The Astrophysical Journal, 891:174 (16pp), 2020 March 10 https://doi.org/10.3847/1538-4357/ab7376

First SETI Observations with China's Five-hundred-meter Aperture Spherical Radio Telescope (FAST)

Zhi-Song Zhang[1,2,3,4], Dan Werthimer[3,4], Tong-Jie Zhang[5], Jeff Cobb[3,4], Eric Korpela[3], David Anderson[3], Vishal Gajjar[3,4], Ryan Lee[4,6,7], Shi-Yu Li[5], Xin Pei[2,8], Xin-Xin Zhang[1], Shi-Jie Huang[1], Pei Wang[1], Yan Zhu[1], Ran Duan[1], Hai-Yan Zhang[1], Cheng-jin Jin[1], Li-Chun Zhu[1], and Di Li[1,2,9]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, People's Republic of China
[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[3] Space Sciences Laboratory, University of California Berkeley, Berkeley, CA 94720, USA
[4] Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA
[5] Department of Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China
[6] Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA
[7] Department of Computer Science, University of California Berkeley, Berkeley, CA 94720, USA
[8] Xinjiang Astronomical Observatory, CAS, 150, Science 1-Street, Urumqi, Xinjiang 830011, People's Republic of China
[9] CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
Received 2019 September 23; revised 2020 January 17; accepted 2020 February 5; published 2020 March 17

Abstract

The Search for Extraterrestrial Intelligence (SETI) attempts to address the possibility of the presence of technological civilizations beyond the Earth. Benefiting from high sensitivity, large sky coverage, and an innovative feed cabin for China's Five-hundred-meter Aperture Spherical radio Telescope (FAST), we performed SETI's first observations with FAST's newly commissioned 19 beam receiver; we report preliminary results in this paper. Using the data stream produced by the SERENDIP VI real-time multibeam SETI spectrometer installed at FAST, as well as its off-line data processing pipelines, we identify and remove four kinds of radio frequency interference (RFI): zone, broadband, multibeam, and drifting, utilizing the Nebula SETI software pipeline combined with machine-learning algorithms. After RFI mitigation, the Nebula pipeline identifies and ranks interesting narrowband candidate ET signals, scoring candidates by the number of times candidate signals have been seen at roughly the same sky position and same frequency, signal strength, proximity to a nearby star or object of interest, along with several other scoring criteria. We show four example candidate groups that demonstrate this RFI mitigation and candidate selection. This preliminary testing on FAST data helps to validate our SETI instrumentation techniques as well as our data processing pipeline.

Unified Astronomy Thesaurus concepts: Astrobiology (74); Sky surveys (1464); Astronomical instrumentation (799)

1. Introduction

The Search for Extraterrestrial Intelligence (SETI), also known as the search for technosignatures (Tarter 2006; Wright et al. 2018b), is a growing field in astronomy. This is partially due to the supercomputer and big data revolution, machine-learning technology, the privately financed Breakthrough Listen Initiative, the thousands of recently discovered exoplanets, as well as the construction of new facilities, including the Five-hundred-meter Aperture Spherical radio Telescope (FAST) (Li et al. 2018).

"The probability of success is difficult to estimate, but if we never search the chance of success is zero" (Cocconi & Morrison 1959). Grimaldi & Marcy (2018) calculate the number of electromagnetic signals reaching Earth based on parameters from the Drake equation. Loeb et al. (2016) calculate the relative formation probability per unit time of habitable Earthlike planets within a fixed comoving volume of the universe. They found that life in the universe is most likely to exist near $\sim$0.1 $M_{\odot}$ stars 10 trillion years from now. Lingam & Loeb (2019) study photosynthesis on habitable planets around low-mass stars to examine if this kind of planet can receive enough photons in an active range waveband of 400–750 nm to sustain Earthlike biospheres.

Radio SETI (Cocconi & Morrison 1959) is an important technique because Earth's atmosphere is relatively transparent at many radio wavelengths, and radio emissions have low extinction through the interstellar medium (Tarter 2001; Siemion et al. 2013). Many radio SETI experiments have been conducted at Green Bank, Arecibo, Parkes, Meerkat, and several other single dish and array telescopes. Some recent examples are Siemion et al. (2013), MacMahon et al. (2018), Price et al. (2018), Chennamangalam et al. (2017), and Enriquez et al. (2017). The first experiment for SETI with the Murchison Widefield Array, one of four precursors for the Square Kilometre Array telescope, has an extremely large field of view (Tingay et al. 2016). The Very Large Array SETI experiment by Gray & Mooley (2017) implemented the search for artificial radio signals from the nearby galaxies M31 (Andromeda) and M33 (Triangulum).

FAST, Earth's largest single-aperture telescope (Nan et al. 2000, 2011), has unique advantages for SETI observations. FAST can observe declinations from $-14^{\circ}.3 < \delta < +65^{\circ}.7$ (versus $+1^{\circ}.5 < \delta < +38^{\circ}.5$ at Arecibo) due to its geographical location and active surface. FAST's sensitivity is $\sim$1800 $m^2\ K^{-1}$ (Arecibo's is about 1100 $m^2\ K^{-1}$).

This paper presents the first results of SETI at FAST. The FAST SETI instrument was installed by the University of California, Berkeley, SETI group in 2018 September, and we have conducted preliminary commensal and targeted observations over the past year. This paper is organized as follows. Section 2 describes the commensal observations at FAST. An overview of the SETI analysis pipeline and SERENDIP data acquisition, reduction, and analysis are described in Section 3.

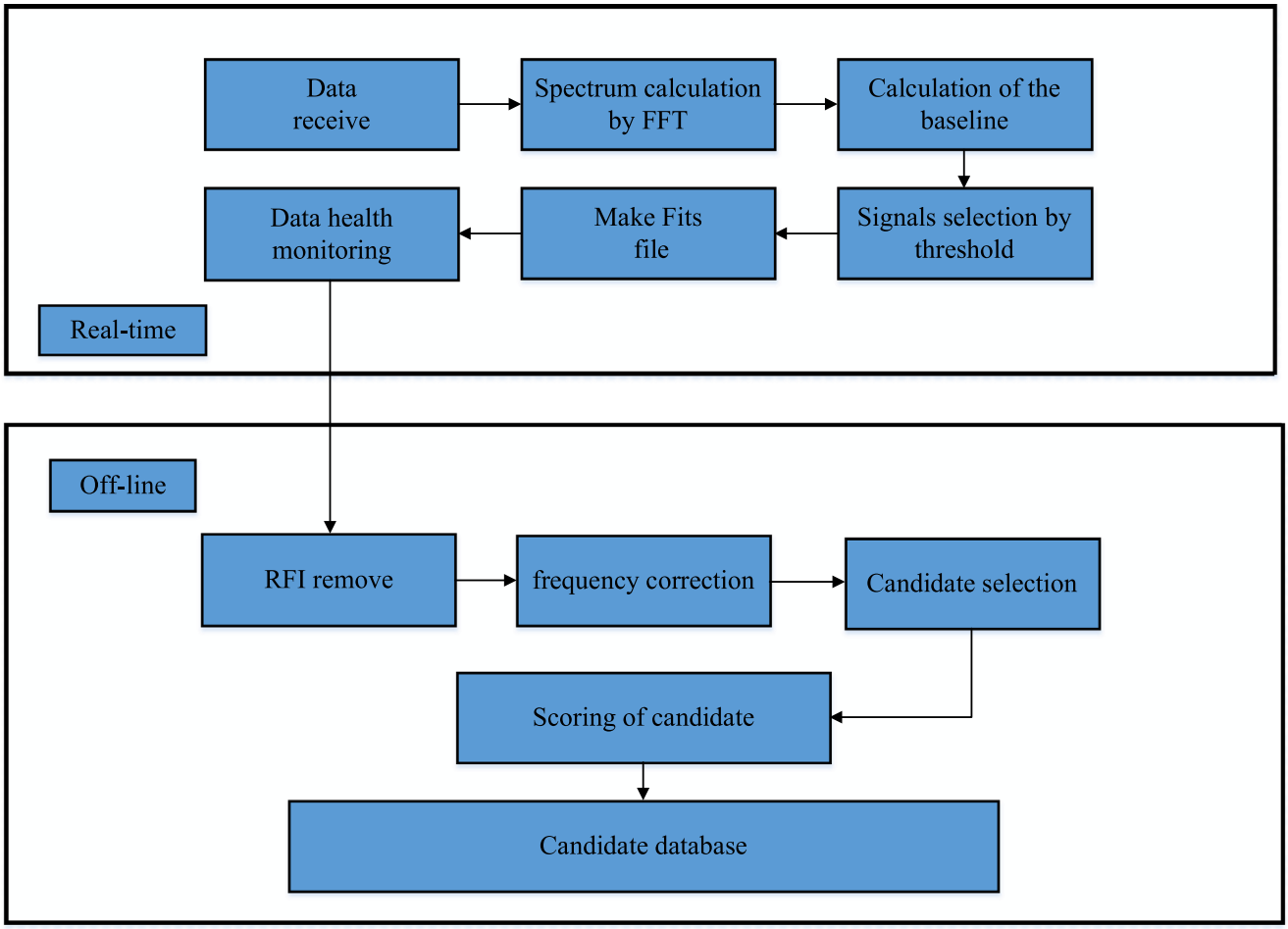

Figure 1. Overall data processing framework. The whole pipeline consists of two main parts, a real-time part and an off-line part. In the real-time part, we reduce the data according to the regulation we set and save into an ETFits file. In the off-line part, we clean the data and select the candidates.

Radio frequency interference (RFI) removal is discussed in Section 4, and machine learning for RFI removal and candidate selection are provided in Section 5. Results are presented in Section 6. A conclusion and future plans are in Section 7.

2. Commensal Observation for SETI at FAST

Radio astronomy is a discipline that relies on observation. However, observing time on large telescopes is typically oversubscribed and often only one source can be observed at a time. To increase sky coverage, commensal observation (Bowyer et al. 1983) is being increasingly employed. During commensal observation, although the direction of the telescope is determined by the primary observer, secondary observers can receive a copy of the raw data in real time. This is the technique used for SETI at FAST.

The Chinese astronomical community has planned a drift-scan program covering the 57% of the celestial sphere ($-14\overset{\circ}{.}3 < \delta < +65\overset{\circ}{.}7$), called the Commensal Radio Astronomy FAST Survey (CRAFTS) (Li et al. 2018). CRAFTS plans to use more than 5000 hr of telescope time. We plan to commensally analyze the sky survey data to find possible ETI candidate targets and to then do follow-up observations on these targets.

3. SERENDIP Data Acquisition, Reduction and Analysis

The amount of raw data produced by these observations will be very large. Using the FAST 19 beam receiver and a sampling rate of 1 billion samples per second (Gsps), the data produced per second will be:

$$1\ \text{Gsps} \times 1\ \text{byte/sample} \times 2\ \text{pol} \times 19\ \text{beams} = 38\ \text{GB s}^{-1}.$$

Such a volume of data is too much for available storage systems. An automated pipeline for data reduction, RFI removal, and candidate selection is very important. We use SERENDIP VI (Cobb et al. 2000; Archer et al. 2016), a real-time data processing system, and Nebula (E. J. Korpela et al. 2020a, in preparation), an off-line data analysis pipeline originally designed for use with SETI@home (E. J. Korpela et al. 2020b, in preparation) together with machine learning to reduce and analyze data. SERENDIP VI is described in this section. Figure 1 shows the overall data processing framework.

The SERENDIP VI system is a 128M channel spectrum analyzer, covering frequency bands from 1000–1500 MHz with a frequency resolution of about 3.725 Hz. The system is composed of a front end, based on field programmable gate array (FPGA) systems, and a back end, based on graphics processing units (GPUs), connected by a 10 Gbps ethernet switch.

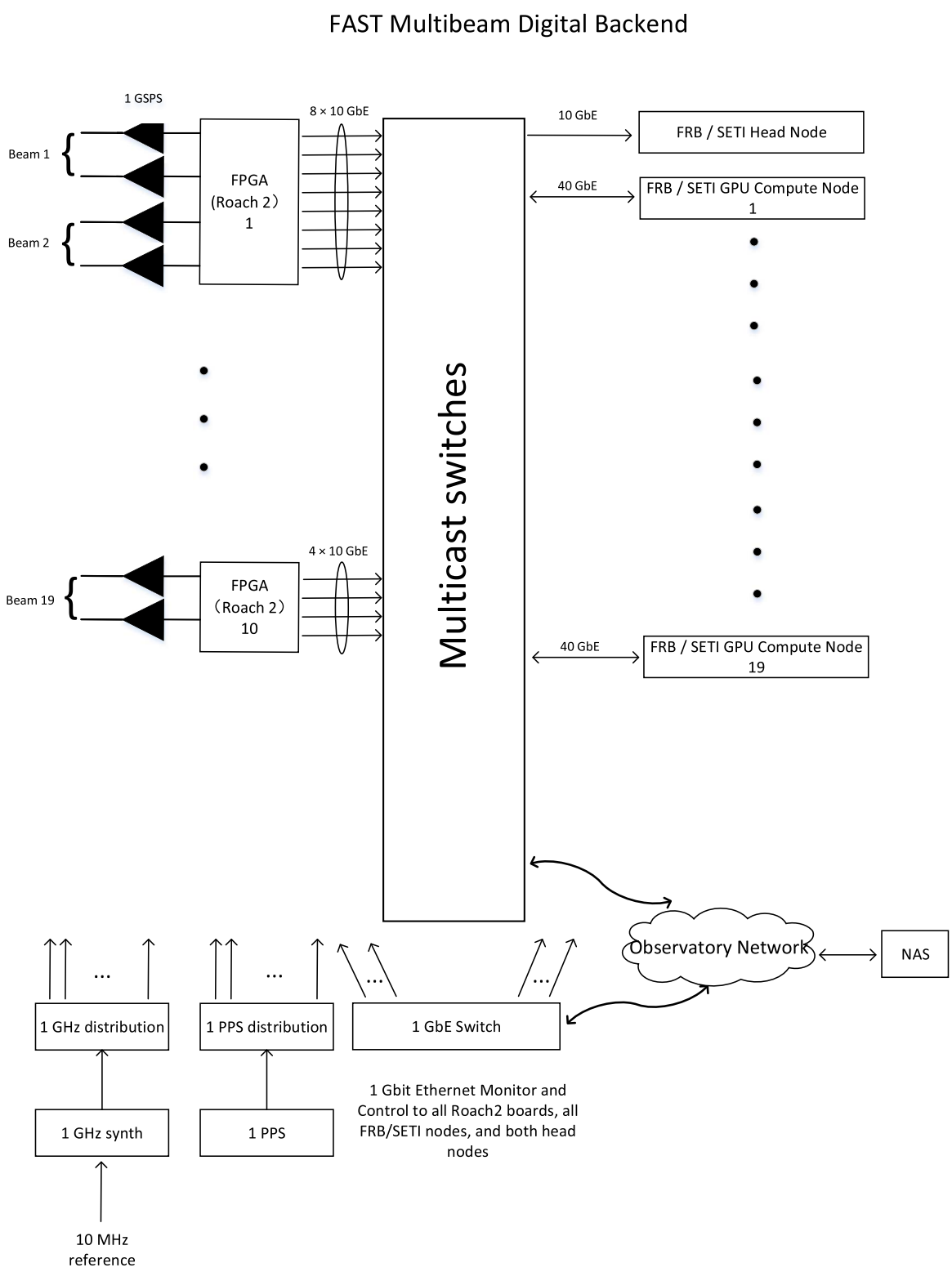

Figure 2. Architecture of the FAST multibeam SETI instrument. Thirty-eight analog signals (19 beams, each with two polarizations) are digitized and the digital samples are packetized by a bank of FPGAs (10 FPGAs are required for 19 beams at two beams per FPGA). Accumulated spectra for the FRB experiment are also formed and packetized at this stage. Packets are then multicast through the 10GbE switch and received by GPU-equipped compute nodes. Each compute node handles one beam, computing high-resolution spectra and identifying narrowband signals with powers significantly above the noise. A head node provides control and monitoring. Precise timing is achieved via a 10 MHz reference and a 1 pulse per second signal. The entire instrument is also attached to the observatory network to allow remote development and operation.

The architecture of the FAST multibeam SETI instrument is shown in Figure 2. The front end performs analog to digital conversion resulting in 8 bit digital samples. These are packetized into 4 KB ethernet packets and multicast across the network to the back end. Each FPGA can process data from two beams, requiring 10 boards for the 19 beam feed array. The

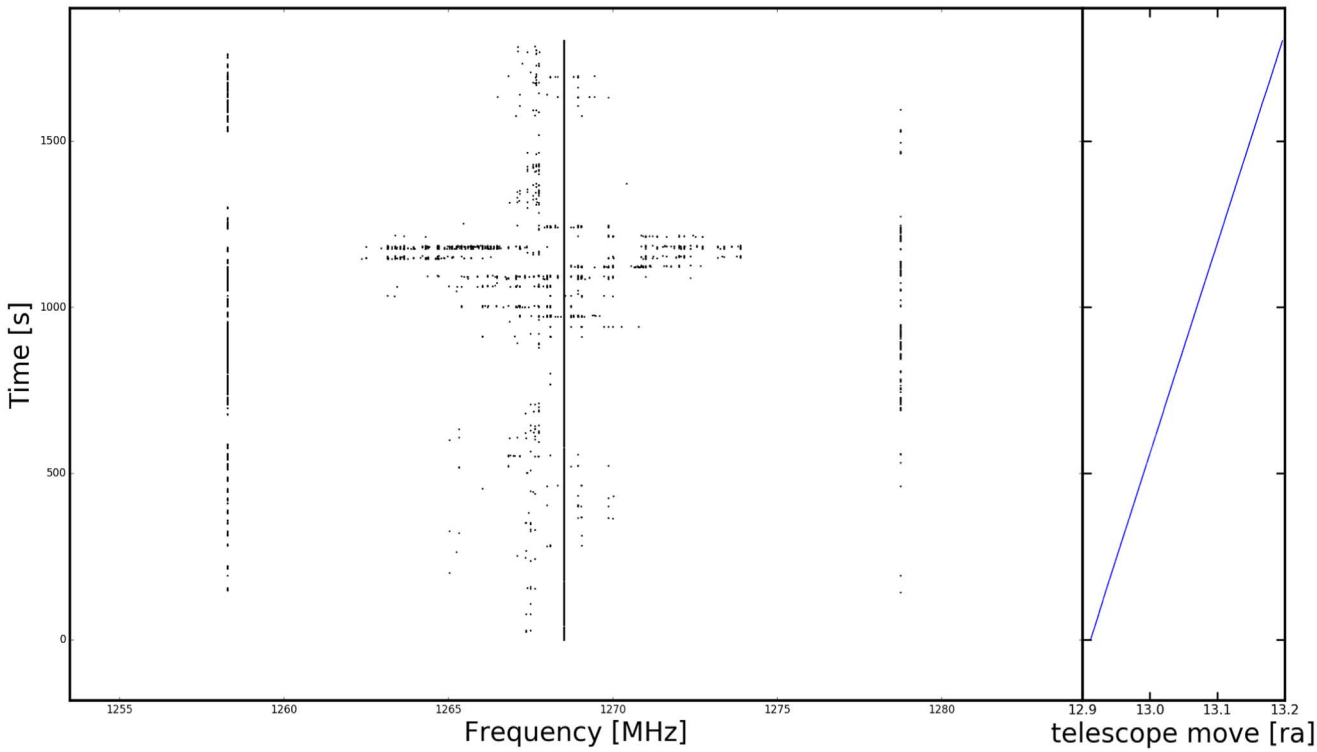

Figure 3. Waterfall plot showing time vs. frequency for a zone RFI and the R.A. of FAST in equatorial coordinates. There is a typical zone RFI around 1268 MHz. Because they often contaminate the entire frequency channel, we call them zone RFI. In this figure, and other waterfall figures in this paper, the R.A. coordinates are not exact (see the position error discussion in Section 4.3). The decl. is constant during these drift-scan observations; for this particular observation, the decl. was 0:42:55.5.

FPGA system employed is CASPER ROACH2 (Hickish et al. 2016). These are widely used and supported within the radio astronomy community. In addition to sending raw samples to the SETI back end, the FPGAs form integrated power spectra for transmission to the fast radio burst (FRB) back end. Multicast is employed so that identical data streams can be received and processed by multiple experiments.

At the back end, data reduction and analysis are performed on the GPU and the results, along with observatory metadata (e.g., time, pointing, receiver status), are stored in files conforming to the Flexible Image Transport System (FITS) standard. Each GPU-based compute node handles one beam for both the SETI and FRB pipelines, requiring 19 compute nodes. In addition to the compute nodes, there is a head node that handles monitor and control functions and hosts a Redis database used for cross-node coordination.

Precise timing is provided by the observatory's 10 MHz reference and a 1 pulse per second signal. Long term data storage and system backups are handled by the observatory's network attached storage.

The GPU processing pipeline consists of the following steps.

(1) The raw time-domain voltage data are copied to GPU memory and transformed into complex frequency domain data via cuFFT (a fast Fourier transform library provided by NVIDIA).

(2) Based on the formula below:

$$\text{Power} = \text{real}^2 + \text{imag}^2, \quad (1)$$

we sum the real part and imaginary part of each channel in the frequency domain to get the power spectrum. This and subsequent steps are coded as calls to Thrust, NVIDIA's C++ template library.

(3) The baseline of the power spectrum is calculated with respect to the local mean utilizing a sliding 8K spectral bin window. The spectrum is then normalized with respect to this baseline.

(4) Finally, the normalized power of each channel is compared to a signal-to-noise ratio (S/N) threshold value. Those channels exceeding the threshold ($S/N > 30$) are recorded in the FITS file, including the signal time, frequency, detection power, mean power, telescope pointing, and other information. Each channel recorded is called a hit.

4. Nebula for RFI Removal

RFI removal has always been a crucial part of radio telescope data analysis. There has been much work done to date on the problem of RFI removal, including the SumThreshold method (Offringa et al. 2010), singular value decomposition, surface fitting and smoothing (Winkel et al. 2007) and sky-subtracted incoherent noise spectra (Wilensky et al. 2019).

Nebula is a complete off-time data analysis system, including data cleaning, RFI removal, candidate selection, and scoring. Here we just give a brief introduction of the RFI removal part using an example from FAST data.

4.1. Narrowband RFI

Narrowband RFI is the most common kind of RFI coming from artificially engineered signals on Earth, especially within the FAST electromagnetic environment. In Nebula, we separate the narrowband RFI into two kinds. Narrowband RFI that is stable in frequency (called zone RFI in Nebula) and narrowband RFI that drifts in frequency (called drifting RFI in Nebula).

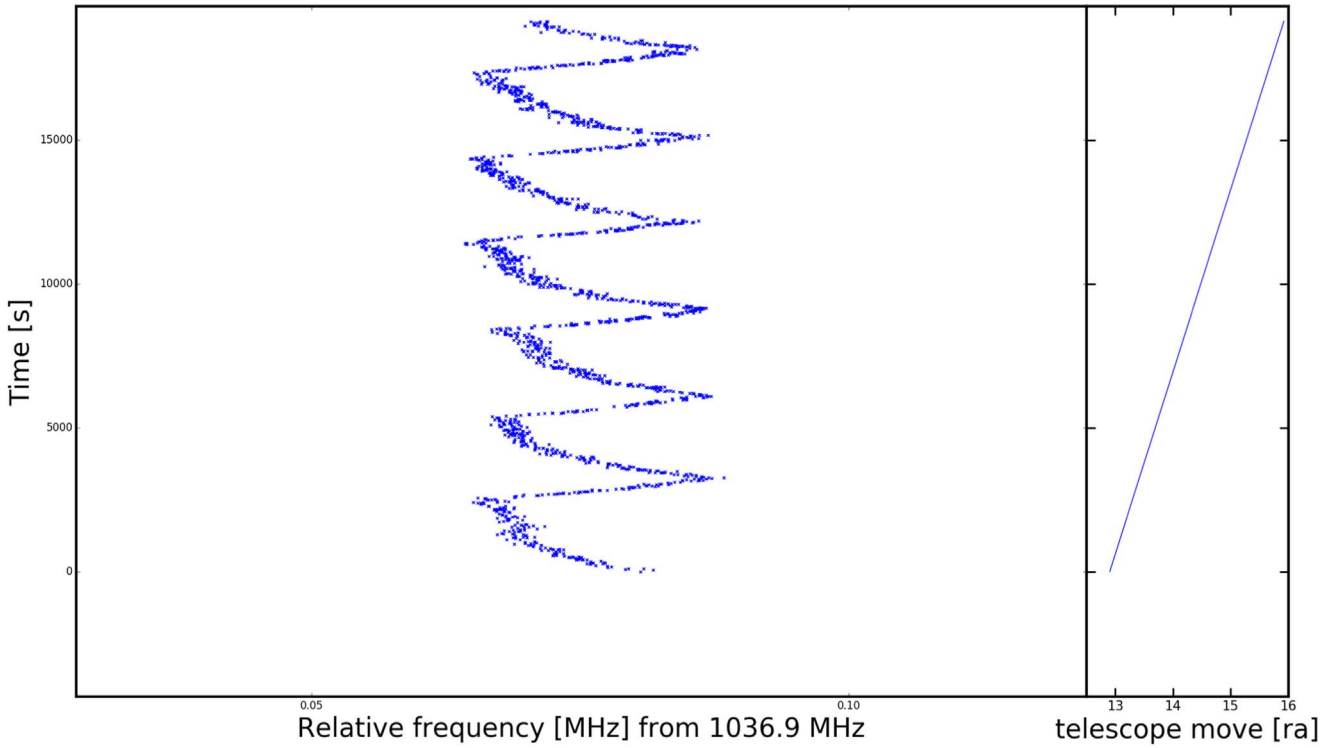

Figure 4. Same as Figure 3 but for a drifting RFI.

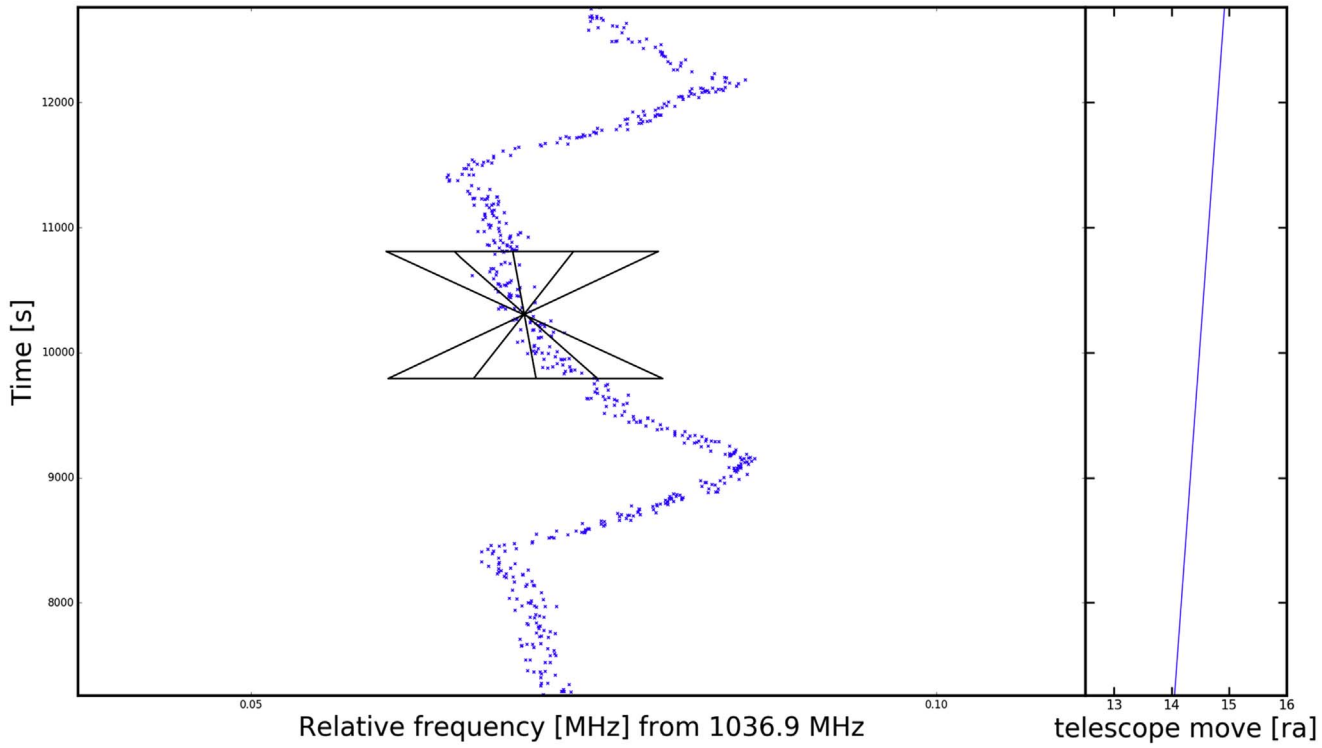

Figure 5. Method of removing drifting RFI in waterfall plot of Figure 4. For simplicity, we only show four bins of the whole 21 bins there.

4.1.1. Zone RFI

Zone RFI is narrowband RFI exhibiting a stable frequency which persists throughout most or all of the total data set. These "zones" become global exclusion filters. There are many sources of such interference, including television and radio broadcasts, and cell phone and satellite signals. As shown in Figure 3, compared with the distribution of the hits we expected, the zone RFI in the middle is more concentrated, forming a vertical narrow band. The existence of such RFI seriously affects the extraction of candidate targets.

For zone RFI, we examine each frequency bin and calculate the number of hits in this frequency bin. If the number reaches above the threshold, all the hits in this bin can be marked as zone RFI and removed. The threshold is set by the Poisson cumulative probability function. The Poisson cumulative distribution is used to model the number of events occurring within a given time interval. The Poisson cumulative distribution probability function can be expressed as:

$$F(x;\lambda) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^{i}}{i!}, \tag{2}$$

where λ is the mean number of hits in a bin, and x is the number of the hits. This function can give the probability of the number of hits in a bin. For the FAST data, we set the probability to be 1×10^{-7}, so we can get a threshold value x from the function.

4.1.2. Drifting RFI

Drifting RFI is narrowband RFI that drifts in frequency, mainly coming from cell phones. We cannot find them with the zone RFI algorithm, because the frequency of the signals is changing over time. Figure 4 is a typical drift RFI from FAST data.

Due to its changing frequency, we cannot simply remove it by frequency channel. In Nebula, we make two symmetrical triangles for each hit. As shown in Figure 5, the shape of the triangle is determined by drift rate and time, which are set empirically. For FAST data, we set the drifting rate to $20\,\mathrm{Hz\,s^{-1}}$ and Δt to 600 s. We can then separate the triangle into 21 bins. If the number of signals in each bin and its opposite three bins is above the threshold, we mark all the signals in the bins as drifting RFI. The threshold is set in the same way as Section 4.1.

4.2. Multibeam RFI

When we use the multibeam receiver, such as for the 19 beams from FAST, we can identify signals that come from nonadjacent beams but with a similar time and frequency. When a signal comes from a point in space, it can be received by one beam and maybe by an adjacent beam. However, terrestrial RFI signals are often picked up in multiple beams simultaneously. In this algorithm, each hit has a time and frequency box. If the box of one hit is in another hit's box and hits are in nonadjacent beams, they are both marked as RFI.

4.3. Pipeline of RFI Mitigation

Sections 4.1 and 4.2 demonstrate three kinds of RFI and their removal methods. In Nebula, we do not remove these three kinds of RFI separately, because narrowband RFI may probably be multibeam RFI. We just mark the hit instead of removing it when the hit is selected by any one of three methods in Nebula. Finally, the hit with one or more RFI labels will be removed after going through all three kinds of methods. Figure 6 shows the processing pipeline of Nebula. When the data was taken for this paper, real time pointing information was not available from the FAST telescope. For this early test data, telescope position information was calculated and merged with the SETI data in postprocessing. We cross-checked our approximate calculated positions with the more accurate positions provided to us later by the FAST telescope pointing system—our position errors are under 0°.014, which is adequate for testing the Nebula data analysis pipeline, although not ideal for the upcoming sky survey. We

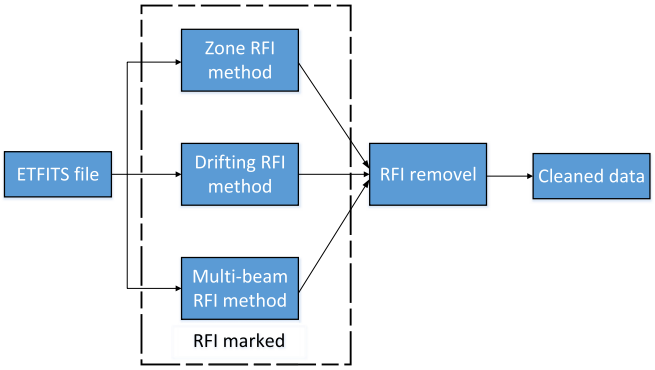

Figure 6. Block diagram of Nebula RFI removal. The full set of raw data go through each of three RFI identification methods. Any hit that is marked by a least one of these methods is removed. All remaining hits constitute the set of clean data that is passed on to candidate selection.

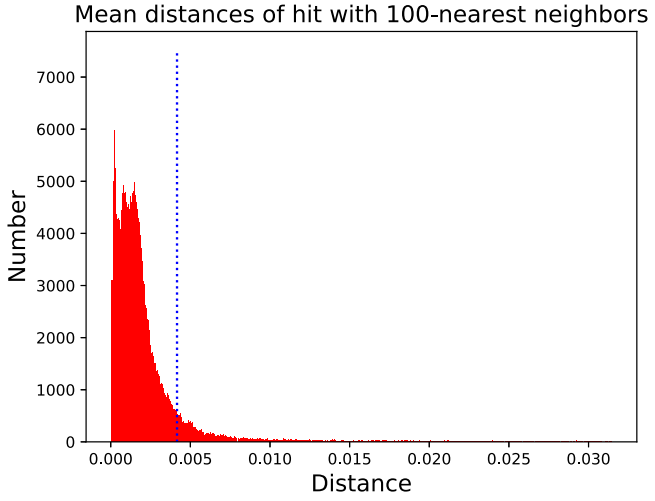

Figure 7. Histogram of the number of hits vs. mean distance. The vertical line is our chosen RFI removal threshold. Most hits are on the left side of the line with low mean distances, which means they come from big clusters, and can be removed. Distance shown in the graph is the mean distance between each hit and its nearest 100 hits in the waterfall plot showed before, expressed by time interval and frequency interval, i.e., $\sqrt{\Delta t^2 + \Delta f^2}$.

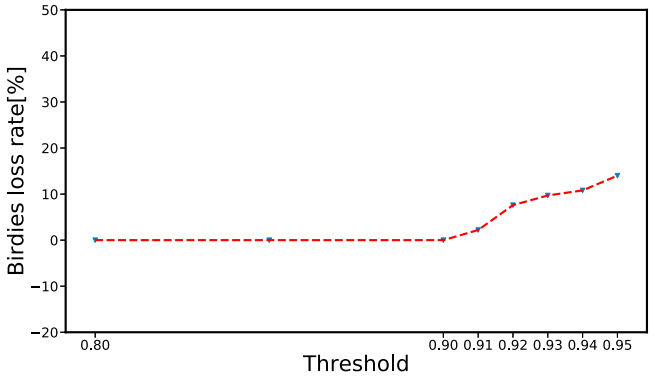

Figure 8. Loss rate curve of birdies. The experiment is based on the FAST data with birdies injected. As shown in the figure, loss of birdies begins at a threshold of 91%, so in order to reduce the possibility of potential signal loss, we choose the threshold of 90%.

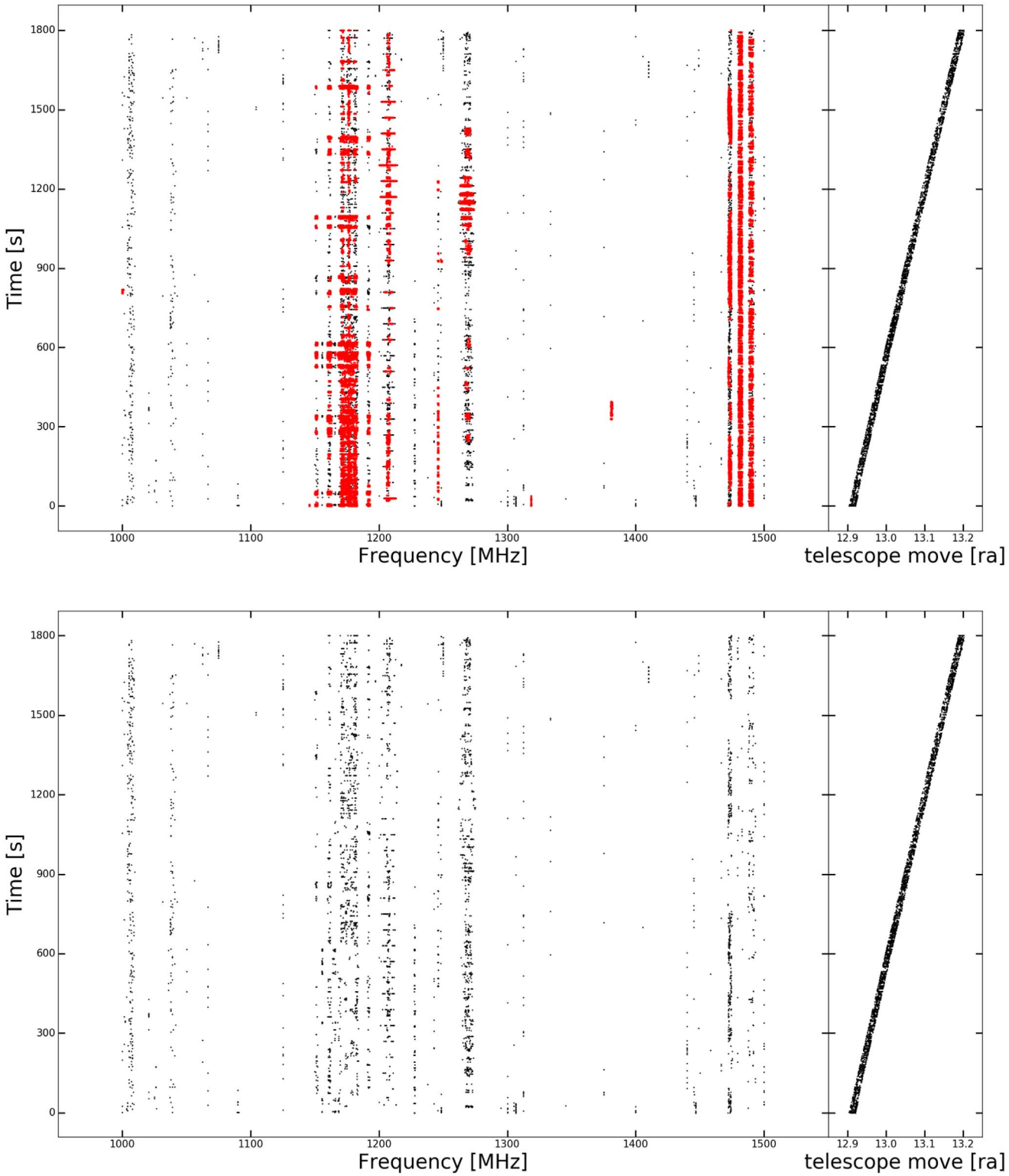

Figure 9. Waterfall plot before (top panel) and after (bottom panel) RFI removal through machine learning. The top panel shows data including RFI, with the RFI marked by the KNN algorithm marked in red. The bottom panel shows hits after KNN RFI removal. Although there is still part of the RFI left in the bottom panel, the KNN algorithm has removed more than 90% of the RFI left by Nebula. The remaining hits are sparse enough not to confuse for our candidate selection pipeline.

expect the pointing errors in the upcoming sky survey will be significantly lower, as our SETI spectrometer will have access to accurate real time telescope pointing data, and the spectrometer will merge this pointing data with the science data before it is recorded to disk.

5. Machine Learning for RFI Removal and Candidate Selection

It should be noted that we employ the traditional assumption that advanced life wishing to be detected at interstellar distances will use narrowband microwave emissions, as narrowband

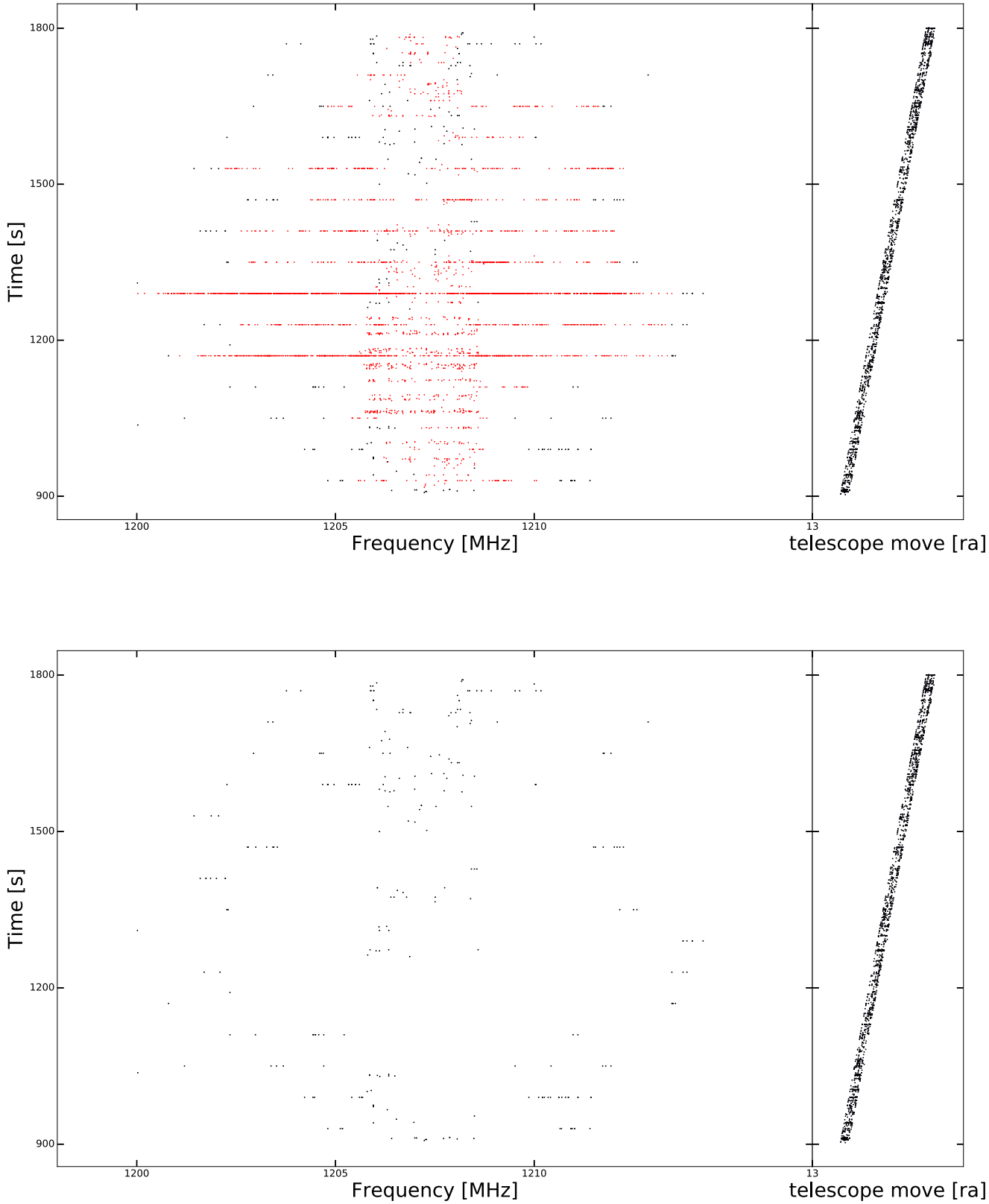

Figure 10. Example of removing broadband RFI using machine learning. Broadband RFI does not often appear, but once present, it can contaminate many frequency bins and signals. It is primarily due to lightning, sparks, or power transmission cables near the Earth's surface. The top panel shows data including broadband RFI, which are in red, while the bottom panel shows data after KNN RFI removal.

signals are most easily distinguished from emissions produced by natural astrophysical sources of microwave emission. The narrowest known natural sources, astrophysical masers, have minimum frequency widths of about 500 Hz (Cohen et al. 1987). Thus we primarily focus on searches for narrowband signals from ETI.

5.1. Machine Learning for RFI Removal

Our Nebula pipeline can remove most of the RFI. Normally, >99% of RFI can be removed, but there are still some atypical RFI left. Two examples are narrowband RFI and broadband RFI. We are unable to detect all narrowband RFI, because at times the power is below our threshold. It is easy to find broadband RFI if the bandwidth is very large. If the bandwidth is less than several MHz, it is much more difficult and our traditional methods do not detect it.

Of course, these RFI signals still have characteristics in common: such emissions cluster in time and frequency on specific scales. The ETI signals typically would not form a large cluster on time and frequency scales, being narrow in frequency and of no longer a duration than that of the telescope observation of a point in the sky in the drift scan. We use the k-nearest neighbor (KNN) algorithm to find the nearest 100 hits for each hit and calculate the mean distance, as was first applied by V. Gajjar et al. (2020, in preparation). Figure 7 plots the histogram of the mean distances for the 100 nearest hits. The blue line is the upper RFI threshold, which is based on removing a specified percentage (90%) of the total hits. Hits below this threshold are presumed to be RFI events. We tested our threshold choice using experiments conducted to test the percentage of simulated "birdie" signals lost due to RFI removal at different threshold values. This can be seen in Figure 8. The threshold value of 90% conserved the most birdies, which maximizes the probability of conserving true ETI signals while minimizing the unremoved RFI. See Section 6.2 for more discussion of birdie generation.

Figure 9 is a waterfall plot before and after using machine learning to remove RFI. Figure 10 is an example of removing broadband RFI. The KNN algorithm can effectively remove much of the broadband and narrowband RFI left by Nebula. Following this removal the data are clean enough for the selection of candidates.

5.2. Machine Learning for Candidate Selection

After RFI removal, we apply a data density clustering algorithm, Density-based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996), to find candidate clusters.

DBSCAN separates the data into three kinds of points: "core points," "adjacent points," and "noisy points." "Core points" are those with more than a predefined number of points within a specified radius, while "adjacent points" have fewer points nearby but belong to a core point. "Noisy points" have neither enough points around nor belong to a core point. DBSCAN does not divide data into parts but identifies tight clusters in any shape against the background. All the noisy points form a background group which then can be discarded. There are two main parameters: eps and Nmin. Nmin stands for minimum cluster population. We set this parameter to 5, because we do not want to miss candidate groups with only a few points. For eps, which stands for maximum distance from the core point, we used a set of experiments to choose the most suitable value. The experiments result are shown in Figure 11. When eps exceeds 140, the loss rate of birdies drops to 0. We set eps to 145, because ETI signals will probably be a smaller cluster than our birdies during a drifting observation.

For each cluster, we apply two rules to distinguish RFI from non-RFI clusters.

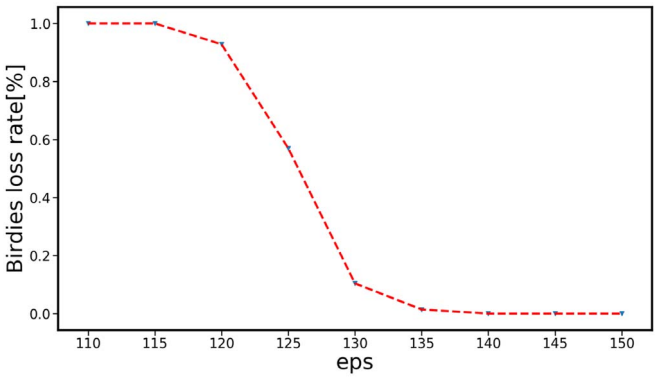

Figure 11. Eps test curve. The experiment is based on the FAST data with birdies injected. As shown in the figure, the loss rate of birdies decreases from when eps is equal to 115 and drops to 0 when eps is equal to 140.

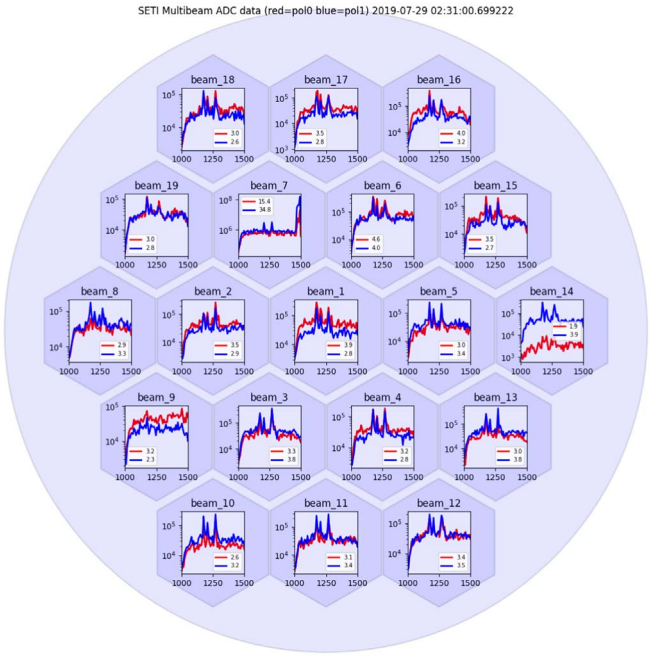

Figure 12. SETI multibeam monitor. There are 19 plots in the figure, standing for 19 beams. Each plot has two lines. The line in red represents the power spectrum of polarization 0 while the line in blue represents the power spectrum of polarization 1.

1. We calculate the sky angle between hits in a cluster, and determine whether the cluster subtends less than 1.5 times the receiver beam's width. We use a larger than unity width because some extraterrestrial signals could be received by a beam and an adjacent beam simultaneously.
2. We calculate the duration and bandwidth of the cluster. We select clusters with narrow bandwidth and a duration less than a few tens of seconds.

These two rules define the expected characteristics of extraterrestrial signals as they would appear during drift-scan sky survey observations. The final step is that we save all the candidates by location pixel. Pixels are defined to be small rectangles that divide the celestial sphere. All the candidates selected by the pipeline will be be saved by pixels in order to further analyze the candidate targets of the same sky position.

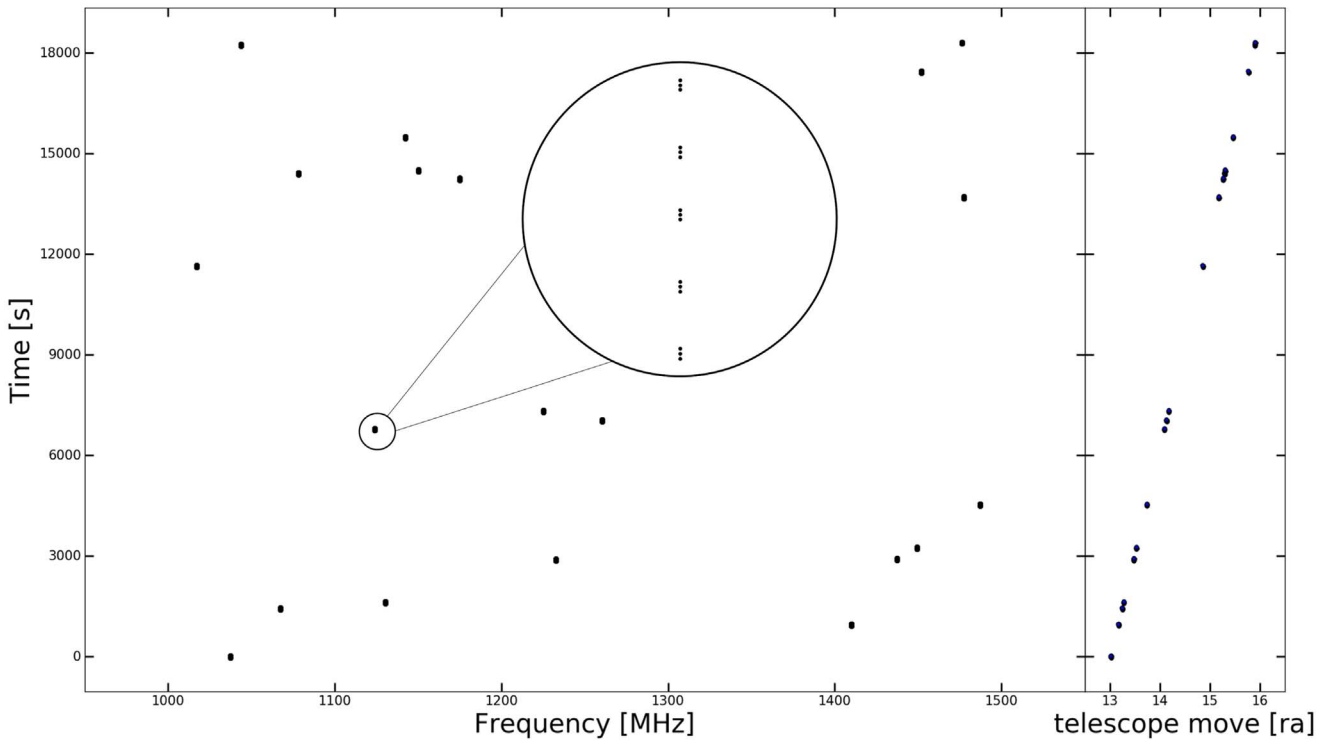

Figure 13. Birdies generated for FAST data. Every black spot is a group of birdies. We zoom in one of them where details are shown in the larger circle. Each group contains several small groups that depend on how many beams go through the birdies.

6. Analysis and Results of FAST Data

6.1. FAST Data

Our data were collected during a drift-scan survey performed by FAST during commissioning in 2019 July. To ensure beam health and data integrity, we employ a system health monitor. The monitor (Figure 12) shows a coarse (2048 point) power spectrum (with rms deviation) for each of 19 beams in both polarization at each beam's location in the focal array. We update the power spectrum and the rms in the plot windows every five minutes to monitor beam health.

The data from each beam are saved into a FITS file separately. The size of each file is roughly 1 GB and records a table of data for each hit received. Each hit contains 14 values, including time, position, beam number, power, S/N, channel number, and frequency. Telescope information can be stored in header fields. This format is known as ETFITS and has been utilized by recent SERENDIP projects.

6.2. Data Processing

In order to verify the validity of our pipeline, some artificial candidate targets, called "birdies," are added to our data. We randomly generate some signals along the moving trajectory of beam one, and if other beams go through the same position, we add more signals with the same frequency into that beam. Birdies generated are shown in Figure 13, which contains 20 groups and 294 signals.

After adding birdies, we use Nebula and the KNN pipeline to remove the RFI. 99.9063% of hits are removed by our pipeline while only 5.1020% of birdies were removed. Note that we make a temporary change to Nebula. To get an ideal velocity, we first remove zone RFI and then simultaneously remove other kinds of RFI. Details are shown in Table 1. Most of the RFI in FAST are removed by the zone RFI algorithm. In the future, we will try to find the source of zone RFI and eliminate them earlier in the data acquisition process. This would make our data set smaller and cleaner, which could increase the probability of finding an ETI signal. We track the birdies removed by Nebula and find that they come from the same birdie group and are heavily polluted by ambient RFI. Figure 14 shows the data before and after our pipeline. We are convinced our pipeline effectively removes most of the RFI and protects most of the birdies and potential candidates.

Table 1
Ratio of Each Kind of RFI

Type of RFI	Zone RFI	Drifting RFI	Multibeam RFI	Total
Number of bytes (MB)	82524.9	735.2	700.9	83961
Percentage (%)	98.1976	0.8748	0.8340	99.9063

6.3. Candidate Selection

With the clean data, we use DBSCAN to search for dense clusters and select clusters with the two rules above. The result is shown in Table 2. Nineteen groups of birdies and 83 groups of candidates are selected. We are very happy to note that our pipeline found 277 birdies, 94.2177% of the total. Only one group of birdies is removed by Nebula. This means that we can successfully find candidate targets that match our desired characteristics. All of the candidates found are shown in Figure 15.

In order to identify whether the candidates we found are unremoved RFI, we examine them with all hits found in the raw data to see if they are associated with RFI features. Four groups of candidates with a raw data background are shown in

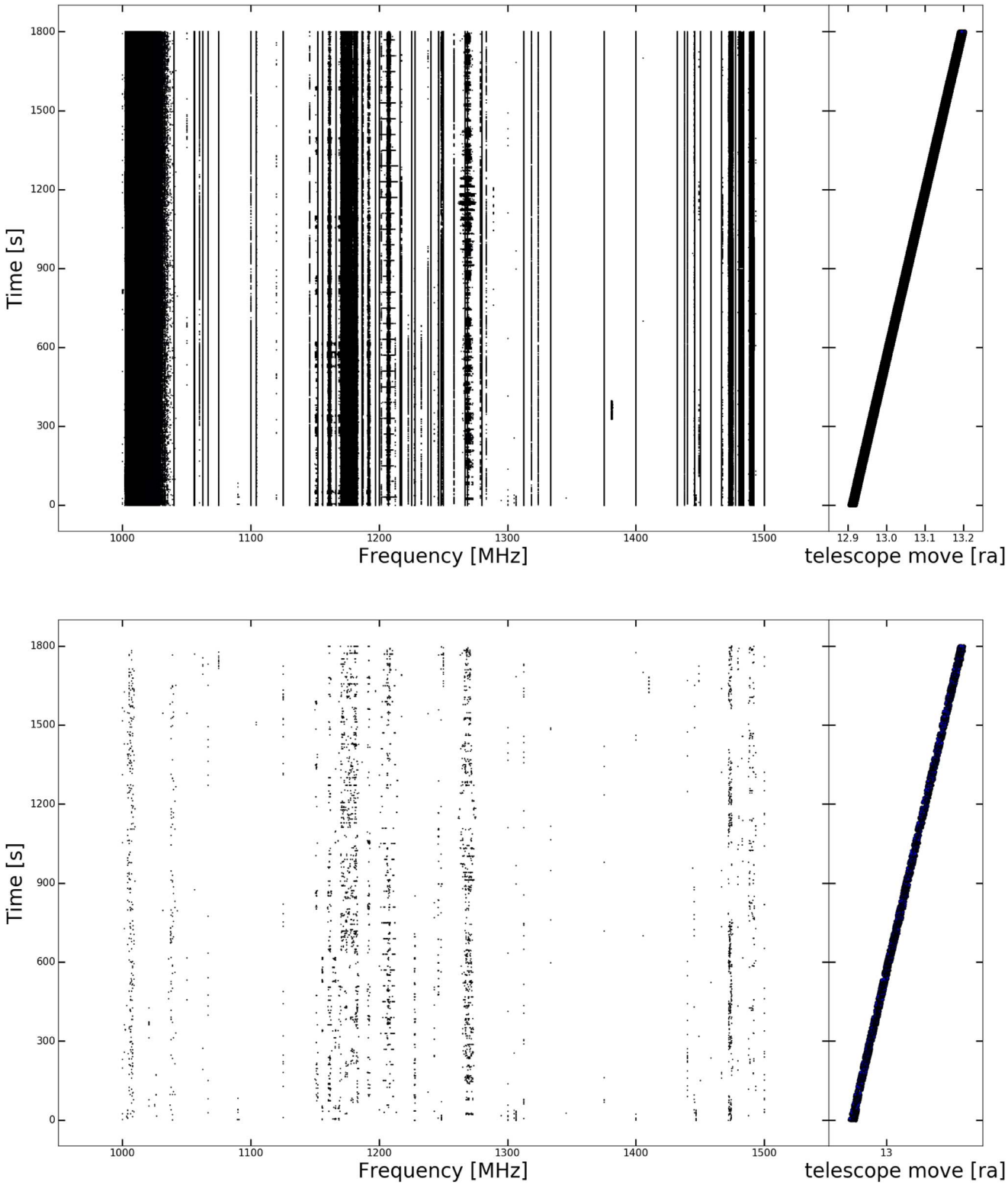

Figure 14. RFI excision of FAST data. The top panel shows unprocessed data while the bottom one shows processed data. We can see directly that most of the RFI are removed. 98.1976% are zone RFI, denoted by the vertical bars in the top panel. Here note that we only show data spanning 1800 s in this figure.

Figures 16 and 17. We examine all of the candidates and find that most of them have an obvious connection to RFI features. This is unsurprising because our RFI removal algorithm cannot completely remove all RFI. The unremoved RFI events match roughly to our definition of a candidate's characteristics. Improvement of the RFI algorithm is an iterative process that we expect to continue as long as this system is used on FAST.

For the 5 hr data, we find two group candidates of interest which have no hits around them. These two group candidates are shown in Figure 17. The red group around 1055 MHz and 4280 s is called Group 1. The red group around 1055 MHz and

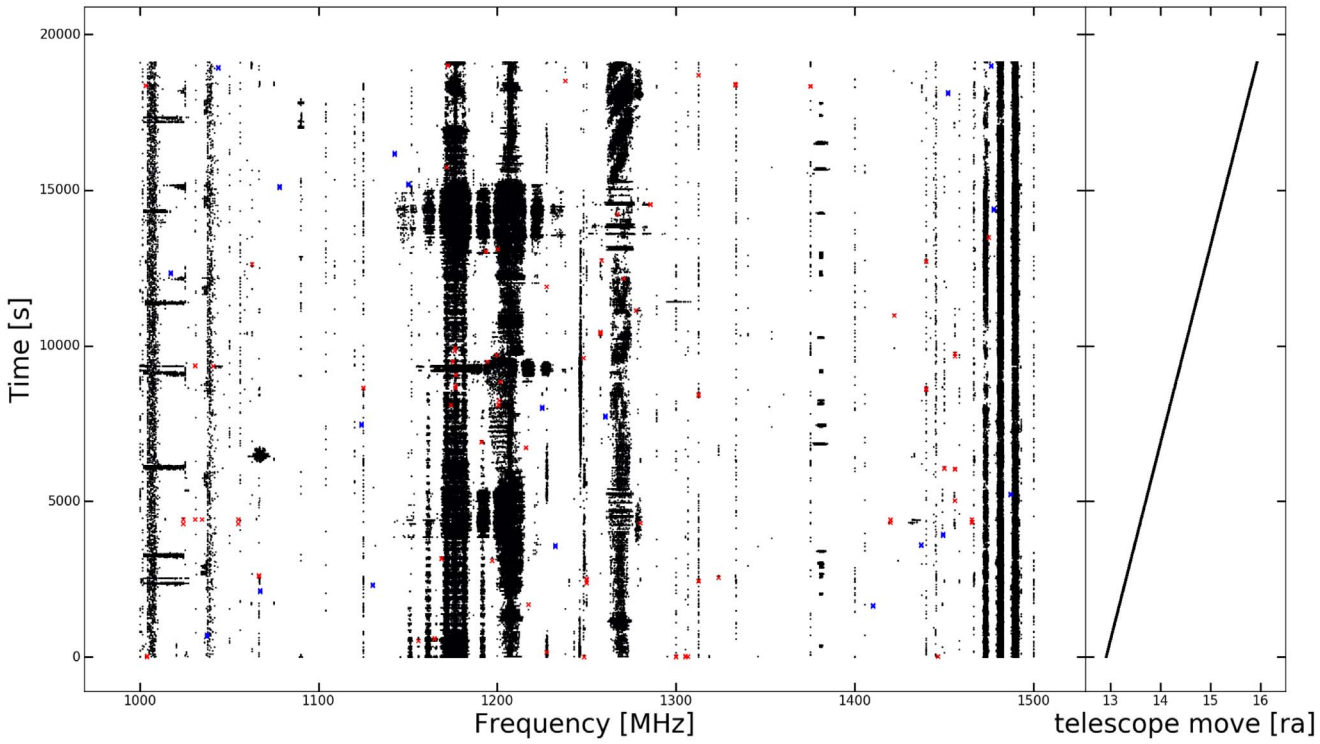

Figure 15. Candidates found in FAST data. In the figure, Blue clusters are 277 birdies while red clusters are other candidates beyond birdies. Each blue or red point represents a group of hits due to the big size of the waterfall plot. Note that the background does not represent the raw data but data after Nebula, because the raw data is too big to plot at the same time.

Table 2
Result of Candidate Selection with DBSCAN

Type of candidate	Birdies	Candidates	Total
Number of hits	277	593	870
Number of groups	19	83	102

4430 s is called Group 2. We zoom into the two groups of candidates in Figure 18. Group 1 is in the top panel of the figure with five hits in the same frequency channel. Group 2 is in the bottom panel of the figure with 80 hits in six successive channels. Group 1 is all from beam 15, and the duration is about 3 s. The duration of Group 2 is about 20 s. Events in beam 14 last 5 s, followed by 15 s of events in beam 15. The position of these two groups of candidates are shown in Table 3.

The two groups of candidates, and even the other candidates which are verified to be part of the RFI, are very consistent with the two rules of ETI assumption. This result indicates that our data processing pipeline can pick out the signals that fit the assumption of ETI. Actually, this paper has presented the effectiveness of our method for RFI removal and candidate selection, which can certainly guarantee the successful implementation of SETI observation with FAST.

6.4. The RFI Environment of FAST

As the largest single-aperture telescope in the world, the electromagnetic environment of FAST is a very important consideration. Due to its extreme sensitivity, it is a big challenge for FAST to mitigate RFI. The solutions for RFI mitigation of FAST include the electromagnetic compatibility measures of the telescope and the maintenance of radio-quiet zones around the site (Zhang et al. 2019). Besides, a lot of RFI monitoring for FAST has been done. Table 4 lists the known RFI sources. From the table, civil aviation occupies 960–1215 MHz and ASIASTAR, a geostationary satellite, occupies 1467–1492 MHz. Some satellites occupy the middle frequency band, which only appears when they pass over or near the telescope. Unfortunately, the RFI monitoring does not have the same high-frequency resolution as our SETI back end. Some of the narrowband RFI is not picked up by the monitor, and we also cannot currently mark the RFI with exact frequency channels due to the low-frequency resolution. Finding the sources of these narrowband RFI as the next step will be important and meaningful work.

7. Conclusion and Future Plans

FAST is the largest single-aperture telescope in the world; its 19 beam receiver allows rapid and sensitive sky surveys with robust RFI rejection, ideal for SETI.

We have developed a SETI signal detection pipeline to operate on FAST's 19 beam SETI instrument. We conducted the first observational test of SETI with FAST's 19 beam receiver in 2019 July. By injecting test signals ("birdies") into our data, we demonstrated the pipeline's RFI removal capability. The ETI signal candidates were selected by the criteria outlined in Section 5, and our data processing pipeline is working well on the preliminary data collected from FAST. We hope that these ETI signal candidates could come from some warm Earth-size planets in the Milky Way, the number of which can be roughly predicted by the Drake equation (Drake 1961; Drake et al. 2010; Grimaldi et al. 2018).

With the SETI capabilities demonstrated in this work, one can estimate the equivalent isotropic radiated power (EIRP) (Enriquez et al. 2017) required of an alien transmitter to be

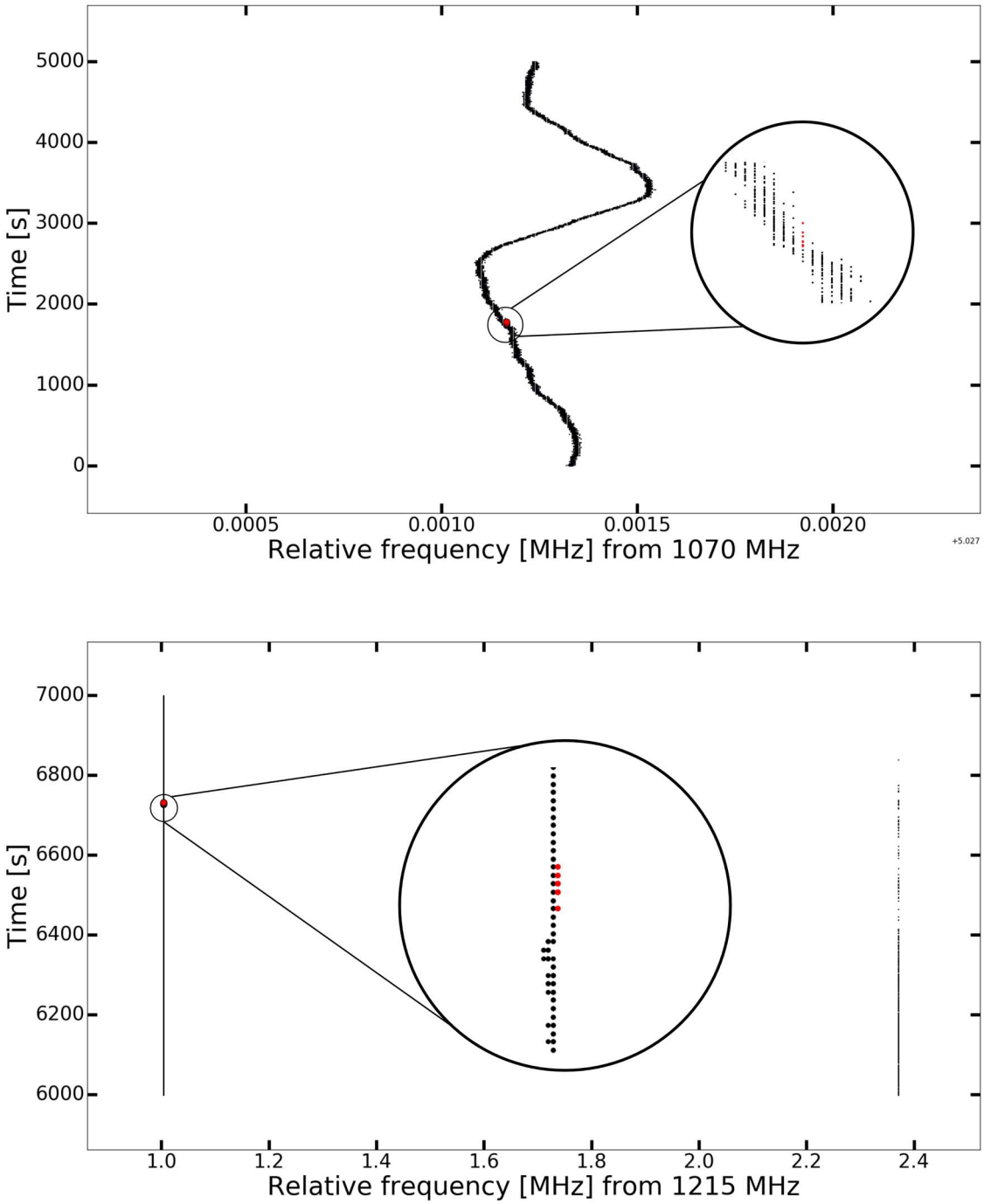

Figure 16. Two groups of candidates within raw data. The red cluster in each figure is candidates while black points are raw data. It is clear that these candidates still belong to part of RFI. The one in the top panel is drifting RFI while one in the bottom panel is zone RFI.

Table 3
Positions and Times of Two Candidate Groups

Group No.	Beam No.	Starting time (JD)	Duration (s)	Starting position (J2000 R.A. J2000 Decl.)	Off position (J2000 R.A. J2000 Decl.)
1	15	2458682.258681	3	21:00:22.51 0:44:04.92	21:00:24.50 0:44:04.92
2	14	2458682.260405	5	21:02:49.20 0:38:13.20	21:02:54.23 0:38:13.20
2	15	2458682.260475	15	21:02:57.94 0:44:04.92	21:03:12.92 0:44:04.92

Note. Telescope coordinates, provided by the FAST pointing team, have errors less than 6″.

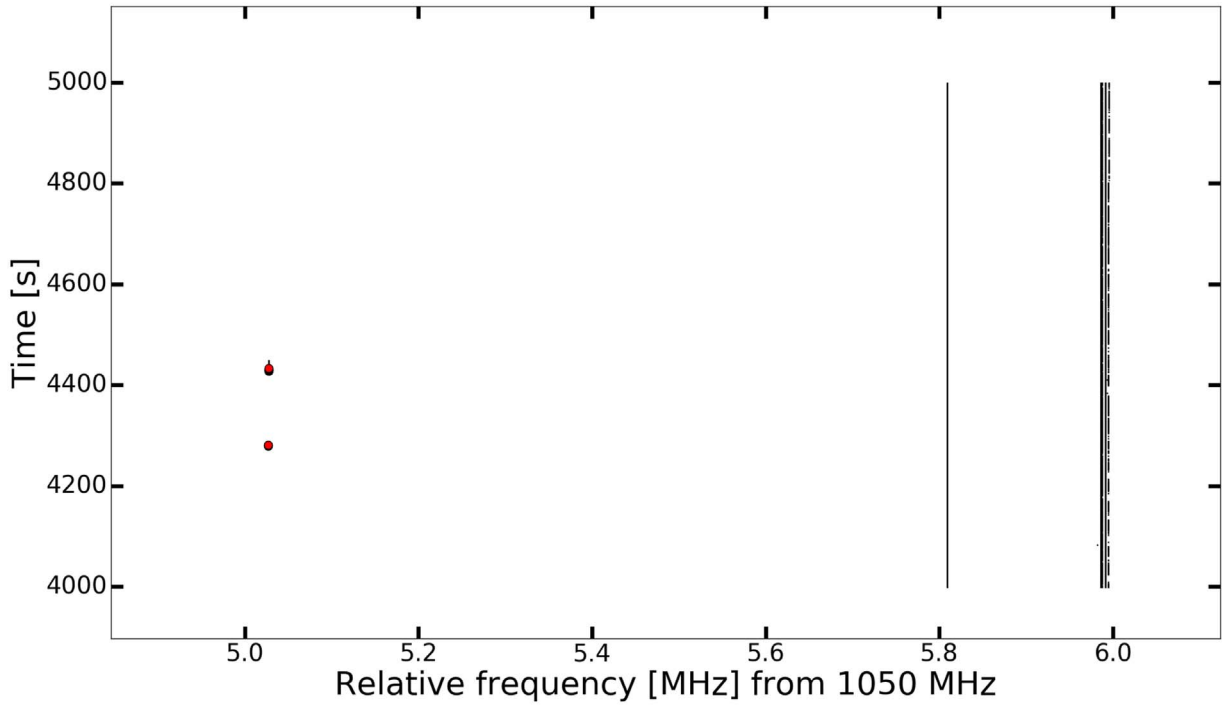

Figure 17. Two groups of interesting candidates. The red group around 1055 MHz and 4280 s is called Group 1. The red group around 1055 MHz and 4430 s is called Group 2. From the figure, we can see that no other hits from nearby frequency channels are in our 5 hr data.

Table 4
Source List of FAST RFI

Frequency band (MHz)	960–1215	1176.45 ± 1.023	1205–1209	1226.6–1229.6	1242–1250	1258–1278	1381–1386	1467–1492
RFI source	Civil aviation	GPS L5, Galileo E5a	BD2 B2, Galileo E5b	GPS L2	GLONASS L2	BD2 B3	GPS L3	ASIASTAR

detected by FAST. For a source of ETI signal candidates at $d_* = 50$ pc, a $\delta\nu = 4$ Hz bandwidth, and a $t_{\rm obs} = 0.25$ s integration, the EIRP limit of FAST is 3.53×10^{13} W, given by

$$\mathrm{EIRP} = 4\pi {d_*}^2 \sigma_{\rm thresh} \mathrm{SEFD} \sqrt{\frac{\delta\nu}{n_{\rm pol} t_{\rm obs}}}, \quad (3)$$

where the system equivalent flux density $\mathrm{SEFD} = 2k_B T_{\rm sys}/A_{\rm eff}$ and $\sigma_{\rm thresh} = \mathrm{S/N_{min}}$ is a S/N threshold value.

We are planning to take data on FAST's commensal drift-scan surveys, such as the multiyear CRAFTS survey; we also intend to observe targets from the *Transiting Exoplanet Survey Satellite*, as well as selected stars within $d_* = 50$ pc from the Sun, e.g., Isaacson et al. (2017).

We are planning several improvements to the multiyear CRAFTS multibeam SETI sky survey on FAST:

To improve FAST's sky survey sensitivity to narrowband signals, we are working on upgrading the spectral resolution of the SERENDIP VI spectrometer to 1 Hz channelization.

We also plan to continue improving our RFI mitigation and candidate selection postprocessing algorithms, which will benefit several other SETI sky surveys as well (SERENDIP6 sky surveys at Arecibo and Green Bank, and the SETI@home multibeam sky survey at Arecibo). The main purpose of our program is to search for extraterrestrial civilizations, but the project also helps study the RFI environment around FAST.

We have begun working on real-time RFI rejection and first level identification of potential candidate signals, which would trigger a 100 s raw voltage dump of time-domain data on all 38 signals from the the multibeam receiver (19 beams and two polarizations) for subsequent off-line analysis. This time-domain data would allow us to cross correlate beams, thereby reducing source position uncertainty, as well as provide more robust RFI mitigation.

We are also considering continuously recording raw time-domain data streams from all 38 signal chains on the FAST multibeam receiver, which will allow us to send out the data to SETI@home volunteers for a more thorough and sensitive analysis. SETI@home is 20 times more sensitive to narrowband signals than SERENDIP, because the enormous computing power provided by the SETI@home volunteers allows the SETI@home screensaver client to compute coherent, very long duration, spectra on tens of thousands of possible signal drift rates. The SETI@home client also searches for pulses, signals with several different bandwidths, and signals that match the telescope beam patterns. An autocorrelation algorithm is used to search for repeating patterns. Although SETI@home is more sensitive than SERENDIP, and searches for a very rich variety of signal types, we probably will not be able to record and process the full bandwidth of all the beams at FAST (recording 80 Gbits s^{-1} for several years is a lot to manage). So SETI@home would likely process a part of the FAST

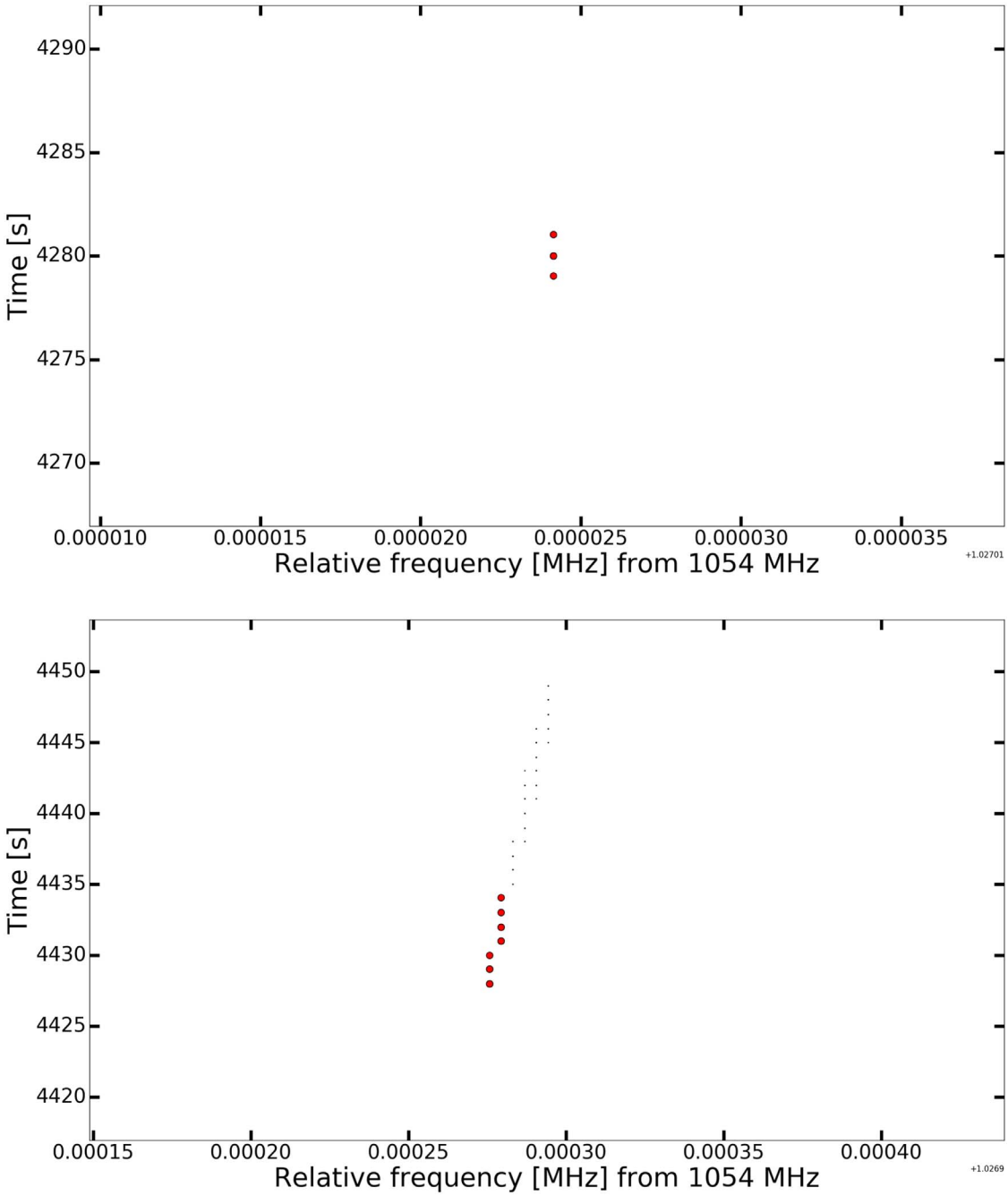

Figure 18. Close-up view of candidates in Figure 17. Group 1 in the top panel only occupies one frequency channel. Group 2 in the bottom panel occupies six successive channels, in total ∼18.6 Hz of bandwidth. Note that Group 2 is in two colors, because only the red points are found by the SETI pipeline while the black points are from the raw data.

multibeam receiver band at very high sensitivity, while SERENDIP6 would process the full band at a reduced sensitivity. Longer term, FAST is planning a sensitive phased array feed, which could provide roughly 100 simultaneous beams; excellent for a next generation SETI sky survey.

More generally, Earthlings are just beginning to learn how we might detect other civilizations if they are out there. We have only had radio technology for a century; that is a blink of the eye in the history of the universe and life on this planet. We are beginning to explore tiny regions of the large parameter space of possible technosignatures from potential extraterrestrial civilizations (Wright et al. 2018a). Even though we are in an infant stage, SETI science and technology is growing exponentially. Radio telescope sensitivity has been doubling every 3.6 yr for the last 60 yr, and SETI spectrometer capabilities have been doubling every 20 months for the last 40 yr. This SETI sky survey commissioning work is a significant step, leading to a powerful new SETI survey on FAST.

We sincerely appreciate the referee's rapid, thorough, and thoughtful response, which helped us greatly improve our manuscript. This work was supported by the National Key R&D Program of China (2017YFA0402600), the National Science Foundation of China (grant Nos. 11929301, 11573006, 11528306, 11803054, 11690024, 11725313), the China

Academy of Sciences International Partnership Program No. 114A11KYSB20160008, Berkeley's Marilyn and Watson Alberts SETI Chair funds, the Berkeley SETI Research Center, and the Radio Astronomy Laboratory at Berkeley. This work was also supported by the National Astronomical Observatories, Chinese Academy of Sciences, and its FAST group. During this work E.J.K. and J.C. were supported in part by donations from the Friends of SETI@home (setiathome.berkeley.edu). Z.-S.Z. sincerely thanks Ling-Jie Kong for her kind help on the response to the referee.

Software: DBSCAN (Ester et al. 1996), scikit-learn (Pedregosa et al. 2011), Python 3 (van Rossum 1995; Millman & Aivazis 2011), SciPy (Virtanen et al. 2020), NumPy (van der Walt et al. 2011), Astropy (Astropy Collaboration et al. 2013, 2018), Matplotlib (Hunter 2007), Nebula (E. J. Korpela et al. 2020a, in preparation).

ORCID iDs

Zhi-Song Zhang https://orcid.org/0000-0001-9294-0363
Tong-Jie Zhang https://orcid.org/0000-0002-3363-9965
Eric Korpela https://orcid.org/0000-0001-8078-9395
Vishal Gajjar https://orcid.org/0000-0002-8604-106X
Di Li https://orcid.org/0000-0003-3010-7661

References

Archer, K., Siemion, A., Werthimer, D., et al. 2016, in 2016 United States National Committee of URSI National Radio Science Meeting (Piscataway, NJ: IEEE), 1
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bowyer, S., Zeitlin, G., Tarter, J., Lampton, M., & Welch, W. J. 1983, Icar, 53, 147
Chennamangalam, J., MacMahon, D., Cobb, J., et al. 2017, ApJS, 228, 21
Cobb, J., Lebofsky, M., Werthimer, D., Bowyer, S., & Lampton, M. 2000, in ASP Conf. Ser. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech (San Francisco, CA: ASP), 485
Cocconi, G., & Morrison, P. 1959, Natur, 184, 844
Cohen, R. J., Downs, G., Emerson, R., et al. 1987, MNRAS, 225, 491
Drake, F. D. 1961, PhT, 14, 40
Drake, F. D., Stone, R. P. S., Werthimer, D., & Wright, S. A. 2010, in Astrobiology Science Conf. 2010, Evolution and Life: Surviving Catastrophes and Extremes on Earth and Beyond, 5211
Enriquez, J. E., Siemion, A., Foster, G., et al. 2017, ApJ, 849, 104
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in KDD'96: Proc. Second Int. Conf. on Knowledge Discovery and Data Mining (Palo Alto, CA: AAAI Press), 226, https://dl.acm.org/doi/10.5555/3001460.3001507
Gray, R. H., & Mooley, K. 2017, AJ, 153, 110
Grimaldi, C., & Marcy, G. W. 2018, PNAS, 115, E9755
Grimaldi, C., Marcy, G. W., Tellis, N. K., & Drake, F. 2018, PASP, 130, 054101
Hickish, J., Abdurashidova, Z., Ali, Z., et al. 2016, JAI, 5, 1641001
Hunter, J. D. 2007, CSE, 9, 90
Isaacson, H., Siemion, A. P. V., Marcy, G. W., et al. 2017, PASP, 129, 054501
Li, D., Wang, P., Qian, L., et al. 2018, IMMag, 19, 112
Lingam, A., & Loeb, M. 2019, MNRAS, 485, 5924
Loeb, A., Batista, R. A., & Sloan, D. 2016, JCAP, 2016, 040
MacMahon, D. H. E., Price, D. C., Lebofsky, M., et al. 2018, PASP, 130, 044502
Millman, K. J., & Aivazis, M. 2011, CSE, 13, 9
Nan, R., Li, D., Jin, C., et al. 2011, IJMPD, 20, 989
Nan, R., Peng, B., Zhu, W., et al. 2000, in ASP Conf. Ser. 213, Bioastronomy 99, ed. G. Lemarchand & K. Meech (San Francisco, CA: ASP), 523
Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Price, D. C., MacMahon, D. H. E., Lebofsky, M., et al. 2018, PASA, 35, 41
Siemion, A. P. V., Demorest, P., Korpela, E., et al. 2013, ApJ, 767, 94
Tarter, J. 2001, ARA&A, 39, 511
Tarter, J. C. 2006, HiA, 14, 14
Tingay, S. J., Tremblay, C., Walsh, A., & Urquhart, R. 2016, ApJL, 827, L22
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
van Rossum, G. 1995, Python Tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Wilensky, M. J., Morales, M. F., Hazelton, B. J., et al. 2019, PASP, 131, 114507
Winkel, B., Kerp, J., & Stanko, S. 2007, AN, 328, 68
Wright, J. T., Kanodia, S., & Lubar, E. 2018a, AJ, 156, 260
Wright, J. T., Sheikh, S., Almár, I., et al. 2018b, arXiv:1809.06857
Zhang, H., Wu, M., Yue, Y., et al. 2019, in 2019 URSI Asia-Pacific Radio Science Conf. (Piscataway, NJ: IEEE), 1