

Empty tables, tested boundaries

A reproducible audit of one Breakthrough Listen radio cadence

Michael Petychakis · Theoxenia · Research report 01 · Version 1.0 · 18 October 2026

Status: prepared for Zenodo; unpublished, no DOI assigned, not externally peer reviewed.

Abstract

We investigate four empty detector tables in a six-scan Green Bank Telescope archive cadence toward HIP544 and three comparison pointings. All six re-fetched 1380-1381 MHz slices exactly reproduce the original local-file SHA-256 checksums. The original turboSETI configuration reproduces 186, 149, 0, 0, 0 and 0 top hits. Independent float64, nearest-channel integration supports the below-threshold result in the four empty scans: their maximum normalized scores remain below 8 against a threshold of 10. Instrumentation also reveals a drift-coverage defect: a requested maximum of 4 Hz/s reaches only 3.826596 Hz/s in magnitude for these 16-sample files. A checked compatibility correction restores all nonzero grid rates inside the requested interval, reaching 3.989864 Hz/s, and finds one additional OFF-scan hit. On identical measured backgrounds, the correction raises recovery of near-boundary injected tracks from 0/12 to 9/12; all 20 controls in the four empty scans are recovered. Remaining control misses are traced to local-maximum selection in crowded data. These results establish a bounded software and data-processing diagnosis, not a survey-completeness measurement or an astrophysical non-detection claim.

1. Question and data

The original EXP-003 search recorded 335 per-scan detections and a clustered table of 273 rows: 149 IN_OFF and 124 SINGLE_POINTING, with no ON_ONLY rows. Empty tables in both later ON scans made broader interpretation premature. We ask whether the emptiness is reproducible, whether the detector is functioning on those backgrounds, and whether its declared search interval is actually covered.

The inputs are six public Breakthrough Listen fine-resolution HDF5 files in the AGBT16B_999_13 archive [1,2]. Their source order is HIP544, HIP1024, HIP544, HIP114, HIP544, HIP1344: ON, OFF, ON, OFF, ON, OFF. The first header starts at MJD 57622.38260416667. Each slice contains 16 integrations, one IF and 357,913 float32 channels. The sampling interval is 18.253611008 s and the channel spacing is -2.7939677238464355 Hz. Each scan contains 292.057776128 s of integrated data; sample-center endpoints are 273.804165120 s apart. Reruns do not add observing exposure.

Bounded HTTPS range reads transferred 517,537,756 bytes in total, below the 268,435,456-byte cap for every file. All six resulting 22,917,760-byte slices exactly match the historical hashes. Their native power values are finite and positive. These are hashes of the local frequency slices, **not** hashes of the full remote archive objects. The original manifest and slice summary contain all source URLs, timestamps, headers, ranges and hashes. Raw telescope files are not redistributed in this package.

2. Methods and controls

The original serial CPU search uses turboSETI 2.3.2, float32 precision, one coarse channel, max_drift=4 Hz/s, min_drift=1e-5 Hz/s, threshold 10, no time-spike flagging and midpoint DC replacement. The locked environment includes Python 3.15.0, NumPy 2.5.4, blimp 2.1.4, h5py 3.16.0 and numba 0.68.0. The measured runs used an Apple Silicon Mac. Package versions, code revisions and the installed upstream source checksum are recorded.

A passive observer records the maximum score, normalization, finite fraction and number of above-threshold values at every hit-search call. It reads the arrays before upstream normalization and then calls the original function. A separate observer inspects the local-maximum selection. Its returned top-hit rows are required to match the unobserved run exactly.

The independent reference sums straight tracks in float64 with nearest-channel offsets $\text{round}(t \cdot k / 15)$, where $t=0, \dots, 15$ and integer endpoint k runs from -391 to $+391$. Tracks leaving the frequency window are excluded; frequencies never wrap. For both methods the normalization is based on the time-integrated spectrum: subtract its median and divide by the population standard deviation after excluding values outside its 5th-95th percentile range. The reference implements this calculation separately from turboSETI. It includes zero drift and uses different paths from the Taylor tree, so identical numerical maxima are not expected. These normalized scores are algorithmic quantities, not calibrated Gaussian significance or flux limits.

The initial protocol fixes three artificial one-channel tracks per scan at 23%, 48% and 73% of the channel range, with slopes -1 , $+1$ and $+10$ channels per integration. Each adds constant power corresponding to 40 normalization-standard-deviation units in the ideal integrated sum. These are synthetic additions to separate copies of real backgrounds. Matching requires frequency within three native channels and drift within 0.03 Hz/s. No amplitude or matching tolerance is adjusted after seeing a miss.

After measuring the coverage gap, the follow-up adds two tracks at 35% and 60% of channels, with slopes $+26$ and -26 channels per integration: approximately -3.979660 and $+3.979660$ Hz/s. Both original and corrected detectors search the same five-track injected copies. This follow-up is documented separately in PROTOCOL.md; it is not represented as part of the initial design.

3. Empty-table results

Scan / role	Original hits	Corrected hits	Corrected maximum	Independent maximum
1 / ON HIP544	186	186	635.244	635.244
2 / OFF HIP1024	149	150	260.186	259.875
3 / ON HIP544	0	0	7.554	7.554
4 / OFF HIP114	0	0	7.678	7.581
5 / ON HIP544	0	0	7.421	7.983
6 / OFF HIP1344	0	0	7.871	7.871

All four empty scans remain below threshold in both calculations. Every recorded upstream trial score is finite. Original per-scan counts reproduce exactly, and an additional cache-based reproduction repeats the original counts, control recoveries and maxima within the declared 0.005 -score tolerance. Valid input power alone would not establish this result; the independent calculation and positive controls provide additional checks.

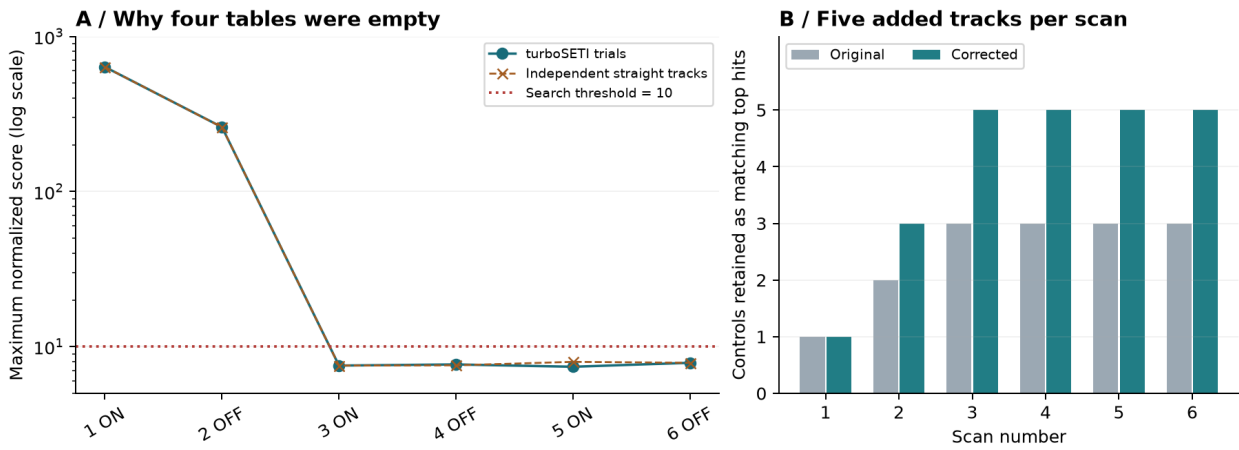


Figure 1. Left: maximum normalized scores and the threshold of 10; logarithmic scale. Right: retained controls on identical five-track injected backgrounds before and after the compatibility correction. Every bar reports measured trials, not a probability estimate.

The corrected run adds one top hit in scan 2, an OFF pointing: 1380.422967 MHz, -3.928638 Hz/s, reported SNR 11.355203. No previous frequency/drift pair is removed. That rate lies outside the original effective interval. Its occurrence in an OFF scan is not a determination of physical origin. The corrected run records 336 per-scan top hits. The original 273-row clustered table is preserved as historical evidence; it is not relabelled as the output of this correction, and no new complete cadence-clustering claim is made here.

4. Drift coverage and compatibility correction

With drift resolution $r = \text{abs}(\text{channel spacing}) / ((N-1) * \text{sampling interval})$, turboSETI's block counter used $\text{floor}(D / (r * N))$, while successive blocks advance by $N-1$ endpoint channels. For $N=16$ and $D=4$, this yields 24 outer blocks and an effective endpoint of 375 channels: only ± 3.826595903 Hz/s. The passive trace contains 798 calls but just 750 distinct nonzero rates; duplicate boundary calls are not independent trials.

The correction consistently uses $\text{floor}(D / (r * (N-1)))$. The final rate mask still rejects values outside the requested maximum. For this dataset, the corrected trace contains 834 calls covering all 782 distinct nonzero grid rates from -3.989863995 to +3.989863995 Hz/s. The next grid point lies outside ± 4 . Zero drift remains deliberately excluded by `min_drift`; the reference includes it as an additional check.

The second small correction converts the one-element total-hit array to a scalar before formatting it. The original NumPy formatting exception occurred after top-hit output, and is not the cause of the four empty tables. Corrected searches complete without that exception.

Theoxenia applies these two exact substitutions only to the hash-verified turboSETI 2.3.2 search function, for the wrapper's serialized CPU search. It restores the original function after success or failure and never edits the installed dependency on disk. Unknown source changes fail closed. Dedicated end-to-end tests place both positive and negative near-boundary tracks in synthetic files: the original search misses the specified drift/frequency matches; the corrected search recovers them. This work has not been submitted or accepted upstream. GPU execution and other upstream versions are not validated by this report.

5. Recovery and crowded-search limits

The initial three-track experiment recovers 15/18 controls, including all 12/12 controls in the four empty scans. The five-track follow-up uses identical inputs for its original/corrected comparison:

Scan	Original: five tracks	Corrected: five tracks
1	1/5	1/5
2	2/5	3/5
3	3/5	5/5
4	3/5	5/5
5	3/5	5/5
6	3/5	5/5
Total	15/30	24/30

Near-boundary recovery changes from 0/12 to 9/12, including all 8/8 boundary controls in the four empty scans. All 20/20 controls in those scans are recovered by the correction. The six remaining misses are in scans 1 and 2. At their injected drift/frequency locations, recorded trial scores are above 10, but stronger nearby per-channel maxima remove them from the final table. The suppression radius is 784 native channels, about 2.19 kHz to either side. In one case a stronger competing drift also replaces the injected drift in the per-channel maximum. The passive selection audit preserves exactly the same returned rows while exposing these internal decisions.

This demonstrates a distinction between crossing a detection threshold and surviving top-hit selection. We retain the misses as a limitation; we do not lower thresholds, move signals or claim perfect recovery. The chosen locations, strengths and idealized one-channel tracks do not represent a randomized population of possible signals.

6. What this establishes, and what it does not

The empty-table question is resolved at the tested configuration: the original result is reproducible; independent tracks remain below threshold; and artificial positive controls are recovered on the same four backgrounds. The drift-boundary defect is separately demonstrated, corrected and checked against real backgrounds. The crowded-data misses are attributable to recorded selection decisions.

This is **not** a statement that HIP544 has no transmitters, an estimate of telescope survey completeness, a calibrated flux/EIRP limit, a physical identification of the strong lines, or a statistical upper limit on civilizations. Only one cadence and one 1 MHz window are studied. Interference, bandpass structure, time variability, drift mismatch, finite resolution, zero-drift exclusion and local-maximum selection all constrain interpretation. Session-wide census and independent external review remain unfinished. Historical recovery-grid results were produced before this compatibility correction; they remain labelled as historical and cannot calibrate the corrected configuration without a new sweep.

7. Reproduce and inspect

The companion README gives one-command reproduction and package verification. A clean checkout needs uv and the locked Python/radio environment. Reproduction fetches the six bounded slices, writes new run manifests, compares the original diagnosis, audits control selection and reruns the corrected configuration. It does not overwrite published records, start a worker or contact Zenodo. Scientific outputs are expected to agree within the stated numerical tolerance, not necessarily byte-for-byte across platforms. Recorded artifact checksums protect the shipped evidence; raw-slice hashes establish input identity.

The figures can be regenerated from the recorded JSON without fetching telescope data. The PDF is rendered from this Markdown with ReportLab 4.4.9. The upload package contains this report, source revision, compact evidence, protocol, metadata and checksum lists. Zenodo publication is a separate owner action; a DOI would provide preservation and citation, not scientific endorsement.

Primary result records:

- Original source: 20261012T203545022355Z-exp-003-cadence.
- Initial diagnosis: 20261018T182502846875Z-detector-diagnosis; code 7241ef8.
- Original reproduction: 20261018T183336278843Z-detector-diagnosis; code b591b72.
- Corrected search: 20261018T183744725763Z-detector-corrected; code 3cf92e1.
- Corrected selection audit: 20261018T183947862671Z-control-selection; code 1fc41a1.

Full commit identifiers, exact processor snapshots, commands, timestamps and output hashes are in each manifest and the packaged source history reference.

8. Credits, licenses and review status

The observations were acquired and released by the Breakthrough Listen team and the Green Bank Telescope, not by Theoxenia. We credit the upstream archive and the turboSETI, blimpy, NumPy, h5py and numba developers. The contribution here is independent reproducibility work, a bounded compatibility correction and inspectable diagnostics; no affiliation or endorsement is implied.

Theoxenia code is Apache-2.0. Authored report text, figures and derived diagnostic summaries are CC BY 4.0. Source observations and third-party components retain their original terms; no raw telescope files or vendored third-party source are included in the evidence archive. Automated analysis, implementation and drafting were assisted by OpenAI Codex. The report is prepared for owner review and external publication; no independent human scientific review or upstream acceptance is claimed.

References

- [1] Lebofsky et al. (2019), The Breakthrough Listen Search for Intelligent Life: Public Data, Formats, Reduction and Archiving. [arXiv:1906.07391](#).
- [2] Berkeley SETI, [Breakthrough Listen public data and participation](#). Exact archive object URLs and local-slice checksums are in the source manifest.
- [3] Enriquez et al. (2017), The Breakthrough Listen Search for Intelligent Life: 1.1-1.9 GHz observations of 692 Nearby Stars. [arXiv:1709.03491](#). Background methodology; this report does not reproduce the full published survey.
- [4] UC Berkeley SETI, [turboSETI source repository](#). The experiments use locked version 2.3.2 and the exact installed source checksum in the manifests, not an unpinned current branch.
- [5] Zenodo, [REST API documentation](#) and [DOI reservation guidance](#). Submission routes checked 18 October 2026.