



A Search for Technosignatures around 31 Sun-like Stars with the Green Bank Telescope at 1.15–1.73 GHz

Jean-Luc Margot^{1,2}, Pavlo Pinchuk², Robert Geil³, Stephen Alexander², Sparsh Arora³, Swagata Biswas⁴, Jose Cebreros⁴, Sanjana Prabhu Desai¹, Benjamin Duclos², Riley Dunne², Kristy Kwan Lin Fu², Shashwat Goel³, Julia Gonzales¹, Alexander Gonzalez², Rishabh Jain³, Adrian Lam⁵, Briley Lewis², Rebecca Lewis², Grace Li⁶, Mason MacDougall², Christopher Makarem³, Ivan Manan⁵, Eden Molina², Caroline Nagib², Kyle Neville⁴, Connor O'Toole², Valerie Rockwell^{2,7}, Yoichiro Rokushima⁸, Griffin Romanek³, Carlyn Schmidgall^{4,9}, Samar Seth³, Rehan Shah⁵, Yuri Shimane⁸, Myank Singhal², Armen Tokadjian², Lizvette Villafana², Zhixian Wang³, In Yun², Lujia Zhu⁴, and Ryan S. Lynch¹⁰

¹Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, Los Angeles, CA 90095, USA; jlmi@epss.ucla.edu

²Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

³Department of Computer Science, University of California, Los Angeles, CA 90095, USA

⁴Department of Mathematics, University of California, Los Angeles, CA 90095, USA

⁵Department of Electrical Engineering, University of California, Los Angeles, CA 90095, USA

⁶Department of English, University of California, Los Angeles, CA 90095, USA

⁷Department of Anthropology, University of California, Los Angeles, CA 90095, USA

⁸Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA

⁹Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095, USA

¹⁰Green Bank Observatory, P.O. Box 2, Green Bank, WV 24494, USA

Received 2020 September 9; revised 2020 November 6; accepted 2020 November 17; published 2021 January 6

Abstract

We conducted a search for technosignatures in 2018 and 2019 April with the L -band receiver (1.15–1.73 GHz) of the 100 m diameter Green Bank Telescope. These observations focused on regions surrounding 31 Sun-like stars near the plane of the Galaxy. We present the results of our search for narrowband signals in this data set, as well as improvements to our data processing pipeline. Specifically, we applied an improved candidate signal detection procedure that relies on the topographic prominence of the signal power, which nearly doubles the signal detection count of some previously analyzed data sets. We also improved the direction-of-origin filters that remove most radio frequency interference (RFI) to ensure that they uniquely link signals observed in separate scans. We performed a preliminary signal injection and recovery analysis to test the performance of our pipeline. We found that our pipeline recovers 93% of the injected signals over the usable frequency range of the receiver and 98% if we exclude regions with dense RFI. In this analysis, 99.73% of the recovered signals were correctly classified as technosignature candidates. Our improved data processing pipeline classified over 99.84% of the ~ 26 million signals detected in our data as RFI. Of the remaining candidates, 4539 were detected outside of known RFI frequency regions. The remaining candidates were visually inspected and verified to be of anthropogenic nature. Our search compares favorably to other recent searches in terms of end-to-end sensitivity, frequency drift rate coverage, and signal detection count per unit bandwidth per unit integration time.

Unified Astronomy Thesaurus concepts: [Search for extraterrestrial intelligence \(2127\)](#); [Technosignatures \(2128\)](#); [Astrobiology \(74\)](#); [Exoplanets \(498\)](#); [Solar analogs \(1941\)](#); [Radio astronomy \(1338\)](#); [Milky Way Galaxy \(1054\)](#)

Supporting material: machine-readable table

1. Introduction

We describe a search for radio technosignatures with the L -band receiver of the 100 m diameter Green Bank Telescope (GBT). We used a total of 4 hr of GBT time in 2018 and 2019 to observe the regions around 31 Sun-like stars near the plane of the Galaxy. We have so far prioritized the detection of narrowband (~ 10 Hz) signals because they are diagnostic of engineered emitters (e.g., Tarter 2001).

Our search builds on the legacy of technosignature searches performed in the period 1960–2010 (Tarter 2001; Tarter et al. 2010, and references therein) and previous searches conducted by our group (Margot et al. 2018; Pinchuk et al. 2019). Other recent searches include work conducted by Siemion et al. (2013), Harp et al. (2016), Enriquez et al. (2017), Gray & Mooley (2017), and Price et al. (2020).

Our choice of search parameters has key advantages compared to the Breakthrough Listen (BL) searches described by Enriquez et al. (2017) and Price et al. (2020), which contend

with much larger data volumes. Specifically, our search provides roughly uniform detection sensitivity over the entire range of frequency drift rates (± 8.86 Hz s^{-1}), whereas the BL searches suffer a substantial loss in sensitivity due to the spreading of signal power across up to 13–26 frequency resolution cells. In addition, we cover a range of frequency drift rates that is 2–4 times wider than the BL searches with a time resolution that is 51 times better.

Our search algorithms are distinct from the BL algorithms in that they alleviate the necessity of discarding approximately kilohertz-wide regions of frequency space around every detected signal. We abandoned this practice in previous work (Pinchuk et al. 2019). In this work, we further refine our algorithm by implementing a candidate signal detection procedure that relies on the concept of prominence and removing the requirement to compute the bandwidth of candidate signals. Our new approach, combined with better end-to-end sensitivity and drift rate coverage, enables a hit rate density or signal detection count

per unit bandwidth per unit integration time that is ~ 200 times larger than that of the BL search described by Price et al. (2020).

A key measure of the robustness and efficiency of a data processing pipeline is provided by the technique of signal injection and recovery (e.g., Christiansen et al. 2013), whereby artificial signals are injected into the raw data and the fraction of signals recovered by the pipeline is quantified. Despite the importance of this metric, we are not aware of an existing tool to quantify the recovery rates of data processing pipelines in radio technosignature searches. We make a first step toward the implementation of this tool and show that our current pipeline detects 93% of the injected signals over the usable frequency range of the receiver and 98% if we exclude regions with dense radio frequency interference (RFI). In addition, our pipeline correctly flagged 99.73% of the detected signals as technosignature candidates. Although our current implementation requires additional work to fully capture the end-to-end pipeline efficiency, it can already illuminate imperfections in our and other groups' pipelines and be used to calibrate claims about the prevalence of other civilizations (e.g., Enriquez et al. 2017).

The article is organized as follows. Our data acquisition and analysis techniques are presented in Sections 2 and 3, respectively. Our preliminary signal injection and recovery analysis is described in Section 4. The main results of our search are outlined in Section 5. In Section 6, we describe certain advantages of our search, including dechirping efficiency, drift rate coverage, data archival practices, candidate detection algorithm, and hit rate density. We also discuss limits on the prevalence of other civilizations, search metrics such as the Drake figure of merit (DFM), and reanalysis of previous data with our latest algorithms. We close with our conclusions in Section 7.

2. Data Acquisition and Preprocessing

Our data acquisition techniques are generally similar to those used by Margot et al. (2018) and Pinchuk et al. (2019). Here we give a brief overview and refer the reader to these other works for additional details.

2.1. Observations

We selected 31 Sun-like stars (spectral type G, luminosity class V) with a median galactic latitude of $0^\circ 85$ (Table 1) because their properties are similar to the only star currently known to harbor a planet with life. We observed these stars with the GBT during two 2 hr sessions separated by approximately 1 yr. During each observing session, we recorded both linear polarizations of the L -band receiver with the GUPPI back end in its baseband recording mode (DuPlain et al. 2008), which yields 2-bit raw voltage data after requantization with an optimal four-level sampler (Kogan 1998). The center frequency was set to 1.5 GHz, and we sampled 800 MHz of bandwidth between 1.1 and 1.9 GHz, which GUPPI channelized into 256 channels of 3.125 MHz each. We validated the data acquisition and analysis processes at the beginning of each observing session by injecting a monochromatic tone near the receiver front end and recovering it at the expected frequency in the processed data.

We observed all of our targets in pairs in order to facilitate the detection and removal of signals of terrestrial origin (Section 3.2). The sources were paired in a way that approximately minimized telescope time overhead, i.e., the sum of the times spent

repositioning the telescope. Pairings were adjusted to avoid pair members that were too close to one another on the plane of the sky with the goal of eliminating any possible ambiguity in the direction of origin of detected signals. Specifically, we required angular separations larger than 1° between pair members, i.e., several times the $\sim 8'.4$ beamwidth of the GBT at 1.5 GHz.

Each pair was observed twice in a four-scan sequence: A, B, A, B. The integration time for each scan was 150 s, yielding a total integration time of 5 minutes per target⁻¹. CoRoT 102810550 and CoRoT 110777727 were each observed for an additional two scans. With 66 scans of 150 s duration each, our total integration time amounts to 2.75 hr.

2.2. Sensitivity

Margot et al. (2018) calculated the sensitivity of a search for narrowband signals performed with the 100 m GBT. Assuming a System Equivalent Flux Density of 10 Jy, integration time of 150 s, and frequency resolution of 3 Hz, they found that sources with flux densities of 10 Jy can be detected with a signal-to-noise ratio (S/N) of 10. The results of that calculation are directly applicable here because our search parameters are identical to that study. Specifically, our search is sensitive to transmitters with the effective isotropic radiated power (EIRP) of the Arecibo planetary radar transmitter (2.2×10^{13} W) located 420 lt-yr from Earth (Margot et al. 2018, their Figure 5). Transmitters located as far as the most distant source (CoRoT 102963038; $\sim 10,407$ lt-yr) and with < 1000 times the Arecibo EIRP can also be detected in this search. Although we selected Sun-like stars as primary targets, our search is obviously sensitive to other emitters located within the beam of the telescope. A search of the Gaia DR2 catalog (Gaia Collaboration 2016, 2018) inspired by Wlodarczyk-Sroka et al. (2020) reveals that there are 15,031 known stars with measured parallaxes within the half-power beamwidths associated with our 31 primary sources. The median and mean distances to these sources are 2088 and 7197 lt-yr, respectively.

2.3. Computation of Power Spectra

After unpacking the digitized raw voltages from 2-bit to 4-byte floating point values, we computed consecutive power spectra with 2^{20} point Fourier transforms, yielding a frequency resolution of $\Delta f = 2.98$ Hz. We chose this frequency resolution because it is small enough to provide unambiguous detections of narrowband (< 10 Hz) technosignatures and large enough to examine Doppler frequency drift rates of up to nearly ± 10 Hz s⁻¹ (Section 2.4). We processed all channels within the operating range of the GBT L -band receiver (1.15–1.73 GHz), excluding channels that overlap the frequency range (1200–1341.2 MHz) of a notch filter designed to mitigate RFI from a nearby aircraft detection radar, for a total processed bandwidth of 438.8 MHz. Although Enriquez et al. (2017) and Price et al. (2020) used the L -band receiver over a larger frequency range (1.1–1.9 GHz), we used a narrower range because we observed serious degradation of the bandpass response beyond the nominal operating range of the receiver.

In order to correct for the bandpass response of GUPPI's 256 channels, we fit a 16-degree Chebyshev polynomial to the median bandpass response of a subset of the processed channels that did not include strong RFI and that were not close to the cutoff frequencies of filters located upstream of the GUPPI back end. After applying the bandpass correction to all

Table 1
Target Host Stars Listed in Order of Observation

Host Star	Spectral Type	Long. (deg)	Lat. (deg)	Parallax (mas)	Distance (lt-yr)	MJD of Scan 1
2018 April 27 20:00–22:00 UT						
TYC 1863-858-1	G0V	185.5	−0.21	1.9547 ± 0.036	1669 ± 31	58,235.84503472
TYC 1868-281-1	G2V	185.3	−0.65	3.8622 ± 0.046	844 ± 10	58,235.84737269
HD 249936	G2V	186.2	−0.80	1.9515 ± 0.0412	1671 ± 35	58,235.85430556
TYC 1864-1748-1	G2V	186.6	+0.87	3.0350 ± 0.043	1075 ± 15	58,235.85671296
HIP 28216	G2V	186.9	−0.93	1.2479 ± 0.0909	2614 ± 190	58,235.86393519
HD 252080	G5V	188.1	+1.18	5.7621 ± 0.0511	566 ± 5	58,235.86640046
HD 251551	G2V	186.5	+1.65	4.5105 ± 0.0755	723 ± 12	58,235.87356481
HD 252993	G0V	186.3	+3.13	6.9544 ± 0.0401	469 ± 3	58,235.87591435
TYC 742-1679-1	G5V	195.8	−2.07	8.4009 ± 0.0360	388 ± 2	58,235.88324074
HD 255705	G5V	196.9	−0.04	6.8001 ± 0.0570	480 ± 4	58,235.88562500
HD 254085	G0V	197.5	−1.96	6.5338 ± 0.0967	499 ± 7	58,235.89270833
HD 256380	G8V	198.0	−0.02	2.3058 ± 0.0398	1415 ± 24	58,235.89505787
TYC 739-1501-1	G2V	198.2	−1.60	58,235.90204861
HD 256736	G2V	198.2	+0.22	6.1808 ± 0.0802	528 ± 7	58,235.90435185
TYC 739-1210-1 ^a	G5V	198.5	−1.16	9.9809 ± 0.0424	327 ± 1	58,235.91119213
2019 April 26 22:00–24:00 UT						
TYC 148-515-1	G5V	212.4	−0.98	5.4947 ± 0.0416	594 ± 4	58,599.92175926
CoRoT 102810550	G2V	211.5	−0.69	1.1333 ± 0.0247	2878 ± 63	58,599.92392361
CoRoT 102830606	G2V	211.4	−0.49	2.2193 ± 0.0357	1470 ± 24	58,599.93030093
TYC 149-362-1	G5V	212.8	+0.69	1.1606 ± 0.0581	2810 ± 141	58,599.93254630
TYC 149-532-1	G2V	213.1	+0.69	7.2260 ± 0.0380	451 ± 2	58,599.93917824
CoRoT 102827664	G4V	211.4	−0.51	2.3394 ± 0.0274	1394 ± 16	58,599.94144676
CoRoT 102936925	G4V	213.6	−0.93	1.0454 ± 0.0223	3120 ± 67	58,599.94826389
CoRoT 110695685	G4V	215.9	−0.83	1.5743 ± 0.0541	2072 ± 71	58,599.95049769
CoRoT 110864307	G2V	216.1	−0.98	1.5985 ± 0.0442	2040 ± 56	58,599.95706019
CoRoT 102951397	G2V	213.6	−0.87	1.1909 ± 0.0249	2739 ± 57	58,599.95931713
CoRoT 102963038	G3V	213.7	−0.85	0.3134 ± 0.0256	10407 ± 850	58,599.96596065
HD 50388	G8V	215.2	−0.75	7.3465 ± 0.0598	444 ± 4	58,599.96820602
TYC 4805-3328-1	G5V	215.4	−0.19	2.5383 ± 0.0455	1285 ± 23	58,599.97480324
CoRoT 110777727	G1V	216.2	−0.90	1.5407 ± 0.0457	2117 ± 63	58,599.97699074
CoRoT 110776963	G4V	215.9	−0.78	2.5207 ± 0.0436	1294 ± 22	58,599.98373843
TYC 4814-248-1	G2V	215.1	+1.32	2.9668 ± 0.0426	1099 ± 16	58,599.98597222

Notes. Successive pairs are separated by a blank line. Spectral types, galactic coordinates, and parallax measurements were obtained from the SIMBAD database (Wenger et al. 2000). Distances in light years were calculated from the parallax measurements. The Modified Julian Date (MJD) refers to the beginning of the first scan.

^a The source paired with TYC 739-1210-1 was observed only once and not analyzed.

channels, we stored the consecutive power spectra as rows in time–frequency arrays (aka time–frequency diagrams, spectrograms, spectral waterfalls, waterfall plots, or dynamic spectra) and normalized the power to zero mean and unit standard deviation of the noise power. The normalized power values reflect the S/N at each time and frequency bin.

2.4. Doppler Dechirping

Due to the orbital and rotational motions of both the emitter and the receiver, we expect extraterrestrial technosignatures to drift in frequency space (e.g., Siemion et al. 2013; Margot et al. 2018; Pinchuk et al. 2019). To integrate the signal power over the scan duration while compensating for Doppler drifts in signal frequency, we used incoherent sums of power spectra, where each individual spectrum was shifted in frequency space by a judicious amount prior to summation. This technique is known as *incoherent dechirping*. Coherent dechirping algorithms exist (Korpela 2012) but are computationally expensive and seldom used.

Because the Doppler drift rates due to the emitters are unknown, we examined 1023 linearly spaced drift rates in increments of

$\Delta\dot{f} = 0.0173 \text{ Hz s}^{-1}$ over the range $\pm 8.86 \text{ Hz s}^{-1}$. To accomplish this task, we used a computationally advantageous tree algorithm (Taylor 1974; Siemion et al. 2013), which operates on the dynamic spectra and yields time integrations of the consecutive power spectra after correcting approximately for each trial Doppler drift rate. The algorithm requires input spectra with a number of rows equal to a power of two, and we zero-padded the dynamic spectra with approximately 65 rows to obtain 512 rows. The output of this algorithm, which was run once for negative drift rates and once for positive drift rates, is stored in 1023×2^{20} drift-rate-by-frequency arrays that are ideal for identifying *candidate signals*, i.e., radio signals that exceed a certain detection threshold (Section 3.1). We quantify the sensitivity penalty associated with the use of the tree algorithm in Section 6.1.

3. Data Analysis

3.1. Candidate Signal Detection

We performed an iterative search for candidate signals on the drift-rate-by-frequency arrays obtained with the incoherent dechirping algorithm. Specifically, we identified the signal with

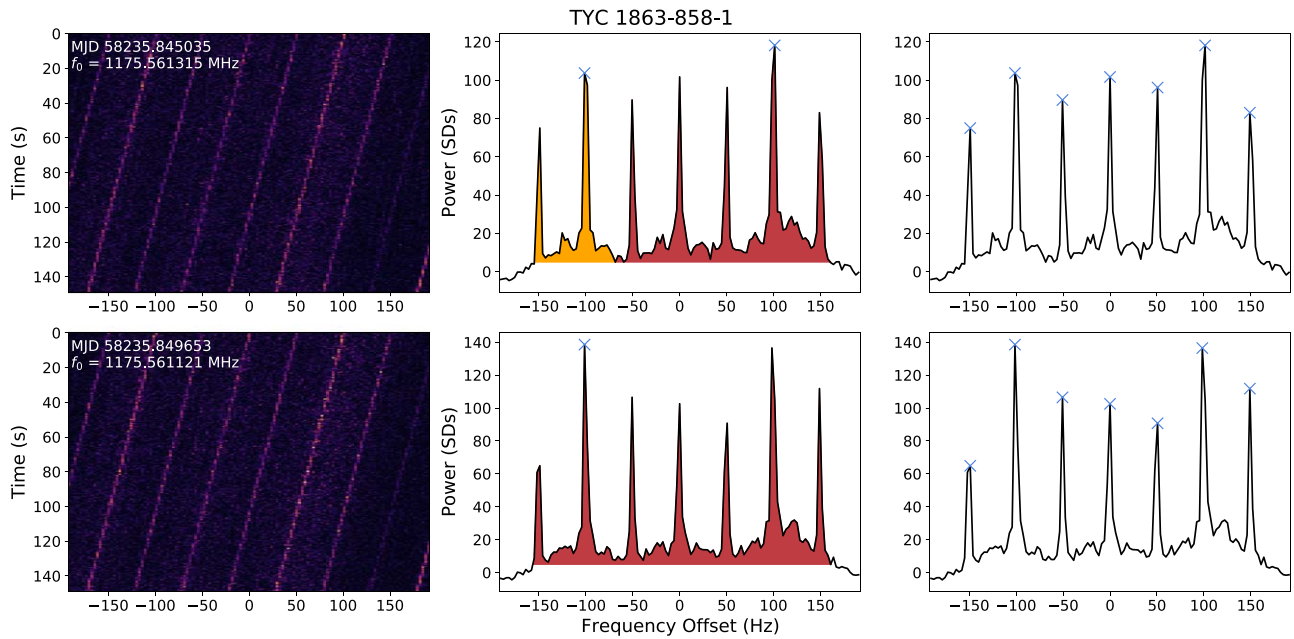


Figure 1. Comparison of signal detection procedures illustrated on an ~ 400 Hz wide region for scans 1 (top) and 2 (bottom) of TYC 1863-858-1. (Left) Dynamic spectra, where pixel intensity represents signal power. (Middle) Integrated power spectra, with blue crosses marking the signals that are detected with the procedure described by Pinchuk et al. (2019). In the first scan, the strongest signal (+100 Hz) is detected, and the corresponding 5σ bandwidth is shown in red. The second-strongest signal (−100 Hz) is then detected, and the corresponding 5σ bandwidth is shown in orange. In the second scan, only the strongest signal, which is now at −100 Hz, is detected. (Right) Integrated power spectra, with blue crosses marking the signals that are detected with the procedure described in this work.

the highest integrated S/N and stored its characteristics in a structured query language (SQL) database, then identified and recorded the signal with the second-highest S/N, and so on. Redundant detections can occur when signals in the vicinity of a candidate signal have large integrated power along similar drift rates. Different data processing pipelines tackle these redundant detections in different ways. Siemion et al. (2013), Enriquez et al. (2017), Margot et al. (2018), and Price et al. (2020) discarded all detections within approximately kilohertz-wide regions of frequency space around every candidate signal detection. This method leaves large portions of the observed frequency space unexamined and biases the results toward high-S/N signals because signals with lower S/N in their vicinity are discarded. Importantly, this method complicates attempts to place upper limits on the abundance of technosignature sources because the pipeline eliminates the very signals it purports to detect (Sections 6.4 and 6.5).

Pinchuk et al. (2019) introduced a novel procedure to alleviate these shortcomings. In order to avoid redundant detections, they imposed the restriction that two signals cannot cross in the time–frequency domain of the scan and discarded detections only in a small frequency region around every candidate signal detection. The extent of this region was set equal to the bandwidth of the candidate signal measured at the 5σ power level, where σ is one standard deviation of the noise. Unfortunately, this bandwidth calculation can cause complications in some situations. For example, small noise fluctuations may result in an unequal number of candidate signals detected in two scans of a source. In the ~ 400 Hz wide region of the spectrum shown in Figure 1, two signals (± 100 Hz) are detected in the first scan, but only one signal (−100 Hz) is detected in the second scan. This incompleteness is detrimental to our direction-of-origin filters (Section 3.2), which rely on accurate signal detection across all scans. Moreover, discarding the region corresponding to a bandwidth measured at 5σ

prevents the detection of at least five other signals per scan in this example (Figure 1).

In this work, we improve on the procedure presented by Pinchuk et al. (2019) in two important ways. First, we identify candidate signal detections on the basis of the topography-inspired concept of *prominence*. The prominence of a signal is defined as the vertical distance between the peak and its lowest contour line, as implemented in the numerical computing package SciPy (Virtanen et al. 2020). Because our integrated spectra are one-dimensional, we take the larger of a peak’s two “bases” as a replacement for the lowest contour line. The high-frequency (low-frequency) base is defined as the minimum power in the frequency region starting on the high- (low-) frequency side of the peak and ending +500 Hz (−500 Hz) away or at the frequency location of the nearest peak with higher (lower) frequency and larger power, whichever results in the smallest frequency interval. While the ± 500 Hz limits are not essential to compute prominences, they do speed up the calculations. Second, we remove the bandwidth dependence of Pinchuk et al.’s (2019) algorithm. Instead, we apply a local maximum filter to the drift-rate-by-frequency arrays in order to remove any points that are not a maximum in their local 3×3 neighborhood. We find that this filter in conjunction with the prominence-based candidate signal detection identifies the signals of interest without introducing redundant detections.

Signals are considered candidate detections if their prominence meets two criteria: (1) it exceeds 10σ , where σ is one standard deviation of the noise in the integrated spectrum, and (2) it exceeds a fraction f of their integrated power. For this analysis, we settled on $f = 75\%$. The second requirement is necessary because power fluctuations superimposed on strong broadband signals that approach or exceed the 10σ detection threshold can yield prominences that exceed 10σ . With this second requirement, a signal with a prominence of 10σ above a 3.0σ baseline would be marked as a detection, but the same

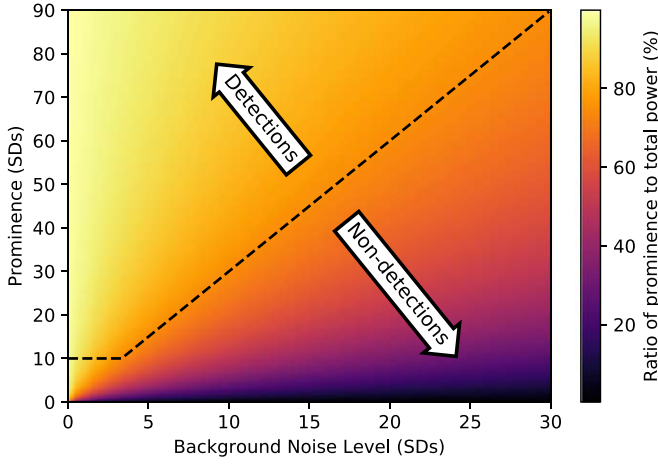


Figure 2. Illustration of detection criteria. Signals above the dashed black line are marked as detections by our pipeline.

signal above a 3.5σ baseline would not. Figure 2 describes the detection space.

As a result of these candidate signal detection improvements, we now detect 1.23–1.75 and ~ 12 times as many signals as we did with the data processing pipelines of Pinchuk et al. (2019) and Margot et al. (2018), respectively (Figure 3). We compare this signal detection performance to that of other searches in Section 6.4.

Once a signal with frequency f_0 and drift rate \dot{f}_0 is detected with the criteria described in this section, we follow the procedure outlined by Pinchuk et al. (2019). Specifically, we eliminate any other candidate signal with frequency f and drift rate \dot{f} if the following inequalities hold true at the start of the scan:

$$\begin{aligned} f_0 < f < f_0 + (\dot{f} - \dot{f}_0)\tau & \quad \text{for } \dot{f} > \dot{f}_0 \\ f_0 > f > f_0 + (\dot{f} - \dot{f}_0)\tau & \quad \text{for } \dot{f} < \dot{f}_0, \end{aligned} \quad (1)$$

where τ is the scan duration. Our candidate detection procedure was applied iteratively until all candidate signals with prominence $\geq 10\sigma$ were identified. Occasionally, signals with prominences $\geq 10\sigma$ but $S/N < 10$ get recorded in the database. This condition tends to occur primarily in regions with dense RFI where the baseline subtraction is imperfect. For this reason, we flagged all signals with $S/N < 10$ and did not consider them to be valid candidates.

3.2. Doppler and Direction-of-origin Filters

After identifying all candidate signals, we applied a Doppler filter and improved variants of our direction-of-origin filters (Margot et al. 2018; Pinchuk et al. 2019) to detect and discard anthropogenic signals in the data.

We began by applying a Doppler filter, which is designed to remove all signals with zero Doppler drift rate, defined here as signals that drift less than one frequency resolution cell ($\Delta f = 2.98$ Hz) over the duration of a scan ($\tau = 150$ s). The signals removed by this filter are of no interest to us because the corresponding emitters exhibit no line-of-sight acceleration with respect to the receiver, suggesting that they are terrestrial in nature.

Next, we applied two direction-of-origin filters, which are designed to remove any signal that is either not persistent (i.e.,

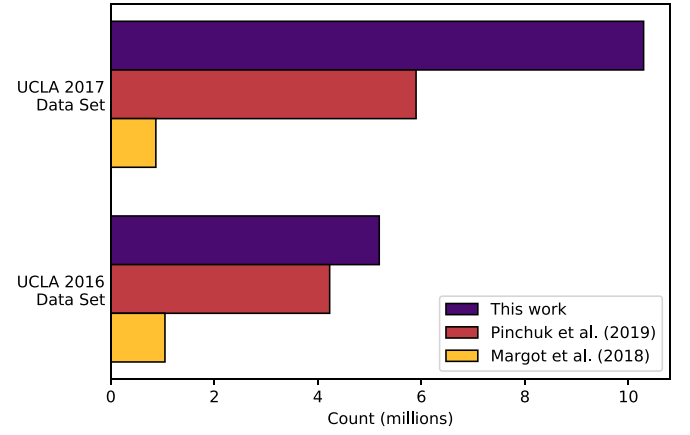


Figure 3. Detection counts obtained with the algorithms presented by Margot et al. (2018), Pinchuk et al. (2019), and this work. Our current pipeline detects 1.23–1.75 as many signals as Pinchuk et al.’s (2019) pipeline and ~ 12 times as many signals as Margot et al.’s (2018) pipeline.

not detected in both scans of its source) or detected in multiple directions on the sky (i.e., also detected in scans corresponding to other sources). Because the largest possible side-lobe gain is approximately -30 dB compared to the main-lobe gain, signals detected in multiple directions on the sky are almost certainly detected through antenna side lobes. The second filter is highly effective at removing such signals.

As explained by Pinchuk et al. (2019), the direction-of-origin filters compare signals from different scans and flag them according to the observed relationships. For example, if a signal from a scan of source A is paired with a signal from a scan of source B, then both signals are removed because they are detected in multiple directions on the sky. In our previous implementation of these filters, two signals were considered a pair if their drift rates were similar and their frequencies at the beginning of each scan were within a predetermined tolerance of a straight line with a slope corresponding to the drift rate. With this definition, it was possible for multiple signals in one scan to be paired with a single signal from a different scan, which is undesirable. For example, a valid technosignature candidate from one of the scans of source A could be labeled as RFI because it was paired with a signal from one of the scans of source B, even if the signal in the scan of source B was already paired with a different (RFI) signal from the scan of source A.

We have redesigned our filter implementation to keep a record of all signals that are paired during filter execution. We use this record to impose the restriction that each signal is allowed to pair with only one other signal in each scan. Additionally, we implemented an improved pairing procedure that is loosely based on the Gale–Shapley algorithm (Gale & Shapley 1962) designed to solve the stable matching problem. Our improved procedure operates as follows. We define the propagated frequency difference $\Delta F(f_i, f_j)$ of two signals from different scans to be

$$\Delta F(f_i, f_j) = |(f_i - f_j) + \overline{\dot{f}_{ij}} \Delta t_{ij}|, \quad (2)$$

where f_i and f_j are the start frequencies of the two signals, $\overline{\dot{f}_{ij}} = (\dot{f}_i + \dot{f}_j)/2$ is the average of the two signal drift rates, and $\Delta t_{ij} = t_j - t_i$ is the time difference between the two scans. Our updated algorithm iterates over all remaining unpaired candidate signals and updates the pairings until $\Delta F(f_i, f_j)$ is

Table 2
Definitions of Operating Regions of Known Anthropogenic Interferers and Associated Signal Counts

Frequency Region (MHz)	Total Detection Count	% of Total Detections	Postfilter Count	Identification
1155.99–1196.91	11,937,074	44.82%	15,034	GPS L5
1192.02–1212.48	135,769	0.51%	276	GLONASS L3
1422.32–1429.99	190,530	0.72%	2945	ARSR products
1525–1559	8,258,612	31.01%	341	Satellite downlinks
1554.96–1595.88	5,016,951	18.84%	19,621	GPS L1
1592.95–1610.48	933,813	3.51%	3569	GLONASS L1

Note. The column labeled “Postfilter Count” lists the number of signals remaining after application of our Doppler and direction-of-origin filters. The time–frequency structure of the RFI labeled as “ARSR products” is similar to that described by Siemion et al. (2013), Margot et al. (2018), and Pinchuk et al. (2019). These products are likely intermodulation products of ARSR.

minimized for all signal pairs. In rare cases, when the minimum value of $\Delta F(f_i, f_j)$ is not unique, multiple pairings are allowed, but no inference about the anthropogenic nature of the signals is made on the basis of these pairings alone.

To ensure that paired signals likely originated from the same emitter, we impose two additional requirements on all signal pairs. First, we require that

$$f_{ij,-} \leq f_j \leq f_{ij,+} \quad \text{or} \quad f_{ji,-} \leq f_i \leq f_{ji,+}, \quad (3)$$

where $f_{ij,\pm} = f_i + (\dot{f}_i \pm \Delta\dot{f})\Delta t_{ij} \pm \Delta f$ represent the propagated frequency bounds and $\Delta\dot{f}$ and Δf are the drift rate and frequency resolution, given by 0.0173 Hz s^{-1} and 2.98 Hz , respectively. This condition places an upper limit on $\Delta F(f_i, f_j)$, and we reject signal pairs whose propagated frequency differences exceed this bound. Second, we require that

$$|\dot{f}_i - \dot{f}_j| \leq 2\Delta\dot{f}, \quad (4)$$

and we reject signal pairs that do not satisfy this criterion. In tandem, these requirements reduce the possibility of pairing two unrelated signals.

To determine if a signal is persistent (first filter), we apply the pairing procedure to candidate signals detected in both scans of a source. Those signals left without a partner are deemed to originate from transient sources and are labeled as RFI. To determine whether a signal is detected in multiple directions of the sky (second filter), we apply the pairing procedure to signals from scans of different sources. In this case, all resulting pairs are attributed to RFI and discarded. Candidate signals remaining after the application of these procedures are marked for further inspection.

3.3. Frequency Filters

A majority of the candidate signals detected in our search are found in the operating bands of known interferers. Table 2 describes the frequency ranges and signal counts associated with the most prominent anthropogenic RFI detected in our data. Candidate signals detected within these frequency regions (except the Air Route Surveillance Radars (ARSR) products region) were removed from consideration because of their likely anthropogenic nature. The combined 2017 and 2018 signal detection counts in the excluded RFI regions ($156,327 \text{ MHz}^{-1}$) are considerably higher than outside of these regions ($20,654 \text{ MHz}^{-1}$) or in the 1400–1427 MHz radio astronomy protected band (6949 MHz^{-1}). The protected band is regrettably polluted, possibly as a result of intermodulation products generated at the telescope (Margot et al. 2018).

The useful bandwidth of our observations $\Delta f_{\text{tot}} = 309.3 \text{ MHz}$ is computed by taking the operational bandwidth of the GBT L -band receiver (580 MHz) and subtracting the bandwidth of the GBT notch filter (141.2 MHz) and the total bandwidth discarded due to known interferers (Table 2; 129.5 MHz).

4. Preliminary Signal Injection and Recovery Analysis

A signal injection and recovery analysis consists of injecting artificial signals into the raw data and quantifying the fraction of signals that are properly recovered by the pipeline (e.g., Christiansen et al. 2013). Although a rigorous injection analysis is beyond the scope of this paper, we performed a preliminary examination by injecting narrowband (2.98 Hz) signals into the dynamic spectra before applying the incoherent dechirping (Section 2.4), candidate detection (Section 3.1), and Doppler and direction-of-origin filtering (Section 3.2) procedures.

4.1. Generation and Injection of Artificial Signals

We selected 10,000 starting frequencies from a uniform distribution over the operating region of the GBT L -band receiver (1.15–1.73 GHz), excluding the frequency region of the GBT notch filter (1.2–1.3412 GHz). For each starting frequency, we also randomly selected a frequency drift rate from the discrete set $\{k \times \Delta\dot{f} : k \in \mathbb{Z}, -510 \leq k \leq 510\}$, with $\Delta\dot{f} = 0.0173 \text{ Hz s}^{-1}$. Each signal was randomly assigned to one of the sources and injected into the first scan of this source. A corresponding partner signal was injected into the second scan of this source. The starting frequency of the partner signal was obtained by linearly extrapolating the frequency of the signal in the first scan, i.e., by adding the product of the artificial drift rate and the known time difference between the two scans. The drift rate of the partner signal was set equal to that of the original signal plus an increment randomly chosen from the set $\{-\Delta\dot{f}, 0, \Delta\dot{f}\}$.

We injected half of the signals at our detection threshold (10σ) to test the limits of our pipeline’s detection capabilities. The remaining signals were injected at an S/N of 20 to test our sensitivity to stronger signals. A total of 20,000 signals were injected into the 2018 April 27 data. A full list of the injected signal properties is available as supplemental online material. Two examples of injected signals are shown in Figure 4 and listed in Table 3.

4.2. Recovery and Classification of Injected Signals

After injecting the signals into the dynamic spectra, we applied our candidate detection procedure (Section 3.1) and stored the output in an SQL database. Signals were considered properly recovered if their properties matched those of the

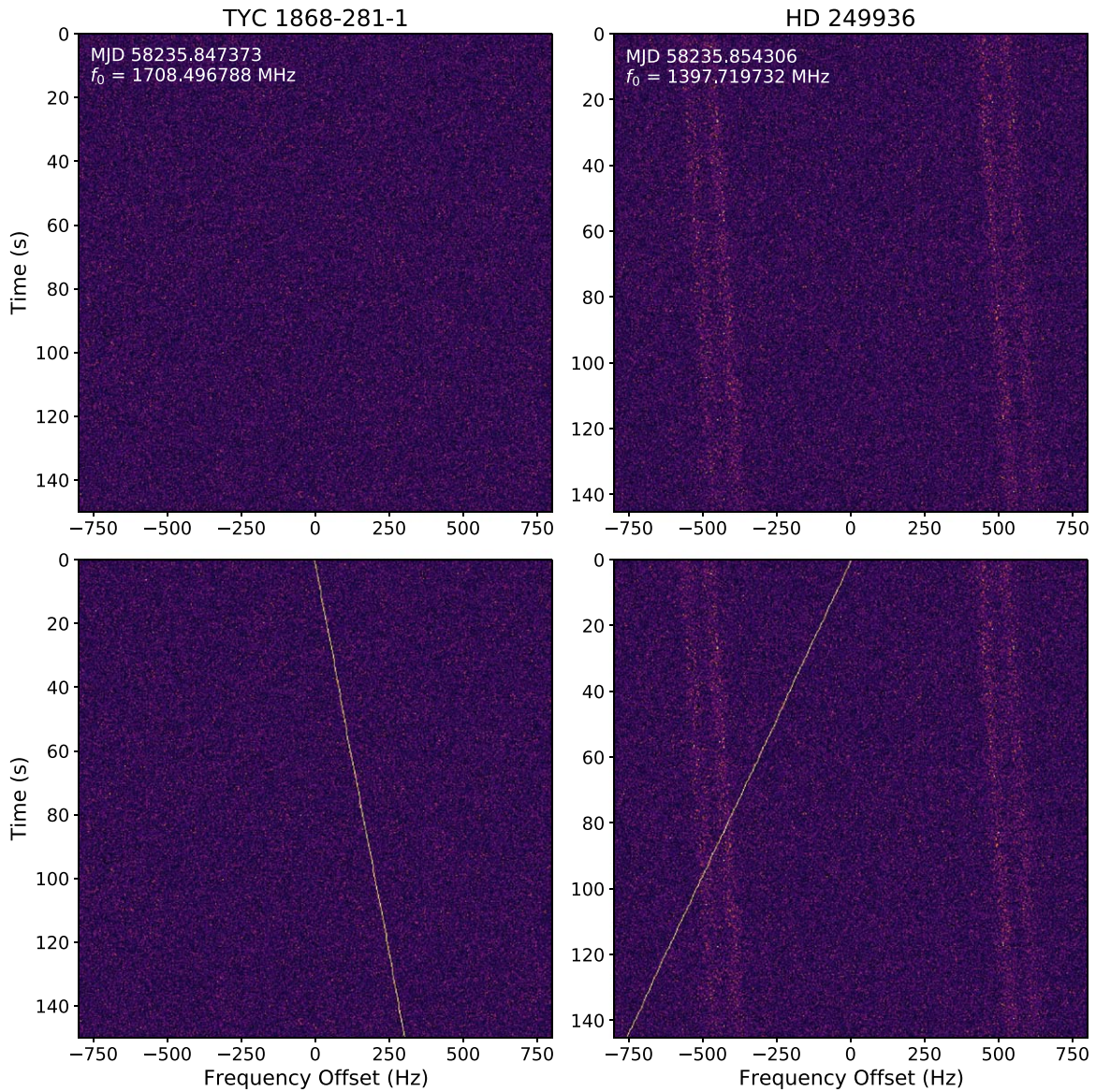


Figure 4. (Top) Time–frequency diagram before signal injection. (Bottom) Time–frequency diagram after signal injection. The injected signal S/N was increased twentyfold to facilitate visualization. The bottom left panel shows a signal that was successfully recovered by our data processing pipeline. The injected signal in the bottom right panel crosses a stronger RFI signal and was missed by our detection algorithm.

Table 3
Properties of Artificial Signals Used for the Signal Injection and Recovery Analysis

Name	Scan	Freq. (Hz)	df/dt (Hz s ⁻¹)	S/N	Detected	DOO_CORRECT
TYC 1868-281-1	1	1708496788.1	2.029626	10	Y	Y
HD 249936	1	1397719731.9	-5.221517	20	N	N/A

Note. Columns show source name, scan number, frequency of injection at the start of the scan, frequency drift rate, S/N, a Boolean indicating whether the signal was recovered by the pipeline, and a Boolean indicating whether the direction-of-origin filter made the correct assignment.

(This table is available in its entirety in machine-readable form.)

injected signals within ± 2 Hz in frequency, $\pm \Delta f$ in drift rate, and ± 0.1 in S/N. We found that our procedure recovered 18,528 (92.64%) of the injected signals. Outside of the regions with dense RFI described in Table 2, our pipeline performs better, with a recovery rate of 97.66%. We observe no significant difference in the recovery rate as a function of drift rate or scan number (Figure 5), but we do notice an $\sim 3\%$ increase in the recovery rate for signals with larger S/N.

We found that most of the signals missed by our pipeline were injected in regions of known RFI (Figure 5). This pattern is a consequence of two known limitations of our candidate detection procedure. First, our algorithms only detect the signal with the highest S/N when two signals intersect in time–frequency space (Pinchuk et al. 2019). Second, signals with a low prominence superimposed on an elevated noise baseline are discarded (Section 3.1). High-density RFI regions such as

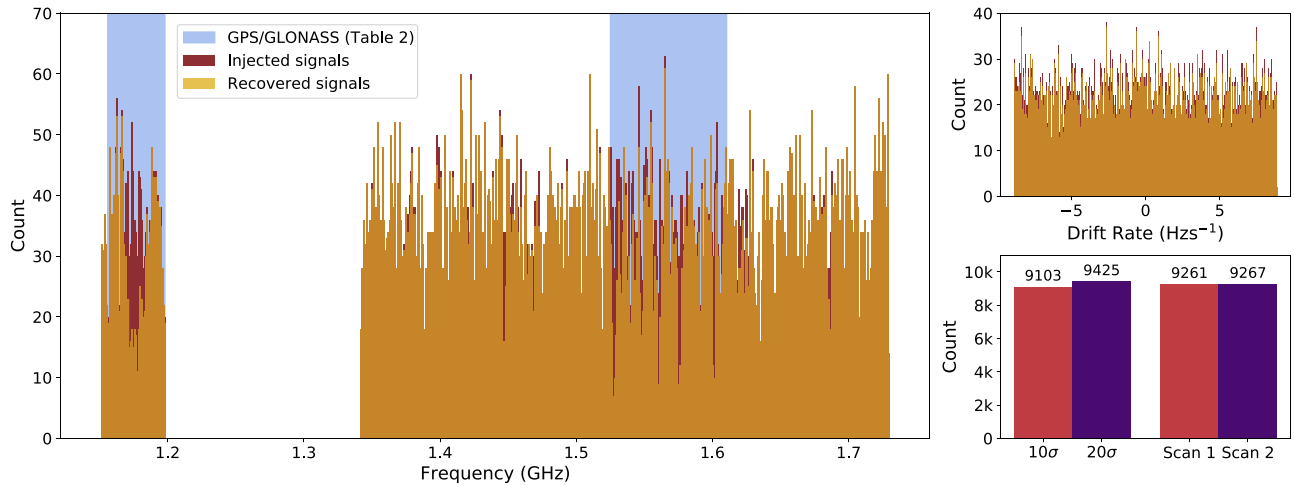


Figure 5. (Left) Frequency distribution of injected and recovered signals. The number of signals recovered within known RFI regions (such as GPS or GLONASS) is substantially lower than in other regions. (Top right) Drift rate distribution of injected and recovered signals. (Bottom right) Signal recovery counts as a function of S/N and scan number. We observe no significant difference in the recovery rate as a function of drift rate or scan number, but we do notice an $\sim 3\%$ increase in the recovery rate of signals with larger S/N.

the ones listed in Table 2 are conducive to both of these conditions, thereby reducing the recovery rate. A cursory analysis suggests that $\sim 70\%$ – 80% of the nondetections are due to the intersecting condition.

In order to quantify the performance of our Doppler and direction-of-origin filters (Section 3.2), we applied our filters to the entire set of detected signals, including the detections resulting from injected signals. To distinguish the performance of these filters from that of our detection algorithm, we removed 414 of the injected signals that were detected in only one scan of a source. Furthermore, we removed 21 signals that were injected with a Doppler drift rate of zero (i.e., stationary with respect to the observer). Of the remaining 18,093 injected signals, 18,044 (99.73%) were flagged as promising technosignature candidates by our Doppler and direction-of-origin filters.

4.3. Performance of Data Processing Pipeline

The preliminary injection and recovery analysis described in this section identified some important limitations of our radio technosignature detection pipeline. Our detection algorithm, which is an improvement over those of Margot et al. (2018) and Pinchuk et al. (2019; Figure 3) and outperforms those of Enriquez et al. (2017) and Price et al. (2020; Section 6.4), experiences degraded performance in regions with dense RFI. In these regions, it is more likely for a technosignature candidate to intersect a strong RFI signal (Figure 4), thereby escaping detection by our pipeline. This limitation could be overcome by using the recorded drift rates and starting frequencies of two signals within a scan to determine whether the signals are predicted to intersect each other in the other scan of the source. If an intersection condition were detected, the known signal could be blanked or replaced with noise, and a new detection procedure could be run to identify previously undetected signals. In the presence of strong RFI, a fraction of the injected signals escape detection because their prominence is below our detection threshold (i.e., prominence $< f \times$ integrated power, with $f = 75\%$). In some situations, a valid technosignature could also be removed if it were detected in a frequency region corresponding to a broadband signal. It may be possible to overcome this limitation in the future by

including a comparison of the properties of the narrowband signal (e.g., drift rate, modulation, etc.) to those of the underlying broadband signal.

Our improved Doppler and direction-of-origin filters performed exceptionally well, only mislabeling 49 of the 18,093 injected signals. The signals that were incorrectly flagged were paired with an RFI signal of similar drift rate in a scan of a different source. This issue can be mitigated by expanding the signal-matching criteria to include signal properties other than starting frequency and drift rate, such as bandwidth or gain ratio.

The results presented in this section provide important insights into the detection capabilities of our current data processing pipeline. In particular, they demonstrate that our pipeline still misses some of the narrowband signals that it is designed to detect. These results are also useful to identify specific areas in need of improvement.

4.4. Limitations of Current Signal Injection and Recovery Analysis

The analysis presented in this section is preliminary because it injects signals into the dynamic spectra and not the raw data. Therefore, the current implementation does not consider certain data processing steps, such as correcting for the bandpass channel response (Section 2.3), calculating the noise statistics and normalizing the power spectra to zero mean and unit variance, or applying the incoherent dechirping procedure (Section 2.4).

In future work, we will implement the ability to inject signals into the raw data. This improved implementation will allow us to quantify the detection performance of the entire pipeline. We anticipate that it will also be helpful in revealing additional areas for improvement.

5. Results

We applied the methods described in Section 3 to the data described in Section 2.1. We detected a total of 26,631,913 candidate signals over both 2018 and 2019 observation epochs. We used the total integration time of 2.75 hr and processed bandwidth of 438.8 MHz to compute a signal detection count

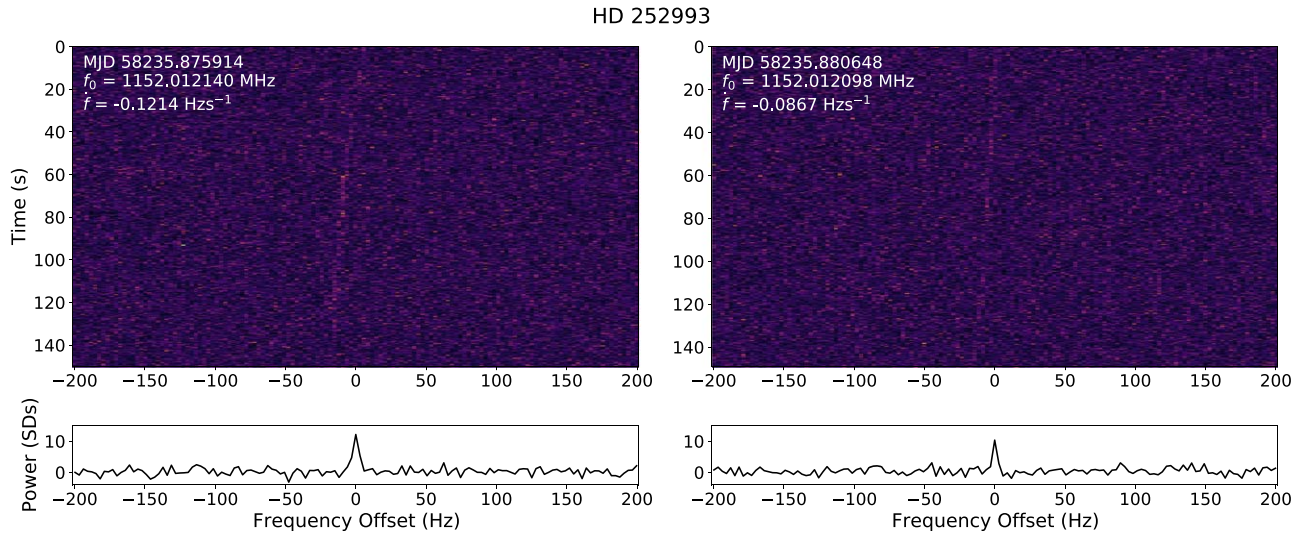


Figure 6. Dynamic spectra (top) and integrated power spectra (bottom) of a final candidate signal that appears in scans 1 (left) and 2 (right) of HD 252993. Although this signal exhibits many of the desirable properties of a technosignature (e.g., narrowband, nonzero Doppler drift rate, persistence), it was ultimately rejected because it was visually confirmed to appear in multiple directions on the sky.

per unit bandwidth per unit integration time. In BL parlance, our detections are referred to as “hits” (Enriquez et al. 2017; Price et al. 2020), and the hit rate density of this search is 2.2×10^{-2} hits hr⁻¹ Hz⁻¹. In comparison, the *L*-band component of Price et al.’s (2020) search with the same telescope and S/N threshold resulted in 37.14 million hits in 506.5 hr over a useful bandwidth of 660 MHz, or a hit rate density of 1.1×10^{-4} hits hr⁻¹ Hz⁻¹, 200 times smaller than ours. We discuss possible reasons for this large differential in Section 6.4.

A complete table of the signal properties of the detected candidates is available in Dryad doi:[10.5068/D1937J](https://doi.org/10.5068/D1937J). Our Doppler and direction-of-origin filters flagged 26,588,893 (99.84%) signals as anthropogenic RFI. A majority of the remaining 43,020 signals were detected within the operating regions of known interferers (Table 2). Candidate signals remaining within these frequency regions were attributed to RFI and removed from consideration.

The remaining 4539 signals were deemed the most promising technosignature candidates. Visual inspection of all of these candidates revealed that they are attributable to RFI. Figure 6 shows an example of a promising signal that was ultimately attributed to RFI.

The vast majority of the most promising candidates were eliminated because they were detected in multiple directions on the sky. These signals escaped automatic RFI classification by our filters for one or more of the following reasons, which are generally similar to the “categories” described by Pinchuk et al. (2019, Section 4).

1. The S/N values of corresponding signals in scans of other sources were below the detection threshold of 10. This difficulty could perhaps be circumvented in the future by conducting an additional search for lower-S/N signals at nearby frequencies.
2. The drift rate of the signal differed from those of corresponding signals in scans of other sources by more than our allowed tolerance ($\pm \Delta \dot{f} = \pm 0.0173$ Hz s⁻¹).
3. The signal was not detected in scans of other sources because it intersected another signal of a higher S/N.

4. The signal bandwidth exceeded 10 Hz, making it difficult to accurately determine a drift rate and therefore link the signal with corresponding signals in scans of other sources.

All of these difficulties could likely be overcome by a direction-of-origin filter that examines the time–frequency data directly instead of relying on estimated signal properties, such as starting frequency and drift rate. We are in the process of implementing machine-learning tools for this purpose.

Because automatic classification and visual inspection attributed all of our candidate signals to RFI, we did not detect a technosignature in this sample. We are preserving the raw data in order to enable reprocessing of the data with improved algorithms in the future, including searches for additional types of technosignatures.

6. Discussion

6.1. Dechirping Efficiency

Over sufficiently short (~ 5 minutes) scan durations, monochromatic signals emitted on extraterrestrial platforms are well approximated by linear chirp waveforms ($f(t) = f(t_0) + \dot{f}(t - t_0)$). Most radio technosignature detection algorithms rely on incoherent dechirping, i.e., incoherent sums of power spectra, to integrate the signal power over the scan duration (Section 2.4). In the context of incoherent sums, the magnitude of the maximum drift rate that can be considered without loss in sensitivity is given by

$$\dot{f}_{\max} = \frac{\Delta f}{\Delta T}, \quad (5)$$

where Δf is the adopted spectral resolution and ΔT is the accumulation time corresponding to one row in the dynamic spectra. If the drift rate of a signal exceeds this maximum drift rate ($\dot{f} > \dot{f}_{\max}$), the signal frequency drift exceeds Δf during ΔT , and power is smeared over multiple frequency channels, resulting in reduced sensitivity.

In this work ($\Delta f = 2.98$ Hz; $\Delta T = 1/\Delta f = 0.34$ s), the maximum sensitivity can be obtained up to frequency drift rates

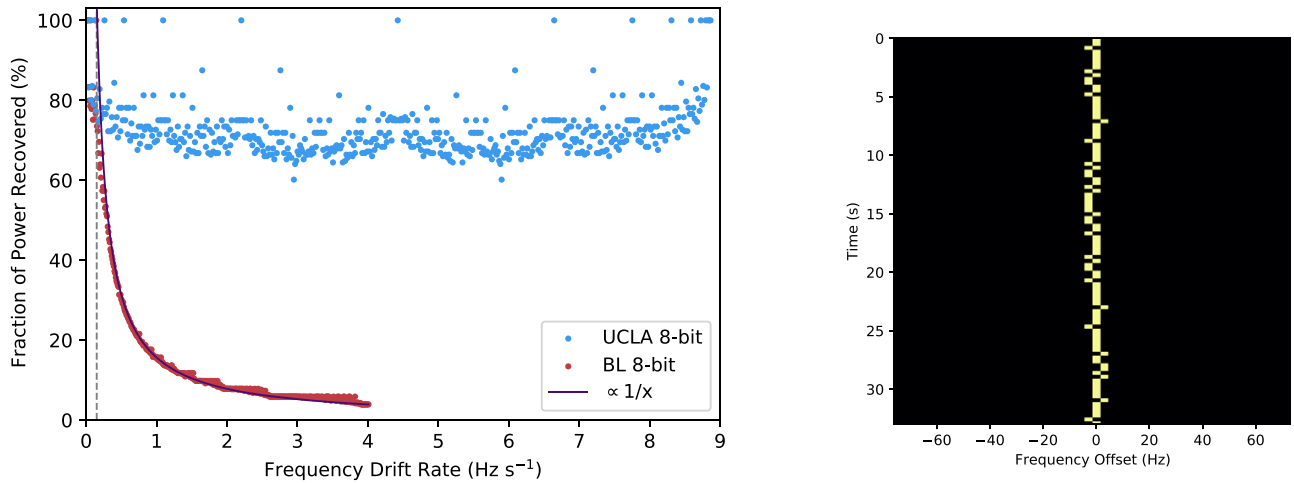


Figure 7. (Left) Dechirping efficiencies of the UCLA (blue) and BL (red) data processing pipelines as a function of Doppler frequency drift rate at nominal frequency resolutions of ~ 3 Hz. Our choices of data taking and processing parameters result in a fairly uniform efficiency ($72.4\% \pm 6.8\%$) across the full range of drift rates considered, with values below 100% due to imperfections of the tree algorithm (see text). The BL choices result in a considerably reduced detection efficiency beyond $\dot{f}_{\max, \text{BL}} = 0.15 \text{ Hz s}^{-1}$ (dashed vertical line), with values as low as 4% due to smearing of the signal power across multiple frequency bins. The performance at frequencies beyond \dot{f}_{\max} is well approximated by a $1/x$ function (purple line), consistent with the inverse bandwidth dependence of the amplitude of a linear chirp power spectrum. (Right) Dynamic spectrum of a linear chirp waveform dechirped imperfectly by the tree algorithm. In this worst-case scenario for $\dot{f} \leq \dot{f}_{\max}$, only 60% of the spectra are shifted by the correct amounts, and only 60% of the power is recovered in the appropriate frequency channel. Only the first 100 rows (~ 30 s) are shown.

of $\dot{f}_{\max, \text{UCLA}} = 8.88 \text{ Hz s}^{-1}$. The BL investigators Enriquez et al. (2017) and Price et al. (2020) used $\Delta f = 2.79 \text{ Hz}$ and $\Delta T = 51/\Delta f = 18.25 \text{ s}$, which yields $\dot{f}_{\max, \text{BL}} = 0.15 \text{ Hz s}^{-1}$. However, these authors conducted searches for signals with drift rates larger than 0.15 Hz s^{-1} , resulting in reduced sensitivity for $>90\%$ of the drift rates that they considered. For instance, at the largest drift rate considered by Price et al. (2020), the frequency drifts by $4 \text{ Hz s}^{-1} \times 18.25 \text{ s} = 73 \text{ Hz}$ (26 channels) during ΔT , and only $\sim 4\%$ of the signal power is recovered in each frequency channel. We express this loss of signal power with a detection efficiency in the range 0%–100% and refer to it as a *dechirping efficiency*.

To confirm the performance of the data processing pipelines, we conducted numerical experiments¹¹ with both our algorithms and BL’s turboSETI package (Enriquez et al. 2017). For the purpose of these simulations, we created noise-free, constant-power dynamic spectra of linear chirp waveforms with the frequency and time resolutions appropriate for the UCLA and BL searches. By considering only integral pixel locations, we simulated frequency drift rates that are exact multiples of the elemental drift rates considered by our respective tree algorithms (0.0173 Hz s^{-1} for the UCLA searches, 0.0096 Hz s^{-1} for the BL searches). We ran the respective tree algorithms on the simulated spectra and recorded the power recovered at each drift rate as a function of total signal power (Figure 7, left). The experiments show that, at nominal frequency resolutions of ~ 3 Hz, dechirping efficiencies of 100% are possible in our and other searches with $\dot{f} \leq \dot{f}_{\max}$, whereas dechirping efficiencies rapidly degrade to values as low as 4% in the BL searches with $\dot{f} > \dot{f}_{\max, \text{BL}}$.

In this experiment, a perfect algorithm would recover 100% of the signal power, as long as $\dot{f} \leq \dot{f}_{\max}$. The tree algorithm (Section 2.4) is not perfect in that it reuses precomputed sums to achieve $N \log N$ computational cost. As a result, the tree algorithm shifts every spectrum by an amount that is not

Table 4

Dechirping Efficiencies Resulting from Incoherent Dechirping of Power Spectra with a Computationally Advantageous but Approximate Tree Algorithm (Section 2.4)

Rows	Min. (%)	Max. (%)	Mean (%)	Median (%)	STD (%)
4	100.00	100.00	100.00	100.00	0.00
8	75.00	100.00	93.75	100.00	11.57
16	75.00	100.00	90.62	93.75	10.70
32	68.75	100.00	85.16	81.25	11.20
64	68.75	100.00	81.64	78.12	9.83
128	64.06	100.00	77.93	75.00	8.88
256	64.06	100.00	75.17	73.44	7.68
512	60.16	100.00	72.42	71.09	6.84
1024	60.16	100.00	70.08	69.14	6.10
2048	56.84	100.00	67.92	66.60	5.53
4096	56.84	100.00	66.01	64.94	5.06

always optimal. In other words, it is unable to perfectly dechirp most linear chirp waveforms. In our simulations of the UCLA pipeline, we do observe 100% of the power recovered for several drift rates (Figure 7, left). On average, the pipeline recovers $72.4\% \pm 6.8\%$ of the signal power. In the worst-case scenario, the fraction of power recovered is 60%. The tree algorithm’s dechirped waveform of this worst-case scenario reveals that 60% of the frequency bins are shifted to the correct locations and 40% are shifted to incorrect locations (Figure 7, right). We quantified the dechirping efficiencies associated with the use of the tree algorithm for a variety of array dimensions (Table 4).

We computed a rough estimate of the mean dechirping efficiency in the search of Price et al. (2020) for the nominal frequency resolution of ~ 3 Hz and a uniform distribution of candidate signals as a function of drift rate. We assumed a generous 100% efficiency between 0 and 0.15 Hz s^{-1} and the $1/x$ trend observed in Figure 7 between 0.15 and 4 Hz s^{-1} . We found a mean efficiency of 16.5%. A weighted mean of the efficiency based on the exact distribution of signals as a

¹¹ Our software is available at https://github.com/UCLA-SETI-Group/dechirping_efficiency.

function of drift rate would provide a more accurate and likely larger value.

We describe two alternate, partial solutions to the loss of sensitivity sustained during incoherent dechirping. The first is to reduce the frequency resolution of the dynamic spectra, thereby increasing the range of drift rates that can be explored without spreading power across multiple channels (e.g., Siemion et al. 2013). However, this solution still results in a loss of sensitivity. For narrowband signals, each doubling of the frequency resolution results in a $\sqrt{2}$ decrease in sensitivity. To reach the maximum drift rates of $\pm 4 \text{ Hz s}^{-1}$ considered by Price et al. (2020), one would have to apply four to five doublings, resulting in frequency resolutions of 45–90 Hz and sensitivity to narrowband signals of 18%–25% of the nominal value. Another, related approach would be to use a drift-rate-dependent boxcar average of the integrated spectra to recover the power that has been spread over multiple channels, e.g., by averaging 26 channels at the maximum drift rates of $\pm 4 \text{ Hz s}^{-1}$ considered by Price et al. (2020). Doing so would degrade the frequency resolution to values up to 73 Hz and the sensitivity to narrowband signals to 20% of the nominal value.

6.2. Extreme Drift Rates

In a recent study¹² of the expected drift rates of a large class of bodies, including exoplanets with highly eccentric orbits and small semimajor axes, Sheikh et al. (2019) recommended searching drift rates as large as $\dot{f}/f_{\text{obs}} = 200 \text{ nHz}$. At the center frequency of our observations (1.5 GHz), this corresponds to a drift rate of 300 Hz s^{-1} . Our data archival policy (Section 6.3) would enable reprocessing of the data with parameters that are more conducive to large drift rates. For example, we could reprocess our data with Fourier transforms of length 2^{17} . This choice would increase our frequency resolution eightfold to 24 Hz and allow us to search for drift rates up to $\sim 570 \text{ Hz s}^{-1}$ without incurring any sensitivity loss due to signal smearing over multiple frequency channels. In contrast, BL archive products include dynamic spectra but do not include most of the raw voltage data (Enriquez et al. 2017; Lebofsky et al. 2019; Price et al. 2020), making it impractical to conduct a search with archival products at drift rates larger than $\sim 1 \text{ Hz s}^{-1}$ with adequate sensitivity (Figure 7).

6.3. Data Requantization and Preservation

Our choice of data recording parameters is largely driven by our dedication to preserve the raw voltage data recorded during our observations. We prefer to archive the raw data as opposed to derived data products such as dynamic power spectra, for four reasons. First, the raw 2-bit data require less storage space than the 32-bit dynamic spectra. Second, the dynamic power spectra can be easily regenerated from the raw data, but the reverse is not true, because phase information is lost in the process of computing power spectra. Third, there are large penalties associated with preserving incoherent averages of individual power spectra. Enriquez et al. (2017) and Price et al. (2020) averaged 51 consecutive spectra to keep the archival volume manageable, which degrades the sensitivity of the search by factors of up to ~ 25 (Section 6.1) and the time resolution by a factor of 51 (Figure 8). As a result, the BL

dynamic spectra would not be useful in confirming or interpreting a signal with 1 Hz modulation, for instance. Fourth, the only way to preserve the ability to conduct novel or improved data analysis with maximum sensitivity and resolution is to preserve the raw data. However, there are penalties associated with storing raw data in 2-bit format as opposed to 8-bit format (e.g., Price et al. 2020).

For this work and previous analyses (Margot et al. 2018; Pinchuk et al. 2019), we selected a data taking mode that yields 2-bit raw voltage data after requantization with an optimal four-level sampler (Kogan 1998). The quantization efficiency, which is the ratio of signal power that is observed with the optimal four-level sampler to the power that would be obtained with no quantization loss, is 0.8825. Price et al. (2020) noted that a consequence of this requantization is that the S/N threshold used in this work (10) would need to be lowered by approximately 12% to detect the same number of candidate signals as 8-bit quantized data. While we agree with this statement, the S/N threshold of radio technosignature searches is somewhat arbitrary, and our choice compares favorably to that of other surveys (Table 5). Should the need ever arise to detect weaker signals, we would simply reanalyze our data with a lower S/N threshold. In addition, the sensitivity enabled by our decision to minimize the accumulation time when computing dynamic spectra (Section 6.1) offsets the losses due to quantization efficiency compared to pipelines with longer accumulation times. Specifically, if we apply the 0.8825 quantization efficiency to the results illustrated in Figure 7, we find that our overall sensitivity surpasses BL’s sensitivity for any drift rate larger than 0.153 Hz s^{-1} and surpasses it by a factor of at least 5 for any drift rate larger than 1.11 Hz s^{-1} .

6.4. Candidate Signal Detection Count

Our results indicate a hit rate density of $2.2 \times 10^{-2} \text{ hits hr}^{-1} \text{ Hz}^{-1}$, whereas Price et al. (2020) obtained a considerably lower value of $1.1 \times 10^{-4} \text{ hits hr}^{-1} \text{ Hz}^{-1}$ with the same telescope and S/N threshold (Section 5). We investigate possible causes for this factor of ~ 200 difference. First, our observing cadence involves two scans of 150 s each per source, whereas Price et al. (2020) used three scans of 300 s each per source. The difference in integration time could perhaps be invoked to explain a factor of up to 3 difference in hit rate density, although a larger number of signals ought to be detectable with BL’s longer scan durations. Second, our processed frequency range extends over 438.8 MHz, whereas Price et al. (2020) used a superset of that range extending over 660 MHz. A nonuniform distribution of dense RFI across the spectrum could perhaps be invoked to explain a factor of up to ~ 2 difference in hit rate density. Third, we examine a range of drift rates that is twice as large as the range used by Price et al. (2020), which may explain a factor of ~ 2 difference in hit rate density if the distribution of hits as a function of drift rate is roughly uniform. These small factors cannot explain the 2 orders of magnitude difference in hit rate density, which must be related to more fundamental effects. We surmise that the two most important factors are the difference in the effective sensitivities of our searches due to different dechirping efficiencies (Section 6.1) and the algorithmic difference in the identification of candidate signals or hits.

As detailed by Pinchuk et al. (2019), the candidate signal detection procedures used in several previous radio technosignature searches (e.g., Siemion et al. 2013; Enriquez et al. 2017;

¹² Sheikh et al. (2019) incorrectly delineated the search parameters of Enriquez et al. (2017) in their Figure 1. The maximum frequency excursion considered in that search is 600 Hz, not 12,000 Hz.

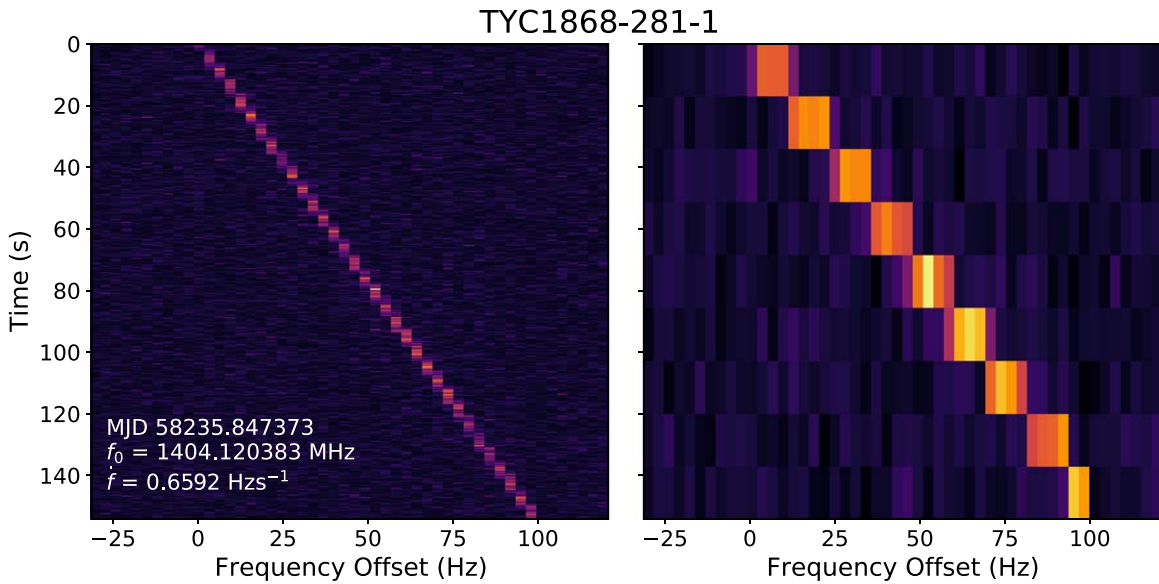


Figure 8. Representative dynamic spectra of a signal shown with the nominal time resolution of $\sim 1/3$ Hz = 0.33 s (left) and the degraded time resolution resulting from time-averaging 51 consecutive spectra (right).

Table 5

S/N Thresholds Used in Recent Searches for Radio Technosignatures

Reference	S/N
Gray & Mooley (2017)	7
Harp et al. (2016)	9/6.5
UCLA SETI searches	10
Price et al. (2020)	10
Enriquez et al. (2017)	25
Siemion et al. (2013)	25

Margot et al. (2018; Price et al. 2020) unnecessarily remove kilohertz-wide regions of frequency space around every signal detection. This practice complicates attempts to place upper limits on the existence of technosignatures, because the algorithms discard many signals that are legitimate technosignature candidates. In addition, this practice leads to slight overestimates of search metrics, such as the DFM (Pinchuk et al. 2019). Here we quantify the number of signals that are unnecessarily discarded by algorithms that remove approximately kilohertz-wide frequency regions around every detection.

To perform this comparison, we used the database of signals detected during the 2018 April 27 observations, and we replicated the procedure described by Enriquez et al. (2017) and Price et al. (2020). The “blinking” procedure used by Price et al. (2020) specifies “Only the signal with the highest S/N within a window $\dots \pm 600$ Hz is recorded as a hit.” To replicate this step, we sorted the signals detected in each scan in decreasing order of S/N and iterated over the sorted lists. At every iteration, we kept the signal with the largest remaining S/N value and eliminated all other signals within ± 600 Hz. The next step described by Price et al. (2020) combines hits that fall within a certain frequency range into groups as long as the signal is detected in every scan of the source. We replicated this step by grouping signals that were present in both scans of each source according to Price et al.’s (2020) prescription for frequency range. The third step of the procedure described by Price et al. (2020) reads: “Additionally, any set of hits for which there is at least one hit

in the OFF observations within ± 600 Hz of the hit frequency from the first ON observation would be discarded.” This elimination seems wasteful because the presence of OFF-scan signals with drift rates that are unrelated to the ON-scan drift rate results in elimination. To replicate this step, we removed all groups of signals for which one or both of the two OFF scans contained an unrelated signal within ± 600 Hz of the detection in the first ON scan. To determine whether the signals were unrelated, we placed the following condition on drift rate,

$$|\dot{f} - \dot{f}_0| > 2\Delta\dot{f}, \quad (6)$$

where \dot{f} is the drift rate of the OFF-scan signal, \dot{f}_0 is the drift rate of the ON-scan signal, and $\Delta\dot{f} = 0.0173$ Hz s^{-1} is the drift rate resolution. Our direction-of-origin filters (Section 3.2) also remove signals if they are found in multiple directions on the sky, but only in conjunction with careful analysis of the drift rates of both signals. Specifically, our filters only remove the two signals if their drift rates are within a tolerance of $2\Delta\dot{f}$. For the purpose of this blanking analysis, we kept signals that satisfy this criterion because both pipelines remove them during subsequent filtering. We found that our pipeline detected 10,113,551 signals, whereas our pipeline with a blanking algorithm modeled after the descriptions given by Enriquez et al. (2017) and Price et al. (2020) detected only 1,054,144 signals. In other words, our pipeline detects ~ 10 times as many signals as the BL-like pipeline over the same frequency range, with a corresponding increase in hit rate density.

To summarize, we found that our algorithmic approach to signal identification explains the largest fraction of the factor of ~ 200 difference in hit rate density between our and the BL searches, likely followed by our better overall sensitivity for $>90\%$ of the frequency drift rates examined by the BL pipeline (Section 6.1), likely followed by our shorter integration times and consideration of a wider range of drift rates. It is also possible that the limited dynamic range of our 2-bit voltage data makes our search susceptible to spurious detections at the

harmonics of strong RFI signals (D. Price 2020, personal communication). We are planning to quantify the importance of this effect in the future.

The differential in hit rate density has implications for the validity of existence limit estimates and figure-of-merit calculations described by Enriquez et al. (2017) and Price et al. (2020).

6.5. Existence Limits

We describe three issues that affect recent claims about the prevalence of transmitters in the Galaxy (Enriquez et al. 2017; Price et al. 2020; Włodarczyk-Sroka et al. 2020).

First, the range of Doppler drift rates considered in these searches has been limited (± 2 and ± 4 Hz s⁻¹), whereas transmitters may be located in a variety of settings with line-of-sight accelerations that would only be detectable at larger drift rates (e.g., Sheikh et al. 2019).

Second, these claims invoke transmitters with certain EIRP values that are calculated on the basis of the nominal sensitivity to nondrifting signals. However, the sensitivity to signals drifting in frequency is demonstrably degraded (Section 6.1) with the incoherent dechirping method used in these searches. The published EIRP values could be erroneous by factors of up to 25 for these searches, depending on the drift rate of the putative signal.

Third, our preliminary candidate signal injection and recovery analysis (Section 4) reinforces the concerns voiced by Margot et al. (2018) and Pinchuk et al. (2019) about Enriquez et al.’s (2017) claims. Pinchuk et al. (2019) argued that an injection and recovery analysis would demonstrate that a fraction of detectable and legitimate signals are not identified by existing pipelines, thereby requiring corrections to the claims. We have shown that our current pipeline misses $\sim 7\%$ of the signals injected into the dynamic spectra (Section 4). We surmise that the BL pipelines used by Enriquez et al. (2017) and Price et al. (2020) miss a substantially larger fraction of signals that they are meant to detect because of reduced sensitivity (Section 6.1), time resolution (Section 6.3), and detection counts (Section 6.4) compared to our pipeline.

In light of these issues, published claims about the prevalence of transmitters in the Galaxy (e.g., Enriquez et al. 2017; Price et al. 2020; Włodarczyk-Sroka et al. 2020) almost certainly need revision. As mentioned in Section 4.4, we are planning improvements to our signal injection and recovery analysis. Until this refined analysis is complete, we will not be in a position to make reliable inferences about the prevalence of radio beacons in the Galaxy.

6.6. Drake Figure of Merit

The DFM (Drake 1984) is a metric that can be used to compare some of the dimensions of the parameter space examined by different radio technosignature searches. It is expressed as

$$\text{DFM} = \frac{\Delta f_{\text{tot}} \Omega}{F_{\text{det}}^{3/2}}, \quad (7)$$

where Δf_{tot} is the total bandwidth observed, Ω is the total angular sky coverage, and F_{det} is the minimum detectable flux. Assuming unit quantization and dechirping efficiencies, our search with an S/N threshold of 10 is sensitive to sources with flux densities of 10 Jy and above (Margot et al. 2018). For

consistency with earlier calculations (Enriquez et al. 2017; Price et al. 2020), we have assumed that the bandwidth of the transmitted signal is 1 Hz, resulting in a minimum detectable flux $F_{\text{det}} = 10^{-25}$ W m⁻². The sky coverage of this search is $\Omega = 31 \times 0.015 \text{ deg}^2 = 0.465 \text{ deg}^2$, i.e., 11 ppm of the entire sky. The useful bandwidth is $\Delta f_{\text{tot}} = 309.3$ MHz (Section 3.3). We used these parameters to calculate the DFM associated with this search and found $\text{DFM} = 1.11 \times 10^{32}$, where we have used units of GHz m³ W^{-3/2} for compatibility with Horowitz & Sagan (1993). We reanalyzed our 2016 and 2017 data sets (Section 6.8) and recomputed DFM values of 5.00×10^{31} and 4.71×10^{31} for these data sets, respectively, with an aggregate DFM for our 2016–2019 searches of 2.08×10^{32} . However, we regard these values and all previously published DFM values with skepticism.

The DFM values published in recent works do not provide accurate estimates of search volume or performance for a few reasons. First, the DFM relies on minimum detectable flux, but authors have ignored factors that can tremendously affect overall search sensitivity, such as quantization efficiency ($\sim 88\%$ for 2-bit sampling) or dechirping efficiency (60%–100% with the tree algorithm and the parameters of this search and as low as $\sim 4\%$ in the recent BL search described by Price et al. 2020). Second, it does not account for the range of drift rates considered in a search, which is clearly an important dimension of the search volume. Third, it ignores the quality of the signal detection algorithms, such that two surveys may have the same DFM even though their data processing pipelines detect substantially different numbers of signals (e.g., the blanking of kilohertz-wide regions of frequency space described in Section 6.4). For these reasons, we believe that the DFM values calculated by authors of recent searches, including our own, are questionable indicators of actual search volume or performance. Horowitz & Sagan (1993) expressed additional concerns, stating that the DFM “probably does justice to none of the searches; it is a measure of the odds of success, assuming a homogeneous and isotropic distribution of civilizations transmitting weak signals at random frequencies.”

In Section 6.1, we showed that the dechirping efficiency degrades rapidly for frequency drift rates larger than f_{max} (Figure 7). As a result, the minimum detectable flux for nondrifting signals, which has been used by Enriquez et al. (2017) and Price et al. (2020) in their DFM estimates (Equation (7)), is not representative of the minimum detectable flux of signals with $>90\%$ of the drift rates that they considered, which can be up to 25 times larger. Given the presence of this flux to the 3/2 power in the denominator of the DFM, we believe that the DFMs of these searches have been inadvertently but considerably overestimated. Other figures of merit, such as Enriquez et al.’s (2017) continuous waveform transmitter rate figure of merit (CWTFM), are also affected by this problem.

We can use our estimates of the mean dechirping efficiencies to quantify plausible errors in DFM estimates. In Section 6.1, we computed a rough estimate of 16.5% for the mean efficiency of the BL search conducted by Price et al. (2020), suggesting that the DFM of their search has been overestimated by a factor of ~ 15 . This value may be revised down once a more accurate estimate of the mean dechirping efficiency becomes available. For the UCLA searches conducted between 2016 and 2019, the mean dechirping efficiency is 72.4%, and

the quantization efficiency is 88.25%, resulting in an overall efficiency of 64% and DFM overestimation by a factor of ~ 2 .

6.7. Other Estimates of Search Volume

The range of drift rates considered in a search program obviously affects the probability of success of detecting a technosignature. For instance, a search restricted to drift rates smaller than $f_{\max, \text{BL}} = 0.15 \text{ Hz s}^{-1}$ could fail to detect the signal from an emitter on an Earth-like planet. The frequency drift rate dimension of the search volume does not appear to have been fully appreciated in the literature. It is distinct from the “modulation” dimension described by Tarter et al. (2010), who focused on “complex ... broadband signals.” It also appears to be distinct from the “modulation” dimension of Wright et al. (2018), who contemplated drift rates on the order of the “Earth’s barycentric acceleration,” i.e., 0.03 Hz s^{-1} at the center frequency of our observations. It is also absent from the CWTFM used by Enriquez et al. (2017), Price et al. (2020), and Włodarczyk-Sroka et al. (2020). The development of an improved figure of merit for radio technosignature searches is beyond the scope of this work. However, we recommend that improved figures of merit include the range of line-of-sight accelerations between emitter and receiver as a dimension of the search volume, as well as explicit guidelines regarding the treatment of quantization and dechirping efficiencies.

6.8. Reanalysis of 2016 and 2017 Data

Margot et al. (2018) presented the results of a search for technosignatures around 14 planetary systems in the Kepler field conducted on 2016 April 15, 16:00–18:00 UT, with the GBT. Pinchuk et al. (2019) presented the results of a similar search conducted on 2017 May 4, 15:00–17:00 UT, that included 10 planetary systems in the Kepler field but also included scans of TRAPPIST-1 and LHS 1140.

We reprocessed these data with our updated algorithms and detected a total of 13,750,469 candidate signals over the 2016 and 2017 epochs of observation. Tables of the signal properties of the detected candidates are available online for both the 2016 (Margot et al. 2020a) and 2017 (Margot et al. 2020b) data sets. We found that 13,696,445 (99.61%) signals were automatically flagged as anthropogenic RFI, and 54,024 signals were labeled as promising. Candidate signals found within operating regions of known interferers (Table 2) were attributed to RFI and removed from consideration. Visual inspection of all of the remaining 4257 candidate signals revealed that they are attributable to RFI. With this improved analysis, we confirm the initial results that no technosignatures were detected in the data obtained in 2016 (Margot et al. 2018) and 2017 (Pinchuk et al. 2019).

7. Conclusions

We described the results of a search for technosignatures that used 4 hr of GBT time in 2018 and 2019. We identified 26,631,913 candidate signals, 99.84% of which were automatically classified as RFI by rejection filters. Of the signals that remained, 4539 were found outside of known RFI frequency bands and were visually inspected. All of these were attributable to RFI, and none were identified as a technosignature.

We presented significant improvements to our signal detection and direction-of-origin filter algorithms. We tested the signal recovery of the updated procedures with a preliminary signal

injection and recovery analysis, which showed that our pipeline detects $\sim 93\%$ of the injected signals overall. This recovery rate increases to $\sim 98\%$ outside of known RFI frequency bands. In addition, our pipeline correctly identified 99.73% of the artificial signals as technosignatures. This signal injection and recovery analysis provides an important tool for quantifying the signal recovery rate of a radio technosignature data processing pipeline. Planned improvements to this tool will further illuminate imperfections in our and other groups’ pipelines and point to additional areas for improvement.

Our search represents only a modest fraction of the BL searches described by Enriquez et al. (2017) and Price et al. (2020) in terms of number of targets and data volume. However, our search strategy has advantages compared to these searches in terms of sensitivity (up to 25 times better), frequency drift rate coverage (2–4 times larger), and signal detection count per unit bandwidth per unit integration time (~ 200 times larger).

We described the limitations of recent DFM calculations in assessing the probability of success of different search programs. These calculations have ignored important factors such as quantization and dechirping efficiencies. In addition, the DFM does not account for the range of drift rates considered in a search or the quality of the signal detection algorithms. As a result, we suggest that recent DFM calculations are questionable indicators of actual search volume or performance. We recommend that improved metrics include the range of line-of-sight accelerations between emitter and receiver as a dimension of the search volume, as well as explicit guidelines regarding the treatment of quantization and dechirping efficiencies.













Our observations were designed, obtained, and analyzed by undergraduate and graduate students enrolled in an annual SETI course offered at UCLA since 2016. The search for technosignatures can be effectively used to teach skills in radio astronomy, telecommunications, programming, signal processing, and statistical analysis. Additional information about the course is available at <https://seti.ucla.edu>.

Funding for the UCLA SETI Group was provided by the Queens Road Foundation, Janet Marott, Michael W. Thacher and Rhonda L. Rundle, Larry Lesyna, and other donors (<https://seti.ucla.edu>). Funding for this search was provided by Michael W. Thacher and Rhonda L. Rundle, Howard and Astrid Preston, K. K., Larry Lesyna, Herbert Slavin, Robert Schneider, James Zidell, Joseph and Jennifer Lazio, and 25 other donors (<https://spark.ucla.edu/project/13255/wall>). We are grateful to a reviewer for useful suggestions. We are grateful to the BL team for stimulating discussions about dechirping efficiency, data requantization, and data archival practices. We are grateful for the data processing pipeline initially developed by the 2016 and 2017 UCLA SETI classes. We thank Smadar Gilboa, Marek Grzeskowiak, and Max Kopelevich for providing an excellent computing environment in the Orville L. Chapman Science Learning Center at UCLA. We are grateful to Wolfgang Baudler, Paul Demorest, John Ford, Frank Ghigo, Ron Maddalena, Toney Minter, and Karen O’Neil for enabling the GBT observations. The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC;

<https://www.cosmos.esa.int/web/gaia/dpac/consortium>). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

Facility: Green Bank Telescope.

ORCID iDs

Jean-Luc Margot  <https://orcid.org/0000-0001-9798-1797>
 Pavlo Pinchuk  <https://orcid.org/0000-0003-4736-4728>
 Sanjana Prabhu Desai  <https://orcid.org/0000-0003-1825-4028>
 Benjamin Duclos  <https://orcid.org/0000-0002-5194-1615>
 Adrian Lam  <https://orcid.org/0000-0002-4688-314X>
 Briley Lewis  <https://orcid.org/0000-0002-8984-4319>
 Rebecca Lewis  <https://orcid.org/0000-0001-7003-0588>
 Mason MacDougall  <https://orcid.org/0000-0003-2562-9043>
 Yuri Shimane  <https://orcid.org/0000-0001-7728-3780>
 Myank Singhal  <https://orcid.org/0000-0002-8793-3246>
 Armen Tokadjian  <https://orcid.org/0000-0002-4675-9069>
 Ryan S. Lynch  <https://orcid.org/0000-0001-5229-7430>

References

- Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2013, *ApJS*, **207**, 35
 Drake, F. 1984, SETI Science Working Group Report, .NASA Tech. Pap. TP-2244
 DuPlain, R., Ransom, S., Demorest, P., et al. 2008, *Proc. SPIE*, **7019**, 70191D
 Enriquez, J. E., Siemion, A., Foster, G., et al. 2017, *ApJ*, **849**, 104
 Gaia Collaboration 2016, *A&A*, **595**, A1
 Gaia Collaboration 2018, *A&A*, **616**, A1
 Gale, D., & Shapley, L. S. 1962, *The American Mathematical Monthly*, **69**, 9
 Gray, R. H., & Mooley, K. 2017, *AJ*, **153**, 110
 Harp, G. R., Richards, J., Tarter, J. C., et al. 2016, *AJ*, **152**, 181
 Horowitz, P., & Sagan, C. 1993, *ApJ*, **415**, 218
 Kogan, L. 1998, *RaSc*, **33**, 1289
 Korpela, E. J. 2012, *AREPS*, **40**, 69
 Lebofsky, M., Croft, S., Siemion, A. P. V., et al. 2019, *PASP*, **131**, 124505
 Margot, J.-L., Greenberg, A. H., Pinchuk, P., et al. 2018, *AJ*, **155**, 209
 Margot, J. L., Greenberg, A. H., Pinchuk, P., et al. 2020a, A Search for Technosignatures from 14 Planetary Systems in the Kepler Field with the Green Bank Telescope at 1.15–1.73 GHz, v4, Dataset, doi:10.5068/D1309D
 Margot, J. L., Pinchuk, P., Greenberg, A. H., et al. 2020b, A Search for Technosignatures from TRAPPIST-1, LHS 1140, and 10 Planetary Systems in the Kepler Field with the Green Bank Telescope at 1.15–1.73GHz, v4, Dataset, doi:10.5068/D1Z964
 Pinchuk, P., Margot, J.-L., Greenberg, A. H., et al. 2019, *AJ*, **157**, 122
 Price, D. C., Enriquez, J. E., Brzycki, B., et al. 2020, *AJ*, **159**, 86
 Sheikh, S. Z., Wright, J. T., Siemion, A., & Enriquez, J. E. 2019, *ApJ*, **884**, 14
 Siemion, A. P. V., Demorest, P., Korpela, E., et al. 2013, *ApJ*, **767**, 94
 Tarter, J. 2001, *ARA&A*, **39**, 511
 Tarter, J. C., Agrawal, A., Ackermann, R., et al. 2010, *Proc. SPIE*, **7819**, 781902
 Taylor, J. H. 1974, *A&AS*, **15**, 367
 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
 Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, **143**, 9
 Włodarczyk-Sroka, B. S., Garrett, M. A., & Siemion, A. P. V. 2020, *MNRAS*, **498**, 5720
 Wright, J. T., Kanodia, S., & Lubar, E. 2018, *AJ*, **156**, 260