VLBA Scientific Memo 43

On scaling of VLBA data

Eric W. Greisen

April 27, 2026

Abstract

Issues with the FITS-IDI data file and the scaling of visibility data by `FITLD` are discussed. Changes to `FITLD` are described and potential calibration issues are also mentioned.

1 Introduction

There appears to be a lot of confusion on the part of VLBA scientists about the contents of DiFX FITS-IDI files and the scaling of data by `FITLD` in $\mathcal{AIPS}$. I hope to clarify the situation with this note. VLBA Sensitivity Upgrade Memo 52 contains a nice explanation of sampling, but only addresses the Van Vleck portion of the scaling that `FITLD` does. The TV018 observations done on 29 July 2025 provide 3 data sets to allow some testing of `FITLD` and subsequent calibration.

2 FITLD

So far as I can tell, the FITS-IDI file does not provide any direct way to tell PFB from DDC from VNDA. Given that PFB is always 16 data channels, corresponding to 8 IFs dual polarization or 16 IFs single polarization with 32 MHz bandwidth each and that DDC will never do those combinations, one can tell PFB from DDC, but not from the more flexible VNDA. VNDA can also do 4- and 8-bit (16- and 256-level) correlation. Thus, when it is in these modes, it can be distinguished. But when VNDA is in the 2-bit mode the FITS-IDI file provides no clue.

I will start by describing what the previous public version of `FITLD` does. Adverb `DIGICOR` was first added to `FITLD` March 1, 1995 by Phil Diamond and Leonia Kogan *before DiFX*. `FITLD` rounds `DIGICOR` is rounded to an integer. If that is negative no "digital corrections" are made. If it is zero, it was set to 1. Value 1 applies all corrections to VLBA only, value 3 applies all corrections to any array. Values 2 and 4 apply the corrections to cross-power and only to total power if zero-padding was used in the correlator. If the keyword `VANVLECK` appears in the FITS-IDI file with a positive value, `FITLD` assumes that the Van Vleck correction was done in the correlator and the Van Vleck factor is set to one. In the previous public version no corrections were made to data where one or both antennas had 256 levels and rather incorrect factors were applied when one or both antennas had 16 levels. In a version of `FITLD` released on 1 April 2026 more appropriate corrections were added for these VNDA data. The Van Vleck factor is not determined for data where one or both antennas have 16 or 256 levels. So far, all data sets of this nature include the `VANVLECK = 1` keyword so a good Van Vleck value has not been needed.

I do not know the origin of these beyond what the source-code comments say. Those comments are:

> *"Applies digital correction for both AUTO and CROSS correlation spectrum. In the CROSS case the correction is applied as a factor because only a low cross-correlation is considered. The program calculates the cross-correlation spectrum's component using the formula:*
>
> $$R(L) = KSI(L) * BFACT$$
> $$BFACT = (GAMMA * *2)/(A * RM * ALFA * H(0))$$
>
> *where $KSI(L)$ and $R(L)$ are the measured and estimated amplitude; A is a coefficient applied to the data at the correlator; RM is the maximum possible value of digital signals' correlation, specified by the types of digitizers; $ALFA$ is a slope of the relation between amplitudes of input analog signals' correlation and normalized digit signals' correlation. $H(0)$ is the value of the normalized auto convolution function of a tapering function at FFT if the argument equals 0. $GAMMA$ is a known factor applied to the digitizers' levels at the VLBA correlator.*
>
> *In the AUTO case a full non linear correction is applied in the delay domain after performing a Fourier transform from a measured spectrum to a correlation function and then the inverse Fourier transform of the corrected correlation function to the spectrum. As of version 8.1 the saturation effect is compensated for in the case of self-spectra."*

The value of $H(0)$ is 0.87890625 for box taper or 0.333 for Hanning. If both levels are 4, the parameters are

$$RM = 4.3048$$
$$ALFAC = 0.882518$$
$$GAMMA = 3.335875 * 64.0/63.0 (= 3.3888254)$$

If both are 2, the parameters are

$$RM = 1.0$$
$$ALFAC = 2.0/PI$$
$$GAMMA = 1.0 * 64.0/63.0 (= 1.0158731)$$

and if one is 2 and the other 4 levels, they are

$$RM = 5.8784$$
$$ALFAC = 0.882518$$
$$GAMMA = 3.335875 * 64.0/63.0 (= 3.3888254)$$

If the `VANVLECK` keyword is greater than 0, $ALFAC = 1.0$. There is also a constant called $COEFA$ set to 5.360.

The scaling factor is then

$$BFACTC = COEFA * RM * ALFAC * H/(GAMMA * GAMMA)$$
$$BFACTA = COEFA * RM * H/(GAMMA * GAMMA)$$

where $BFACTC$ is divided into cross-correlations and $BFACTA$ to auto-correlations. $BFACTC$ and $BFACTA$ are the same when the Van Vleck correction has been done at the correlator. The scaling factors are

$levels$	$BFACTA$	$BFACTC$
$2x2$	4.5648699	2.9060888
$2x4$	2.4113936	2.1280982
$4x4$	1.7658831	1.5584236

The data are then divided by $BFACTC$ for cross-correlations and for RL and LR auto-correlations.

For RR and LL auto-correlations with number of levels 2 or 4, the spectrum is first dealt with by a subroutine whose precursor comments say:

> *"Routine performs the digital correction of auto correlation spectrum transforming a measured spectrum to the delay domain, applying a digital correction to the obtained auto correlation function and transforming the corrected autocorrelation function to frequency domain."*

This correction involves using 1000-value tables which are different for the 2x2, 2x4, and 4x4 cases. A saturation parameter is then computed for the old VLBA correlator (pre DiFX) only

$$SATUR \quad = \quad 0.125 * NSSAT * WT + 1.0$$

where WT is the weight of the data sample before scaling and $NSSAT$ is 2 for full polarization data and 1 otherwise. The auto-correlation samples are then divided by $BFACTA$ and multiplied by $SATUR$.

A new public version of `FITLD` has been released on April 23, 2026. In that version, if antennas have more than 4 levels, $BFACTC$ is changed by dividing by 3.5 for any 16-level antenna and by 3.8 for any 256 level antenna. Thus:

$$\begin{aligned}
4-16 \quad BFACTC &= SQRT(BFACTC)/3.5 \\
16-16 \quad BFACTC &= (1.0/3.5)**2 \\
4-256 \quad BFACTC &= SQRT(BFACTC)/3.8 \\
16-256 \quad BFACTC &= 1.0/(3.5*3.8) \\
256-256 \quad BFACTC &= (1.0/3.8)**2
\end{aligned}$$

These values are simply empirical and are based on our current knowledge of visibility amplitudes from Emmanuel's early test observations. These are described in some detail below. The complicated auto-correlation corrections involving FFTs are not done when the antenna has more than four levels and such data are simply scaled by the same factor used for cross correlations.

in the April 23 version `FITLD` was also changed so that `DIGICOR = 0` means that the scaling described above is done but the FFT operation on auto-correlation data is omitted. If the Van Vleck correction is done in the correlator, the FFT operation is omitted.

3 Calibration

Later calibration steps are also uncertain. `ACCOR` always uses all channels. The bandpass procedure `VLBABPSS` used by itself or in the `VLBARUN` and `VLBAPIPE` pipeline scripts always normalizes the bandpass amplitudes and zeros the average phases using all channels in powers rather than voltages. `ACSCL` uses the `ICHANSEL` adverb when run directly. The `VLBAUTIL` procedure `VLBAAMP` uses the default value of this adverb which selects the inner three quarters of the channels. The bandpass shape of VNDA looks much nicer than the current ones, with little roll off other than very close to the edges. But apparently the channels in the roll-off regions are mysteriously noisy, while the source remains detectable in the roll-off regions with RDBE. This may change what is appropriate for these calibrations.

4 Testing

TV018 comes in three parts and three antennas are missing. The normal mode (RDBE) is used for stations FD, KP, LA, NL, and SC (antennas 2, 4, 5, 7, and 10) and VNDA is used for OV and PT (antennas 8 and 9). The data were taken with multiple interleaving scans with VNDA in 2-bit, 4-bit, and 8-bit (4-, 16-, and 256-level) modes. The 4 and 8 bit data sets have `VANVLECK = 1`, but, curiously, the 2-bit data set has no `VANVLECK` keyword.

The scaling factors computed and used by the new version of FITLD were[1]

data set	levels	Cross c	Auto C
2B	4-4	0.64167	0.56629
4B	4-4	0.56629	0.56629
	16-16	12.25	12.25
8B	4-4	0.56629	0.56629
	256-256	14.44	14.44

The visibility amplitudes are then scaled by task ACCOR by factors such that the average auto-correlation is 1.0. With pre-VNDA data sets and DIGICORR = 1, this factor is normally rather close to one. The ACCOR voltage gains for these data sets retain that property. The ACCOR *voltage* gains for antenna 4 (RDBE) and antenna 8 (VNDA) are listed below for both values of DIGICOR. Note that the RDBE antennas (including 4) use 4 levels in all 3 data sets.

data set	DIGICOR	gain(4)	gain(8)
2B	+1	0.993	0.993
4B	+1	0.993	1.004
8B	+1	0.993	1.02
2B	-1	0.7422	0.7422
4B	-1	0.7422	3.513
8B	-1	0.7422	3.888

Note that the gains for all antennas are near 1.0. The scaling for the RDBE antennas has been used for a long time and hence this is expected. The gains for the VNDA antennas are close to one in the 4- and 8-bit modes. They were made to be that way with the empirical 3.5 and 3.8 factors. The 8-bit mode in particular may be used with lower gains allowing more head-room for RFI, so the 3.888 factor may not always be appropriate. ACCOR will correct this in any case.

The plots on the following six pages show the resulting visibilities. The data are averaged over all spectral channels and IFs. The outermost spectral channels (numbers 1 and 80) in each IF were flagged to avoid odd edge effects. Due to the large delays in this data set, FRING was required as well as ACCOR. VLBAMPCL was used on a single scan average to eliminate the gross delays. Then VLBAFRNG was used over the whole data sets with SOLINT 30 seconds and only spectral channels 11-70. The colors in all plots indicate the data set with yellow for 8B, green for 4B, and pink for 2B. Figure 1 shows auto-correlation data from a VNDA antenna calibrated by ACCOR and FRING with DIGICOR set to +1. Figure 2 shows the same except DIGICOR was -1. The agreement is pretty good, but given that ACCOR is supposed to force these data to 1.0, that is not surprising. Figure 3 shows the data from non-VNDA baseline 2-4, calibrated by ACCOR and FRING with DIGICOR set to +1. The agreement is quite good as one would expect from this much used mode. Figure 4 shows the same thing but with DIGICOR set to -1. The 2-bit data clearly need the Van Vleck digital correction. Figure 5 shows the VNDA baseline after ACCOR and FRING with DIGICOR set to +1 while Figure 6 is the same except DIGICORR was -1. The agreement with DIGICOR = 1 is quite good and again the 2-bit data really need the Van Vleck correction done by DIGICOR.

Unfortunately, things are not quite this simple. The Fourier process applied to 4-level autocorrelations appears to introduce an oscillation not present in the data with DIGICOR set to -1. Figures 7 and 8 show the autocorrelation spectra from antenna 4 (always 4 level) and and antenna 8 (4 and 16 level). The DIGICOR = +1 data are plotted in yellow and the DIGICOR = −1 data are plotted in green. They agree rather closely only in the data from antenna 8 and the 4B data set. Antenna 8 does not have the FFT operation performed on it in this data set.

In case the number of spectral channels affects this result, VNDA data set TV106G was read with the new version of FITLD. This data set has all antennas at 4 levels, so the FFT is done for all antennas when DIGICOR = +1. Figure 9 shows that the oscillations are present whenever the FFT is done.

[1]We do not consider baselines in which the antennas have a different number of bits (other than older 2 levels with 4). They exist in the current data but are unlikely to be relevant in the long run.

The obvious correction for this problem would be to skip the FFT step always. Unfortunately, with other data sets one would conclude that the FFT improves things. Figure 10 shows spectra for antennas 4 and 8 with the yellow for `DIGICOR = +1` and green for `DIGICOR = -1`. The oscillations in the spectrum for antenna 4 are pretty bad in both. However, for antenna 8 the FFT clearly improves things. The `FITLD` help file advises the use of the FFT in all cases except when very strong spectral lines are present and no zero padding was used in the correlator. Zero padding was not used in all of the data sets examined here.

The FFT subroutine used by `FITLD` is called `FOURG` and is based on a published algorithm by Norman Brenner (MIT). I was not able to access the reference in IEEE Audio and Electroacoustics from June 1967 and June 1969. I was able to get the code of another version of `FOURG` by Brenner in Methods Of Experimental Physics, Volume 12, Part C. This `FOURG` is quite different from the one now used by *AIPS*. I got it to work (sort of) but it does not appear useful. `FOURG` is used widely in *AIPS* and generally appears to work. It would seem that we should simply omit the FFT in `FITLD` as is done by the revised `DIGICOR = 0` default. Users who care about the shape of the auto-correlation spectra can examine their 2-bit data and decide whether the FFT correction is needed.

5 Summary

The scaling in `FITLD` does two unique things: (1) apply the Van Vleck correction if it was not done by DiFX and (2) make a subtle change to the shape of auto-correlation functions. It is not clear why the latter is done, whether it should also be done for 16- and 256-level data, and whether it is being done correctly. The shape of the auto-correlation spectrum is used in the rare case that `BPASS` fails on cross-correlation data. Other than that I can't think of any use for it so long as the average is not changed (which seems to hold true). All the rest of the various scaling done in `FITLD` is simply to allow `ACCOR` to have gains near 1.0. That really is cosmetic and simply a matter of user training.

I believe that VNDA is not fully ready to offer to outside users but it may be close. We do not understand how to scale these data. Furthermore, we cannot as yet tell if any of these data are correctly calibrated. The source observed, while strong, was not a point source. Emmanuel has been awarded test time to observe a small diameter source with the VLA (24 antennas) and the VLBA plus Y3. The VLBA antennas will record with RDBE and, where available, simultaneously with VNDA. That should allow us to study the calibration issues in detail. The effects of the difference in bandpass shape from DDC to VNDA may also cause issues in absolute calibration.

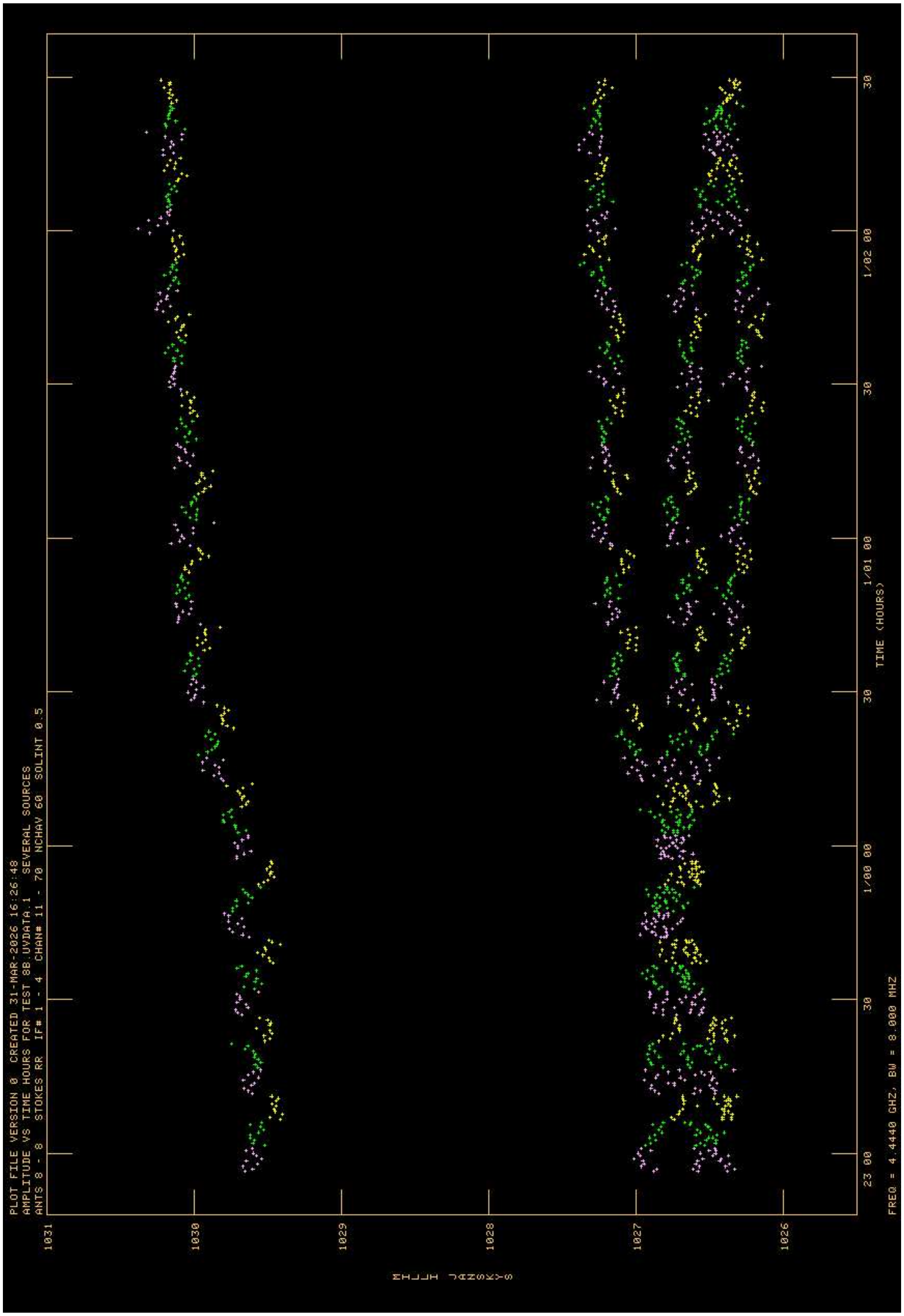

Figure 1: Auto-correlation visibilities from antenna 8 after calibration with `ACCOR` and `FRING`. `DIGICOR` was +1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

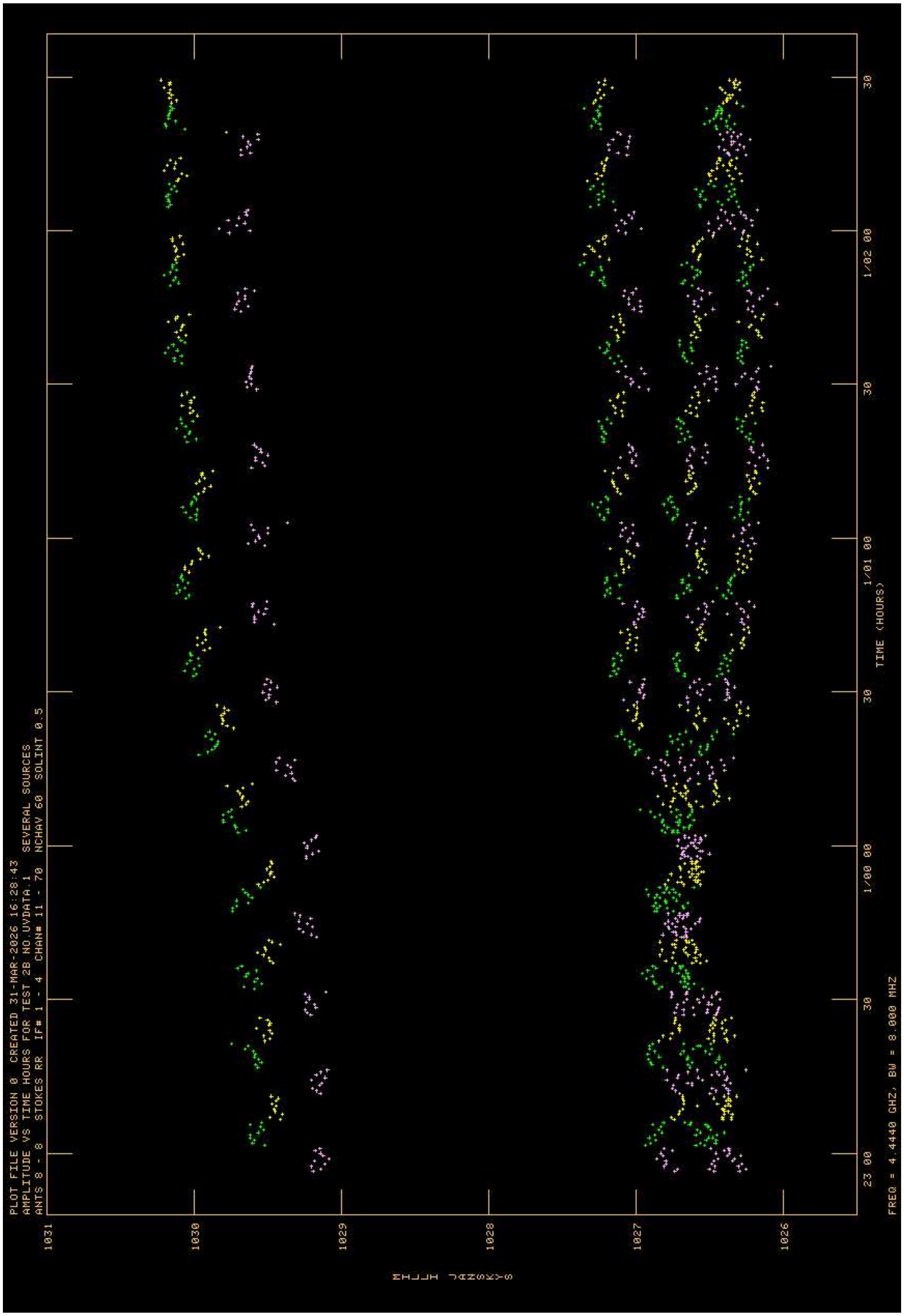

Figure 2: Auto-correlation visibilities from antenna 8 after calibration with `ACCOR` and `FRING`. `DIGICOR` was -1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

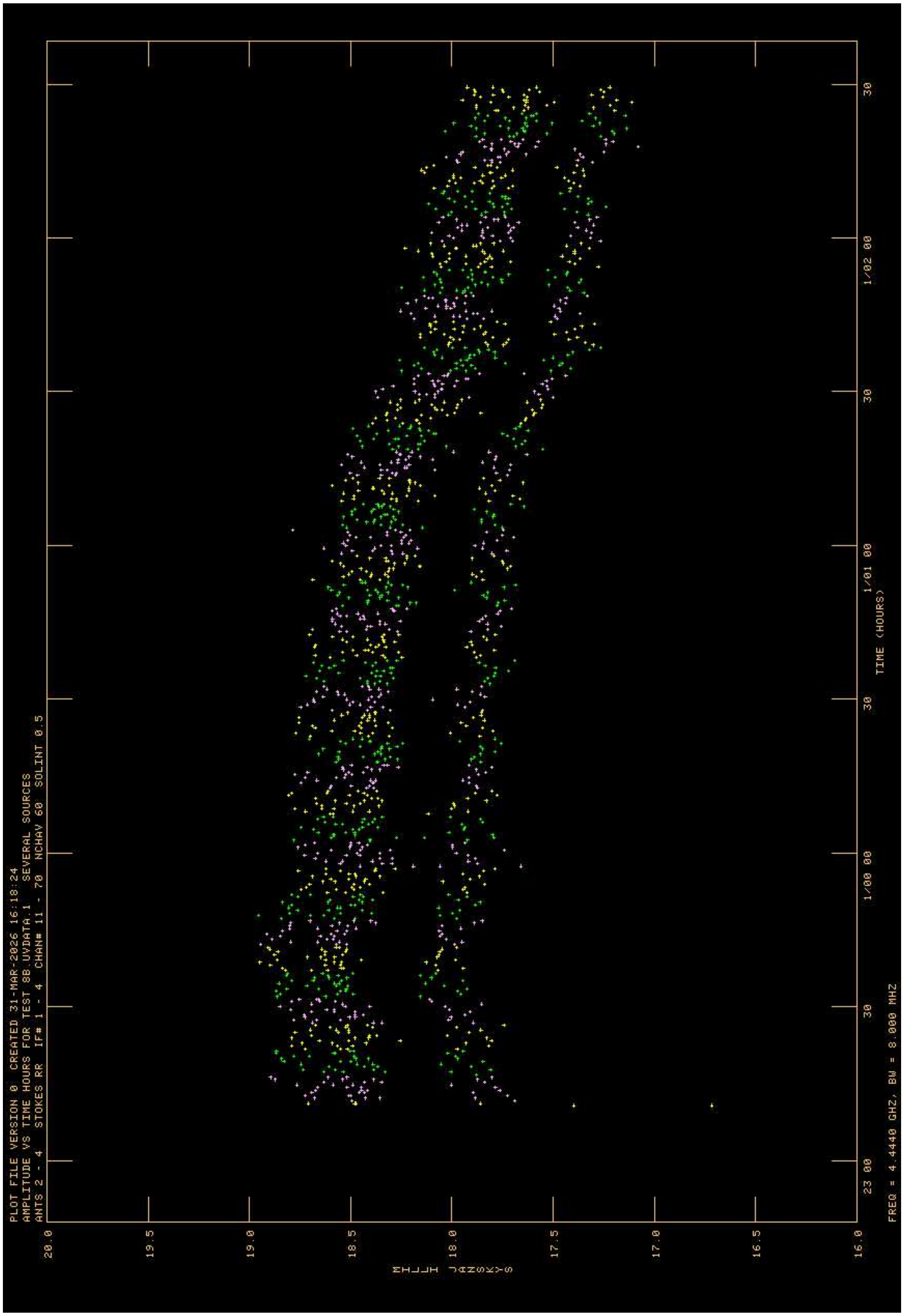

Figure 3: Visibilities from non-VNDA baseline 2 - 4 after calibration with `ACCOR` and `FRING`. `DIGICOR` was +1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

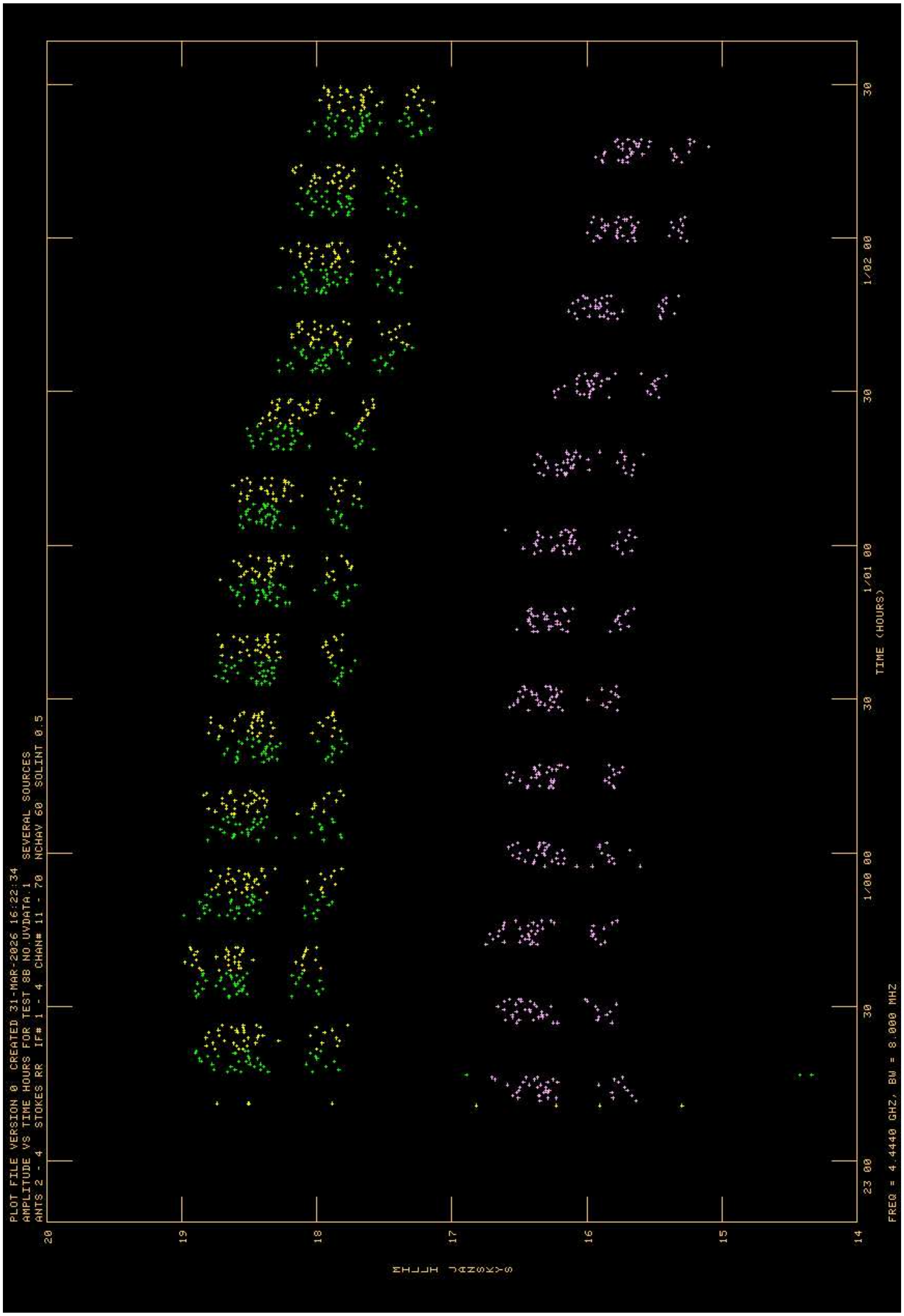

Figure 4: Visibilities from non-VNDA baseline 2 - 4 after calibration with `ACCOR` and `FRING`. `DIGICOR` was -1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

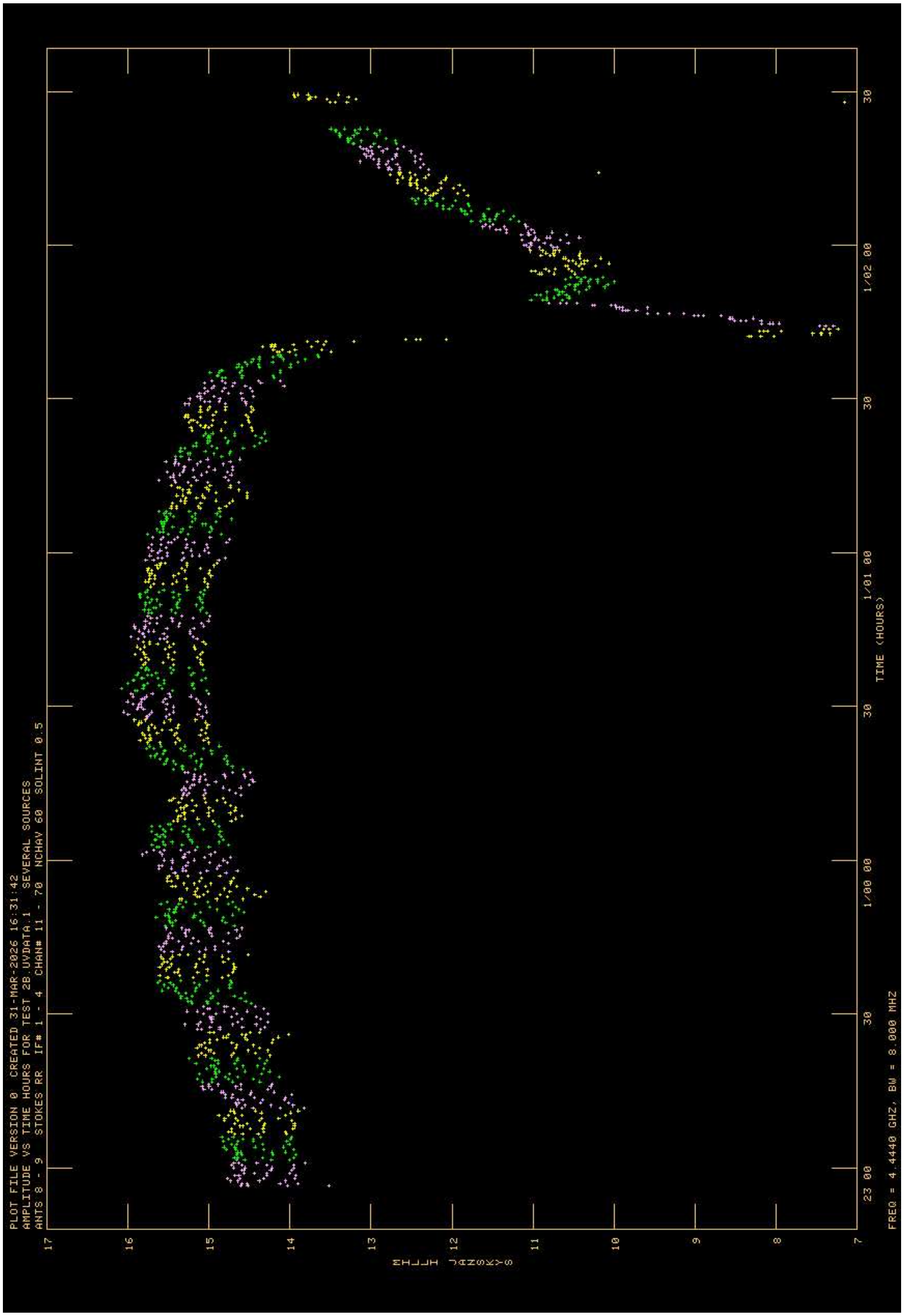

Figure 5: Visibilities from VNDA baseline 8 - 9 after calibration with `ACCOR` and `FRING`. `DIGICOR` was +1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

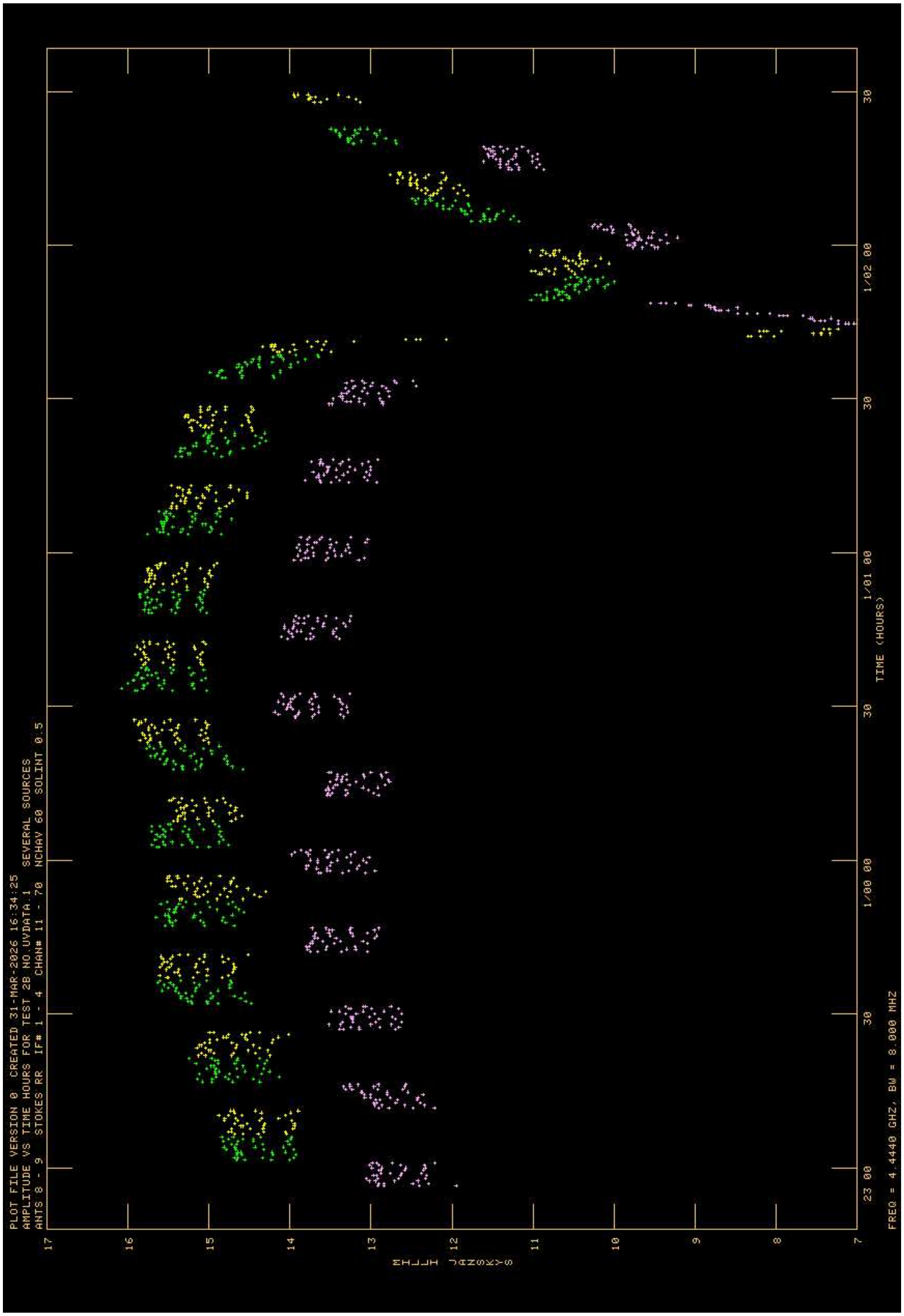

Figure 6: Visibilities from VNDA baseline 8 - 9 after calibration with `ACCOR` and `FRING`. `DIGICOR` was -1. Yellow data are from 8B, green are from 4B, and pink are from 2B.

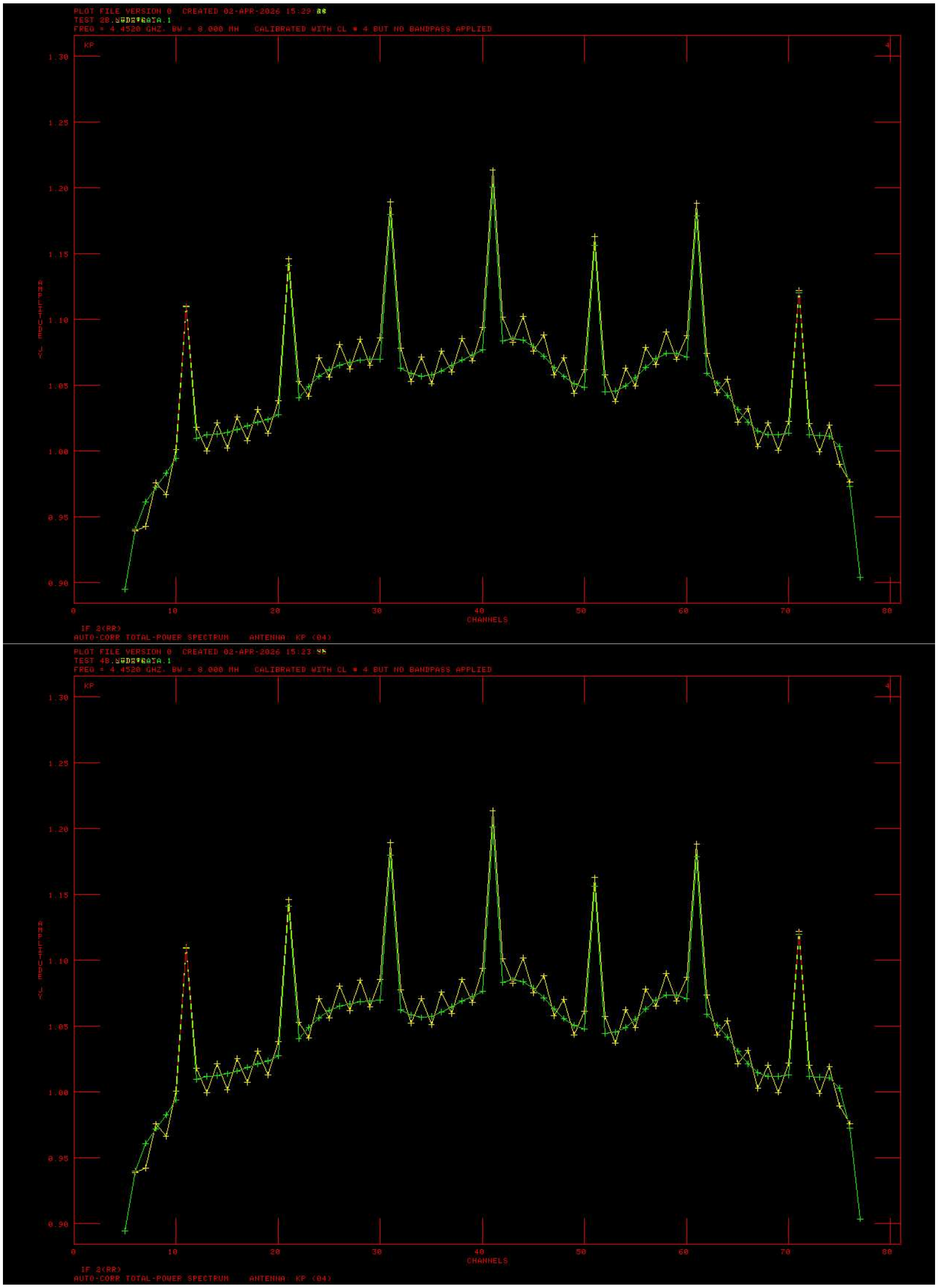

Figure 7: Autocorrelation spectra for antenna 4. Yellow data are are from `DIGICOR = +1` while green are from `DIGICOR = -1`. The top plot is from the 2B data set while the bottom is from the 4B data set. Since antenna 4 is always 4-level, the FFT is always applied when `DIGICOR = +1`.

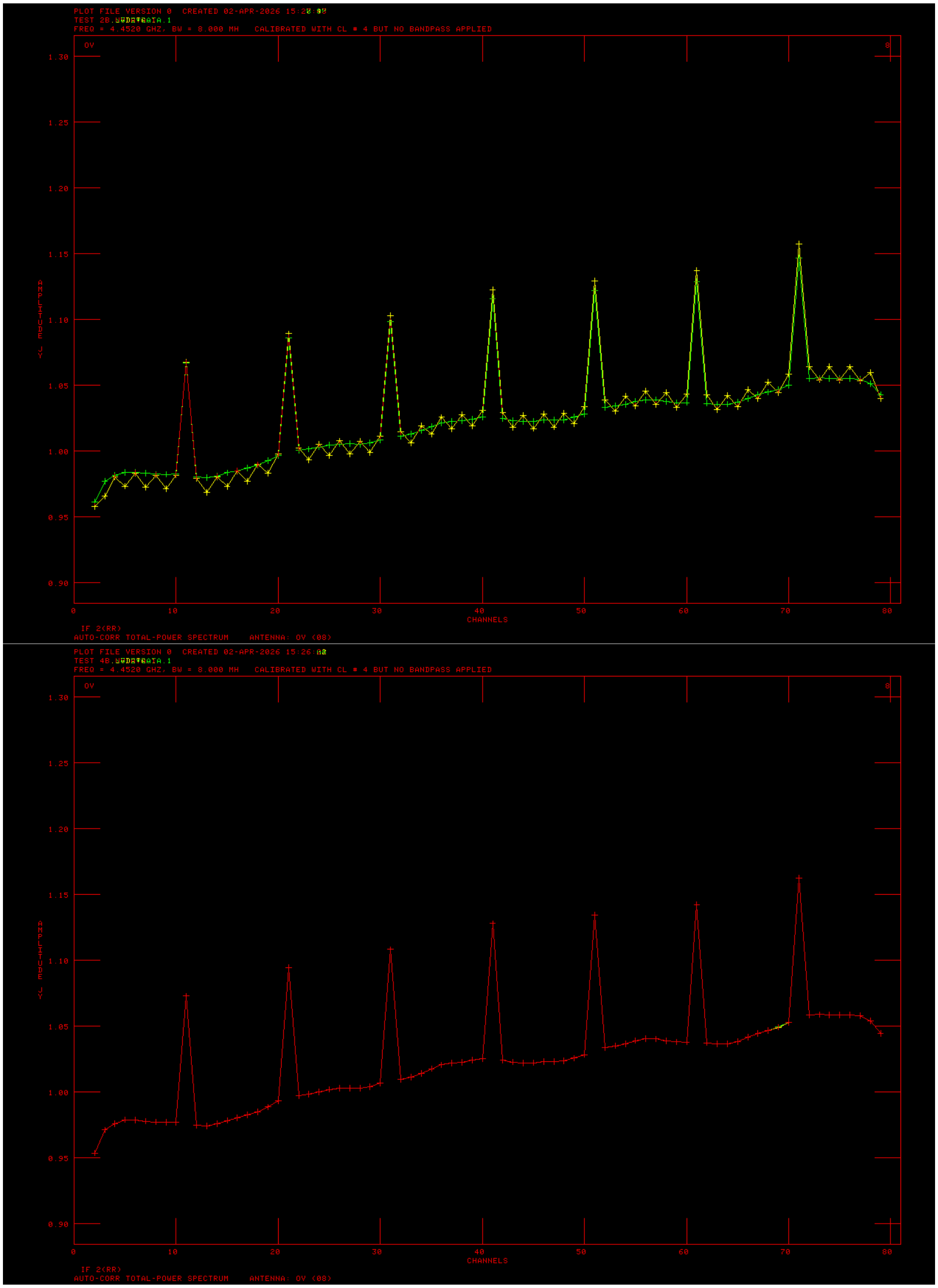

Figure 8: Autocorrelation spectra for antenna 8. Yellow data are are from `DIGICOR = +1` while green are from `DIGICOR = -1`. The top plot is from the 2B data set while the bottom is from the 4B data set. Since antenna 8 is 4-level in 2B, the FFT is applied when `DIGICOR = +1`. The FFT is not applied when antenna 8 is in 16-level.

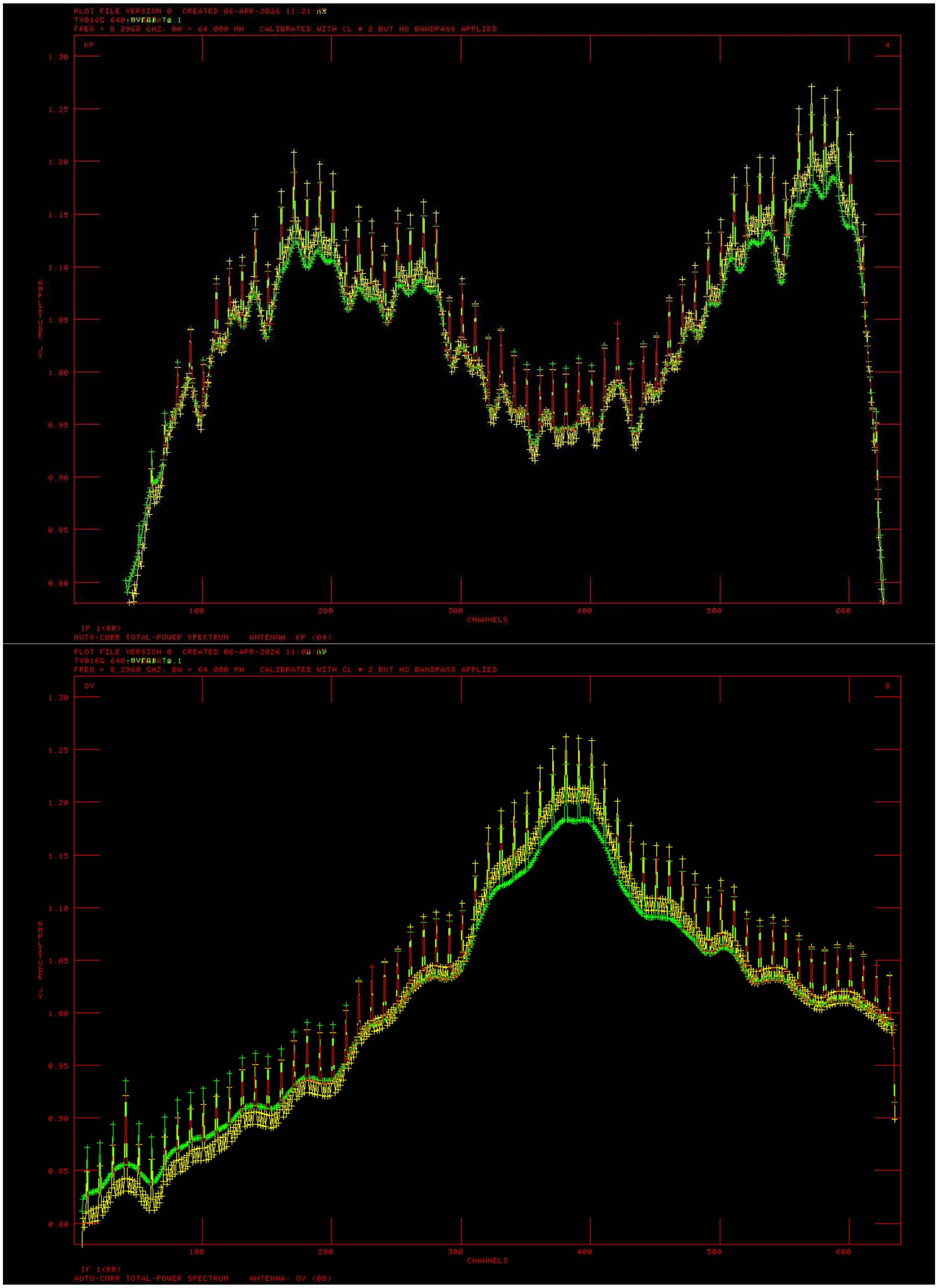

Figure 9: Autocorrelation spectra from TV016G observed 6 April 2026 for antennas 4 (top) and 8 (bottom). Yellow data are are from `DIGICOR = +1` while green are from `DIGICOR = -1`. The FFT causes oscillations since both antennas are at 4 levels. The number of channels does not affect this result apparently.

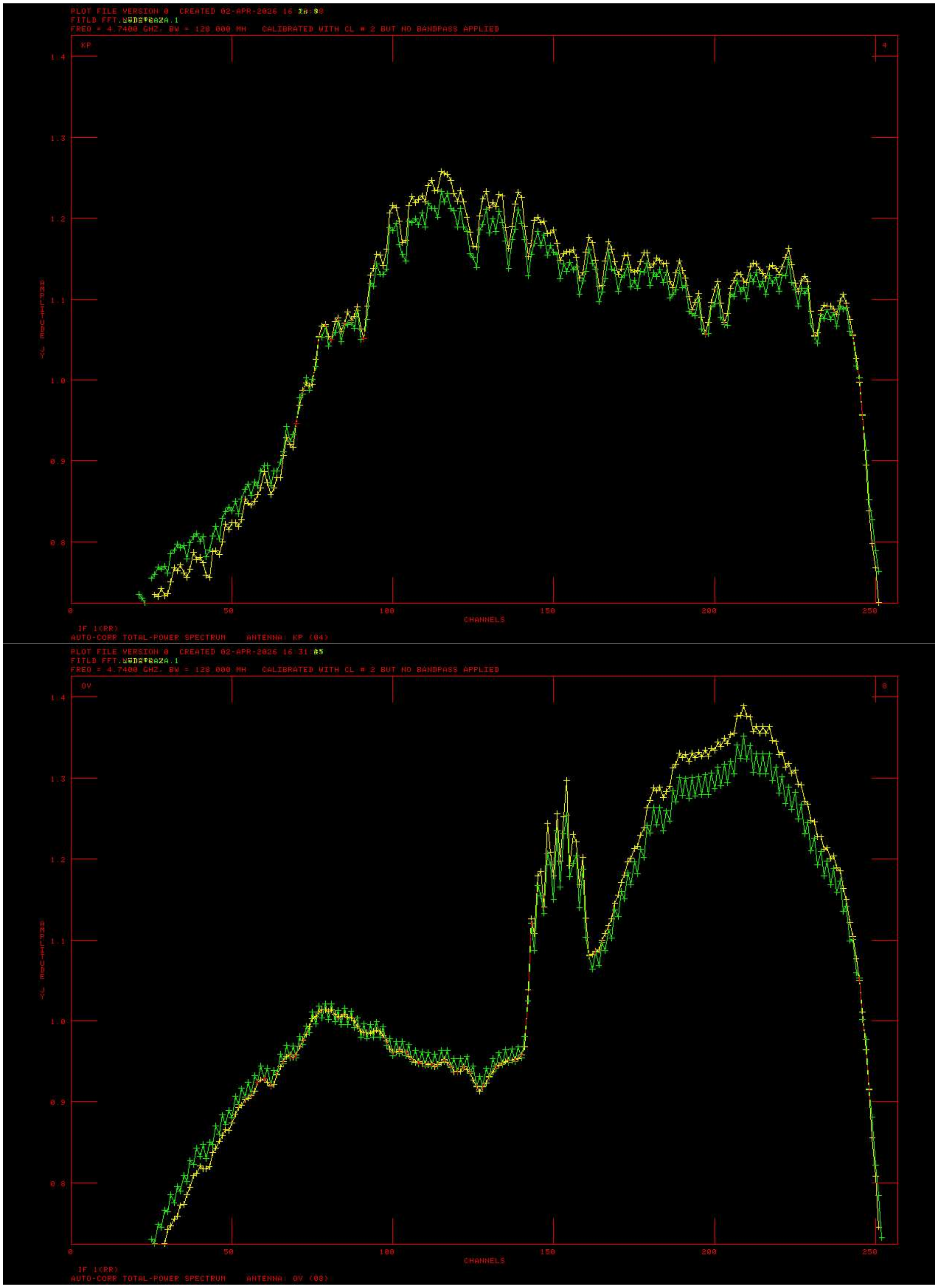

Figure 10: Autocorrelation spectra from BF138 observed 12 March 2025 for antennas 4 (top) and 8 (bottom). Yellow data are are from `DIGICOR = +1` while green are from `DIGICOR = -1`. The FFT causes changes in the spectral shape and appears to improve the oscillations at least for antenna 8.