UIT THE ARCTIC UNIVERSITY OF NORWAY
PANSY RECEIVER MEMO SERIES
TROMSØ, NORWAY

PANSY MEMO #013
2026-10-20

To: PANSY data users
From: Juha Vierinen
Subject: **Sparse complex spectra downloads: file contents and data layout**

1 Scope and archive structure

The *Download spectra* button at `https://juha.no/rvd/` returns a TAR archive of the selected UTC day's available native HDF5 records, together with `dmd_properties.h5`. The download is not restricted to the clicked echo, displayed receiver beam, range gate, or browser imaging window. Changing the web imaging algorithm or its regularization does not change these data.

This memo describes the `pansy_pmse_sparse16_v2` product, whose complex spectra are formed from 16 ms complementary-pair samples. New production uses rectangular weighting: no Hann taper and no artificial endpoint zeros. Historical Hann records remain identifiable until reprocessed from available raw recordings; dewindowing cannot convert them. The saved `look_window` is authoritative; new `config_json` includes `window_kind=rectangular`. Do not infer weighting from the schema alone. Readers must inspect each record's `schema_version`, `config_json` and quality indicators. Legacy formats and incomplete or insufficient-coverage records must not be assumed to have the same usable scientific arrays.

The relative directory structure is

```
pansy-sparse-spectra-YYYY-MM-DD.tar
  dmd_properties.h5
  YYYY-MM-DDTHH-00-00/
    pmse@<600-second Unix-time bucket>.h5
      /<mode-start sample>/
        complete
        schema_version
        sparse_index
        spectra
        ... coordinates, statistics and provenance
```

The directories are hourly; the files use 600-second buckets. Numeric HDF5 group names are mode-start sample indices at 1×10^6/s, not Unix seconds. The properties file defines the Digital Metadata channel. Keep it with the directory tree when using `digital_rf.DigitalMetadataReader`; direct HDF5 inspection with `h5py` does not require Digital RF.

The handler copies existing HDF5 payloads unchanged, rather than generating a spectra-only export. Large scientific arrays are losslessly compressed within HDF5; the enclosing TAR is not gzip-compressed. Today's archive grows as records arrive and is not a frozen final-day release. A mode spanning midnight is not split between two downloads: its file belongs to its start-time bucket. Always skip groups without `complete=1`. If a currently written file is not readable, allow the producer to finish and retry the download.

The archive includes decoded complex spectra, dense noise estimates, summary statistics, coordinates, calibration inputs, and processing provenance. It does *not* contain raw ADC recordings, browser plots, reconstructed images, or separately archived channel-covariance spectra. The web page's separate derived-plot HDF5 link is a different product. The inverse-FFT recipe in Section 6 preserves the retained complex spectrum to floating-point precision. It cannot restore coefficients discarded by sparse selection.

2 Sparse-cell indexing and physical quantities

The principal array is

$$S_{\ell cn} = \texttt{spectra}[\ell, c, n], \qquad \text{shape}(S) = (K, 7, N), \tag{1}$$

stored as `complex64`: two float32 components, eight bytes per complex sample. Here K is the number of coherent looks and N the number of retained cells. The row `sparse_index`[n] contains three zero-based integers (b, f, r): beam index, full shifted FFT-bin index, and range-axis index. For this cell, its coordinates are

$$R_n = \texttt{slant_range_km}[r], \qquad v_n = \texttt{radial_velocity_mps}[f], \tag{2}$$

$$h_{\mathrm{nom},n} = \texttt{nominal_altitude_km}[b, r]. \tag{3}$$

Beam order is 0 Zenith, 1 North, 2 East, 3 South, 4 West. Receiver-channel order is retained in `rx_channels` and `module_names`. All seven channels and every valid coherent look are saved for each selected cell, including weak individual channel/look values. Selection is not applied independently to each receiver or look, which would bias interferometric correlations.

The spectral covariance at a retained cell is recoverable directly:

$$C_{ij,n} = \frac{1}{K} \sum_{\ell=0}^{K-1} S_{\ell in} S^*_{\ell jn}. \tag{4}$$

There is no need to store the 21 distinct cross-correlations and seven autocorrelations redundantly. The stored spectra have *not* had the interferometer phase snapshot applied. The historical correction is $C_{ij} \exp[+\mathrm{i}(\phi_i - \phi_j)]$; its sign, direction convention and contemporary calibration validity require an independent known-source check.

Power is decoded ADC-count squared per spectral bin under the stored normalization, not absolute reflectivity or temperature. Doppler velocity obeys $f_D = 2 f_{\mathrm{radar}} v/c$ algebraically; a physical up/down interpretation must not be inferred without validating the convention. Nominal altitude is beam-center height, not the true height of an echo arriving through a sidelobe.

3 Scientific array dictionary

In Tables 1 and 2, F is the paired-spectrum FFT length, R the saved range-axis length, and F_{noise} the pre-pair noise FFT length. Dimensions are read from each record, not hard-coded. Floating-point coordinate axes are generally float64; inspect the stored dtype.

Table 1: Scientific arrays in a normal complete v2 record.

Field	Layout / type	Meaning
`spectra`	$K \times 7 \times N$; complex64	Decoded, windowed complex Fourier coefficients; receiver phases preserved.
`sparse_index`	$N \times 3$; int32	Columns are beam, full FFT bin and range-axis index.
`mean_power`	$7 \times N$; float32	Mean $\lvert S\rvert^2$ across looks; recomputable.
`power_variance` `power_standard_error`	$7 \times N$; float32	Sample variance of power and standard error of the mean, respectively.
`noise_power` `noise_power_standard_error`	$7 \times 5 \times R$; float32	Dense noise power and uncertainty, retained even when no cells are selected.
`detector_snr_db`	N; float32	Excess signal/noise in dB; read `snr_definition`.
`detector_standard_error` `null_probability`	N; float32 / float64	Error of the normalized total-power detection statistic and approximate null tail probability. The error is not in dB.
`high_velocity_candidate`	N; bool	Screening mask for $\lvert v\rvert \geq 12.5\,\mathrm{m/s}$, not a physical event classification.
`slant_range_km` `range_sample`	R; float64 / int32	Slant range in km and decoded sample index.
`nominal_altitude_km` `altitude_valid`	$5 \times R$; float64 / bool	Beam-center height in km and valid-altitude mask.
`doppler_frequency_hz` `radial_velocity_mps`	F; float64	Full shifted FFT coordinates in Hz and m/s. Sparse indices reference this full axis.

`beam_azimuth_deg`
`beam_elevation_deg`
`rx_channels`
	5 / 5 / 7	Beam angles in degrees and original receiver-channel names.

`noise_doppler_frequency_hz`
`noise_radial_velocity_mps`
`noise_guard_mask`
	F_{noise}	Pre-pair noise axes and guard mask, distinct from the paired signal axis.

The detector's dB quantity is

$$\mathrm{SNR}_{\mathrm{dB},n} = 10 \log_{10} \left[\frac{1}{7K} \sum_{c,\ell} \frac{|S_{\ell cn}|^2}{N_{cbr}} - 1 \right], \tag{5}$$

where N_{cbr} is `noise_power`[c,b,r]. The standard archive applies a 3 dB excess-SNR gate and statistical selection, with the record's configuration being authoritative. Power variances have squared-power units; power standard errors have power units. These differ from uncertainty in the normalized detection statistic.

4 Timing, geometry and reconstruction inputs

The timing and calibration inputs are part of the download, not extra browser products. Complete records can still have `quality=partial_coverage` or insufficient usable looks. Elapsed mode duration and decoded coverage differ.

Table 2: Timing, receiver geometry and complementary-pair reconstruction inputs.

Field	Layout	Meaning
`mode_start_sample` `mode_stop_sample_exclusive`	scalar int64	Actual half-open observing interval in 1 MHz sample coordinates.

`mode_duration_seconds`
`coverage_seconds`
`coverage_fraction`
	scalar	Elapsed duration, decoded support in seconds and coverage ratio.

`look_bounds_sample`	$K \times 2$; int64	Inclusive start and exclusive stop of each coherent look.

`n_coherent_looks`
`pulses_per_beam_per_look`
`scheduled_pulses_per_beam_per_look`
	scalar / K / K	Number of looks, used pulses and scheduled pulses, including partial-look accounting.

`look_coherent_seconds`
`look_coherent_signal_power_gain`
`look_equivalent_noise_bandwidth_hz`
	K	Scheduled coherent duration (s), coherent signal power gain and equivalent noise bandwidth (Hz).

`look_window`	$K \times F$; float32	Pair-domain windows padded to the paired FFT length.
`pairs_per_beam_per_look` `look_pair_time0_sample_by_beam`	K / $K \times 5$	Used pair count and first pair phase-reference sample for each beam.
`effective_ipp_seconds` `pair_voltage_doppler_response`	scalar / F	Pair sampling interval (s) and pairing voltage response over Doppler.
`antenna_positions_m` `module_names`	7×3 / 7	Receiver-module phase-center positions (m) and names.

`phasecal_snapshot_rad`
`phasecal_sha256`
`geometry_sha256`
	5×7 / 5 / scalar	Phase snapshot (rad), per-beam phase-file hashes and geometry hash.

`phasecal_applied`
`phasecal_status`
`interferometer_calibration`
	scalar / text	Normally phase correction is not applied; status and conventions are explicit.

`tx_reference_template`
`complementary_code_family`
`complementary_code_family_similarity`
	$20 \times$ `tx_samples` / 5×4 / 5×4	Measured TX reference and complementary-family assignments/similarities.

Zero-filling omitted Fourier bins and taking an inverse shifted FFT recovers only a retained, windowed complementary-pair echo sequence. It does not recover discarded Fourier content or the original ADC samples. Read `reconstruction`, `pair_normalization` and `pair_phase_reference` before interpreting the reconstructed amplitudes or phases.

5 Additional metadata and overhead

The following field families are also included unchanged. Some are conditional on processing mode or record status.

Quality and schema
`complete`, `schema_version`, `quality`, `skipped_looks_json`. Completion is a write-publication flag, not a promise of full observing coverage; an empty retained-cell array may be valid.

Configuration and conventions
`config_json`, `normalization`, `per_pulse_tx_normalization`, `pair_normalization`, `pair_phase_reference`, `reconstruction`, `spectra_axis_order`, `sparse_index_columns`, `radial_velocity_sign`, `physical_height_convention`.

Detection and noise diagnostics
`snr_definition`, `detection_statistics`, `n_tested_cells`, `noise_gamma_shape_conservative`, `noise_guard_effective_bins`, `null_power_relative_standard_deviation`, `null_power_degrees_of_freedom_conservative`, `noise_estimator_domain`. These describe an approximate statistical model, not an empirical bound on false atmospheric events.

TX timing and validation
`tx_cycle_start_sample`, `tx_cycle_marker_observed`, `tx_cycle_similarity`, `tx_cycle_verified`, `rx_cycle_used` are T-element cycle arrays. Related scalar counts are `n_tx_cycles`, `n_tx_metadata_markers`, `n_recovered_cycle_markers` and `total_pulses_per_beam`. `tx_mean_min_max_power` is $K \times 3$. Recovered timing is not interpolated received voltage; validation flags and actual receiver support are retained.

Processing provenance
`generator`, `code_sha256`, `source_commit`, `raw_source`, `tx_source`, `processing_wall_seconds`, `peak_rss_bytes`. The archive preserves original source-path strings, runtime and peak-memory diagnostics. Those paths can reveal local directory names; this memo does not reproduce host paths or network access details.

Mean powers, variances and standard errors are recomputable from the stored complex spectra. Geometry, configuration, and reference metadata repeat across records for reproducibility. Their storage cost and HDF5/TAR overhead are extra to the complex-spectra payload, whose uncompressed size is

$$B_{\text{spectra}} = 8K\,7N \quad \text{bytes}. \tag{6}$$

Each dense float32 noise array adds $4\,7\,5R$ bytes before compression. A full dense spectral-grid comparison uses $K\,7\,5FR$ complex samples; a comparison restricted to the tested domain instead uses $K\,7\,$`n_tested_cells`. Neither ratio is a byte-compression ratio.

6 Recovering the complex time-domain signal

The inverse FFT is reversible for the retained complex spectrum, not for the original full-band waveform. One must inverse-transform the complex coefficients separately for every look and receiver, not the mean powers, SNR values, or reconstructed image. Combining looks incoherently into a power spectrum before inverse transformation would destroy the voltage phases.

Let d_{2m} and d_{2m+1} be the two measured-TX-decoded complex echoes in a complementary pair, for one receiver, look, beam and range. The forward stages are

$$p_m = (d_{2m} + d_{2m+1})/\sqrt{2}, \tag{7}$$

$$q_m = w_m p_m, \qquad 0 \le m < J, \qquad q_m = 0 \quad (J \le m < F), \tag{8}$$

$$X = \mathrm{fftshift}\big(\mathrm{FFT}_F(q)\big), \qquad \widetilde{X} = MX, \tag{9}$$

where J is the scheduled pair count, w is the saved unit-L^2 weighting including validity masking, and M is the archived sparse-bin mask. For new rectangular records, $w_m = 1/\sqrt{N_{\mathrm{valid}}}$ at every valid pair and zero only at rejected/padded samples. This constant normalization preserves the spectral noise-power units; it is not a taper. Historical records instead contain a normalized Hann taper. The forward FFT is unnormalized. The exact retained-signal inverse is

$$q_{\mathrm{ret}} = \mathrm{IFFT}_F\left(\mathrm{ifftshift}(\widetilde{X})\right). \tag{10}$$

NumPy's default inverse FFT includes the required $1/F$ factor. Do not divide by F a second time, and do not inverse-transform only the compact list of retained bins: first place them at their original full-FFT indices.

The following example uses only NumPy and h5py after extracting the TAR. It selects one beam/range combination present in a complete record, reconstructs all looks and all seven receivers at that gate, and explicitly checks the complex FFT round trip. Other beam/range gates are reconstructed independently.

```
from pathlib import Path
import h5py
import numpy as np

root = Path("extracted-day")
for path in sorted(root.glob("*T*/pmse@*.h5")):
    with h5py.File(path, "r") as h:
        keys = sorted(k for k in h if k.isdigit())
        for key in keys:
            g = h[key]
            if "complete" not in g or int(g["complete"][()]) != 1:
                continue
            if "spectra" not in g or g["spectra"].shape[2] == 0:
                continue
            index = g["sparse_index"][()]
            b, _, r = map(int, index[0])  # choose a retained gate
            cells = np.flatnonzero((index[:, 0] == b)
                                   & (index[:, 2] == r))
            F = len(g["doppler_frequency_hz"])
            K = g["spectra"].shape[0]
            X = np.zeros((K, 7, F), dtype=np.complex64)
            X[:, :, index[cells, 1]] = g["spectra"][:, :, cells]
            kept = np.zeros(F, dtype=bool)
            kept[index[cells, 1]] = True
            q = np.fft.ifft(np.fft.ifftshift(X, axes=2), axis=2)
            check = np.fft.fftshift(np.fft.fft(q, axis=2), axes=2)
            atol = 1e-6 * max(1., float(np.max(np.abs(X))))
            np.testing.assert_allclose(check, X, rtol=1e-6, atol=atol)
            amplitude = np.abs(q)       # [look, channel, pair_time]
            phase_rad = np.angle(q)
```

6.1 Amplitude, phase, timing and window handling

The returned q_{ret} is a complex amplitude-domain signal. Its envelope is $|q_{\mathrm{ret}}|$, its phase is $\arg q_{\mathrm{ret}}$, and its power is $|q_{\mathrm{ret}}|^2$. No magnitude, conjugation, phase-calibration factor, or extra FFT normalization is applied before the inverse. It remains in the measured-TX-normalized complementary-pair domain with the saved coherent weights applied, not original ADC voltage units.

The normal full look uses $F = 128$ complementary pairs (256 decoded pulses), 16 ms pair spacing, and 2.048 s coherent duration. Its FFT-bin spacing is 0.488 281 25 Hz, or approximately 1.5573 m/s at 47 MHz. A partial look is padded to the same grid but has shorter actual duration and poorer resolving power. Rectangular weighting has a different leakage response from historical Hann weighting; do not confuse equal FFT-bin spacing with equal spectral response.

The nominal sample separation is `effective_ipp_seconds`, normally 0.016 s. Pair times refer to the midpoint of the two pulses. `look_pair_time0_sample_by_beam` gives each look/beam's first reference sample. The example's uniform time grid is nominal: for sample-level clock tracking use `tx_cycle_start_sample`, `look_bounds_sample`, the IPP from `config_json`, and cycle-validation masks. The last look may be shorter than F samples. Do not concatenate looks through gaps as one continuous measured record. For nominal absolute times use $t_{\ell m} = s_{0,\ell b}/f_s + m\Delta t$, where $s_{0,\ell b}$ is the stored first-pair reference sample and f_s comes from `config_json`. The scheduled pair count is `scheduled_pulses_per_beam_per_look`/2. Keep all F output samples for the FFT round trip; padding and gaps are not additional measured samples.

Do not reapply a window, divide pointwise by saved zeros, crop the IFFT, or discard its zero-padding interval before checking the round trip. Sparse-bin masking is a Fourier projection, so the retained-signal IFFT may be nonzero even at original window zeros, invalid input times or padded times. These values are not newly recovered measurements. Historical Hann endpoints cannot be recovered by division, and small Hann weights amplify errors. New rectangular weights have no taper zeros: their single nonzero value is a safely reversible constant.

6.2 Removing the rectangular normalization

For a rectangular look let $a_\ell = \max_m w_{\ell m}$. The retained-bin projection in decoded pair-voltage units is simply $p_{\mathrm{ret},\ell} = q_{\mathrm{ret},\ell}/a_\ell$. With all Fourier bins retained, this recovers every valid original pair sample, including both endpoints. Sparse selection still omits frequencies, and values at rejected/padded times are not recovered measurements. The following fragment uses q from the preceding example inside the same open-record context and rejects historical tapers rather than silently dividing by Hann:

```
w = g["look_window"][()].astype(np.float64)
a = np.max(w, axis=1)
if (not np.isfinite(w).all() or np.any(w < 0) or np.any(a <= 0)
        or not np.all((w == 0) | (w == a[:, None]))):
    raise ValueError("Not rectangular: cannot undo historical Hann")
pair_voltage = q / a[:, None, None]
```

The repository reader provides this as `reconstruct(record,beam,restore_pair_scale=True)`, or the command-line option `python -m pmse.reconstruct_echoes -pair-voltage` with the usual root, key, beam and output arguments. Its default output retains the original stored-FFT normalization for backward compatibility. Keep the saved validity weights alongside either output and do not combine separate looks across gaps.

6.3 What can and cannot be reversible

1. **Retained complex spectrum: reversible.** Applying a forward FFT and shift to Eq. (10) returns every saved complex coefficient and the zero-filled spectrum, within floating-point error. This is not a bit-identical promise after float32 arithmetic or serialization conversions.
2. **Full pre-selection windowed pair signal: not recoverable from the current sparse archive.** Exact recovery requires the omitted complex coefficients as well, or a lossless complementary residual. Two full signals can differ only in discarded bins yet produce the same sparse archive.
3. **Unwindowed pair signal: rectangular normalization is reversible.** With a full Fourier grid, division by the one nonzero constant restores all valid pair samples. Rejected samples remain absent. Historical Hann records additionally lose endpoint samples; their taper cannot be safely undone from sparse spectra.
4. **Individual decoded pulses or raw ADC: further information is absent.** The pair sum does not determine its two inputs; matched filtering, range selection and discarded raw samples cannot be undone by an inverse FFT.

Consequently, full-waveform reversibility would require an explicitly changed archive contract and additional stored data. That is distinct from the currently verified retained-spectrum FFT round trip. Switching to rectangular weighting removes the taper loss but does not change sparse retention or undo the complementary-pair sum. Replacing historical Hann records requires reprocessing the original raw recordings, not manipulating old coefficients.

6.4 Changing the FFT length or window after downloading

For rectangular records, first recover the retained pair-voltage projection as above. A different analysis window or FFT length can then be applied to that complex time series. This is a reanalysis of retained information, not an exact recreation of an alternative processing chain applied to the full original decoded signal. Frequency coefficients discarded during detection remain unknown. Zero-padding interpolates the spectrum but does not improve true resolving power. Cropping or shortening looks changes their response; longer coherent integrations require phase-consistent, correctly timestamped samples and explicit handling of gaps and observing-mode boundaries.

The 128-bin grid, rectangular constant normalization and retained-cell mask must therefore be kept as provenance even if a derived product uses another grid or window. For a noise-filled reanalysis, distinguish measured retained coefficients from synthetic omitted-bin coefficients as in Section 7.

6.5 Historical Hann replacement

The requested rectangular backfill reprocesses available raw recordings oldest-first, using a separate one-core worker. Each new native record is written and read back before publication. Historical Hann groups are saved in a separate recovery location, then replaced atomically in the live archive; other records in the same file are preserved. Recovery copies are not part of the public download. While reprocessing is in progress, a day may contain both weightings: inspect each record's actual `look_window`. Expired raw recordings cannot be regenerated, so Hann records without available raw support cannot be converted by this procedure.

7 Recovering noise information

The complex noise present in the retained bins is already part of the measured `spectra`, and therefore already part of q_{ret}. It is not stored as a separate noise realization. Adding another noise draw to the retained coefficients would double-count noise, not reconstruct the observation.

The dense `noise_power`[c,b,r] estimates the mean noise power in a decoded spectral bin for each receiver, beam and range. Its standard error describes uncertainty in that estimate; neither field contains the phases of discarded noise samples. **Actual noise in omitted bins cannot be recovered.** What can be reconstructed is a specified statistical noise model, with a reproducible random seed if desired. This is a synthetic realization, not the original missing data.

7.1 A model consistent with the saved coherent weights

Let $N_c = \texttt{noise_power}[c, b, r]$ for the chosen beam and range. Under a white, proper-complex Gaussian receiver-noise model, draw independent real and imaginary components with variance $N_c/2$:

$$\eta_{\ell cm} = \sqrt{N_c/2}\,(u_{\ell cm} + \mathrm{i}v_{\ell cm}), \qquad u, v \sim \mathcal{N}(0, 1). \tag{11}$$

Generate the noise *before* the coherent window, then form $q^N_{\ell cm} = w_{\ell m}\eta_{\ell cm}$ and transform it with the same unnormalized FFT and shift as the data. Because $\sum_m |w_{\ell m}|^2 = 1$, this gives

$$\mathbb{E}|X^N_{\ell cf}|^2 = N_c, \qquad \mathbb{E}|q^N_{\ell cm}|^2 = N_c|w_{\ell m}|^2. \tag{12}$$

This preserves the correlation induced by any saved gaps, padding, or historical Hann weighting. For a complete, gap-free rectangular look the modeled Fourier bins are independent under the white Gaussian assumption. Drawing independent coefficients in every bin would miss gaps, partial looks and historical taper correlations. The complementary pair sum has unit white-noise gain because of its $1/\sqrt{2}$ normalization.

The following helper reads the saved windows and noise powers from an open complete v2 record. The boolean `kept` mask comes from the inverse-FFT example in Section 6; it is a full F-element frequency mask, not a mask over the compact N retained cells.

```
def synthetic_noise(g, b, r, kept, seed=2026):
    rng = np.random.default_rng(seed)
    K = g["spectra"].shape[0]
    F = len(g["doppler_frequency_hz"])
    w = g["look_window"][()].astype(np.float64)  # [K, F]
    N = g["noise_power"][:, b, r].astype(np.float64)
    assert w.shape == (K, F) and kept.shape == (F,)
    np.testing.assert_allclose(np.sum(w*w, axis=1), 1., atol=2e-6)
    shape = (K, 7, F)
    eta = (rng.normal(size=shape) + 1j*rng.normal(size=shape))
    eta *= np.sqrt(N[None, :, None] / 2.)
    q_noise = w[:, None, :] * eta
    X_noise = np.fft.fftshift(np.fft.fft(q_noise, axis=2), axes=2)
    masked = np.where(kept[None, None, :], X_noise, 0.)
    q_noise_retained = np.fft.ifft(
        np.fft.ifftshift(masked, axes=2), axis=2)
    return X_noise, q_noise, q_noise_retained
```

The returned arrays have shape $(K, 7, F)$. The second return is modeled full windowed noise; the third is its projection onto the same retained-frequency mask as the measured signal. Their complex envelopes, phases and powers are obtained using `np.abs`, `np.angle` and squared absolute value, respectively. For F_M retained bins, Parseval's identity gives the model's expected retained-noise energy per look/channel as $N_c F_M/F$, not N_c per output time sample.

7.2 Optional synthetic filling of discarded bins

If a complete, noise-filled Fourier grid is needed for a simulation or a visualization, preserve the observed coefficients and fill *only* omitted bins. The following fragment belongs inside the same open-record context as the inverse-FFT example:

```
X_noise, _, _ = synthetic_noise(g, b, r, kept, seed=2026)
X_filled = X.copy()
X_filled[:, :, ~kept] = X_noise[:, :, ~kept]
q_filled = np.fft.ifft(np.fft.ifftshift(X_filled, axes=2), axis=2)
np.testing.assert_array_equal(X_filled[:, :, kept], X[:, :, kept])
```

This augmentation preserves the retained complex data but does not restore discarded weak signals or the original full-band waveform. It also need not restore exact window zeros, because observed retained bins and simulated omitted bins are mixed. Label it as synthetic noise filling, and keep the original retention mask and random seed with any derived result.

The model assumes independent receiver noise, independent looks, stationary per-record noise power and white pair-domain noise before windowing. Real array noise can have inter-receiver, temporal or range correlation that cannot be inferred from scalar noise powers. Detection-conditioned noise statistics, colored interference and estimator uncertainty are not supplied by this simulation. A repeated seed reproduces the synthetic model draw, not the original discarded noise realization.

8 Interpretation limits and provenance of this memo

A missing sparse cell is not zero power: it may be outside the selection domain or fail the archive's selection tests. Missing or incomplete records may instead reflect coverage or delivery. Doppler moments computed only from retained bins are censored estimates. Discarded cells cannot be recovered by lowering a threshold after the raw recordings have been deleted.

The dense per-channel noise powers do not supply a measured full spatial noise covariance. Historical phase snapshots are not an independent validation of contemporary calibration. A high radial velocity alone does not establish a vertical draft. Browser reconstruction images, including their fit residuals and regularization settings, are not added to this native archive.

The field catalogue was checked against a complete native v2 record and the producer/download-handler source. Source documentation is maintained at `https://github.com/jvierine/pansy_receiver/tree/main/pmse`. This memo uses the unmodified `uitmemo.sty` and UiT seal from `https://github.com/jvierine/uit-memo-style`, revision `06dd4b15d32ddeaf06f6c9df583079b84f494c87`. Private template sources are build dependencies, not bundled public web assets.

Script: `scripts/build_pmse_download_guide.sh`.