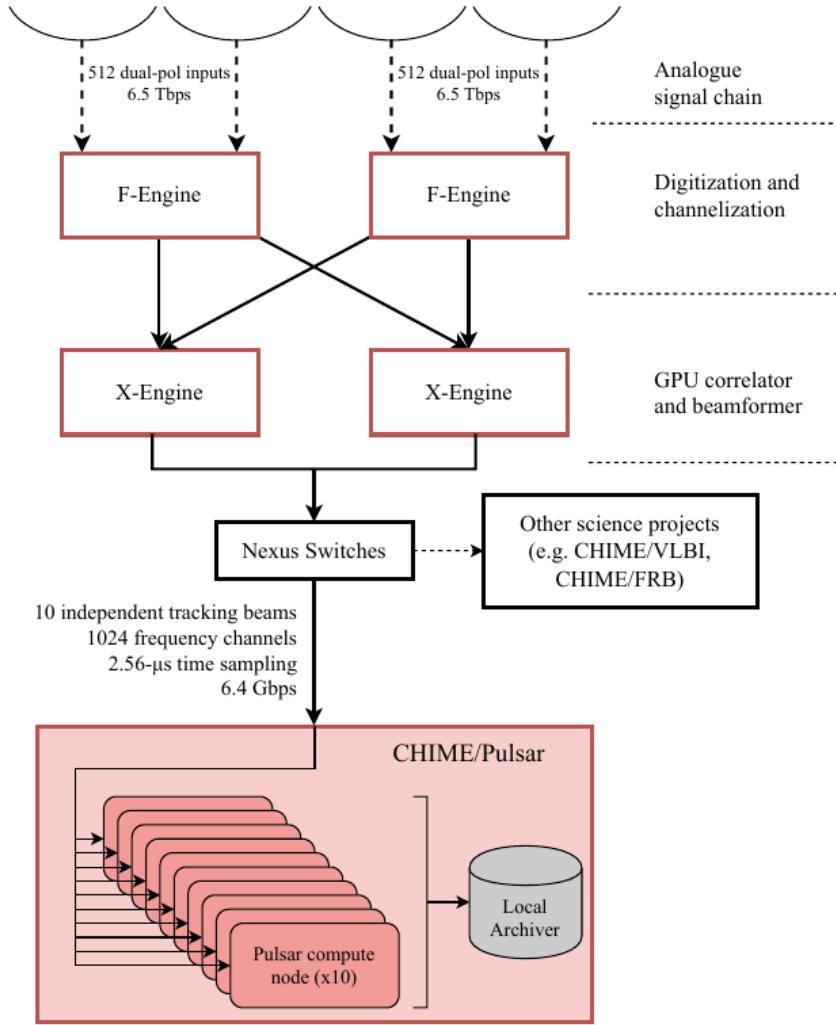


Figure 3.1: From [10]. Schematic of the CHIME telescope signal path extended to show CHIME/Pulsar backend. As mentioned in Chapter 2, the F-Engine handles the digitization and channelization of the analog signal inputs and the X-Engine handles the correlation and beamforming. The data are then sent to a Nexus switch which passes them to their respective project backend. In this case, the data are sent to the 10 pulsar compute nodes and eventually end up in the local archiver.

3.3 Software

The crux of the CHIME/Pulsar instrument is its 10 dual-polarization tied-array beams. These beams are formed digitally within the CHIME X-engine by summing all 1024 inputs phased to specified celestial coordinates. These 10 beams can follow 10 independent sky positions. Although these beams can track a source across the sky, sensitivity decreases drastically when the source is outside CHIME’s primary beam. Thus, tracking is only performed on sources when they are within the primary beam.

3.3.1 Data Types

Scientifically, the main purpose of CHIME/Pulsar is to monitor radio pulsars with regular cadence. Such observations are taken in "fold mode" where baseband data are dedispersed and manipulated to give time-averaged pulse profiles using a set of pulsar rotation parameters for coherent averaging of the signal. Observations can also be taken in "filterbank mode," where the dedispersed baseband data are downsampled and converted into a Stokes-I data stream. Finally, in exceptional cases, raw baseband data are collected and saved.

Filterbank mode is especially useful for detecting non-periodic transients such as FRBs. The format of it begins with a header containing information about the data such as the date it was taken, which telescope it came from, the bandwidth of the telescope, sampling time, etc. What follows this header is a 2D stream of frequency channels vs time. This stream is usually visualized by chopping the file into a series of frequency vs time plots.

These are referred to as waterfall plots and can be seen in Figure 2.3 (note that the integrated pulse profile in **(a)** and **(b)** of this figure is not saved in the filterbank file format). If an FRB is known to repeat, CHIME/Pulsar can collect filterbank data at its location for offline dedispersion and searching for bursts. This is what was done for the 15 repeaters searched with the convolutional neural network discussed in Chapter 4.

The data packets are received by the CHIME/Pulsar compute nodes, rearranged into a polarization-time-frequency series, coherently dedispersed at a specified DM, and downsampled in time. The data is then manipulated into the community-established filterbank format and written to disk.

3.3.2 Tracking Beam Scheduler

The 10 available beams of CHIME/Pulsar allow for multiple sources to be observed at once. It is often the case however that more than 10 interesting sources are within the CHIME primary beam. To decide which sources get observation priority I developed an automated scheduling algorithm. If more than 10 sources are available to be scheduled at one time, this scheduler draws 10 from a probability distribution of the sources based on a set of user defined weights. These weights create a priority system for the sources. There are five base weight values, each with double the probability of the previous level. As sources are observed, these values are gradually decayed based on observation frequency to ensure that lower priority sources are eventually observed. This system can also be entirely overridden, allowing for target-of-opportunity observations of immediately interesting sources. This algorithm does not however retain a history of past observations, and

has until recently been relatively simplistic. The recent upgrades I made to the algorithm are discussed in Chapter 5.

3.4 Scientific Intent

The main science driver for CHIME/Pulsar has been the long-term timing of pulsars, with a focus on Pulsar Timing Array (PTA) targets. The motivation for such a long-term study is that nanohertz-frequency gravitational waves using a PTA depend on the robustness of the MSP sample. Naturally, CHIME/Pulsar has thus been primarily observing NANOGrav sources³ and will play a key role in their data releases of the stochastic GW background at nanohertz frequencies.

CHIME/Pulsar also observes any newly discovered pulsars found by the Green Bank North Celestial Cap (GBNCC) [43] and PALFA [25] surveys to help confirm and characterise the sources. All observable, published pulsars (as listed in the ATNF pulsar catalog⁴) are also observed.

Another interesting field is the study of plasma structures in the Galaxy. Effects of such structures can be seen in pulsar data via temporal variation of DM [23, 24], frequency-dependent DM [12, 14], scintillation [5, 49], and multi-path scattering/pulse broadening [4, 33]. Timing pulsars with daily cadence can illuminate these features and help in small scale plasma structure studies.

With the recording of baseband data, CHIME/Pulsar has access to full Stokes information. This allows for polarization studies to be performed.

³<http://nanograv.org/>

⁴www.atnf.csiro.au/research/pulsar/psrcat/

Ng et al. 2020 has already published many of the RMs for the detectable pulsars in the northern hemisphere [37]. This has helped probe the magneto-ionic medium surrounding the pulsar and in the interstellar medium.

Also thanks to CHIME/Pulsar's near daily observation cadence, glitch monitoring of pulsars will be very powerful. Over 200 pulsars have been shown to "glitch," (where the rotation period abruptly changes) and studying them post-glitch provides information of the interior structure of neutron stars [17, 18]. CHIME/Pulsar has the ability to search for timing variations consistent with pulsar glitches and give nearly realtime alerts to the community.

Most relevant to this thesis, and the study of FRBs, is CHIME/Pulsar's ability to search for non-periodic transients such as FRBs and RRATs. CHIME/Pulsar can observe at a much higher time resolution ($2.56\ \mu s$ vs 1 ms) and coherently dedisperses the data while CHIME/FRB does so incoherently. Moreover, if a tracking beam is placed on a source, it will not miss detection by falling between beams, as is sometimes the case with CHIME/FRB. Altogether, these aspects make CHIME/Pulsar an ideal instrument for repeating FRB follow-up. The team simply adds the repeater to the scheduling algorithm and searches the resulting filterbank data for bursts. However, until recently efforts to search these terabytes of data for bursts using conventional single pulse search algorithms such as PRESTO had been unsuccessful. A new method was developed using a convolutional neural network. This network and its results are discussed in the next Chapter.

Chapter 4

Machine Learning Search for FRB Repeaters

In this Chapter I will outline a convolutional neural network (CNN) that was developed to search terabytes of archived CHIME/Pulsar observations of FRB repeaters. I will discuss its architecture, how it was trained, its performance statistics, and finally its results.

4.1 Neural Networks

4.1.1 The Basics

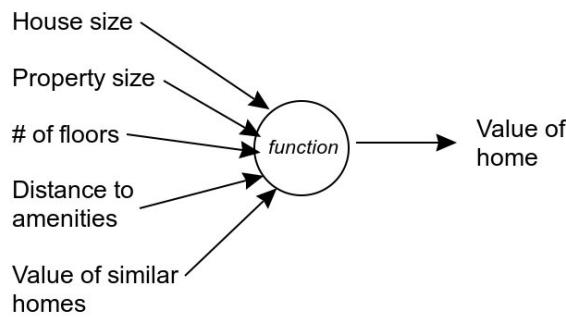
The majority of the information from this section comes from the excellent online resource Neural Networks and Deep Learning¹ by Michael Nielsen.

At their core neural networks (NNs), and machine learning (ML) models in general, are made to predict an outcome based on a set of input data.

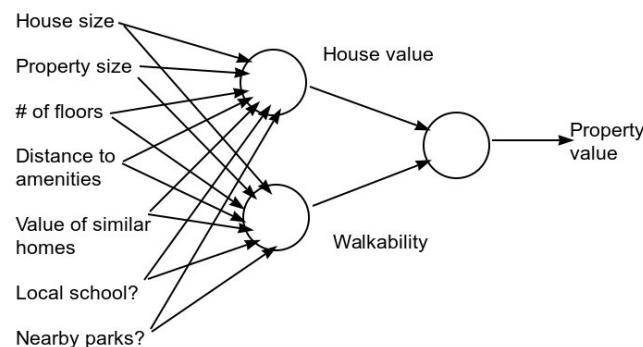
¹<http://neuralnetworksanddeeplearning.com>

These data could be as simple as household income used to predict mortgage rates, to as complex as realtime video and radar information used to drive a car autonomously. The simplest mathematical tool used for prediction is a linear function, $y = mx + b$. That is, for an x data point, the slope and intercept of the linear fit can be tuned such that the prediction, y , is accurate. Now lets expand this idea to multiple input data points. Let's say you're trying to value a home. Some things you may consider in your valuation are its size, number of floors, property size, distance to amenities such as groceries stores, and the value of similar homes in the area. All these things can be thought of as the input variables into our function which outputs a housing price prediction. Figure 4.1a shows a conceptual drawing of this idea. The input data are on the left, the circle represents the function with its tuneable parameters, and the prediction is output on the right. The circle containing the tuneable function is often called a neuron and the function is called the activation function. This example is the most basic 1-neuron neural network.

Now lets expand on this. Say you now want to predict intermediate characteristics of this house and have them contribute to the value of the property. This could be something like the value of the house itself, as well as the walk-ability or the quietness of the neighbourhood. These aspects are likely to be the result of the previously mentioned input data. This would result in a network similar to Figure 4.1b. The middle set of neurons is known as a hidden layer and is kept completely "under the hood". This is considered a fully connected hidden layer since each input neuron connects to each hidden neuron. Each neuron has its own activation function, and



(a)



(b)

Figure 4.1: Conceptual examples of a single neuron neural network (a) and a two neuron, single hidden layer NN (b). Both of these are examples of linear neural networks. Each neuron (circle) houses an activation function that is trained to help predict home or property value based on a set of input data (house size, property size, etc.).

the output of these functions are used as the input to the final layer neuron. As NNs get more complex they can contain multiple hidden layers with many neurons in each.

4.1.2 Learning

But how do we tune the parameters of the neurons such that we get an accurate prediction? This is where the training of the NN comes in. In order to train the network we need to pass it a set of labelled data. Going back to $y = mx + b$ this would mean that we tune m and b by passing the network many examples of where x and y are given. This extends to larger networks where you train on a data set that consists of the types of input data the network will encounter and the expected prediction.

First lets start by distancing ourselves from $y = mx + b$. Although it is a great function, the problems NNs face can rarely be solved by a combination of linear functions. We will still use two parameter activation functions and we refer to those parameters as *weights* and *biases*. The weight is commonly multiplied by the input data and the bias is added to it. This value is then passed through the activation function.

The way that a network is trained is by systematically making small adjustments to the weights and biases of the neurons which propagate through the network and result in a more accurate network prediction. The accuracy of the network on a specific set of input data is usually quantified by a cost function. The most conceptually understandable cost function is the

squared-cost-function:

$$C(w, b) \equiv \frac{1}{2n} \sum_x ||y(x) - a||^2, \quad (4.1)$$

where w is the set of all weights in the network, b is the set of all biases, n is the total number of training inputs, a is a vector of outputs of the network with x input, $y(x)$ is the expected prediction, and the sum is over all training inputs, x . If our network is accurate, C will be small, and large if it is inaccurate. The idea now is to make adjustments to w and b such that we get a decrease in C . This is done via gradient descent.

Considering the general cost function $C(v)$ where $v = v_1, v_2, \dots$, any change in v_i gives a change in C of

$$\Delta C = \frac{\partial C}{\partial v_1} \Delta v_1 + \frac{\partial C}{\partial v_2} \Delta v_2 + \dots \quad (4.2)$$

or

$$\Delta C = \nabla C \cdot \Delta v. \quad (4.3)$$

Since we want to decrease C , ΔC should always be negative. To achieve this we choose

$$\Delta v = -\eta \nabla C. \quad (4.4)$$

This gives

$$\Delta C = -\eta \|\nabla C\|^2. \quad (4.5)$$

Here, η is defined as the non-negative *learning rate* of the network, which ensures that ΔC is always negative (towards a more accurate prediction).

We can loop the process of updating $v \rightarrow v' = v - \eta \nabla C$, decreasing C each time, resulting in increasingly accurate predictions, until we reach a global minimum for C (as accurate as the network can get).

Thinking back to our weights and biases, we may have many variables to update considering our network's architecture. If we denote w_k as the k -th weight in the network and b_l as the l -th bias, the update rule becomes

$$w_k \rightarrow w'_k = w_k - \eta \frac{\partial C}{\partial w_k} \quad (4.6)$$

$$b_l \rightarrow b'_l = b_l - \eta \frac{\partial C}{\partial b_l}. \quad (4.7)$$

In order to speed up learning, an approach called *Stochastic gradient descent* is used. Here, instead of trying to calculate ∇C for each training input and averaging across the entire data set, ∇C is estimated by calculating ∇C_{x_j} for a subset of randomly chosen training inputs from the larger training sample. This subset is known as a mini-batch. If we denote the training sample in a mini-batch as X_j and have m mini-batches, Equations 4.6 and 4.7 become

$$w_k \rightarrow w'_k = w_k - \frac{\eta}{m} \sum_j \frac{\partial C_{X_j}}{\partial w_k} \quad (4.8)$$

$$b_l \rightarrow b'_l = b_l - \frac{\eta}{m} \sum_j \frac{\partial C_{X_j}}{\partial b_l}. \quad (4.9)$$

Thus, for the network to learn, we first select a mini-batch from the training data set, calculate each weight and biases contribution to the cost function and update via Equations 4.8 and 4.9, then move on to the next mini-batch. Completion of training on a mini-batch is called an *epoch* of training.

4.1.3 Convolutional Neural Networks

Until now the type of NN considered has been linear. That is, each layer is a linear set of neurons. In this section I will outline *Convolutional Neural Networks* (CNNs) and how they are excellent at image recognition. There are three main ideas to CNNs: local receptive fields, shared weights, and pooling.

Local Receptive Fields

In the previous example, each layer of the network was a line of fully connected neurons. In a CNN, we replace this 1-dimensional line with a 2-dimensional plane of neurons (see Fig. 4.2). In the case of images, each input neuron usually corresponds to a single pixel in the image. Instead of connecting each input neuron to each neuron in the next layer, we will only connect a subset to a single neuron in the next layer. This connection scheme can be seen in Figure 4.3. In the case of Figure 4.3, this is taken as a 5×5 grid of input neurons. This grid is known as the *local receptive field* (LRF) with a kernel size of 5. The neuron in the next layer can conceptually be thought to trace the characteristics of this region of the original image, or the importance of this section to the classification of the image. We can then step the LRF over one or more pixels at a time, connecting to the next neuron in the hidden layer, until the entire image is scanned. This process can then occur again for the hidden layer into a second hidden layer, and so on until the end of the CNN architecture.

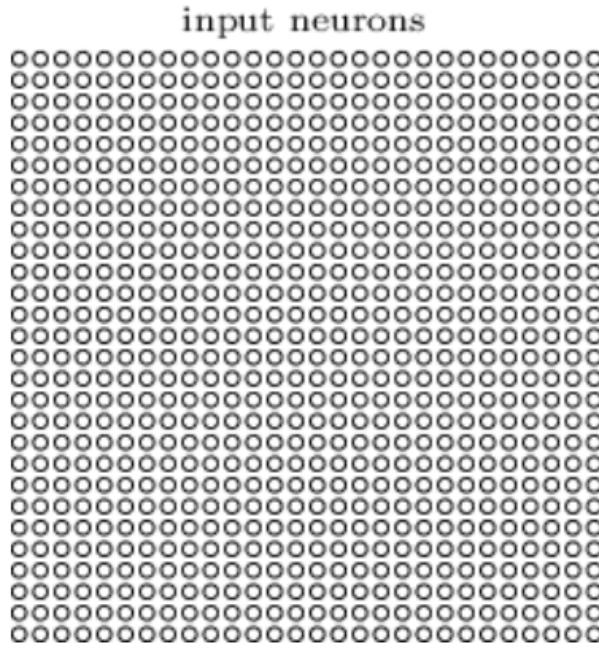


Figure 4.2: Conceptual diagram of the input data to a convolutional neural network. Instead of the linear input data seen in Figure 4.1a, the CNN input layer is a 2-dimensional grid of inputs, usually the pixels of an image.

Shared Weights and Biases

What is commonly used in CNNs is the idea of shared weights and biases. This means that the weights and biases of the LRF remain constant throughout its scan of the image/layer. The specific parameters in the LRF need not be the same, but must remain constant throughout the scan. This means that all the neurons in the first hidden layer detect exactly the same feature, but they search in different locations of the image. This is one of the fundamental reasons for why CNNs are excellent at image recognition. It does not matter where in the image the item of interest is, the network will still detect the key features it has been trained to find.

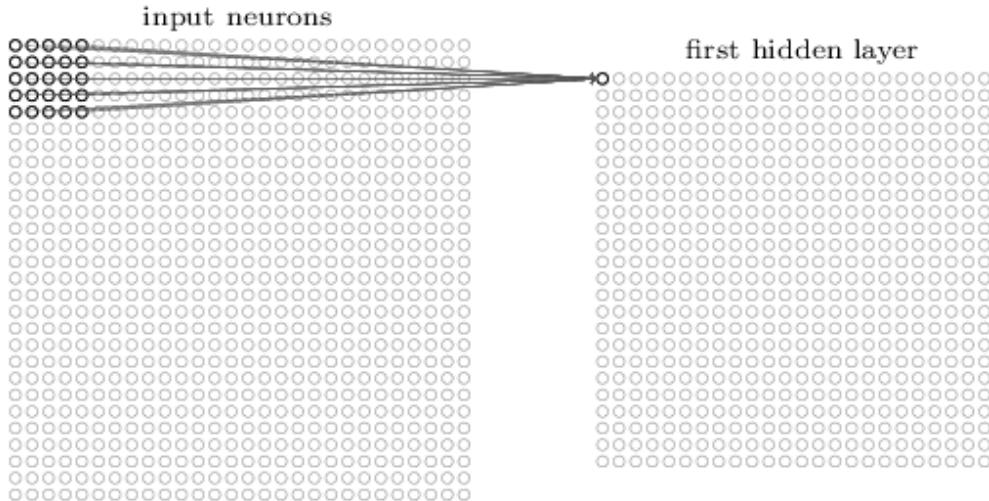


Figure 4.3: The connection of a 5x5 kernel of input neurons to the first neuron in the first hidden layer. This kernel will then be stepped a configurable number of neurons to the right, and will connect the new 5x5 input neurons to the second neuron in the first hidden layer. This process will continue until the entire input layer has been scanned.

For the CNN to search for multiple features in its first pass of the original image, multiple LRFs are used to create multiple first-hidden-layers (visualized in Figure 4.4). These layers are commonly referred to as *feature maps*. The more features that need to be searched for, the more feature maps are needed. These maps are then searched again for more abstract features if a second hidden layer is constructed.

Pooling

Although I've stated it, the process of input image to first hidden layer, to second hidden layer to third, etc, isn't exactly correct. There is often an intermediate step between hidden layers known as pooling. In brief,

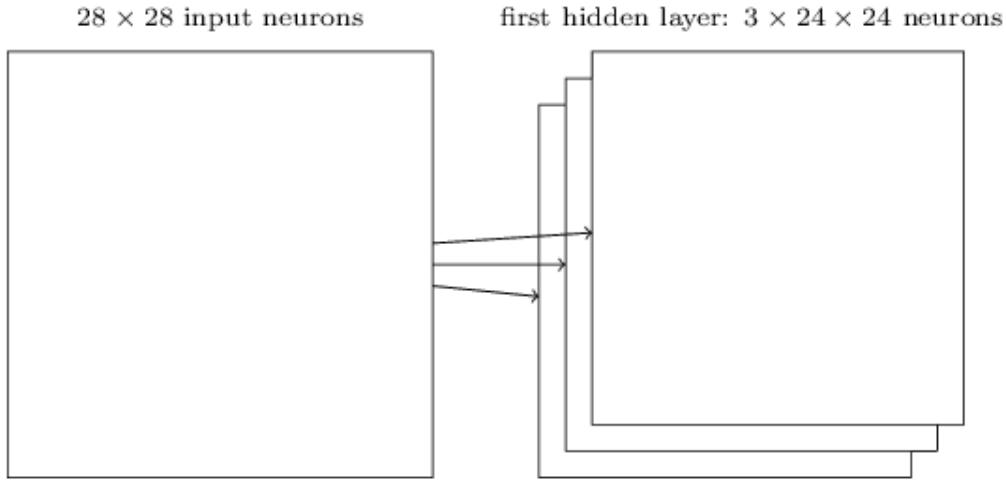


Figure 4.4: Conceptual image showing that the input layer can be scanned multiple times to produce multiple feature maps in the first hidden layer. Here the 2D grid of circles seen in Figure 4.2 is replaced with a square. These three feature maps in the first hidden layer are the 2D analog to the 2 neurons in the first hidden layer of Figure 4.1b. Each feature map corresponds to the scanning of the input layer with a fixed LRF. Thus, they search the input image for specific key features.

each pooling layer takes the output from a feature map and creates a condensed feature map. This is achieved by taking either the maximum (for max-pooling) or average (for average-pooling) of a small grid of the feature map. Figure 4.5 shows an example of a 2×2 max-pooling layer. For each convolutional layer we add a pooling layer before the next hidden convolutional layer. The motivation for this is that we are only interested in finding a particular feature anywhere in the image, not necessarily in specific location. If said feature is in the image, it will be found by the feature map and

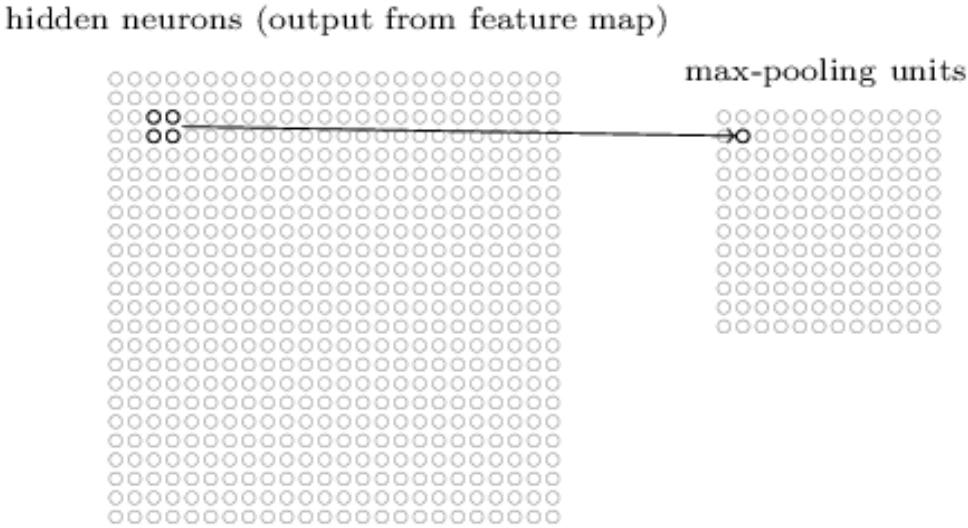


Figure 4.5: Similar to Figure 4.3, this figure shows the maximum pooling of a convolutional layer. Pooling is usually done in small kernel sizes which take either the maximum or average of the neurons in the kernel.

emphasised by the pooling layer. This process also results in many fewer parameters needed in the later layers.

Finally, for classification purposes, it is common to fully connect the final pooled layer to a linear layer which outputs the prediction. In the next section I will outline the architecture of a CNN developed to search CHIME/Pulsar data of FRB-repeaters for bursts.

4.1.4 A CNN for FRB Repeater Searching

Architecture

I will now outline the architecture of the CNN used to search the CHIME/Pulsar data for FRB repeaters. The data passed to the network consists of a set of

256x256 pixel images with a 7 pixel header. Each pixel is 32 bits in length. Aside from the header, which contains information about the sources name, timestamp, DM, and more, each input to the network is the 256x256 image corresponding to a frequency vs time plot (waterfall plot) similar to those seen in Figure 2.3. The goal of the network is to search such images for the characteristic vertical streak of a dedispersed FRB.

The network used in this search consists of 8 layers, 5 convolutional and 3 linear. Each convolutional layer neuron uses the Rectified Linear Unit (ReLU) function as its activation function. An example plot of the ReLU function can be seen in Figure 4.6. This function was chosen due to it being the default choice in the deep learning field as well as its high gradient at the positive extreme of the function. This yields to fast gradient descent and fast learning.

The first layer of the network takes a 256x256 chunk of the filterbank file as input and creates 64 feature maps, each formed from a LRF with a kernel size of 7 and a stride length of 1, padded with 3 rows/columns of zeros on all sides of the input data. This gives each feature map a size of 256x256. These maps are then average pooled with a kernel size of 2 and stride of 2, leaving 64 128x128 feature maps. Layer 2 takes these 64 maps and scans them with a kernel size of 3 and stride of 1, with padding of 1, and creates 128 maps of size 128x128. This layer is then also average pooled in the same way as layer 1, resulting in 128 pooled layers of size 64x64. Layers 3, 4 and 5 all use a kernel size of 3, padding of 1, and max pooling with a kernel size of 2 and stride of 2. They take the 128 feature maps from layer 2 and create 256, then 512, then another 512 maps, respectively with sizes 32x32, 16x16,

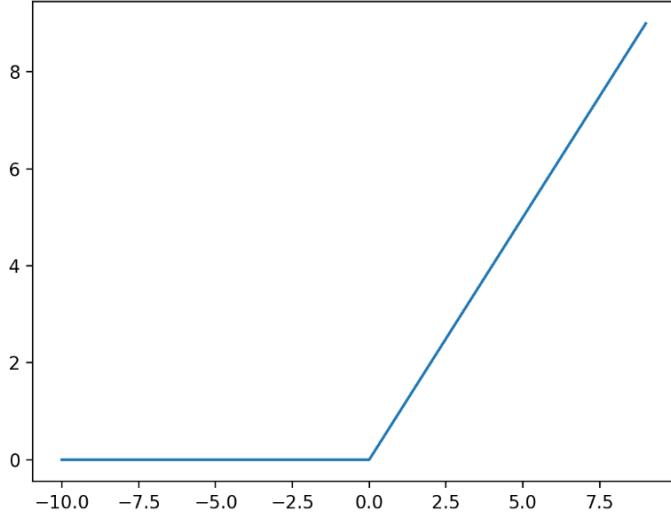


Figure 4.6: The rectified linear unit function (ReLU) define as $f(x) = \max(0, x)$. This function is often used in CNNs as the activation function due to it maintaining a high gradient when x is large. Another common activation function is $f(x) = \tanh(x)$, but this suffers from low gradients at large value of x . Maintaining a high gradient ensures faster training of the CNN model.

and 8x8, respectively. Next is the first of 3 linear layers. This first linear layer consists of $8 * 8 * 512 = 32768$ neurons which fully connect to the next linear layer of 4096 neurons. This layer then connects to the final layer of 128 linear neurons which have an output of the form (a,b). The max value of this output is taken and everything with a max value > 0 is classified as a potential FRB. The learning rate η was set to 0.006 and training occurred over 5 epochs. This learning rate was found though trial and error until a rate giving a sufficiently high accuracy was found.

Data Preparation

A program for data preparation is needed as the data from CHIME/Pulsar comes as a single dispersed filterbank file of minutes in length. This file needs to be dedispersed and separated into the 256x256 + 7 waterfall plot and header information. A custom program named *gen_samples* was written in C++ to handle this. The following list is a brief summary of the steps taken by *gen_samples* to prepare the data.

1. Filename, DM, and RFI mask file are passed to *gen_samples*
2. Filterbank header information is skipped over to get the frequency vs time data
3. The temporal shift for each frequency channel given the DM of the repeater is calculated
4. A block of data 4096-ms long is read (allows for downsampling to 16-ms)
5. The data are downsampled into 1-ms wide time bins, down from CHIME/Pulsar's native $2.56 \mu\text{s}$
6. For each block the data are dedispersed using the DM shifts from point 3
7. This block is then looped over for temporal downsampling of 1-ms, 2-ms, 4-ms, 8-ms, and 16-ms, creating a 256x256 plot each loop
8. These plots are saved as .plt type files to be read by the CNN

Training

The key to a successful NN is the robustness of its training data. Theoretically, the training of this model to detect FRBs seems quite simple. There are two classes needing to be identified, FRBs and non-FRBs. So the training set must consist of a large sample of labelled FRBs and a large sample of labelled non-FRBs. These non-FRBs can consist of general noise, or common RFI patterns like those seen in Figure 2.3. A network such as the one described requires many thousands of training examples in order to get a $> 99\%$ accuracy. This might seem like a problem considering there are only a little over 500 FRBs published to date. To overcome this, extensive data augmentation was performed.

Augmentation of the non-FRB training set was fairly trivial. This involved simply recreating similar .plt files consisting of Gaussian noise. Although including common RFI signatures into the non-FRB sample would be desired, an initial model was trained using only these noise samples. The performance of this model was deemed acceptable (mentioned later in this Chapter), so the analysis of the CHIME/Pulsar data began, as opposed to further improvements of the CNN. Including RFI signatures in the non-FRB training sample is the obvious next step if improvements to the CNN are desired. It should be noted that downsampling of the training data was not necessary as the CNN does not recognize the downsampling factor, it simply sees the 256x256 pixel image. In total 3832 non-FRB training samples were used in training.

The augmentation of the FRB training set was slightly more involved. First, an initial version of the network was trained by plotting entire filter-

bank files known to contain bursts, isolating the plots that contained these bursts, and training on those. This initial model was expectedly poor and had an exceptionally high false positive rate. However, it did prove useful for finding further training bursts. This initial model was run on the filterbank data of the first 500 FRBs detected by CHIME/FRB. Once the false positives were sifted through and the bursts were identified they were augmented in 3 ways.

First the bursts were augmented in time. The 256x256 plots containing the burst found by the initial CNN were usually obtained by getting the number of seconds between the peak of the burst and the start of the filterbank file. If one then wants to plot the burst in the centre of the 256x256 image, they simply take the chunk of data that starts at $t_{burst} - 128t_{width}$ where t_{burst} is the location of the burst and t_{width} is the temporal width of each x-pixel (1-ms, 2-ms, 4-ms, 8-ms or 16-ms). To augment the bursts temporally, instead of subtracting $128t_{width}$ a number n was selected randomly between -128 and 128 and nt_{width} was subtracted. The bursts were also augmented in DM. This was done by first calculating the maximum DM offset that can be applied to the burst while still having it fully contained in the time window

$$DM_{max} = \frac{\Delta t}{4.15 \times 10^6 \left[\frac{1}{f_{min}^2} - \frac{1}{f_{max}^2} \right]} 100 \text{ pc cm}^{-3} \quad (4.10)$$

where Δt is the time between the start of the window and the location of the burst, $f_{min} = 400$ MHz, and $f_{max} = 800$ MHz. The DM offset applied to augment the burst is then randomly chosen between 0 and DM_{max} . The

final step in augmentation is the addition of Gaussian noise to the entire image. This noise is drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 0.0001. This standard deviation was chosen through trial and error after noticing that higher values tended to "wash out" many of the augmented bursts. A tag of "1" was entered into the 7 part header of the images to denote it as an FRB. Augmented RFI were tagged with 0.

Once augmentation of the bursts was complete, the entire set was inspected by eye to ensure that all washed out bursts were removed. The final set consisted of 3832 training bursts. Figures 4.7a and 4.7b show examples of the augmented training and RFI images, respectively.

Testing

Testing of the NN consisted of three stages. In the first stage, the network was simply tested on a subset of its training data set aside for testing. In the second stage the network was passed all the non-augmented filterbank files of the first 500 FRBs detected by CHIME/FRB. Although these were the bursts that the network was trained on, it was a good test to determine if the network could sufficiently ignore all the portions of the filterbank files where there were no bursts. That is, it was expected that the network would find all the bursts, but what was of particular interest was its false positive rate. The third test consisted of running the model on CHIME/Pulsar data of FRB 180916 (the third repeater found by CHIME/FRB, hereafter named R3). The data chosen for this test were from days where CHIME/FRB detected R3, so there was expected to be a burst in the CHIME/Pulsar data.

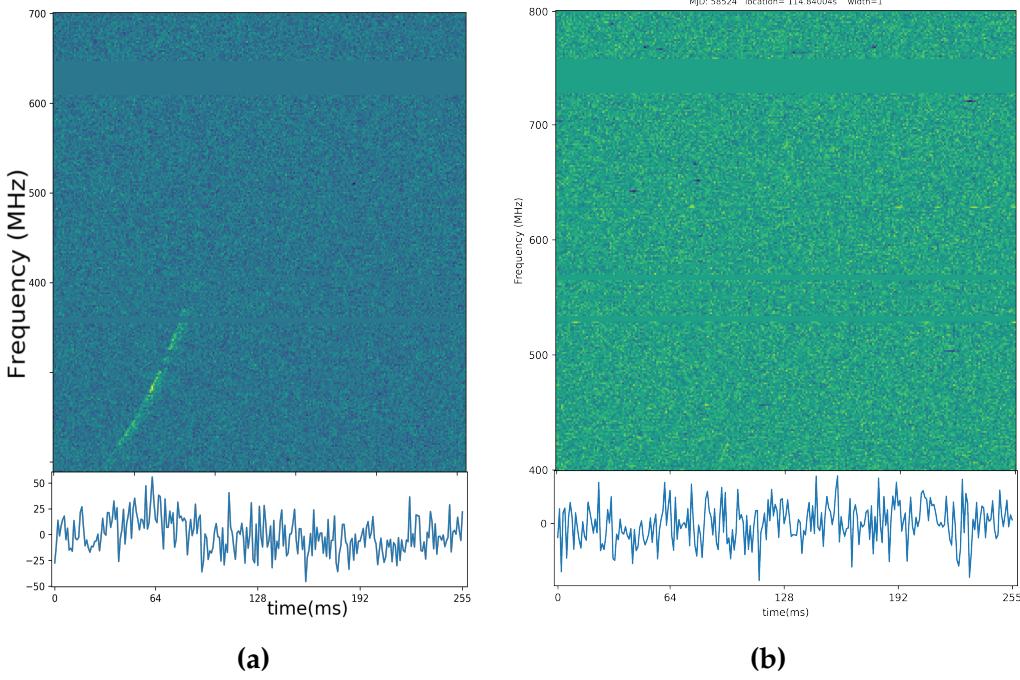


Figure 4.7: Examples of training data passed to the CNN to learn bursts (a) and RFI (b). RFI training data was derived from Gaussian noise and sifted through by eye to ensure any images with burst-like structures were removed. Burst training data were augmented in time, DM, and had low levels of Gaussian noise added. The time and DM augmentation can be clearly seen in **a** by the non-centring of the burst in time and the clear sub-optimal dedispersion (non-vertical burst). The entire training set totalled to 7664 images with equal parts bursts and non-bursts.

Since the network had not seen these data before, these tests were particularly good for determining a false negative rate.

The first stage of training yielded a 99.8% accuracy on the training data. In test 2, the model was tested on 496 filterbank files confirmed to contain bursts. The network successfully detected all but one burst, giving a false negative rate of 0.202%. The network was then set on all CHIME/Pulsar observations of R3. There are a total of 334 CHIME/Pulsar observations of R3, 21 of which were on days where CHIME/FRB detected a burst. The network successfully found all 21 bursts as well as 3 bursts not detected by CHIME/FRB. These three bursts will be discussed in the Results section of this Chapter. In total 1327 potential bursts were identified by the network. Although this number of false positives may seem high, only \sim 30 minutes of time was needed to sort through them by eye. This time commitment seemed adequate considering the \sim 190 hours of R3 observations that were analysed.

4.1.5 Results

The following sections will cover the results of analysing the entire record of CHIME/Pulsar data taken of FRB repeaters. Each repeater will be discussed individually and a summary table of all the bursts found by the CNN that were missed by CHIME/FRB will be provided at the end of the Chapter (Table 4.1).

The naming scheme developed by the CHIME/FRB team to denote repeaters, R#, where numbers correspond to successive repeater discoveries will be used. The repeaters where CHIME/Pulsar data were available were

R2-R17 excluding R8. Priority analysis was given to repeaters R7, R9, R10, and R13 since these were repeaters with only 2 recorded bursts and finding a third would significantly increase their claim as a true repeater (as opposed to the chance coincidence of two FRBs occurring at similar DMs at the same part of the sky).

R7, R9, R10, R13

For the two burst repeaters there are \sim 24.7 hours of R7 data (296 filterbank files times 5 minutes per observation), \sim 56.3 hours for R9 (260 times 13 min), \sim 216 hours for R10 (72 times 3 hours), and 14.7 hours for R13 (40 times 22 min). Unfortunately, for these \sim 312 hours of observations no additional bursts were found.

R2

The second CHIME discovered FRB repeater, R2, has amassed 21 bursts since its discovery on August 14, 2018. After searching the \sim 371 hours of CHIME/Pulsar data with the CNN, an additional 4 bursts can be added to this total. These bursts were detected on Feb 14 2019, Feb 21 2019, April 1 2019, and Sept 3 2019. They have SNR values of 6.6, 7.5, 9.4, and 7.0, respectively. These bursts can be seen in Figure 4.8.

R3

One of the most prolific repeaters, and one of only two repeating FRBs found to have a clear period [7], is FRB 180916, or R3 by the naming scheme

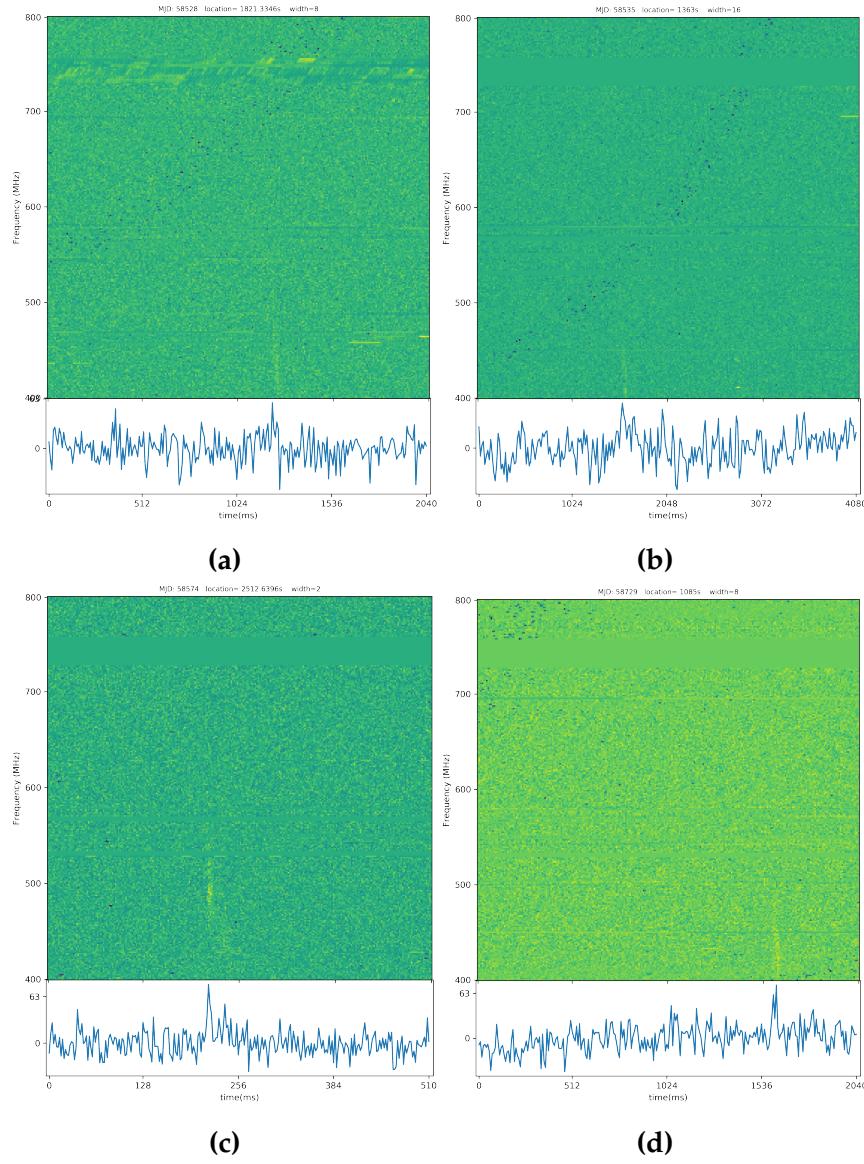


Figure 4.8: The 4 bursts found by the CNN for the repeater R2. SNRs for these bursts are 6.6, 7.5, 9.4, and 7.0 for **a**, **b**, **c**, and **d** respectively. They were detected on Feb 14 2019, Feb 21 2019, April 1 2019, and Sept 3 2019, respectively. These add to the 21 R2 bursts detected by CHIME/FRB.

used here. In total CHIME/FRB has detected 71 bursts since its initial detection on Sept 16, 2018. In total CHIME/Pulsar recorded \sim 189.8 hours of R3 observations from Nov 21, 2018 to Mar 16, 2020. In addition to the 71 CHIME/FRB bursts, the CNN found 3 bursts in the CHIME/Pulsar data. These bursts can be seen in Figure 4.9. Prior to being sent to the CNN all R3 data were dedispersed to $DM=350.0 \text{ pc cm}^{-3}$. The first of these three was from observations on Nov 14, 2019, and the latter two were both from Jan 4, 2020, separated by \sim 130 seconds. Their signal-to-noise ratios (SNRs) are 10.9, 8.8 and 13.7, respectively.

R5

Of the \sim 122 hours of observations of R5, only one burst was found by the CNN that was not detected by CHIME/FRB. This burst can be seen in Figure 4.10. This adds to the only 6 bursts detected thus far by CHIME/FRB of R5. This burst comes from data taken on July 24, 2019 and has an SNR of 6.12.

R6

CHIME/Pulsar observed R6 286 times between March 10, 2019 and March 17, 2020. Combined this gives \sim 60 hours of observations of R6. To date, CHIME/FRB has detected 9 bursts from R6. In the CHIME/Pulsar data, the CNN detected 4 additional bursts not detected by CHIME/FRB. This 44% increase in total R6 bursts will significantly aid in any future statistical analyses for the repeater, such as periodicity searches. The bursts can be seen in Figure 4.11. The first burst was detected in an observation from

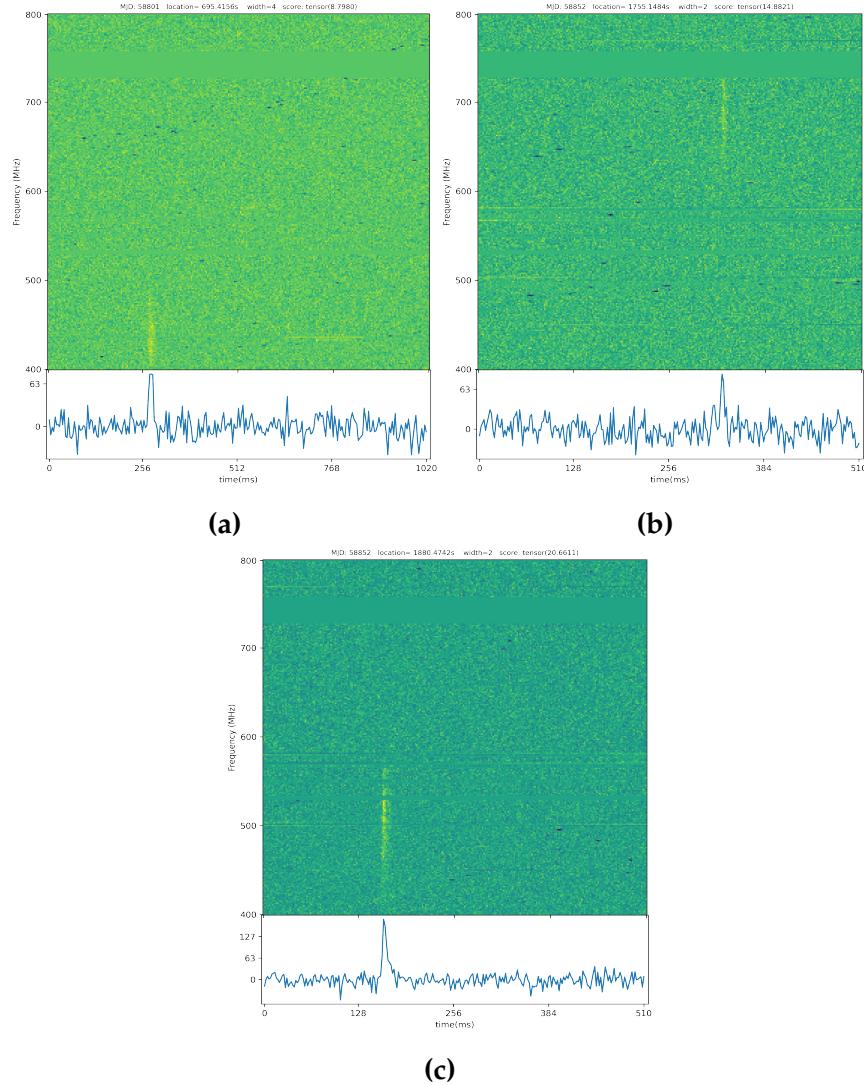


Figure 4.9: Waterfall plots of the three R3 bursts found by the CNN that were not found by CHIME/FRB. These bursts come from observations on Nov 14, 2019 and Jan 4, 2020. Note that **b** and **c** occur on the same day. These burst of SNRs of 10.9, 8.8 and 13.7, respectively. The title space of each plot contains the modified Julian date (MJD) of the observation, the time corresponding to the start of this image relative to the beginning of the observation, the time downsampling used (in ms), and the score given by the CNN. Higher scores relate to a higher confidence by the classifier that the image contains an FRB.

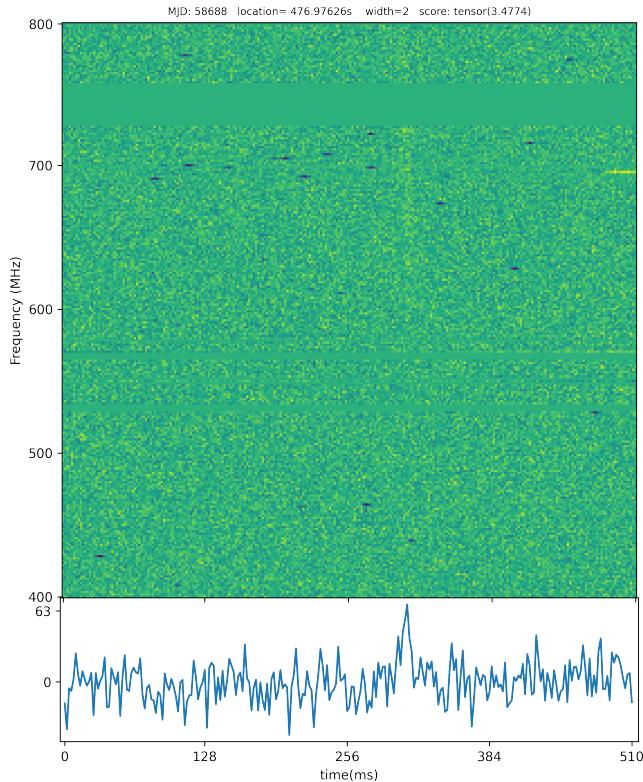


Figure 4.10: Frequency vs time waterfall plot of the R5 burst found by the CNN that was not detected by CHIME/FRB. This burst has an SNR of 6.12 and comes from an observation on July 24, 2019. As in Figure 4.9, the title space the modified Julian date (MJD) of the observation, the time corresponding to the start of this image relative to the beginning of the observation, the time downsampling used (in ms), and the score given by the CNN.

Repeater Name	Hours of Observations	CHIME/FRB Detections	New CNN Detections
R2	370.8	21	4
R3	189.8	71	3
R4	280.5	9	0
R5	121.9	6	1
R6	53.9	8	4
R7	24.7	2	0
R9	53.6	2	0
R10	216	2	0
R11	14.7	2	0
R12	110.2	8	0
R13	14.9	2	0
R14	57.6	10	0
R15	120.2	5	0
R16	28.0	6	0
R17	38.6	21	0

Table 4.1: Summary table of the repeaters searched by the CNN, total number of observation hours for each repeater, current number of CHIME/FRB detections as of June 18 2021, and the number of CNN detections missed by CHIME/FRB.

June 22, 2019 and has a SNR of 25.0, the highest detection by the CNN. The second burst comes from Dec 5, 2019 and has a SNR of 5.35. The third and fourth bursts come from Jan 13, 2020 and Feb 10, 2020 and have SNRs of 8.0 and 7.2, respectively.

The locations in time of these 4 addiditional bursts relative to the CHIME/FRB detections can be seen in Figure 4.12

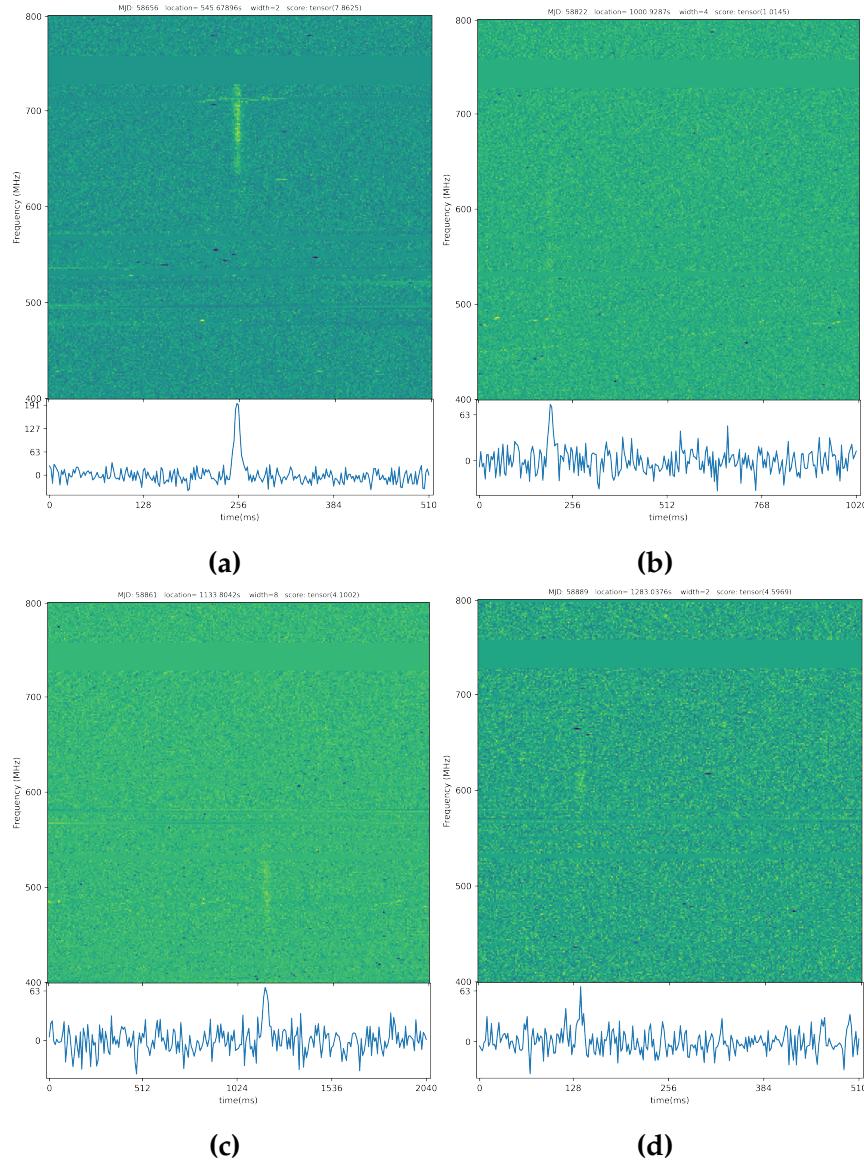


Figure 4.11: The 4 bursts not detected by CHIME/FRB that were found by the CNN for R6. This set increases the existing 9 burst set by 44%. These additional bursts were detected on June 22 2019, Dec 5 2019, Jan 13 2020, and Feb 10 2020, and have SNRs of 25.0, 5.35, 8.0 , and 7.2 respectively. The titles spaces are similar to Figures 4.9 and 4.10.

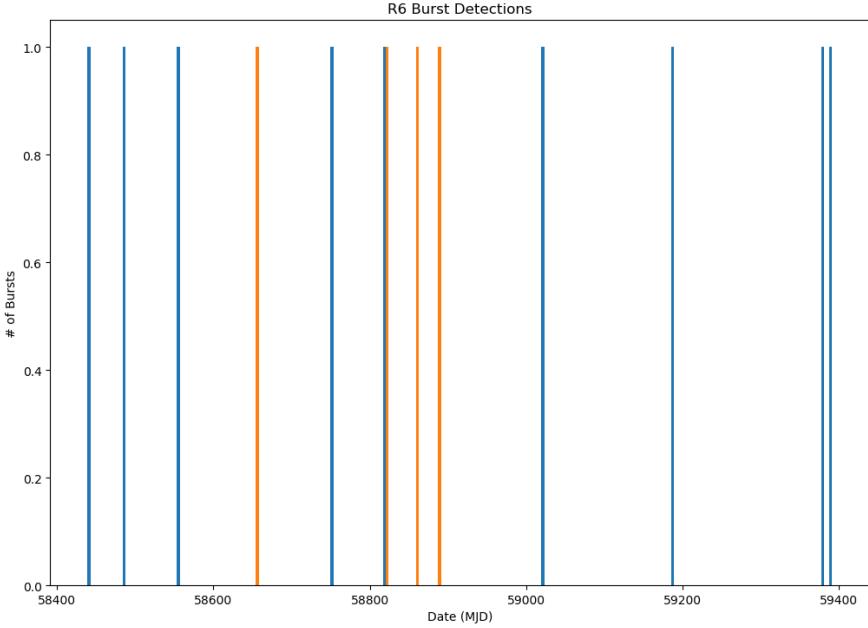


Figure 4.12: Visualization of the dates when R6 was detected by CHIME/FRB (blue) and CHIME/Pulsar using the new CNN (orange). Note that the last observation of R6 with CHIME/Pulsar occurred on MJD 58925 (March 17, 2020).

4.1.6 Conclusion

In total, 12 bursts were found by the CNN in the archived CHIME/Pulsar data. This has proven that the observation of FRB repeaters with CHIME/Pulsar should be continued. The detection of bursts missed by CHIME/FRB can help characterize CHIME/FRBs systematic effects. Scientifically, additional bursts are useful for better determining source burst rates, enabling more meaningful repeater searches, as well as searches for DM variations. They will also aid in the determination of luminosity functions and burst width

distributions. All of these are important for helping to determine if in fact repeating FRBs come from a different progenitor to their non-repeating counterparts, and eventually, what exactly causes the FRB phenomena.

Chapter 5

An Upgraded Scheduling Algorithm for CHIME/Pulsar

5.1 Motivation

As mentioned in Chapter 3, the CHIME/Pulsar instrument is tasked with tracking interesting sources on the sky while they are within the CHIME primary beam. These sources could include pulsars, FRBs, soft gamma repeaters (SGRs), and many more. A problem arises when we realise that CHIME/Pulsar receives only 10 tracking beams from the correlator, and there are often more than 10 interesting sources on the sky that we would like to observe. For example, Figure 5.1 shows a snapshot of everything in the primary beam at a certain time. Most of these points are previously observed FRBs, but when the Galactic centre passes over the telescope there are often many more than 10 pulsars within the beam.

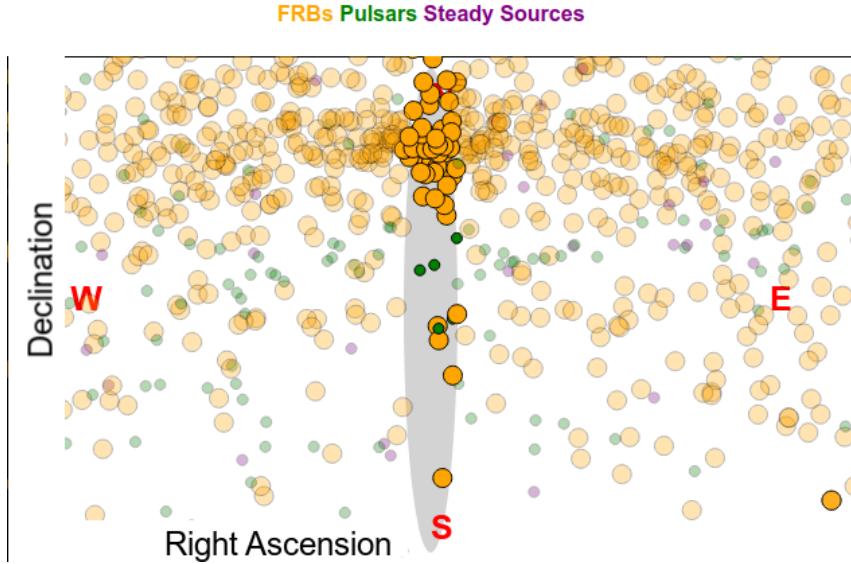


Figure 5.1: Snapshot of the sources in the primary beam at ~11:00am, June 3, 2021. Yellow circles are locations of previously detected repeating and non-repeating FRBs, green circles are for known pulsars, and purple circles mark steady sources. The grey shaded region marks the CHIME primary beam. It is clear that there are often more than 10 interesting sources in the primary beam. Thus a scheduling algorithm for the 10 CHIME/Pulsar beams needed to be developed.

Moreover, the level of interest of certain sources may be higher for some groups as opposed to others. For example, the CHIME/FRB group might put much higher priority on observing FRB repeaters than pulsars, but CHIME/Pulsar's science goals must still be respected.

In order for CHIME/Pulsar to begin operations an interim solution was needed. This formalized into a basic scheduling algorithm to organize the tracking of sources by CHIME/Pulsar. The algorithm was based on a priority system, where each source in the list of desirable sources was given a priority on a 4-point scale, either 16, 32, 64, or 128. Each step up in priority

would result in the source being twice as likely at being selected for observation. There was also functionality for the overriding of the schedule for Target of Opportunity (ToO) events.

Among other things, one large drawback to this initial algorithm was that it did not save a long term history of the observations it made. Of course the data were saved, but it became desirable to store various metrics of the observations in a database connected to the scheduler itself. One particular metric of interest was the tracking of which project wanted to observe which sources. For CHIME/FRB and CHIME/Pulsar this may have been obvious based on which sources where being observed (CHIME/Pulsar would be unlikely to want FRB-repeater data). But with the addition of the new CHIME/Outriggers project, which requires regular observations of pulsars for calibration purposes, it was clear that project tracking of the sources was needed.

Furthermore, since project tracking was to be implemented, in an effort to ensure that the CHIME/Pulsar team retained majority observation time, the addition of project quotas was desired. These quotas would limit the amount of observation time a project other than CHIME/Pulsar could use in a given amount of time (say, a month). Altogether these upgrades required significant improvements to the current scheduling algorithm. With the potential of this algorithm to be connected to the CNN mentioned in the previous chapter to schedule observations of, observe, and search FRB-repeater data for bursts, I took on the responsibility of upgrading the scheduler.

The scheduler was developed making use of the MongoDB¹ database system and GraphQL Application Programming Interface (API) query language. The code was developed for asynchronous operation with the Python driver Motor for MongoDB. The entirety of the scheduler was written in Python.

5.1.1 Outline of Logic

Much of the logic of the upgraded scheduler was taken from the initial version. This included, but was not limited to, the code to calculate accurate sky positions given RA and Dec from the JD2000 epoch, selecting sources based on their priority, and the general structure. The flow of logic is visualized in Figure 5.2.

The algorithm begins by calling `scheduler.py` which first queries a source table. This table consists of python dictionary entries constructed from the source list text files for each of the CHIME projects, CHIME/FRB, CHIME/Pulsar, and CHIME/Outiggers. A description of the keys of source table entries can be found in Table 5.1. The scheduling algorithm then queries an observation table. This table contains the history of all observations planned by the scheduler. An outline of the keys in an observation table entry can be found in Table 5.2. In order for the scheduler to operate efficiently even after many months of observations table entries have been saved, only the past 24 hours of observations are queried.

The end time of the last observation for each of the 10 beams is then found from this observation table and used as the earliest time a new source

¹www.mongodb.com

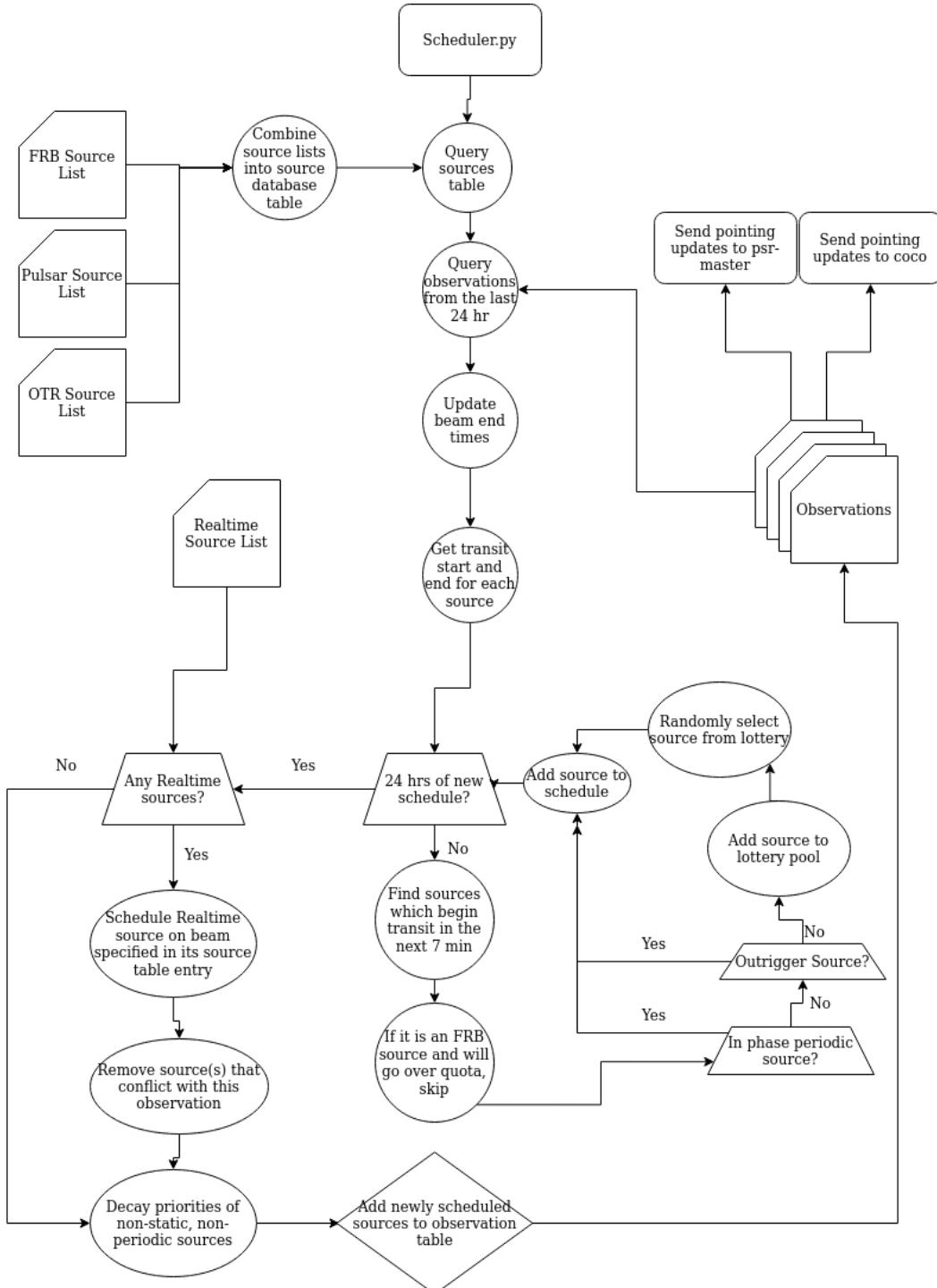


Figure 5.2: Diagram outlining the logic of the upgraded CHIME/Pulsar scheduling algorithm.

can be scheduled on each beam. The transit start and end times for each source in the source table are then calculated. The main scheduling loop is then initiated.

First, there is a check to see if there is a full schedule for the next 24 hours. If not, the algorithm loops through each source in the source table. If a source's next transit time is not within the next 7 minutes of the earliest available beam it is skipped. There is then a check to see that if this source was scheduled, would it put its corresponding project over their observation time quota. If so, it is skipped. Then it is checked if the source is periodic and in its active phase, such as a periodic FRB like FRB 180916. If it meets both of these criteria it is added to the schedule and the end time for that beam is updated. If the source is periodic and not in phase, it is skipped. It is then checked if the source is a CHIME/Outrigger calibrator. These sources need to be observed with daily cadence, so if the source meets this criteria it is added to the schedule. Finally, if the source has not been scheduled but still meets scheduling criteria it is added to a lottery. All sources that begin their transit in the next 7 minutes of the earliest ending beam and make it to this stage without being skipped are added to this lottery.

The lottery consists of a list of the names of each source added to it. Each source name is added n times, where n is the priority of the source. Initial source priorities are either 16, 32, 64, or 128. So a priority 16 source is added to the lottery 16 times and a priority 32 source is added 32 times. This makes each increase in priority level twice as likely to be scheduled. A

random index is selected from this list of source names and that source is added to the schedule.

It is then again checked if the schedule has been written for the next 24 hours. If it has not, then the loop repeats. If it has, a separate source list of realtime sources is checked. These sources include anything where a Target of Opportunity (ToO) observation is needed and the current schedule needs to be overridden. If there is a source in this list, its transit time is calculated, and if its transit is within the next 2 hours it is scheduled on the beam specified in its source table entry. If there are any conflicting observations they are removed.

The algorithm then decays the priorities of all probabilistic sources based on their number of observations in the last 15 days. The updated priorities are given simply by

$$P_n = P_i \frac{1}{n+1}, \quad (5.1)$$

where P_n is the new priority (truncated to an integer), P_i is the original priority of the source, and n is the number of observations of the source in the last 15 days. If the priority of the source ever drops to < 1 it is reset to P_i . An entry for each observation is then constructed according to Table 5.2 and added to the observation table in the database.

A separate program runs asynchronously with `scheduler.py` which communicates the transit time and beam pointing information to the CHIME correlator. This process reads from the observation table and is in charge of tracking the status of the observations and changing the `observation_status` key in Table 5.2 to either `SCHEDULED`, `OBSERVING`,

Key	Description
source_name	name of the source
num_obs	total number of observations of this source to date
data_collected	total size of all data collected of this source
on_sky_time	total on sky time of source to date
previous_obs	date of most recent observation
scheduling_mode	Mode of scheduling. Static, periodic, probabilistic
period	Period of source, if periodic
periodicity_reference	Reference date for periodic sources
active_phase_window	Phase window if periodic source
rajd2000	right ascension coordinate according to JD2000 epoch
decjd2000	declination coordinate according to JD2000 epoch
ra	Current right ascension
dec	Current declination
dm	Dispersion measure of source
scaling	Scaling
observing_mode	Mode of observation, gated or otherwise
projects	List of projects that wish to observe this source
beam	A default initial beam selection.
priority	Initial priority of the source
current_priority	Current (decayed) priority of the source
notes	Any other notes on the source

Table 5.1: A list of all the keys in a source table entry, with descriptions for each key.

COMPLETE, or MISSED. The MISSED tag is only used for error tracking purposes in cases where a source's transit time has come and gone but it was never set to OBSERVING. The sending of the transit information completes the logic flow of the upgraded scheduling algorithm.

5.1.2 Testing

Aside from ensuring that the database queries and mutations (adding observations to the table, updating number of observations and on sky time,

Key	Description
observation_id	Unique observation ID
project_name	Project requesting observation
source_name	Name of source
date_of_observation	Date of observation
scheduling_status	Whether source has been accepted, rejected, or knocked off schedule
observation_status	Status of observation (scheduled, observing, complete, missed)
data_status	Status of data (acquired, transferred, transfer failed)
local_data_copy	If there is a copy of the data locally
data_location	location of the data
unix_utc_start	unix start time of observation
unix_utc_end	unix end time of observation
beam	beam number observation occurred on
scaling	scaling
dm	Dispersion measure of source
mode	mode of observation (periodic, static, probabilistic)
nbin	number of bins
rajd2000	right ascension coordinate according to JD2000 epoch
decjd2000	declination coordinate according to JD2000 epoch
ra	Current right ascension
dec	Current declination
jd	Julian data of observation

Table 5.2: A list of all the keys in an observation table entry, with descriptions for each key.

etc.) where working as expected, there were 4 main tests performed on the scheduler. These checked if the `observation_status` tag was being updated correctly, that the source priorities were being updated, that higher priority sources were being observed more often than lower priority ones, and that the ToO overriding system was working as expected.

The first test was on the `observation_status` tag. The scheduling algorithm was simulated such that \sim 24 hours of schedule were created. The system was then left for a few hours so that a sufficient number of simu-

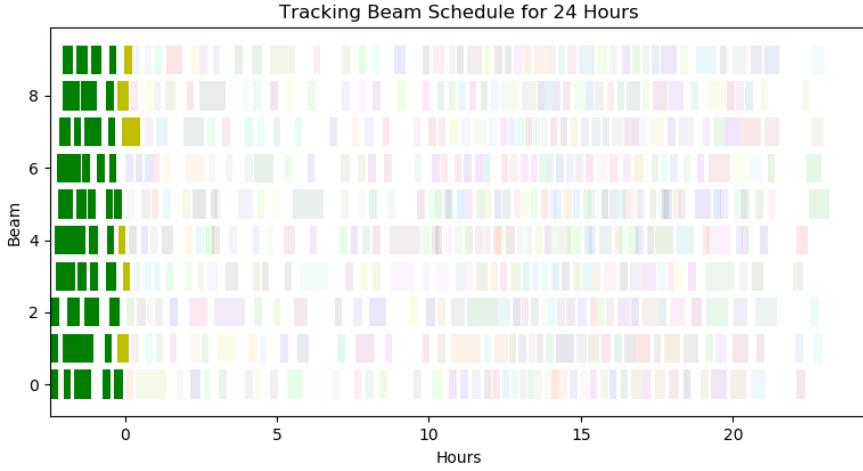


Figure 5.3: A visualization of ~ 24 hours of schedule. Each block corresponds to a source's transit. The y-axis is the beam number (0-9) and the x-axis is time marked in hours from now. The green coloured blocks have their observation status set to COMPLETE, OBSERVING observations are in yellow, and the faded multi-colour blocks are labelled SCHEDULED. The change in colour of the faded sources is simply due to their colours being randomized.

lated transits would have occurred and have their status's changed. The results of this test can be visualized in Figure 5.3. Each block in this figure represents the duration of a source's transit on one of the 10 beams (labelled 0-9). Green corresponds to an `observation_status` of COMPLETE, yellow for OBSERVING, and faded multi-colour for SCHEDULED. There are no MISSED observations. This visualization makes it clear which sources are currently transiting, which have been observed, and which are scheduled to be observed. This test was successful.

Figure 5.4 shows the results of the priority tests. This histogram shows the number of observations of 8 sources near the Galactic center over many

days. For this test, 2 sources for each level of priority were chosen that are within ± 4 degrees of RA=266.375 (the Galactic centre). This location was selected to ensure that more than 10 sources would be in the primary beam, meaning the priority system would need to be used. It is clear from the figure that higher priority sources are observed more often than lower priority ones. It is also clear that the priorities are being updated correctly. If they weren't it would appear in the plot as each increase in priority level would be observed exactly twice as much as the previous level.

Finally the ToO functionality was tested. To do this, first a source was removed from the regular pulsar source list. The schedule was then created for ~ 10 days. This pulsar was then added to the realtime source list. The scheduler immediately recognized the addition of the source and added it to the already created schedule 2 hours before its transit time. A visualization of 24 hours of the schedule before and after the addition of the realtime source can be seen in Figure 5.5. The first image shows the schedule before the addition of the realtime source and the second shows the source added to the schedule and marked in cyan. This was interpreted as a successful implementation of this ToO system.

5.1.3 Current Status

Currently the new scheduler has been integrated into the CHIME/Pulsar system and waits to take over for the old scheduler. It is currently running in a simulated state. That is, it is operating as it would but the code which sends the beam pointing information to the CHIME correlator has been disabled. This is to allow for continued testing of its operation while we await

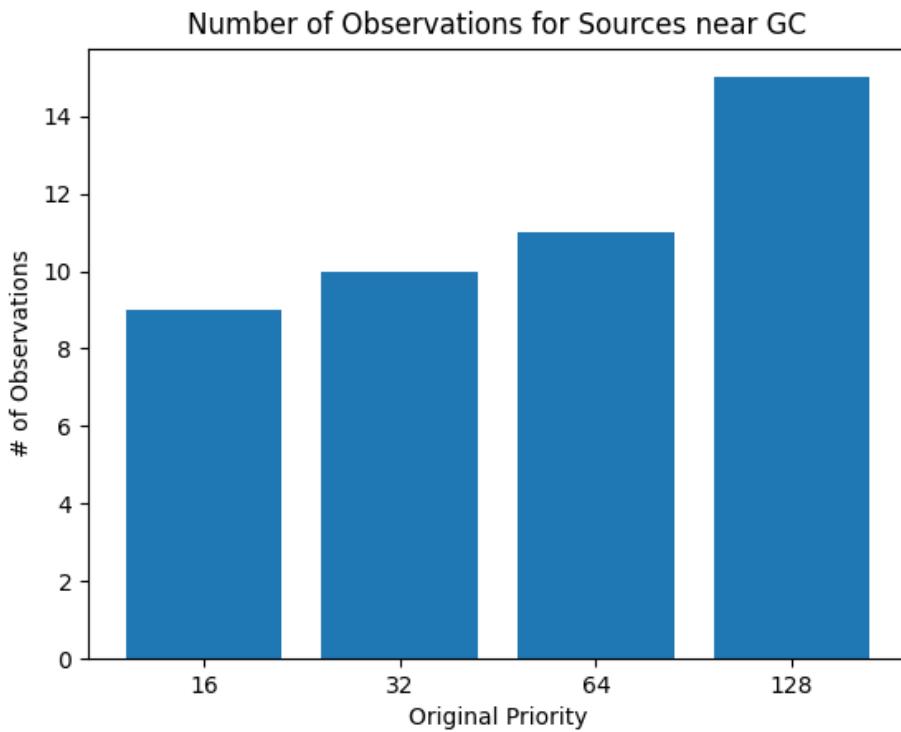
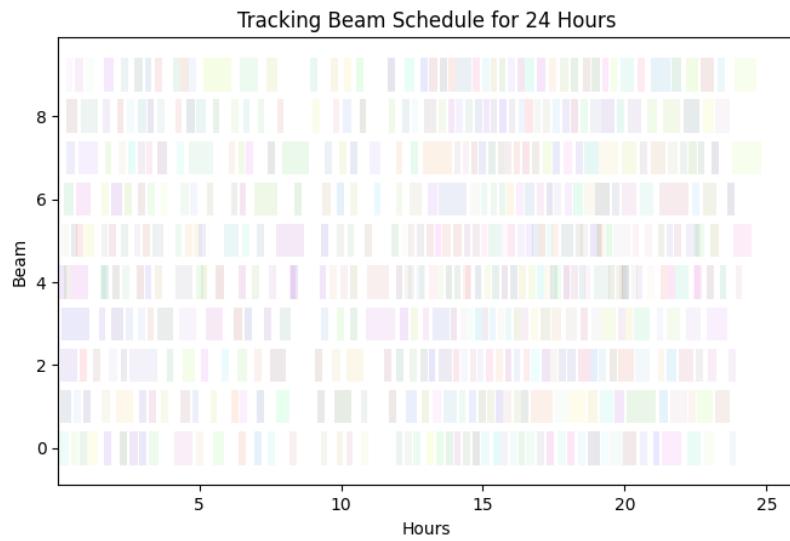
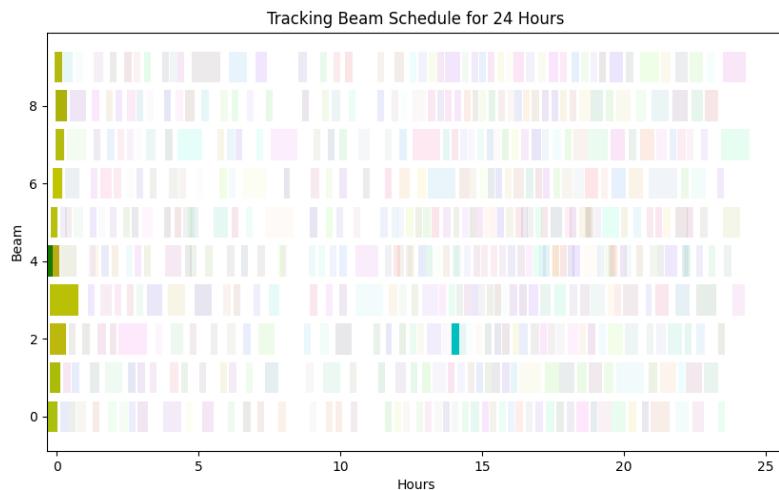


Figure 5.4: Histogram of the number of observations of 8 sources near the Galactic centre after 15 days of simulated schedule. Each bin corresponds to one of the 4 original priority values. Two sources for each level were selected. It can clearly be seen that higher priority sources are observed more often than lower priority ones. Note that even though *initially* priority 128 sources are 8 times as likely to be observed as priority 16 sources, as the priorities decay the distribution in the number of observations smooths out.



(a)



(b)

Figure 5.5: Before (a) and after (b) the addition of a source to the realtime source list to test ToO functionality of the scheduling algorithm. The real-time source is marked in cyan.

the next upgrade cycle of CHIME/Pulsar when it will be implemented. This is likely to occur in the coming weeks.

Chapter 6

Conclusion

The mystery of FRBs is still far from solved, but recent advancements continue to drive the field further. Currently, majority effort in the field is placed on cataloging as many bursts as possible, to aid in constraining theoretical models. The recent publication of over 500 FRBs in the first CHIME/FRB catalog [9] is but the beginning of such catalogs from the group, increasing the number of published FRBs by an order of magnitude. As more bursts are discovered so too are more repeating FRBs. This leads to interesting ideas of two classes of FRBs, repeating and non-repeating. As mentioned, periods have only been teased out of 2 repeaters so far, FRB 180916 [7] and FRB 121102 [40]. This could lead to an idea of two populations of repeaters, periodic and non-periodic, but more data of each type are needed.

CHIME's FRB searches have been quite fruitful largely due to the vast amount of data the telescope processes. So much so that an "embarrassment of riches" has come as a result. The motivation of this thesis was

to help tackle this exact problem. CHIME/Pulsar had taken data of FRB repeaters but due to the hundreds of hours of observations, and standard single pulse search tools such as PRESTO producing many false positives, a new solution was needed to search the data. These false positives are due to a high amount of RFI in the data. It was believed that a CNN would be better at ignoring this RFI and could detect real, faint bursts. The development of this CNN is shown to have been necessary, otherwise the bursts found in Chapter 4 may have gone undetected. For CHIME/FRB, burst candidates are found using the real time search software `bonsai` which are then classified by a ML algorithm. Positive candidates are then confirmed by eye through the visual inspection of a waterfall plot. Since a CNN could be used to search the many hours of CHIME/Pulsar data (instead of just a single image output from the real time search pipeline), and CNNs are excellent at image recognition, this made a convolutional neural network the favoured ML algorithm for this repeater search. In total, of the 15 repeaters that were observed by CHIME/Pulsar, 12 bursts were found using this CNN that were not detected by CHIME/FRB. These were missed by CHIME/FRB most likely due to them being below the detection threshold at the time of observation. Another possibility is that the bursts occurred on the edge or even between formed CHIME/FRB beams. These beams can be seen in Figure 6.1 from [20].

This shows that although CHIME/FRB is excellent at FRB detection, there could be a small subset of bursts that have gone undetected. Thus, further use of this CNN is crucial for collecting as many FRBs as possible.

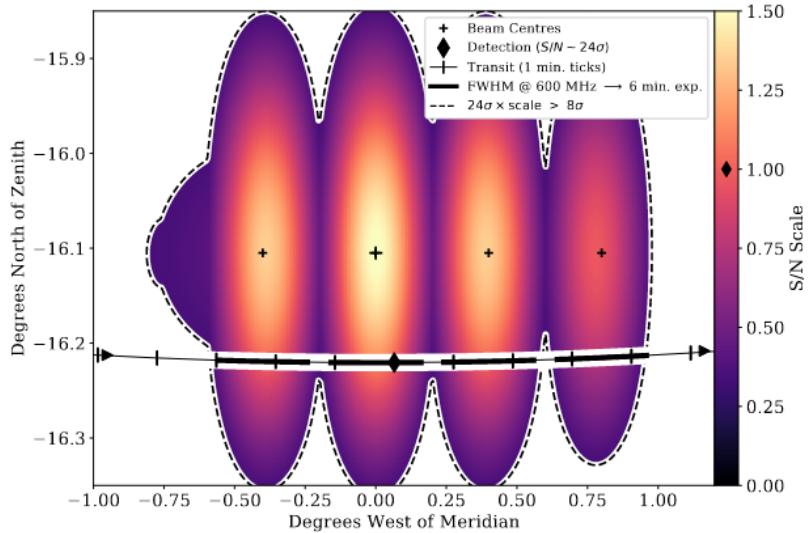


Figure 6.1: From [20]. Detection location and transit sensitivity of a CHIME/FRB detection of FRB 121102. The four ovals represent the four detection beams with the color contours visualizing their sensitivity. It is clear that sensitivity decreases further from the center of the beam. This decrease in sensitivity could be the cause of the missed CHIME/FRB detections for the repeaters searched in Chapter 4.

The detection of additional bursts is important for many aspects of understanding the FRB phenomena. Firstly, they will help to determine burst rates. Burst rates help to determine how many of these sources exist and significantly aid in periodicity searches. More bursts will also help in statistical studies to determine whether or not all FRBs are in fact repeaters. This would mean that the apparent non-repeater population is the same as repeaters, but with such low burst rates that we have yet to detect additional bursts from them. As for specific repeaters, additional bursts allow for more robust searches for DM variations. These variations would help to understand the environment of the FRB. Variations in scattering time could also

be searched for. More bursts will also help with determinations of luminosity functions for individual sources. This would be helpful for planning observations with other telescopes and for understanding the population as a whole.

Along with the development of this CNN, upgrades to the scheduling algorithm for tracking sources with CHIME/Pulsar were done. Of these upgrades, a quota system was implemented for each of the projects utilizing CHIME/Pulsar's tracking capabilities. This new scheduling algorithm along with the use of the CNN described in Chapter 4, should allow for the continued observation of FRB repeaters with CHIME/Pulsar, while ensuring that CHIME/Pulsar science is respected, and that the data recorded are analysed in an efficient manner. I will now outline what the next steps and future development of these projects might look like.

6.0.1 Next steps: Scheduler

Currently, the upgraded scheduler for the CHIME/Pulsar instrument has completed testing and awaits integration into the system. Once this occurs the scheduler will maintain a constant 24 hours of schedule and save the entire history of all observations. Quota tracking of CHIME/FRB sources will also be monitored. In the future, it could be of interest to use the scheduler and CNN in tandem to target 2-burst repeaters in an effort to quickly find a third burst, thus validating it as a true repeating FRB. This will be of increasing importance as more FRBs are discovered, and thus the chance probability of two unassociated bursts appearing to come from the same source increases. Further development will also be needed to complete the tracking

of all desired metrics. For example, the `data_status`, `local_data_copy` and `data_location` keys in the observation table entries are currently left blank.

6.0.2 Next steps: CNN

As mentioned, the CNN developed in this thesis will find its most scientifically impactful use through the confirmation of 2-burst repeaters. This should also be performed alongside regular followup of known repeaters to assist in increasing the bursts' datasets. The detection of bursts missed by CHIME/FRB will also be useful in understanding CHIME/FRB systematic effects as most of the events above the CHIME/FRB threshold should be detected by both instruments. This thesis has shown that CHIME/FRB is capable of missing relatively high SNR bursts, which could be the case for both repeating and non-repeating FRBs. Although the number of bursts missed is relatively small, this problem is still of interest in order to maximize CHIME/FRB's scientific output. As previously mentioned, some possibilities for why these bursts were missed could be that they occurred between formed beams, the combination of a band-limited spectrum and location in the formed beam could have resulted in lower SNRs, overly aggressive RFI filtering in the CHIME/FRB real time system, or simply that the CNN is at a distinct advantage over the real time system since it knows the burst DM in advance.

If further improvement of the CNN is desired, reducing false positives while maintaining the low false negative rate could be possible. Since the non-FRB training set consisted of only Gaussian noise, augmentation of

common false positives (bright single channel RFI, sweeping DM=0 RFI, etc) could be done and added to this set. This would be particularly useful when searching for low DM bursts as many of the false positives for these repeaters are sweeping RFI signatures at DM=0, but have not been dedispersed enough to be completely washed out. Thus, although the current CNN has been successful at detecting many bursts missed by CHIME/FRB, there is still room for its improvement.

Bibliography

- [1] AMIRI, M., BANDURA, K., BERGER, P., BHARDWAJ, M., BOYCE, M. M., BOYLE, P. J., BRAR, C., BURHANPURKAR, M., CHAWLA, P., AND ET AL. The chime fast radio burst project: System overview. *The Astrophysical Journal* 863, 1 (Aug 2018), 48.
- [2] BARSDELL, B. R., BAILES, M., BARNES, D. G., AND FLUKE, C. J. Accelerating incoherent dedispersion. *Monthly Notices of the Royal Astronomical Society* 422, 1 (Mar 2012), 379–392.
- [3] BELOBORODOV, A. M. A flaring magnetar in FRB 121102? *The Astrophysical Journal* 843, 2 (July 2017), L26.
- [4] BHAT, N. D. R., CORDES, J. M., CAMILO, F., NICE, D. J., AND LORIMER, D. R. Multifrequency observations of radio pulse broadening and constraints on interstellar electron density microstructure. *The Astrophysical Journal* 605, 2 (apr 2004), 759–783.
- [5] BHAT, N. D. R., RAO, A. P., AND GUPTA, Y. Long-term scintillation studies of pulsars. i. observations and basic results. *The Astrophysical Journal Supplement Series* 121, 2 (apr 1999), 483–513.

- [6] BHATTACHARYA, D. Formation and evolution of binary and millisecond radio pulsars. *Physics Reports* 203, 1–2 (May 1991), 1–124.
- [7] CHIME/FRB COLLABORATION, AMIRI, M., ANDERSEN, B. C., BANDURA, K. M., BHARDWAJ, M., BOYLE, P. J., BRAR, C., CHAWLA, P., CHEN, T., CLICHE, J. F., CUBRANIC, D., DENG, M., DENMAN, N. T., DOBBS, M., DONG, F. Q., FANDINO, M., FONSECA, E., GAENSLER, B. M., GIRI, U., GOOD, D. C., HALPERN, M., HESSELS, J. W. T., HILL, A. S., HÖFER, C., JOSEPHY, A., KANIA, J. W., KARUPPUSAMY, R., KASPI, V. M., KEIMPEMA, A., KIRSTEN, F., LANDECKER, T. L., LANG, D. A., LEUNG, C., LI, D. Z., LIN, H. H., MARCOTE, B., MASUI, K. W., MCKINVEN, R., MENA-PARRA, J., MERRYFIELD, M., MICHILLI, D., MILUTINOVIC, N., MIRHOSSEINI, A., NAIDU, A., NEWBURGH, L. B., NG, C., NIMMO, K., PARAGI, Z., PATEL, C., PEN, U. L., PINSONNEAULT-MAROTTE, T., PLEUNIS, Z., RAFIEI-RAVANDI, M., RAHMAN, M., RANSOM, S. M., RENARD, A., SANGHAVI, P., SCHOLZ, P., SHAW, J. R., SHIN, K., SIEGEL, S. R., SINGH, S., SMEGAL, R. J., SMITH, K. M., STAIRS, I. H., TENDULKAR, S. P., TRETYAKOV, I., VANDERLINDE, K., WANG, H., WANG, X., WULF, D., YADAV, P., AND ZWANIGA, A. V. Periodic activity from a fast radio burst source. 582, 7812 (June 2020), 351–355.
- [8] CHIME/FRB COLLABORATION, ANDERSEN, B. C., BANDURA, K., BHARDWAJ, M., BOUBEL, P., BOYCE, M. M., BOYLE, P. J., BRAR, C., CASSANELLI, T., CHAWLA, P., CUBRANIC, D., DENG, M., DOBBS, M., FANDINO, M., FONSECA, E., GAENSLER, B. M., GILBERT, A. J.,

GIRI, U., GOOD, D. C., HALPERN, M., HILL, A. S., HINSHAW, G., HÖFER, C., JOSEPHY, A., KASPI, V. M., KOTHES, R., LANDECKER, T. L., LANG, D. A., LI, D. Z., LIN, H. H., MASUI, K. W., MENA-PARRA, J., MERRYFIELD, M., MCKINVEN, R., MICHILLI, D., MILUTINOVIC, N., NAIDU, A., NEWBURGH, L. B., NG, C., PATEL, C., PEN, U., PINSONNEAULT-MAROTTE, T., PLEUNIS, Z., RAFIEI-RAVANDI, M., RAHMAN, M., RANSOM, S. M., RENARD, A., SCHOLZ, P., SIEGEL, S. R., SINGH, S., SMITH, K. M., STAIRS, I. H., TENDULKAR, S. P., TRETYAKOV, I., VANDERLINDE, K., YADAV, P., AND ZWANIGA, A. V. CHIME/FRB Discovery of Eight New Repeating Fast Radio Burst Sources. *885*, 1 (Nov. 2019), L24.

- [9] COLLABORATION, F., AMIRI, M., ANDERSEN, B. C., BANDURA, K., BERGER, S., BHARDWAJ, M., BOYCE, M. M., BOYLE, P. J., BRAR, C., BREITMAN, D., CASSANELLI, T., CHAWLA, P., CHEN, T., CLICHE, J. F., COOK, A., CUBRANIC, D., CURTIN, A. P., DENG, M., DOBBS, M., FENGQIU, DONG, EADIE, G., FANDINO, M., FONSECA, E., GAENSLER, B. M., GIRI, U., GOOD, D. C., HALPERN, M., HILL, A. S., HINSHAW, G., JOSEPHY, A., KACZMAREK, J. F., KADER, Z., KANIA, J. W., KASPI, V. M., LANDECKER, T. L., LANG, D., LEUNG, C., LI, D., LIN, H.-H., MASUI, K. W., MCKINVEN, R., MENA-PARRA, J., MERRYFIELD, M., MEYERS, B. W., MICHILLI, D., MILUTINOVIC, N., MIRHOSSEINI, A., MÜNCHMEYER, M., NAIDU, A., NEWBURGH, L., NG, C., PATEL, C., PEN, U.-L., PETROFF, E., PINSONNEAULT-MAROTTE, T., PLEUNIS, Z., RAFIEI-RAVANDI, M., RAHMAN, M., RANSOM, S. M., RENARD, A., SANGHAVI, P., SCHOLZ, P., SHAW,

J. R., SHIN, K., SIEGEL, S. R., SIKORA, A. E., SINGH, S., SMITH, K. M., STAIRS, I., TAN, C. M., TENDULKAR, S. P., VANDERLINDE, K., WANG, H., WULF, D., AND ZWANIGA, A. V. The first chime/frb fast radio burst catalog, 2021.

- [10] COLLABORATION, P., AMIRI, M., BANDURA, K. M., BOYLE, P. J., BRAR, C., CLICHE, J. F., CROWTER, K., CUBRANIC, D., DEMOREST, P. B., DENMAN, N. T., DOBBS, M., DONG, F. Q., FANDINO, M., FONSECA, E., GOOD, D. C., HALPERN, M., HÖFER, C., KASPI, V. M., LANDECKER, T. L., LIN, H. H., MASUI, K. W., MCKEE, J. W., MENA-PARRA, J., MEYERS, B. W., MICHILLI, D., NAIDU, A., NEWBURGH, L., NG, C., PATEL, C., PINSONNEAULT-MAROTTE, T., RANSOM, S. M., RENARD, A., SCHOLZ, P., SHAW, J. R., SIKORA, A. E., STAIRS, I. H., TAN, C. M., TENDULKAR, S. P., TRETYAKOV, I., VANDERLINDE, K., WANG, H., AND WANG, X. The chime pulsar project: System overview, 2020.
- [11] CORDES, J. M., AND LAZIO, T. J. W. Ne2001. i. a new model for the galactic distribution of free electrons and its fluctuations. *arXiv preprint astro-ph/0207156* (2002).
- [12] CORDES, J. M., SHANNON, R. M., AND STINEBRING, D. R. FREQUENCY-DEPENDENT DISPERSION MEASURES AND IMPLICATIONS FOR PULSAR TIMING. *The Astrophysical Journal* 817, 1 (jan 2016), 16.

- [13] DENG, W., AND ZHANG, B. COSMOLOGICAL IMPLICATIONS OF FAST RADIO BURST/GAMMA-RAY BURST ASSOCIATIONS. *The Astrophysical Journal* 783, 2 (Feb. 2014), L35.
- [14] DONNER, J. Y., VERBIEST, J. P. W., TIBURZI, C., OSŁOWSKI, S., MICHILLI, D., SERYLAK, M., ANDERSON, J. M., HORNEFFER, A., KRAMER, M., GRIESMEIER, J.-M., KÜNSEMÖLLER, J., HESSELS, J. W. T., HOEFT, M., AND MISKOLCZI, A. First detection of frequency-dependent, time-variable dispersion measures. *Astronomy & Astrophysics* 624 (Apr. 2019), A22.
- [15] ESTER, M., KRIEGEL, H.-P., SANDER, J., XU, X., ET AL. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (1996), vol. 96, pp. 226–231.
- [16] FONSECA, E., ANDERSEN, B. C., BHARDWAJ, M., CHAWLA, P., GOOD, D. C., JOSEPHY, A., KASPI, V. M., MASUI, K. W., MCKINVEN, R., MICHILLI, D., PLEUNIS, Z., SHIN, K., TENDULKAR, S. P., BANDURA, K. M., BOYLE, P. J., BRAR, C., CASSANELLI, T., CUBRANIC, D., DOBBS, M., DONG, F. Q., GAENSLER, B. M., HINSHAW, G., LANDECKER, T. L., LEUNG, C., LI, D. Z., LIN, H. H., MENAPARRA, J., MERRYFIELD, M., NAIDU, A., NG, C., PATEL, C., PEN, U., RAFIEI-RAVANDI, M., RAHMAN, M., RANSOM, S. M., SCHOLZ, P., SMITH, K. M., STAIRS, I. H., VANDERLINDE, K., YADAV, P., AND ZWANIGA, A. V. Nine New Repeating Fast Radio Burst Sources from CHIME/FRB. 891, 1 (Mar. 2020), L6.

- [17] HASKELL, B., AND MELATOS, A. Models of pulsar glitches. *International Journal of Modern Physics D* 24, 03 (Feb. 2015), 1530008.
- [18] HASKELL, B., AND SEDRAKIAN, A. *Superfluidity and Superconductivity in Neutron Stars*. Springer International Publishing, Cham, 2018, pp. 401–454.
- [19] HESSELS, J. W. T. A radio pulsar spinning at 716 hz. *Science* 311, 5769 (Mar 2006), 1901–1904.
- [20] JOSEPHY, A., CHAWLA, P., FONSECA, E., NG, C., PATEL, C., PLEUNIS, Z., SCHOLZ, P., ANDERSEN, B. C., BANDURA, K., BHARDWAJ, M., AND ET AL. Chime/frb detection of the original repeating fast radio burst source frb 121102. *The Astrophysical Journal* 882, 2 (Sep 2019), L18.
- [21] KASPI, V. M. Grand unification of neutron stars. *Proceedings of the National Academy of Sciences* 107, 16 (Apr 2010), 7147–7152.
- [22] KEANE, E. F., KRAMER, M., LYNE, A. G., STAPPERS, B. W., AND McLAUGHLIN, M. A. Rotating radio transients: new discoveries, timing solutions and musings. *Monthly Notices of the Royal Astronomical Society* 415, 4 (June 2011), 3065–3080.
- [23] LAM, M. T., CORDES, J. M., CHATTERJEE, S., JONES, M. L., McLAUGHLIN, M. A., AND ARMSTRONG, J. W. SYSTEMATIC AND STOCHASTIC VARIATIONS IN PULSAR DISPERSION MEASURES. *The Astrophysical Journal* 821, 1 (apr 2016), 66.

- [24] LAM, M. T., McLAUGHLIN, M. A., CORDES, J. M., CHATTERJEE, S., AND LAZIO, T. J. W. Optimal frequency ranges for submicrosecond precision pulsar timing. *The Astrophysical Journal* 861, 1 (jun 2018), 12.
- [25] LAZARUS, P. The palfa survey: Going to great depths to find radio pulsars. *Proceedings of the International Astronomical Union* 8, S291 (Aug 2012), 35–40.
- [26] LORIMER, D. R., BAILES, M., McLAUGHLIN, M. A., NARKEVIC, D. J., AND CRAWFORD, F. A bright millisecond radio burst of extragalactic origin. *Science* 318, 5851 (2007), 777–780.
- [27] LORIMER, D. R., AND KRAMER, M. *Handbook of pulsar astronomy*. Cambridge University Press, 2012.
- [28] LYUBARSKY, Y. A model for fast extragalactic radio bursts. *Monthly Notices of the Royal Astronomical Society: Letters* 442, 1 (Apr. 2014), L9–L13.
- [29] LYUTIKOV, M. Radius-to-frequency mapping and FRB frequency drifts. *The Astrophysical Journal* 889, 2 (Jan. 2020), 135.
- [30] MANCHESTER, R. N., HOBBS, G. B., TEOH, A., AND HOBBS, M. The australia telescope national facility pulsar catalogue. *The Astronomical Journal* 129, 4 (Apr 2005), 1993–2006.
- [31] MANCHESTER, R. N., HOBBS, G. B., TEOH, A., AND HOBBS, M. "the atnf pulsar catalogue". *Astronomical Journal* (2005), 129.

- [32] MARCOTE, B., NIMMO, K., HESSELS, J. W. T., TENDULKAR, S. P., BASSA, C. G., PARAGI, Z., KEIMPEMA, A., BHARDWAJ, M., KARUPPUSAMY, R., KASPI, V. M., ET AL. A repeating fast radio burst source localized to a nearby spiral galaxy. *Nature* 577, 7789 (Jan 2020), 190–194.
- [33] MCKEE, J. W., LYNE, A. G., STAPPERS, B. W., BASSA, C. G., AND JORDAN, C. A. Temporal variations in scattering and dispersion measure in the Crab Pulsar and their effect on timing precision. *Monthly Notices of the Royal Astronomical Society* 479, 3 (07 2018), 4216–4224.
- [34] MC LAUGHLIN, M. A., LYNE, A. G., LORIMER, D. R., KRAMER, M., FAULKNER, A. J., MANCHESTER, R. N., CORDES, J. M., CAMILO, F., POSSENTI, A., STAIRS, I. H., HOBBS, G., D’AMICO, N., BURGAY, M., AND O’BRIEN, J. T. Transient radio bursts from rotating neutron stars. 439, 7078 (Feb. 2006), 817–820.
- [35] METZGER, B. D., MARGALIT, B., AND SIRONI, L. Fast radio bursts as synchrotron maser emission from decelerating relativistic blast waves. *Monthly Notices of the Royal Astronomical Society* 485, 3 (Mar. 2019), 4091–4106.
- [36] MURASE, K., KASHIYAMA, K., AND MÉSZÁROS, P. Erratum: A burst in a wind bubble and the impact on baryonic ejecta: high-energy gamma-ray flashes and afterglows from fast radio bursts and pulsar-driven supernova remnants. *Monthly Notices of the Royal Astronomical Society* 467, 3 (Feb. 2017), 3542–3543.

- [37] NG, C., PANDHI, A., NAIDU, A., FONSECA, E., KASPI, V. M., MASUI, K. W., MCKINVEN, R., RENARD, A., SCHOLZ, P., STAIRS, I. H., TENDULKAR, S. P., AND VANDERLINDE, K. Faraday rotation measures of Northern hemisphere pulsars using CHIME/Pulsar. *496*, 3 (Aug. 2020), 2836–2848.
- [38] PETROFF, E., BARR, E. D., JAMESON, A., KEANE, E. F., BAILES, M., KRAMER, M., MORELLO, V., TABBARA, D., AND VAN STRATEN, W. FRBCAT: The Fast Radio Burst Catalogue. *33* (Sept. 2016), e045.
- [39] PETROFF, E., HESSELS, J. W. T., AND LORIMER, D. R. Fast radio bursts. *The Astronomy and Astrophysics Review* *27*, 1 (May 2019).
- [40] RAJWADE, K. M., MICKALIGER, M. B., STAPPERS, B. W., MORELLO, V., AGARWAL, D., BASSA, C. G., BRETON, R. P., CALEB, M., KARASTERGIOU, A., KEANE, E. F., ET AL. Possible periodic activity in the repeating frb 121102. *Monthly Notices of the Royal Astronomical Society* *495*, 4 (May 2020), 3551–3558.
- [41] SPITLER, L. G., CORDES, J. M., HESSELS, J. W. T., LORIMER, D. R., MC LAUGHLIN, M. A., CHATTERJEE, S., CRAWFORD, F., DENEVA, J. S., KASPI, V. M., WHARTON, R. S., ALLEN, B., BOGDANOV, S., BRAZIER, A., CAMILO, F., FREIRE, P. C. C., JENET, F. A., KARAKO-ARGAMAN, C., KNISPEL, B., LAZARUS, P., LEE, K. J., VAN LEEUWEN, J., LYNCH, R., RANSOM, S. M., SCHOLZ, P., SIEMENS, X., STAIRS, I. H., STOVALL, K., SWIGGUM, J. K., VENKATARAMAN, A., ZHU, W. W., AULBERT, C., AND FEHRMANN, H. Fast Radio Burst Discovered in the Arecibo Pulsar ALFA Survey. *790*, 2 (Aug. 2014), 101.

- [42] SPITLER, L. G., SCHOLZ, P., HESSELS, J. W. T., BOGDANOV, S., BRAZIER, A., CAMILO, F., CHATTERJEE, S., CORDES, J. M., CRAWFORD, F., DENEVA, J., FERDMAN, R. D., FREIRE, P. C. C., KASPI, V. M., LAZARUS, P., LYNCH, R., MADSEN, E. C., McLAUGHLIN, M. A., PATEL, C., RANSOM, S. M., SEYMOUR, A., STAIRS, I. H., STAPPERS, B. W., VAN LEEUWEN, J., AND ZHU, W. W. A repeating fast radio burst. *531*, 7593 (Mar. 2016), 202–205.
- [43] STOVALL, K., LYNCH, R. S., RANSOM, S. M., ARCHIBALD, A. M., BANASZAK, S., BIWER, C. M., BOYLES, J., DARTEZ, L. P., DAY, D., FORD, A. J., FLANIGAN, J., GARCIA, A., HESSELS, J. W. T., HINOJOSA, J., JENET, F. A., KAPLAN, D. L., KARAKO-ARGAMAN, C., KASPI, V. M., KONDRATIEV, V. I., LEAKE, S., LORIMER, D. R., LUNSFORD, G., MARTINEZ, J. G., MATA, A., McLAUGHLIN, M. A., ROBERTS, M. S. E., ROHR, M. D., SIEMENS, X., STAIRS, I. H., VAN LEEUWEN, J., WALKER, A. N., AND WELLS, B. L. THE GREEN BANK NORTHERN CELESTIAL CAP PULSAR SURVEY. i. SURVEY DESCRIPTION, DATA ANALYSIS, AND INITIAL RESULTS. *The Astrophysical Journal* 791, 1 (jul 2014), 67.
- [44] TAN, C. M., BASSA, C. G., COOPER, S., DIJKEMA, T. J., ESPOSITO, P., HESSELS, J. W. T., KONDRATIEV, V. I., KRAMER, M., MICHELLI, D., SANIDAS, S., ET AL. Lofar discovery of a 23.5 s radio pulsar. *The Astrophysical Journal* 866, 1 (Oct 2018), 54.
- [45] TAYLOR, J. H. A Sensitive Method for Detecting Dispersed Radio Emission. 15 (June 1974), 367.

- [46] TENDULKAR, S. P., BASSA, C. G., CORDES, J. M., BOWER, G. C., LAW, C. J., CHATTERJEE, S., ADAMS, E. A. K., BOGDANOV, S., BURKE-SPOLAOR, S., BUTLER, B. J., ET AL. The host galaxy and redshift of the repeating fast radio burst frb 121102. *The Astrophysical Journal* 834, 2 (Jan 2017), L7.
- [47] TENDULKAR, S. P., GIL DE PAZ, A., KIRICHENKO, A. Y., HESSELS, J. W. T., BHARDWAJ, M., ÁVILA, F., BASSA, C., CHAWLA, P., FONSECA, E., KASPI, V. M., KEIMPEMA, A., KIRSTEN, F., LAZIO, T. J. W., MARCOTE, B., MASUI, K., NIMMO, K., PARAGI, Z., RAHMAN, M., PAYÁ, D. R., SCHOLZ, P., AND STAIRS, I. The 60 pc Environment of FRB 20180916B. 908, 1 (Feb. 2021), L12.
- [48] THORNTON, D., STAPPERS, B., BAILES, M., BARSDELL, B., BATES, S., BHAT, N. D. R., BURGAY, M., BURKE-SPOLAOR, S., CHAMPION, D. J., COSTER, P., D'AMICO, N., JAMESON, A., JOHNSTON, S., KEITH, M., KRAMER, M., LEVIN, L., MILIA, S., NG, C., POSSENTI, A., AND VAN STRATEN, W. A population of fast radio bursts at cosmological distances. *Science* 341, 6141 (July 2013), 53–56.
- [49] WANG, N., MANCHESTER, R. N., JOHNSTON, S., RICKETT, B., ZHANG, J., YUSUP, A., AND CHEN, M. Long-term scintillation observations of five pulsars at 1540 MHz. *Monthly Notices of the Royal Astronomical Society* 358, 1 (03 2005), 270–282.
- [50] YAO, J., MANCHESTER, R. N., AND WANG, N. Ymw16: Electron-density model. *Astrophysics Source Code Library* (2019), ascl-1908.

- [51] ZHANG, B. FRB 121102: A repeatedly combed neutron star by a nearby low-luminosity accreting supermassive black hole. *The Astrophysical Journal* 854, 2 (Feb. 2018), L21.
- [52] ZHANG, Y. G., GAJJAR, V., FOSTER, G., SIEMION, A., CORDES, J., LAW, C., AND WANG, Y. Fast radio burst 121102 pulse detection and periodicity: A machine learning approach. *The Astrophysical Journal* 866, 2 (Oct 2018), 149.