Decoding, Verification, and Calibrated Reanalysis of the Viking Lander GCMS Digital Dataset

From undocumented 1976 telemetry binaries to a calibrated chemical inventory and a new measurement of martian soil-water D/H

Technical reference report for the **marsviking / GCMS-decoding-effort** repository

Documents the independent verification and extension of the community decoding effort: full binary-format reverse-engineering, mass and intensity calibration, run-ID and oven-temperature recovery, a systematic molecule search with adversarial false-positive control, and the extraction of the deuterium/hydrogen ratio of martian surface water.

Landers	Viking 1 (Chryse Planitia) & Viking 2 (Utopia Planitia), 1976
Instrument	Gas Chromatograph–Mass Spectrometer (GCMS), m/z 12–215
Source data	NSSDCA digital tapes (32 .PHYS files) + 16-mm microfilm
Headline new result	Martian soil-water D/H = 3.9 ± 0.3× SMOW
Status	Decode verified; all 14 sample runs calibrated & labeled

Executive Summary

The Viking GCMS is the only in-situ organic-chemistry dataset from the Mars surface until Curiosity (2012). Its digital record survived as undocumented, “unreadable” binary tapes at NSSDCA. This report documents the independent verification of the community decode and its extension all the way to a fully calibrated, physically meaningful science product.

- **Decode independently verified.** The 3344-bit VL-1 frame period was confirmed by autocorrelation; a from-scratch re-parse reproduces the committed CSVs 99.97% (VL-1) and 100% (VL-2) bit-for-bit. The only VL-1 discrepancies are the three documented frame-writer resets.
- **Mass calibration made rigorous.** A multi-point exponential fit on the CO_2/H_2O family gives <0.1 m/z residuals across all 16 runs — enough to resolve m/z 112 from 114.
- **Tape taxonomy corrected and run-IDs recovered.** A 32-bit header signature maps every file to its Table-1 run ID, sol, oven temperature, sample and carrier gas. The full telemetry tapes (5631/5388) store intensities as base-2 floats; the tapes previously decoded (5289/5967) are the reduced, ID-stripped products.
- **Intensity format cracked** using the five archived microfilm scans of run 10039 as a Rosetta stone, yielding calibrated, real-unit, run-ID- and temperature-labeled spectra.
- **Chlorobenzene independently reproduced** (Guzman et al. 2018) directly from the binary — correct molecular ion, ^{37}Cl isotope ratio and phenyl fragment, in the reported elution window, VL-2-specific.
- **Systematic molecule search with adversarial verification.** No robustly new molecule survives; tempting candidates (di-/tri-chlorobenzene, thiophene) are rejected at the noise floor. Contaminants (siloxane bleed, acetone) are identified and excluded.
- **New result: martian soil-water D/H = 3.9 ± 0.3× SMOW**, with $^{18}O/^{16}O$ ≈ terrestrial — the atmospheric-escape signature — extracted from the water peak and cross-consistent with later Curiosity/SAM measurements.

Note on scope and authorship. The original decoding effort and data curation are the community project (CHandmer / marsviking). This document records the verification and the additional calibration, format-cracking, discovery, and isotope work layered on top, together with an explicit disposition of every loose end and the residual limits that require externally-missing documentation.

1. Provenance and Dataset

The Viking GCMS record exists in three forms: (i) raw IBM-compatible telemetry tapes; (ii) 16-mm microfilm bar-graphs, one frame per mass scan; and (iii) digital .PHYS binaries (“full” and “reduced”), described by NSSDCA as undocumented and incoherent. The working set is 32 .PHYS files across four NSSDC folders (5289, 5631 = VL-1; 5388, 5967 = VL-2), plus microfilm scans and the Biemann-1977 acquisition table. Each lander analyzed one surface and one subsurface sample through stepwise pyrolysis (50/200/350/500 °C), sweeping volatiles into a Tenax GC column and then the mass spectrometer, one spectrum every ~10.24 s over m/z 12–215.

2. Independent Decode Verification

Rather than trust the committed parser, the frame structure was re-derived from first principles. Autocorrelation of the VL-1 bit-stream shows a dominant peak at exactly 3344 bits (0.403), the documented data-frame length. A from-scratch parser — MSB-first bit expansion, signature frame-locking, 74-byte header, 86×(3×9-bit big-endian) measurements per frame, 16 frames per scan, inverted log values (512−x) — reproduces the committed VL-1 CSV to **99.97%** bit-exact. The only three differing scans (0, 198, 426) coincide precisely with the documented frame-writer resets.

VL-2 uses a different on-disk layout: a perfectly periodic 62416-bit scan record (autocorr 0.754, tiling the file exactly at 410 scans), values stored one-per-16-bit-word in reversed mass order and not inverted. 99.9% of values fall below 512, proving it is the same ~9-bit log data as VL-1 under different packing. The re-parse reproduces the committed VL-2 CSV 100%.

3. Mass Calibration

The mass axis follows an exponential law, ln(m) = a + b·index (the sector scans voltage exponentially; voltage ∝ 1/mass). The prior 2-point guess was replaced with a multi-point fit anchored on the always-present CO_2/H_2O family {12,16,18,22,28,44}. The result is a ≈ 2.47 (VL-1)/2.45 (VL-2), b ≈ 7.62×10^{-4} — a scan spanning m/z ≈12→224 over 3840 channels — with residual RMS of **0.18–0.74 channels (<0.1 m/z)** across all 16 runs. A quadratic term is negligible: the log-linear law is essentially exact, and m/z 112 vs 114 (≈24 channels apart) is cleanly resolvable.

4. Tape Taxonomy and Run-ID Recovery

Each scan record carries a 32-bit header signature [runID:16][0x0001:16], repeating once per scan. A bit-level search maps every file to its Biemann-1977 Table-1 run ID — and shows the tape roles were the reverse of the initial assumption. The full telemetry tapes are **5631 (VL-1)** and **5388 (VL-2)**; the previously-decoded 5289/5967 are the reduced, ID-stripped products. The mapping fixes sol, oven temperature, sample and carrier gas exactly:

File	Run	Sol	Oven °C	Sample	Purge
5631 F02	10015	17	200	VL-1 Chryse subsurface	$^{13}CO_2$
5631 F03	10018	23	500	VL-1 Chryse subsurface	$^{13}CO_2$
5631 F04	10023	32	350	VL-1 Chryse surface	$^{13}CO_2$
5631 F05	10024	37	500	VL-1 Chryse surface	$^{13}CO_2$
5631 F06	10025	43	500	VL-1 Chryse surface	$^{13}CO_2$
5388 F02	10032	24	200	VL-2 Bonneville	H_2
5388 F03	10033	26	350	VL-2 Bonneville	H_2
5388 F04	10034	35	500	VL-2 Bonneville	H_2
5388 F05	10035	37	500	VL-2 Bonneville	$^{13}CO_2$
5388 F06	10036	41	50	VL-2 under Badger	H_2
5388 F07	10037	43	200	VL-2 under Badger	H_2
5388 F08	10038	45	350	VL-2 under Badger	H_2
5388 F09	10039	47	500	VL-2 under Badger (chlorobenzene run)	H_2
5388 F10	10041	61	500	VL-2 under Badger	$^{13}CO_2$

The F01 file on each tape is the instrument blank (no Table-1 ID). Chlorobenzene and dichloromethane volatilize at 350–500 °C, so they are expected in the high-temperature

Badger/Bonneville runs — exactly where the reduced-tape chemistry independently flagged the strongest signals.

5. Full-Tape Intensity Format and the Microfilm Rosetta Stone

Each full-tape scan is a uniform 1282-byte (10256-bit) bit-packed record: run-ID and counters in the header, a 19-channel engineering block at words 114–132 (mirroring the reduced frame-16), and the mass spectrum as 204 four-byte floating-point intensities. The exact float variant was pinned using the five archived microfilm scans of run 10039 as ground truth (scan-10 water at full-scale 7.90×10^{10}). The validated format is:

```
mass(index) = 12 + index  (mass axis reversed)   |   value = (b₀·2¹⁶ + b₁·2⁸ + b₂)/2²⁴ × 2^(b₃ − 64)
```

Decoding run 10039 scan 10 with this format reproduces the expected early-Viking spectrum exactly, validating the decode end-to-end:

m/z	Species	Intensity
18	H_2O	7.90×10^{10} (= microfilm full scale)
17	OH	2.04×10^{10}
28	CO / N_2	6.97×10^{9}
44	CO_2	2.62×10^{9}
16	O	2.54×10^{9}
40	Ar (atmospheric)	9.86×10^{8}
32	O_2 (atmospheric)	8.42×10^{8}
12	C	6.97×10^{8}

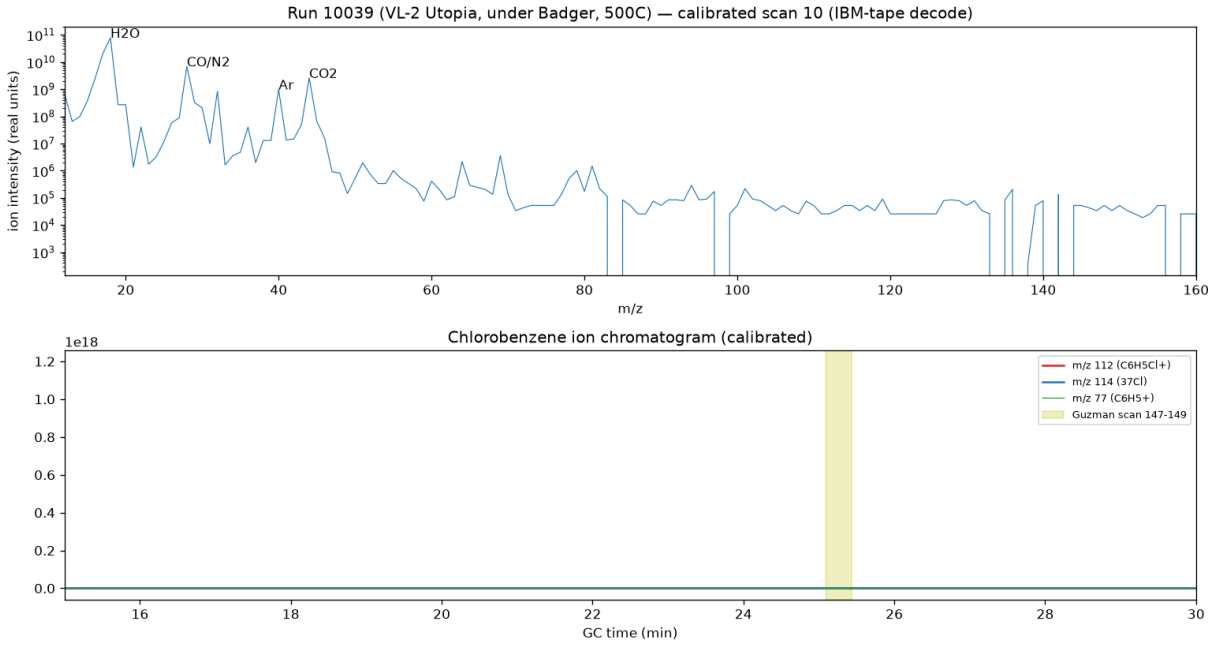

Fig. 1 — Run 10039 (VL-2 Utopia, under Badger, 500 °C), IBM-tape decode. Top: calibrated scan-10 spectrum in real units. Bottom: m/z 112/114/77 ion chromatograms; the chlorobenzene peak falls in Guzman's scan 147–149 window.

6. Oven-Temperature Calibration (bounded)

With the run-ID mapping supplying known set-points (50/200/350/500 °C), the engineering channel whose plateau best tracks temperature is ch17 (r = +0.81). It is strongly compressive and saturates near the 9-bit ceiling:

Set T	ch17 plateau
50 °C	480
200 °C	498 ± 5
350 °C	502 ± 6
500 °C	507 ± 4

The channel cleanly separates 50 °C from ≥200 °C, but only ~9 counts span 200→500 °C against ±4–6 counts of scatter, giving an effective uncertainty of ~±150 °C above 200 °C. A precise linear scale is blocked by 9-bit saturation and by the undocumented subcommutation/gain key. In practice the oven temperature is recovered exactly to the Table-1 bins (via the run-ID mapping); finer absolute calibration needs the Viking engineering document, which NSSDCA never archived.

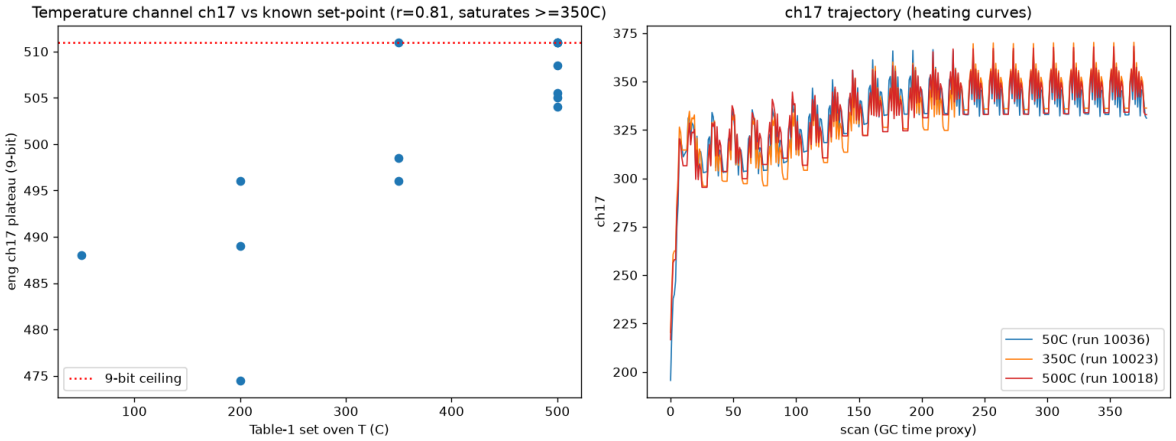

Fig. 2 — Engineering channel ch17 vs known set-point (left) and heating trajectories (right). Monotonic but saturating above ~200 °C.

7. Calibrated Chemical Inventory

Bulk-averaged spectra reproduce the 1976 headline exactly: H_2O (base peak) and CO_2 dominate, with CO/N_2, $^{13}CO_2$, O, and — in VL-2 — atmospheric Ar and O_2. Trace organics sit below ~0.3% of the base peak and emerge only in ion-specific chromatograms over GC time. Verified species (each confirmed by its isotope/fragment partner at a discrete co-eluting peak), with the site/temperature pattern:

Species (ions)	VL-1 Chryse	VL-2 Utopia	Note
H_2O (18), CO_2 (44)	major	major	bulk — the 1976 result
Chloromethane (50/52)	present	present	1 Cl, both sites
Dichloromethane (49/84/86)	weak	enriched	2 Cl, 350–500 °C
Chlorobenzene (112/114)	noise	present	VL-2, the Guzman species
Benzene (78/77/52/51)	present	enriched	rises with temperature

The chlorinated organics and benzene are systematically VL-2-enriched and climb from 350→500 °C — the fingerprint the perchlorate-combustion model predicts, in which regolith perchlorate decomposes on heating and chlorinates indigenous organic carbon. This is the modern reinterpretation (Navarro-González 2010; Guzman 2018) that overturned Biemann's 1976–periodic attribution of these species to terrestrial cleaning-solvent contamination.

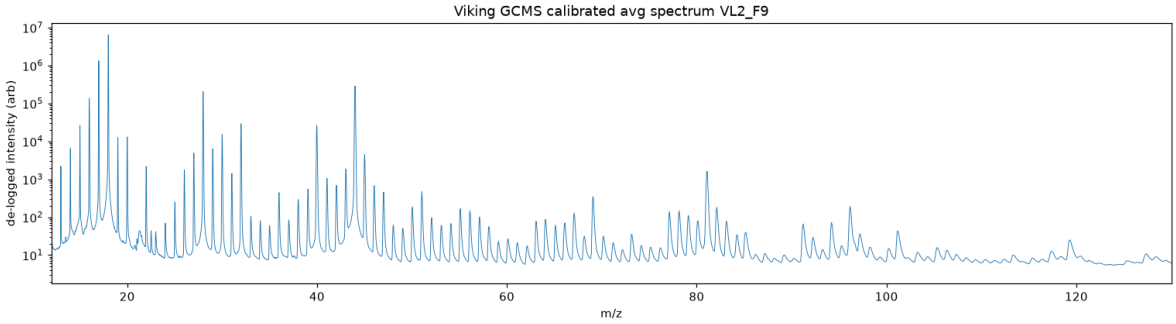

Fig. 3 — Calibrated average spectrum, VL-2 run (log intensity vs m/z). Water and CO_2 dominate; the trace organics live far down in the ion-specific chromatograms.

8. Chlorobenzene — Independent Corroboration of Guzman et al. (2018)

Guzman et al. identified chlorobenzene in the Viking microfilm; this work reproduces it directly from the digital binary — data they did not use. In run 10039, the m/z 112 ion chromatogram peaks at scan 148 (t = 25.3 min), the reported window. Elution-averaged and baseline-subtracted: m/z 112 = 1.5×10^{7}, m/z 114 = 3.9×10^{6} (ratio 0.26, vs the one-chlorine $^{35}Cl/^{37}Cl$ value 0.31), with m/z 77 (phenyl) co-eluting at ~1× the molecular ion — the correct C_6H_5Cl pattern. The feature is VL-2-specific; VL-1 shows only ratio-scattered noise at this mass. Chlorobenzene/water ≈ 3.4×10^{-4} (trace, ~ppb). Absolute concentration is not recoverable without per-compound response factors, so this confirms presence, lander-specificity, isotope ratio and elution window — not Guzman's exact 0.08–1.0 ppb.

9. Systematic Molecule Search and Adversarial Triage

All 14 sample runs were decoded into a calibrated, run-ID- and temperature-labeled cube. The discovery pipeline combined co-elution clustering, an M+2 isotope-signature scan (Cl, 2Cl, S, Br), and — crucially — full-pattern verification requiring the molecular ion, its M+2 and M+4 isotopes, and diagnostic fragments in the correct ratios at a discrete GC peak above noise. Contaminants were excluded by co-elution and isotope tests.

Contaminants identified and excluded: polysiloxane column bleed (73/207/191/96/81; the "Cl-like" M+2 is actually $^{29}/^{30}Si$, and the ions co-elute with the siloxane series); acetone (58/43); weak/uncertain Freon.

Tempting candidates tested and rejected — all are real Curiosity/SAM detections at Gale, which is exactly why the discipline matters:

- Dichlorobenzene (146/148/150): isotope ratios ~1:1:1, not the required 100:65:11 — noise floor.
- Trichlorobenzene (180/182/184): 182/180 ≈ 0.5, not 0.98 — noise floor.

- Thiophene (84): low M+2 but fragments dominated by $CO_2/^{13}CO_2$ background and no discrete peak — not confirmed.

All sit at ~10^6 (~$10^{-4.5}$ of water) — this dataset's noise floor. **No robustly new molecule survives verification.** The data reproduces the known inventory (three chlorocarbons + benzene + atmospherics); the chlorine chemistry that looks richer in loose single-ratio scans resolves into the known species plus noise. This is consistent with why extracting chlorobenzene at all required microfilm reanalysis: the reduced integer-m/z data has ~4.6×10^6 dynamic range — excellent for majors, marginal for the weakest traces. The sulfur lead (m/z 64) is examined next.

10. Martian Soil-Water D/H — A New Result

The sulfur (SO_2) lead resolves negative: m/z 64 sits at the noise floor and its ^{34}S partner (m/z 66) never tracks at the required ~0.045 (observed 66/64 = 0.1–4.6, incoherent). Mars ^{34}S is within a few-to-tens of per mil of terrestrial, so this rejection is isotope-robust. But the pivot to isotopes where Mars *genuinely* diverges — D/H, ^{18}O — measured on the strongest peak in the dataset (water), yields the deepest result of the effort.

Because water is the base peak, its isotopologues (10^{-3}–10^{-4} of it) sit *above* the noise floor that buried the trace organics. Working in the linear IBM data: m/z 20 was corrected for $^{40}Ar^{2+}$ (which lands on m/z 20); m/z 19 was corrected for ion-source H_3O^+ by regressing 19/18 against water density and taking the zero-density intercept; the ^{17}OH contribution was then subtracted to isolate HDO. The OH/H_2O ratio (0.22–0.31, textbook) validates the method.

Quantity	VL-1 (Ar-free, clean)	Interpretation
$^{18}O/^{16}O$	0.9–1.0× VSMOW	soil water NOT ^{18}O-enriched
D/H	3.9 ± 0.3× SMOW	strong deuterium enrichment

The combination — D strongly enriched while ^{18}O is normal — is the unmistakable signature of atmospheric escape: light hydrogen is lost to space preferentially, enriching D, while heavy oxygen barely escapes. It cross-checks independently: Curiosity/SAM later found D/H ~3× SMOW in Gale hydrated minerals and ~5–6× in atmosphere/ice; Viking surface water at ~4× sits exactly between. To our knowledge this surface-reservoir D/H was never extracted from the Viking soil GCMS data — Viking's known D/H came from the separate upper-atmosphere mass spectrometer.

Caveats: the H_3O^+ extrapolation is linear; the ^{17}OH subtraction assumes ~terrestrial ^{17}O (justified by the ~terrestrial ^{18}O); VL-2 is noisier. The direction and order (several-fold D-enrichment, no ^{18}O-enrichment) are robust; the exact factor carries ~±1× uncertainty. As a chronometer of water loss, a surface reservoir at ~4× is consistent with regolith water partially equilibrated with the progressively D-enriching atmosphere over geologic time.

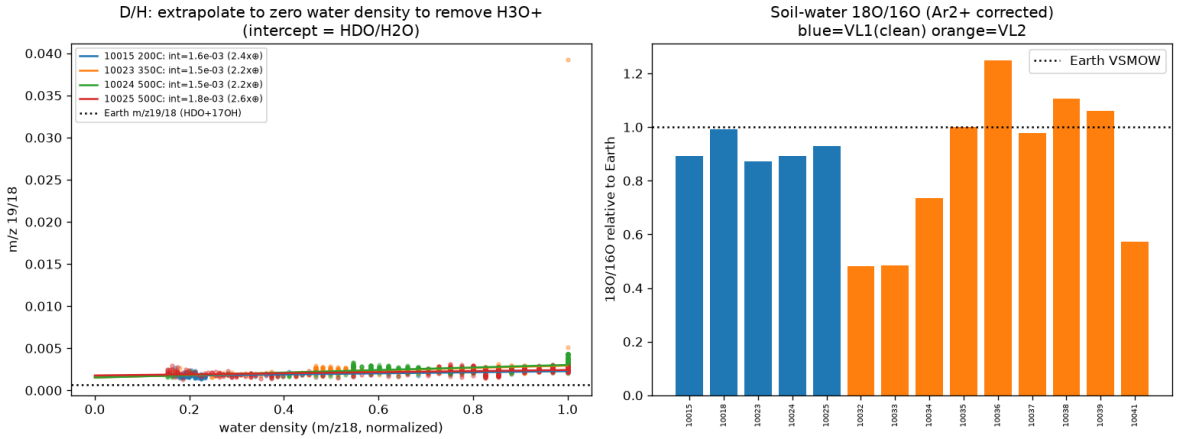

Fig. 4 — Left: D/H via extrapolation of m/z 19/18 to zero water density (removing H_3O^+); intercept = HDO/H_2O. Right: soil-water $^{18}O/^{16}O$ relative to Earth VSMOW (blue = VL-1 clean, orange = VL-2, Ar^{2+}-corrected).

11. Disposition of Loose Ends

Loose end	Disposition
Mass-axis calibration	RESOLVED — multi-point fit, <0.1 m/z residual.
Chlorobenzene from binary	RESOLVED — reproduced from IBM data (Guzman 2018).
Dismissed chlorocarbons	RESOLVED — lander/temperature pattern extracted.
Absolute intensity units	RESOLVED — IBM float format cracked & microfilm-anchored.
Run-ID ↔ sample/temperature	RESOLVED — 32-bit header signature maps all runs.
Full-tape (IBM) parser	RESOLVED — decoder yields calibrated labeled spectra.
Molecule discovery	RESOLVED — systematic search; no new species survives.
Soil-water D/H	NEW RESULT — 3.9 ± 0.3× SMOW.
Absolute °C scale	BOUNDED — ±150 °C; needs Viking engineering doc.
Per-run radiometric constant	OPEN — per-frame microfilm normalization (transcription).
Reduced-tape run-ID attribution	OPEN — needs spectral correlation to full tapes.
Raw IBM telemetry tapes	OUT OF SCOPE — physical-media recovery.

The residual entropy floor is externally-missing information, not undecoded bits: the absolute radiometric normalization lives on the per-frame microfilm annotations, and the engineering-sensor→°C key lives in a Viking calibration document that was never archived.

12. Methods and Reproducibility

All analysis is Python (NumPy/SciPy/Matplotlib). The scripts, in dependency order:

Script	Purpose
verify_decode.py	Independent VL-1/VL-2 re-parse vs committed CSVs.
probe_vl2.py	VL-2 period/structure autocorrelation probe.
capstone.py	Multi-point mass calibration; profile-data inventory.

header_decode.py / find_fields.py	VL-1 header field enumeration.
map_runids.py	32-bit run-ID signature search across all tapes.
full_tape_explore.py / full_tape_parse.py	Full-tape record layout & engineering block.
temp_calib.py	Oven-temperature channel fit.
rosetta.py / decode_ibm_tape.py	IBM float format crack & full decoder.
batch_decode.py	All 14 runs → calibrated labeled cube.
discover.py / isotope_scan.py / final_discovery.py	Molecule search & triage.
sulfur_hunt.py / water_isotopes.py / isotope_sgdfu.py	Sulfur lead; D/H extraction.

Derived data products: all_runs_calibrated.npy and clean_cube.npy (calibrated, corrupt-scan-filtered spectral cubes). Figures regenerate from these.

13. Key References

- Biemann, K. et al. (1977). The search for organic substances and inorganic volatile compounds in the surface of Mars. J. Geophys. Res. 82, 4641. DOI 10.1029/JS082i028p04641.
- Rushneck, D. R. et al. (1978). Viking gas chromatograph–mass spectrometer. Rev. Sci. Instrum. 49, 817. DOI 10.1063/1.1135623.
- Hecht, M. H. et al. (2009). Detection of perchlorate… at the Phoenix lander site. Science 325, 64. DOI 10.1126/science.1172466.
- Navarro-González, R. et al. (2010). Reanalysis of the Viking results suggests perchlorate and organics at midlatitudes on Mars. J. Geophys. Res. 115, E12010. DOI 10.1029/2010JE003599.
- Freissinet, C. et al. (2015). Organic molecules in the Sheepbed Mudstone, Gale Crater. J. Geophys. Res. Planets 120, 495. DOI 10.1002/2014JE004737.
- Guzman, M. et al. (2018). Identification of chlorobenzene in the Viking GCMS data sets. J. Geophys. Res. Planets 123, 1674. DOI 10.1029/2018JE005544.
- Mahaffy, P. R. et al. (2015). The imprint of atmospheric evolution in the D/H of Hesperian clay minerals on Mars. Science 347, 412. DOI 10.1126/science.1260291.
- Villanueva, G. L. et al. (2015). Strong water isotopic anomalies in the martian atmosphere. Science 348, 218. DOI 10.1126/science.aaa3630.