Draft version December 4, 2019
TEX is now
Typeset using LATEX default style in AASTeX62

Spectral structure in the HERA prototype.

Aaron Ewall-Wice[1]

[1] *Jet Propulsion Laboratory*
4800 Oak Grove Ave.
Pasadena, CA 91109 USA

ABSTRACT

We investigate spectral structure in uncalibrated cross and auto-correlations in the new HERA RF chain from observations on **JD 2458465**. We find that two classes of artifacts may limit the ability of the new system to achieve its science goals. **(a)** super-horizon structure that appears most severe and is likely associated with a divot at 55 MHz. **(b)** Sharp, spike-like artifacts caused by the charge pump in the RF over-fiber system and appear in the cross correlations. These structures should be understood and brought below a carefully justified, validated, and documented acceptable level before the build out of the 350-element HERA system in order to guarantee a detection of redshifted 21 cm fluctuations in the early Universe.

1. EXECUTIVE SUMMARY

We investigate spectral structure in the *uncalibrated* auto and cross-correlations of HERA-6 using data from day **2458465** which was used since it was the only day in which all six antennas were present in the data. Our investigation reveals that

1. There exists intrinsic high-delay structure (not originating from side-lobes) that is observable in both the auto and cross correlations. This structure is at roughly **30 dB at 750 ns** and could severely hamper a detection of 21 cm fluctuations.

2. **In autocorrelations** This structure is observed primarily when the "divot" at 55 MHz is present in the band.

3. **In autocorrelations** When the "divot" is not present, the structure dissapears below a systematic floor set up by charge pump artifacts in the autocorrelations.

4. **In the cross-correlations**, similar super-horizon power is present. However, this power does not change significantly when the divot is excluded, leading me to believe that it may be caused by a different problem, independent of the "divot". Residuals in cross correlations excluding the divot appear to be maximized at roughly 70 MHz. We dub the collective large-delay features in the cross correlations as the "cross wings". Super-horizon emission might be caused by

   (a) Dish-dish reflections (over the air cross-talk).

   (b) A resonance in the 40-meter power cable.

   (c) (RFI flagging) Sidelobes of a 2600 ns reflection (we eliminate this possibility).

   We will show in a comparison between a long coherent and incoherent ingetration that the "cross-wings" fring slowly.

5. There is some evidence that the sharp frequency spikes originating in the charge pump driving the RF over fiber laser are present in the cross correlations. This could happen if

Corresponding author: Aaron Ewall-Wice
aaronew@berkeley.edu

(a) signal generating the spikes is able to leak into other signal chains through over the air or electronic cross-talk.

(b) The signal generating the spikes (unlike the sky-signal) has a phase coherence time that is greater than the integration time in the correlator.

Further investigation is needed to determine if the charge pump spikes integrate down in the autocorrelations.

We enumerate super-horizon features in Table 1.

In the remainder of this memo, we will demonstrate these take-aways using inspections of data from the new HERA-6 system. The organization of this memo is as follows. In § 2, we introduce our linear delay-filter and demonstrate the efficacy of our analysis technique for inspecting spectral structure on simulated data. In § **??**, we inspect autocorrelations and discuss evidence for the charge-spike floor and spectral structure arising primarily from the "divot". In § **??**, we inspect cross-correlations and find that structures hidden in the autocorrelations (potentially due to the enhanced amplitudes of the charge spikes) are present in the cross-correlations at identical frequencies to the auto-correlations. We summarize our conclusions in § **??**.

2. DELAY-FILTERING AND VALIDATION WITH SIMULATIONS

2.1. *The Linear Delay Filter*

In this analysis, we use a new method for removing foregrounds that we dub a "linear delay filter". This filter is implemented as a matrix which is applied to our data through a matrix-multiplication

$$\mathbf{z} = \mathbf{R}\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is a raw visibility, $\mathbf{R}$ is our filter matrix, and $\mathbf{z}$ is our filtered data. We will implement a delay-filter in this way by setting

$$R_{ij}^{-1} = F_i F_j \left[\epsilon \frac{\sin\left[2\pi(\nu_i - \nu_j)\tau_f\right]}{2\pi(\nu_i - \nu_j)\tau} + \delta_{ij}\right] \tag{2}$$

where τ_f is a delay that we manually set along with a regularization factor, ϵ. F_i is a vector that is equal to unity at unflagged channels and equal to 0 at flagged channels. For a fixed baseline, we typically choose $\tau_f = \tau_b + \tau_D$ where τ_b is the baseline light crossing time and τ_D is the dish-diameter light-crossing time. Note that if we were to interpret this as a data-covariance, it would be a matrix that is diagonal in delay-space and very small for delays outside of the wedge. One can show (by approximating δ_{ij} as a Dirac-Delta function and performing the continuous equivalent of matrix multiplication), that when no RFI flagging is present, this yields

$$R_{ij} \approx \delta_{ij} - 2\tau_f \frac{\sin\left[2\pi(\nu_i - \nu_j)\tau_f\right]}{2\pi(\nu_i - \nu_j)\tau_f} + \epsilon 4\tau_f^2 \frac{\sin\left[2\pi(\nu_i - \nu_j)\tau_f\right]}{2\pi(\nu_i - \nu_j)\tau_f} \tag{3}$$

We can choose ϵ to be very small and ignore the last term in equation 3 in our analytic calculations (for pedagogical purposes). If we wanted to re-normalize, as is done in the final step of a power-spectrum estimate, we would keep track of this term.

$$R_{ij}x_j \approx \int d\nu \left[\delta_D(\nu_i - \nu) - 2\tau \frac{\sin\left[2\pi(\nu_i - \nu_j)\tau\right]}{2\pi(\nu_i - \nu_j)\tau}\right] x(\nu) \tag{4}$$

Thus the action of R_{ij} is to subtract a copy of x_j convolved with wedge kernel – i.e., to subtract off the wedge! The action of this filter is very simple to the delay-filter discussed in Parsons & Backer (2009) except that it is completely linear and as a result is more stable and errors can more easily be propagated through any analysis pipeline that uses it, perhaps even analytically. The downside to this filter is that it does not restore clean components and rather only leaves us with the residual of structure subtracted from within the filter width, τ_f.

2.2. *Validation with Simulations*

We validate the efficacy of our delay-filter by applying it to simulations of a well-understood sky with flagging identical to our data set.

To simulate ideal visibilities; for each redundant baseline group we draw realizations of a complex Gaussian random field with a Covariance matrix of an isotropic sky dominated by point sources.

$$C_{ij}^{pt} = \sigma_s^2 \int \frac{d\Omega d\nu d\nu'}{\sin\theta} \Phi_i(\nu)\Phi_j(\nu')A(\hat{s},\nu)A^*(\hat{s},\nu')e^{-2\pi I b\cdot\hat{s}(\nu-\nu')/c} e^{-\frac{1}{2}\sigma_\alpha^2 \log^2(\nu\nu'/\nu_0^2)} \left(\frac{\nu\nu'}{\nu_0^2}\right)^{-\alpha_0} \tag{5}$$

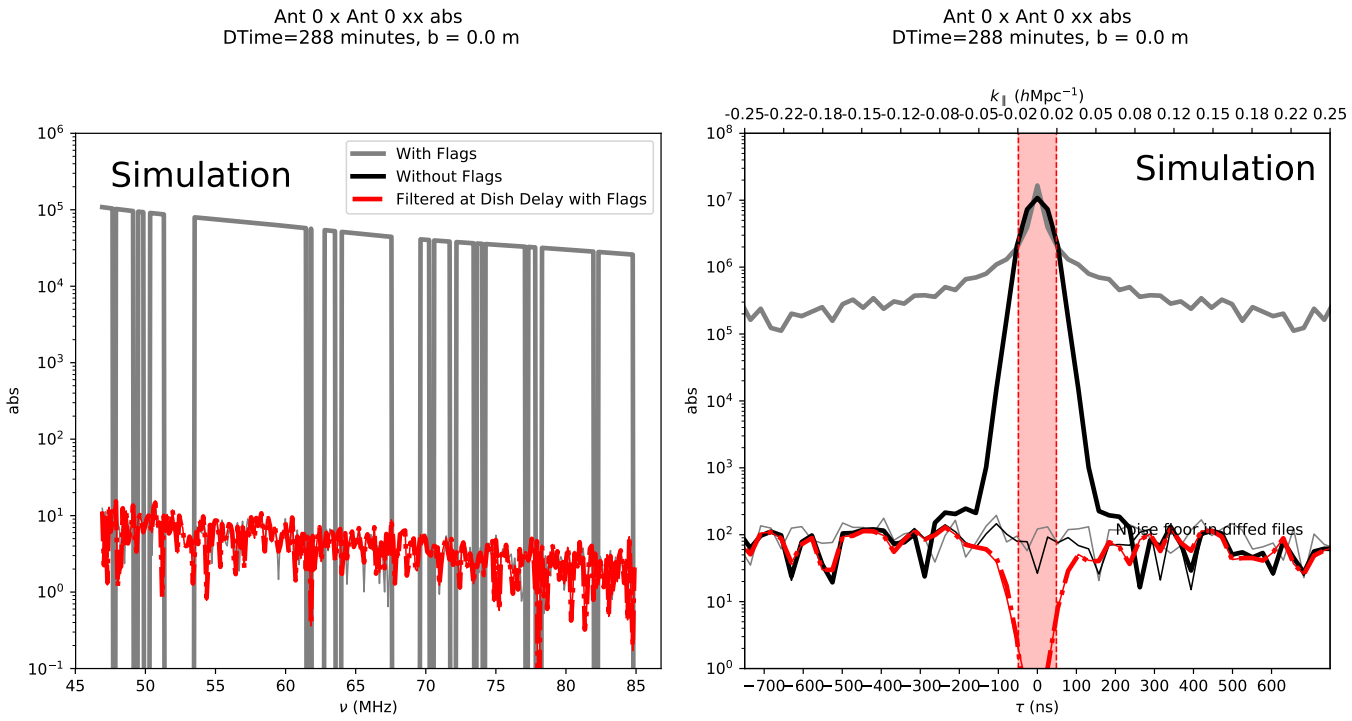

Figure 1. We verify our analysis framework using a simulated data set where autocorrelations are drawn from a Gaussian distribution with a covariance given by equation 5 and mean given by equation 7. A direct Fourier transform of the autocorrelations with RFI flagging (thick grey line) yields high side-lobes which are effectively removed when applying the delay filter in equations 1

where

$$\sigma_s^2 = \int_0^\infty \frac{dn}{dSd\Omega} S^2 dS \tag{6}$$

The mean of each visibility is given by

$$\langle V_i \rangle = \int dS \frac{dN}{dSd\Omega} \int d\Omega A(\hat{s}) e^{2\pi I \mathbf{b}\cdot\hat{\mathbf{s}}/\lambda_i} \tag{7}$$

All time steps and channels have thermal noise added with a covariance of

$$C_{ij}^N = \delta_{ij} \left(\frac{2k_B}{\lambda_i^2} \frac{T_{rx} + T_{sky}(\lambda)}{\sqrt{2\Delta\nu\tau_{\rm acc}}} \int d\Omega A_i(\hat{s}) \right)^2 \tag{8}$$

where $\tau_{\rm acc}$ is the accumulation time in each channel and $\delta\nu$ is the channel width.

We add a noise bias to the mean of our autocorrelations using the square root of equation 8 with $\sqrt{2\Delta\nu\tau_{\rm acc}} = 1$

For this ideal data set, we assume Φ_i and Φ_j, our "channelization functions" are given by Dirac Delta-functions. For beams, we simulate Airy-beams with diameters of 14.6 meters.

We demonstrate the application of our filtering pipeline to a simulated autocorrelation in Fig. 1. With RFI flagging present, the unweighted delay-transform has large side-lobes that are two orders of magnitude above the noise floor (thick grey lines). We compare the noise floor with filtering of the RFI flagged simulation (thick red line) with the Fourier transform of the simulation with no flagging (thick black line) and find that they are in decent agreement at the noise floor.

3. DATA

Having validated our analysis on simulations, we now turn out attention to the data.

We focus on four hours of data from the Julien day 2458465 between LSTs $\approx$ 5 to $\approx$ 9 hours. For the sake of simplicity, we focus on two antennas (0 and 1). Since there are substantial changes in visibilities over this time-period, we integrate our data incoherently by taking even and odd conjugate products between adjacent 10 second time-steps.

We show illustrative waterfall plots of our auto and cross correlations in Fig. 2.

3.1. *Noise from Diff Files*

The ultimate goal of our analysis is to determine in which parts of delay-space our data is consistent with thermal noise. To estimate the thermal noise level, we use diffed data from the HERA correlator which subtracts even and odd data integrations at high cadence and should be entirely noise like. We demonstrate that the highest delays of our auto and cross correlations are consistent with the noise levels in the diff files in Fig. 3 where we compare the cleaned delay-transform of autos/cross correlation from a single 10 second time-step with data from the equivalent diff file. At high delays, the overall amplitude of the diffed data agrees with the non-diffed data, indicating that the residuals are indeed noise-like.

If we average incoherently over four hours, to beat down the noise, the situation changes somewhat. We show the averaged delay transformed visibilities in Fig. 4. In this figure, the autocorrelations never reach down to the level of the integrated diff files, indicating the presence of a biasing systematic. The cross correlation does reach down to the noise, but only after ≈ 1000 ns. Anything that can be detected by a single baseline after four hours outside of $k \approx 0.15 h\text{Mpc}^{-1}$ ($\lesssim 300$ ns for these frequencies) is a problem for 21 cm. The rest of this memo will focus on better understanding the sources of these super-horizon biases in our correlation products.

3.2. *AutoCorrelations*

We demonstrate the delay-structure of an auto-correlation in Fig. 5. We apply our filter to the autocorrelation out to 300 ns (purple dashed lines) which is roughly 0.15 $h\text{Mpc}^{-1}$ at $z \approx 15$ and compare to an `aipy` delay-clean within the same region (black solid line). The reason behind our 300 ns filter-width choice (rather then simply the horizon delay) is that Side-lobes from the significant RFI flagging causes power from super-horizon emission to contaminate greater delays then the intrinsic power. We arrived at 300 ns by testing for convergence of large-delay ($\gtrsim 1000$ ns) power after applying filters of increasing width and found 300 ns to be a reasonable width where super-horizon power converges and is not be dominated by low-delay side-lobes.

Both our linear delay filter and the 1d clean remove power to a similar level and are in decent agreement outside of the filter region.

Inspecting Fig. 5, there appears to be a floor at high-delays beyond $\approx 750\,ns$. What is the source of this intrinsic floor, which is roughly an order of magnitude above the noise level, and roughly 40 dB down from the DC? From the frequency domain, we can see that this structure is caused by spike-like features every ~ 2.5 MHz which are known to be caused by the switched power supplies for lasers responsible for converting RF signals into optical over fiber. We confirm that these spiked features in the residuals of Fig. 5 are indeed sourced by the RFoF switching noise bias through careful comparison of the non RFI flagged residuals with the un-flagged auto-correlations (Fig. 6). Each spike in the auto residual corresponds with a spike in the total auto-correlation.

While the systematic floor clearly prohibits a detection of 21 cm fluctuations, there is also decaying structure that rises above the systematic floor between $\sim 100 - 750$ ns (out to $k_\parallel \sim 0.2 h\text{Mpc}^{-1}$. At this LoS scale, power is down by ~ 35 dB which is a factor of ~ 100 above the acceptable level level for a 21 cm detection. In the frequency domain, the residuals are maximized at the location of the divot, suggesting that it is associated with the large-delay wings in the autos.

We verify that the super-horizon structure is indeed sourced by the divot by excluding that part of the band from our delay transform (Fig. 7). When the divot is excluded, the large-delay structure disappears below the charge spikes.

We show a direct comparison between a delay-transformed and 1d cleaned auto-correlation in Fig. 8. Wings clearly extend out to high delays when the divot is present while they vanish when the divot is excluded.

Our take-away from investigating the auto-correlations is that there is physical large-delay structure out to at least $\gtrsim 750$ ns which is at 35 dB below the foregrounds – roughly two orders of magnitude too large for a viable EoX/H detection. When the divot is excluded from the band, this structure is dramatically reduced but to what extent is still unknown due to the switching noise emitted by the RF over fiber laser power supply. The contamination from the divot in antenna 1 is worse than antenna 0.

4. CROSS-CORRELATIONS

We next inspect residual structure in cross-correlations. We plot filtered visibility (incoherently averaged over ~ 5 hours) in Fig. 9. The residual structure present in the autocorrelations is also present in the crosses and peaks at the

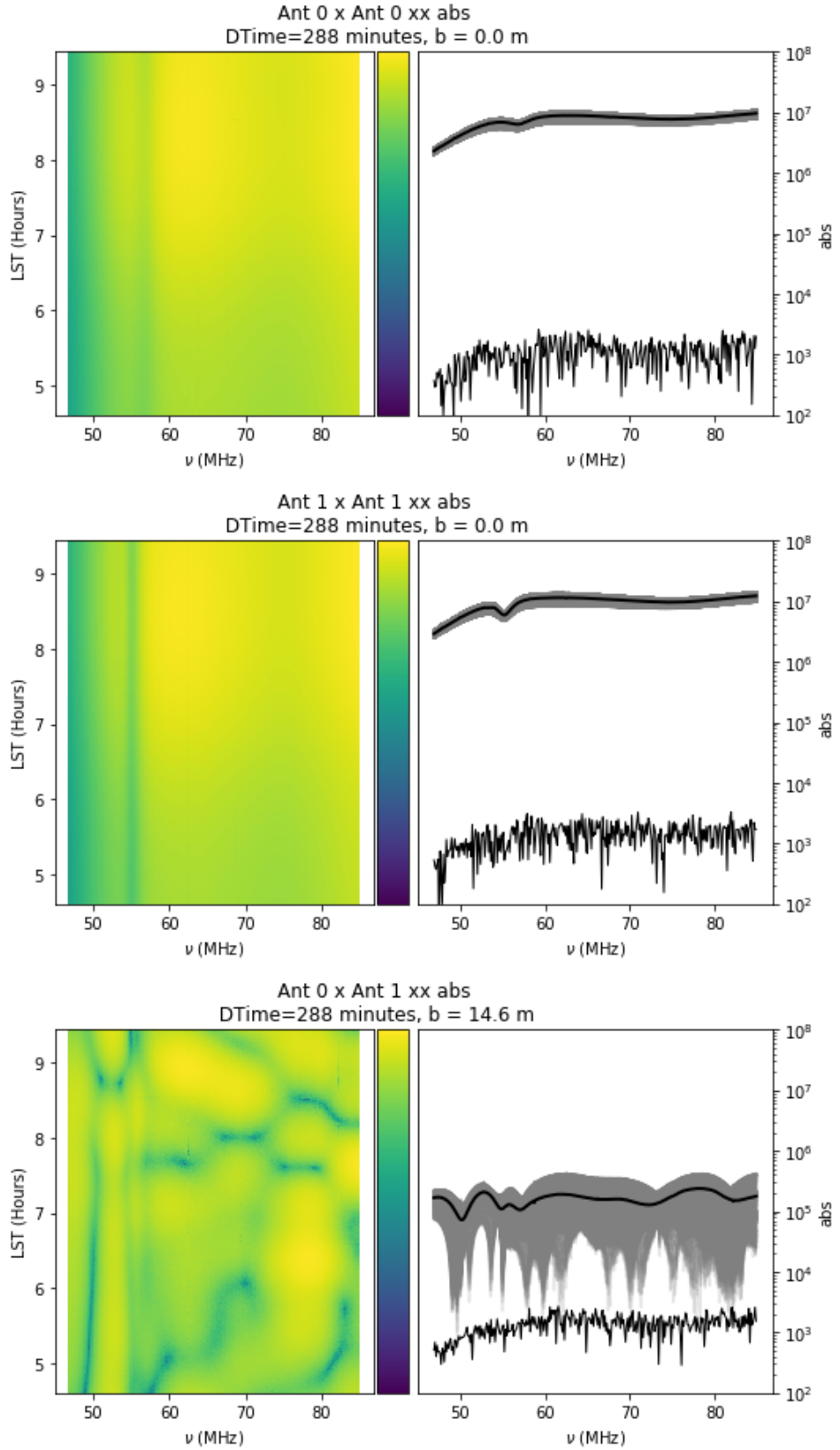

Figure 2. Waterfall plots illustrating the time evolution of auto and cross correlations over the four hours of LST studied in this memo. While auto-correlations are stable in shape and roughly constant in amplitude at the 10 % level, the cross-correlation pass through a large number of fringes, necessitating incoherent averaging. Diffed noise levels are thin black curves.

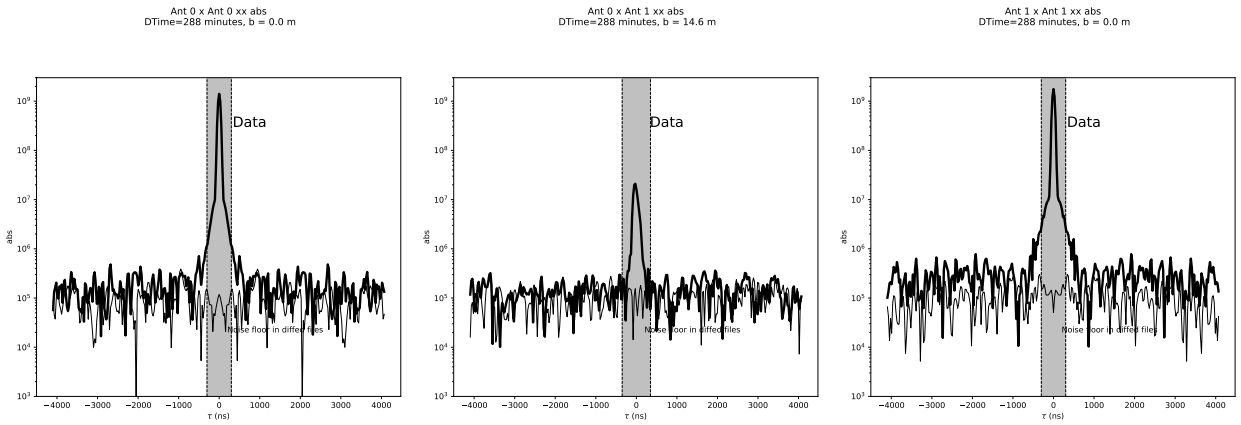

Figure 3. Delay-cleaned and restored correlation products from two antennas over a single 10 second integration (thick black curves) and the noise-dominated "diffed" files (thin black curves). At large delays the diffed files agree very well with the non-diffed data – indicating that the non-diffed data is noise-like in these regions of delay space. Grey shaded regions indicate where cleaning has been performed.

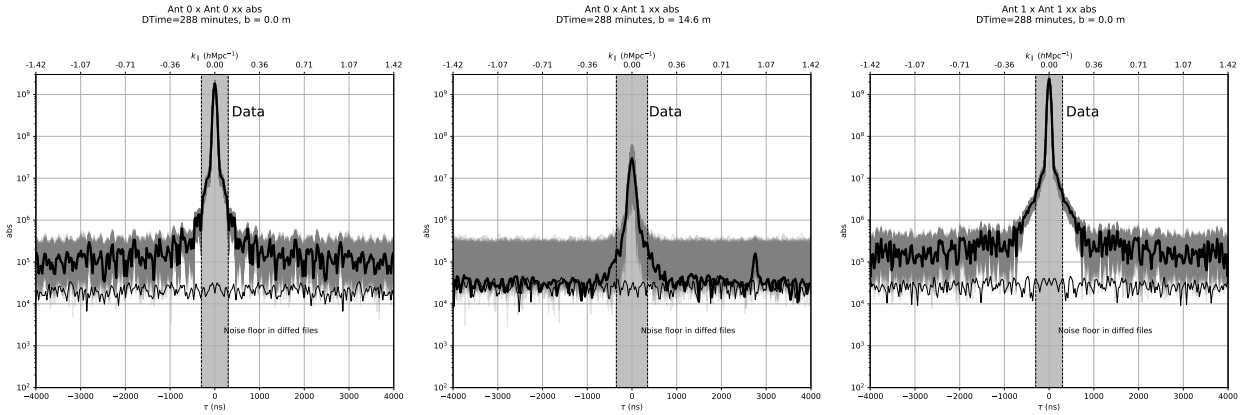

Figure 4. The same as Fig. 3 except now the thick and thin black curves denote data averaged over 288 minutes (4 hours and 40 minutes). Light grey curves are correlation products for each 10 second time-step. The grey shaded region indicates the delays in which cleaning has been performed.

55 MHz location of the divot. Decaying structure similar to what was observed in the autocorrelations is visible above the thermal noise out to $\sim$ 1000 ns (purple dashed lines and black lines at high delays). At 750 ns, the large-delay structure is roughly $\sim$30 dB down from peak power – which is similar to what appears in the autocorrelations.

However, when we exclude the divot, super-horizon structure remains at nearly the same level as when the divot is not included (Fig. 11). We also note that in the frequency domain, the residuals outside of the divot are most significant at $\approx$ 70 MHz (Fig. 10).

We conclude that this super-horizon cross correlation structure could be

(a) be a completely independent issue from the divot that is hidden by the charge-pump noise floor in auto-correlations but will appear once the charge-pump spikes are improved.

(b) Is associated with the divot (though this may be unlikely given that the structure does not change at all when we exclude/include the divot).

4.1. *Things the super-horizon emission in cross correlations is not caused by.*

I eliminate several sources of the super-horizon structure in the crosses. These include

1. It is not associated with side-lobes of the 2600 ns reflection peak discussed in AP's extrapolated calibration memo. We note here that the reflection only appears at positive delay which suggests it may only be present in

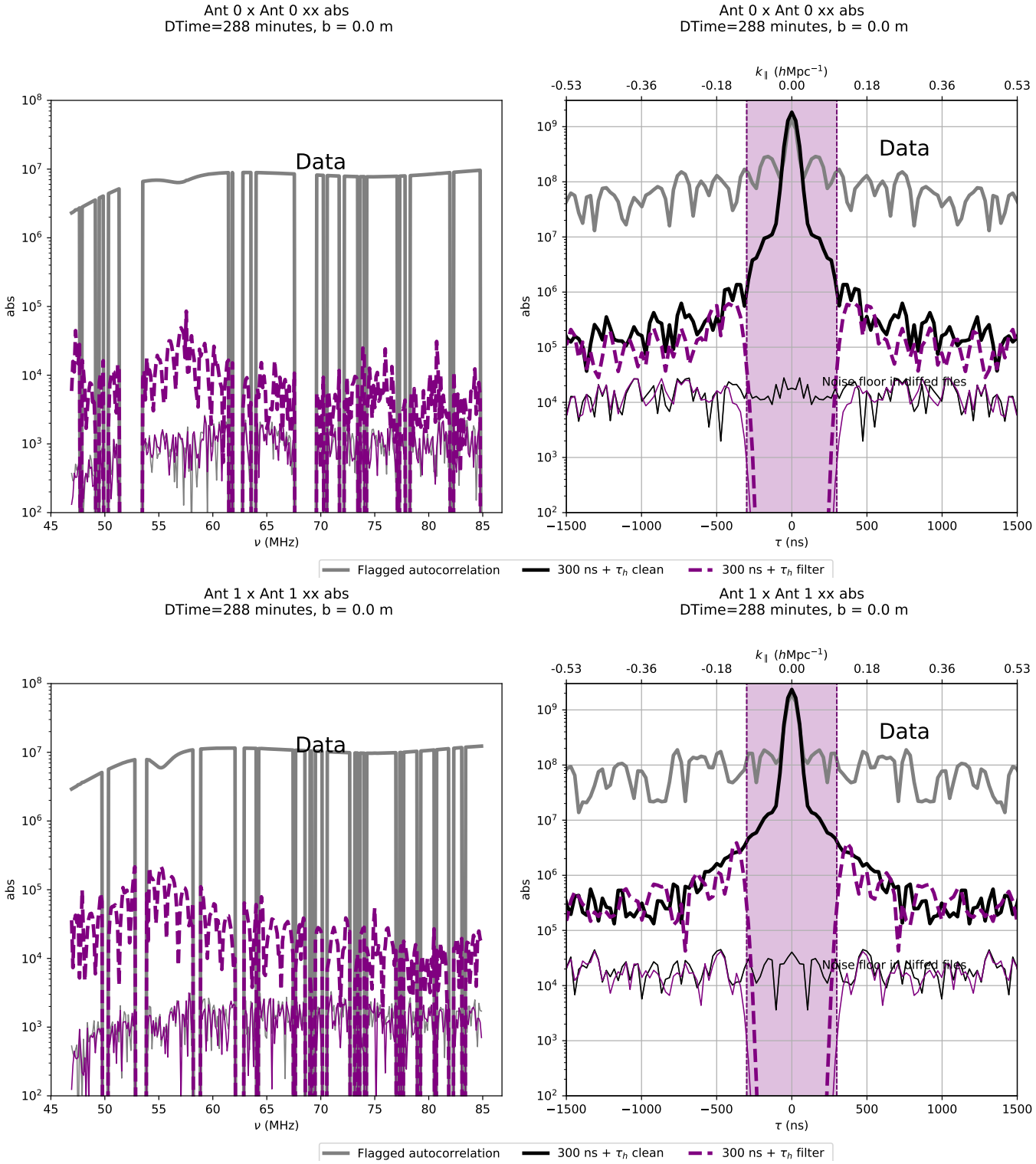

Figure 5. XX auto-correlation of antenna 0 (top) and antenna 1 (bottom). In the delay domain (right), side-lobes of the flagged auto-correlation (thick grey lines) dominate power. Purple dashed lines are filtered data at 300 ns. Black lines are residuals plus a restored model from a 1d `aipy` clean. The intrinsic systematic floor exists in the delay domain 40 dB (and 10 dB above the thermal noise floor). This is caused by switching noise in the RF-over fiber laser (spikes in the purple dot-dashed line on the left hand plot).

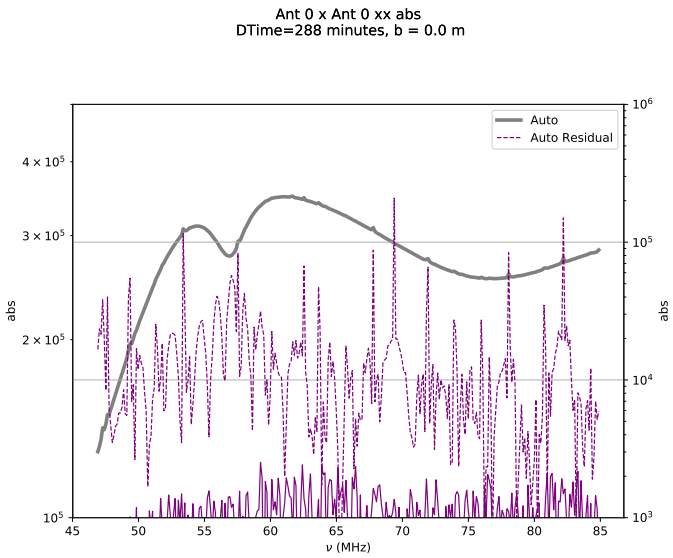

Figure 6. A comparison between autocorrelation residuals from 45-85 MHz and the auto itself. We see that the spike-like features that dominate the residual correspond to the RFoF charge-pump spikes. Note that the residuals are on a different y-scale than the autocorrelation. In addition, to maximize spike-visibility, the autocorrelation has been flattened by dividing out a best-fit first-order polynomial.

one of the antennas on the $0-1$ baseline. I determine this by filtering the fiber reflection delays and do not find any significant difference (Fig. 12 left).

2. It is not caused by the ≈ 70 MHz RFI spike. We conclude as such by flagging 10 channels around 70 MHz and still obtain similar residuals (Fig. 12 right).

4.2. *Close Inspection of Residuals*

4.2.1. *Coherent versus Incoherent*

We zoom in on frequency domain residuals of baseline $0x1$ Fig. 13 after four hours of incoherent integration (black line). I note that, at least within the low band, charge-pump spikes do not distinctively though they could be a part of the systematics floor. Instead, the residuals are dominated by the divot below 60 MHz and rapidly oscillating residuals above 60 MHz. The residuals of the incoherent integration approaches the noise level above 60 MHz, motivating us to see what happens with a coherent integration, plotted in red. The coherent integration noise is a factor of a few below the level of the coherent residual, wich agrees remarkably well with the incoherent average over the regions of frequency where it exceeds the noise level. The close agreement between the coherent and incoherent integration leads us to believe that the residuals are slowly fringing, if at all.

In delay space, we see that for the coherent integration, things are very dire. Structure remains above the thermal noise level out to $\gtrsim 4000$ ns and in frequency space the residuals clearly contain abundant fine-scale, high-delay frequency structure. We furnish additional evidence that the residuals are slowly fringing by comparing two coherent integrations – one between 5 and 7 hours of LST and the other between 7 and 9 hours. The frequency domain residuals are highly correlated.

We wrap up our inspection by investigating whether the charge-spikes are present in cross correlations. In the low band, it does not appear that charge-spikes are a limiting systematic (although isolated spikes are present; potentially due to RFI). We inspect the residuals in Fig. **??** where we average data over 4 hours. A 1 μs ripple dominates the residuals over the divot and also pops up at ≈ 75 MHz.

4.2.2. *Polarizations*

4.2.3. *Additional Baselines*

5. CHARGE SPIKES IN THE CROSS CORRELATIONS?

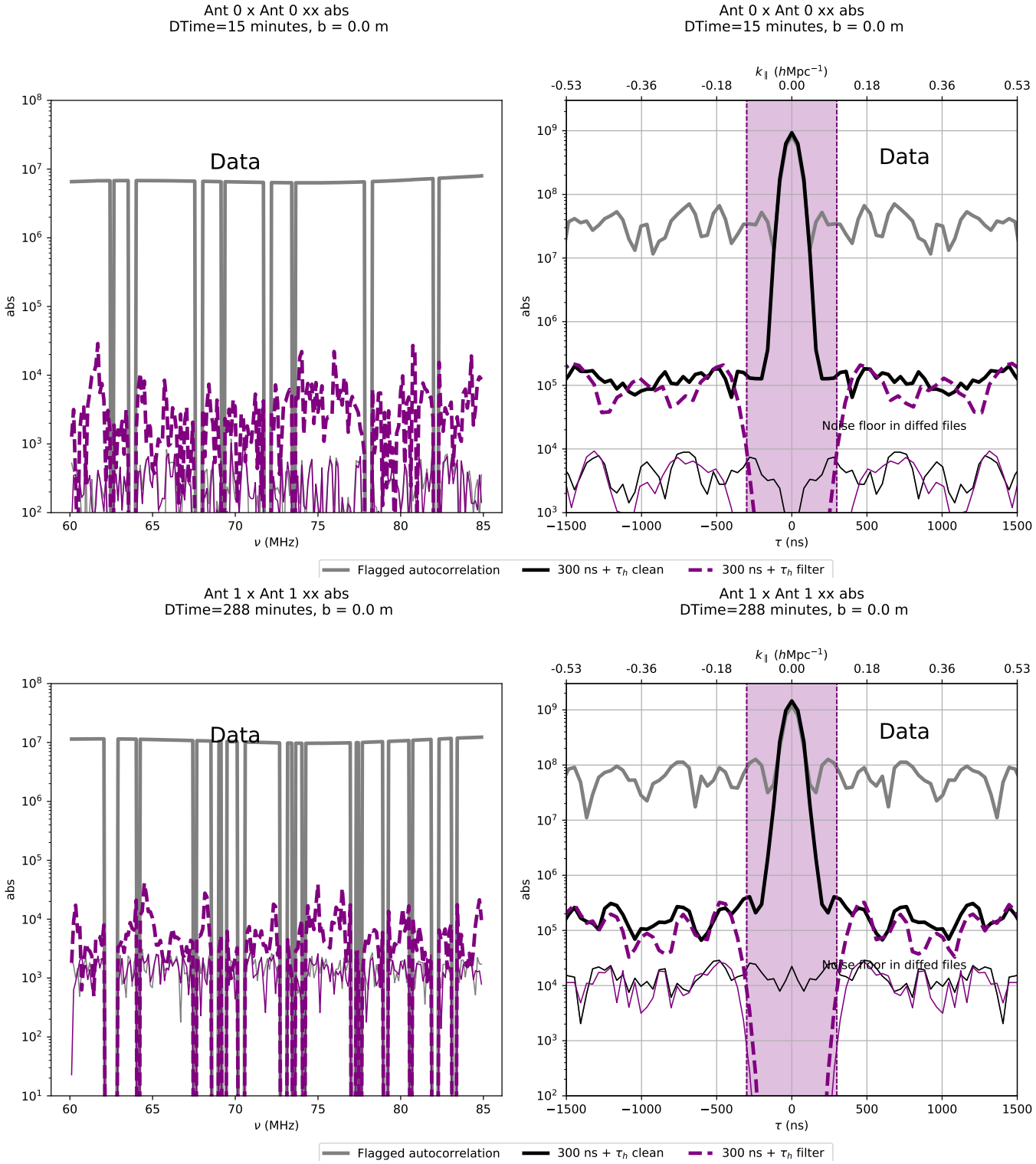

Figure 7. The same as Fig. 5 except now the bandwidth is limited to between 60 and 85 MHz. While power-supply switching noise remains a significant issue (setting up a systematics floor roughly 10 dB up), there is no longer any evidence for super-horizon emission associated with the "divot".

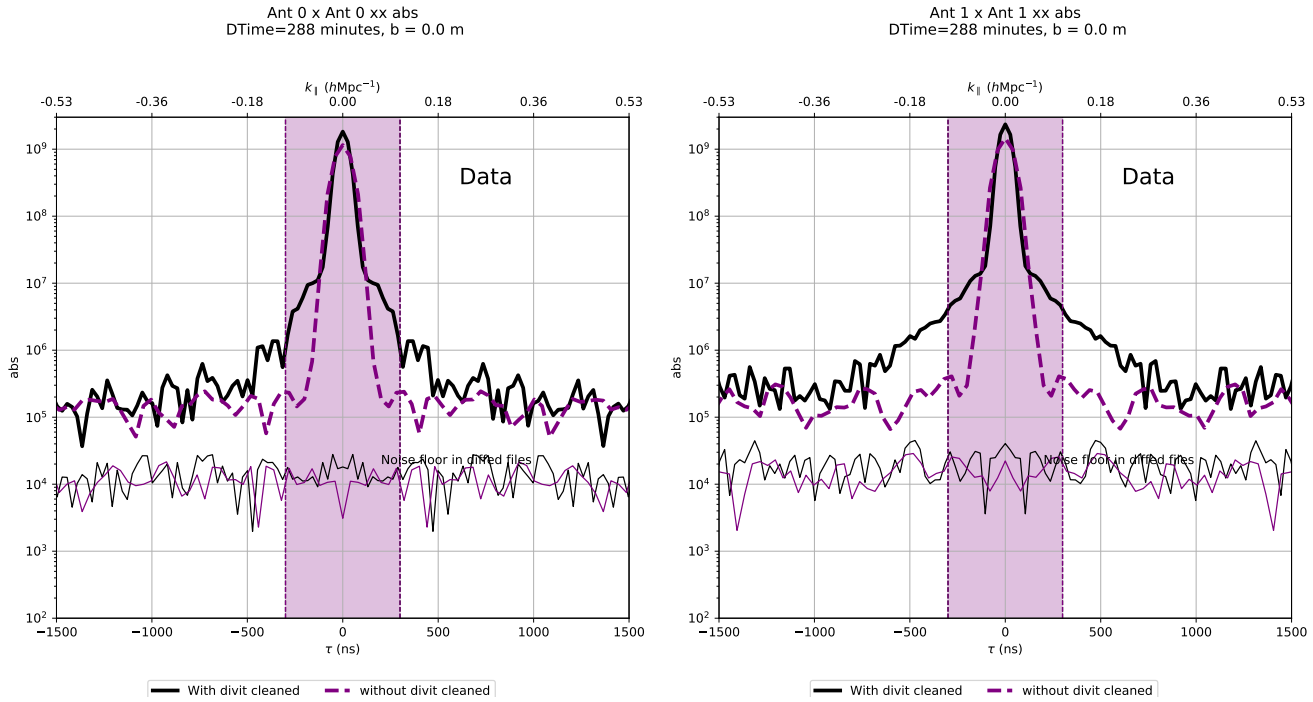

Figure 8. Direct comparison between `aipy` 1d cleaned auto-correlations with (45-85 MHz) (black solid) and without (60-85 MHz) (dashed purple) the divot included in the band. Left is antenna 0. Right is antenna 1.

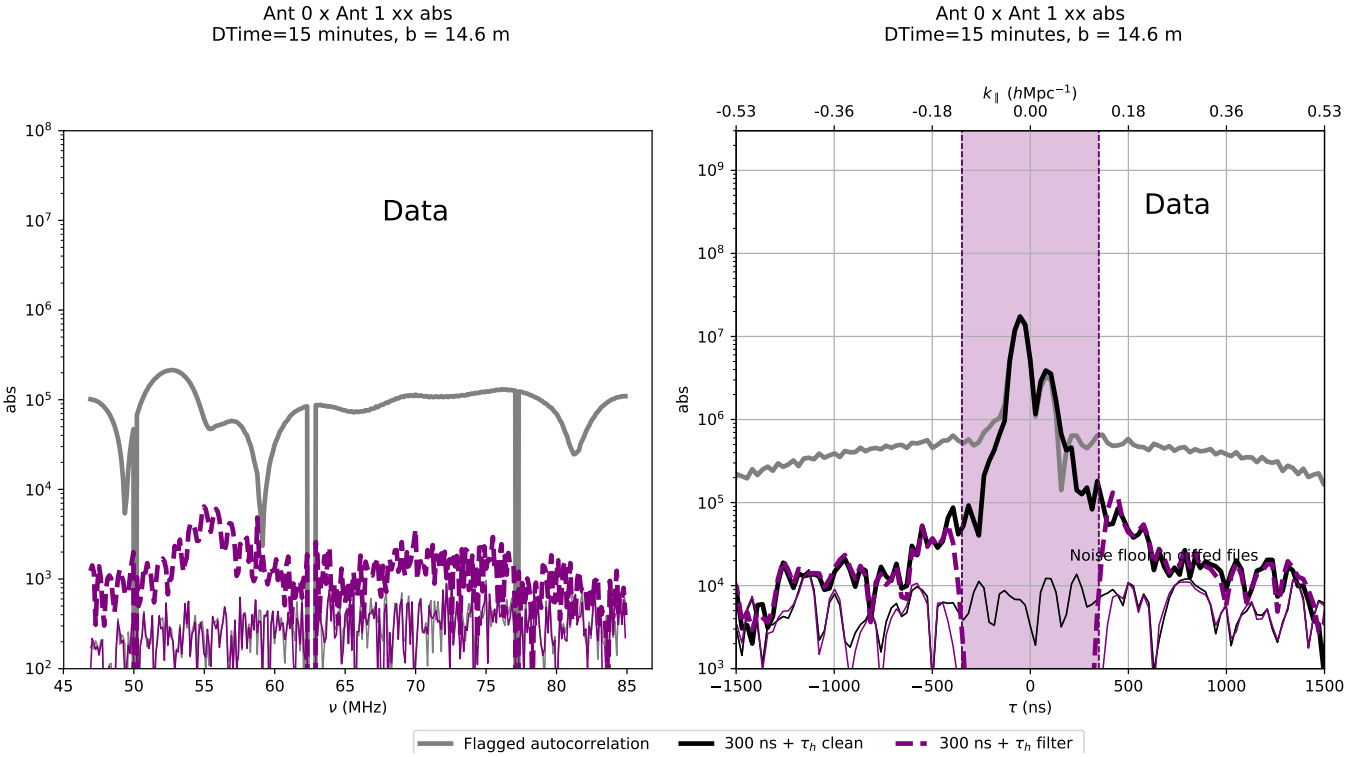

Figure 9. Same as Fig. 5 but now focusing on cross-correlations.

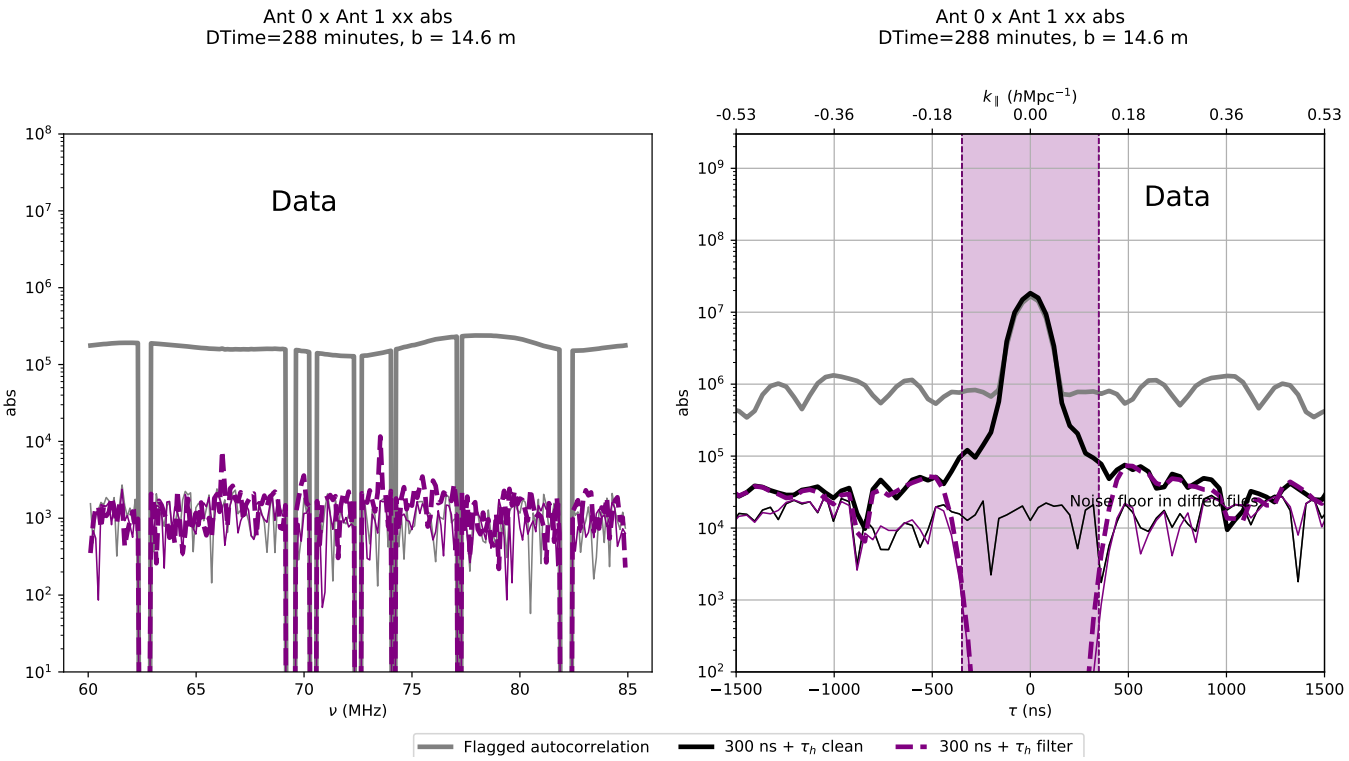

Figure 10. Same as Fig. 7 but now focusing on cross-correlations. Excluding the divot in cross correlations does not appear to have a significant impact on super-horizon power.

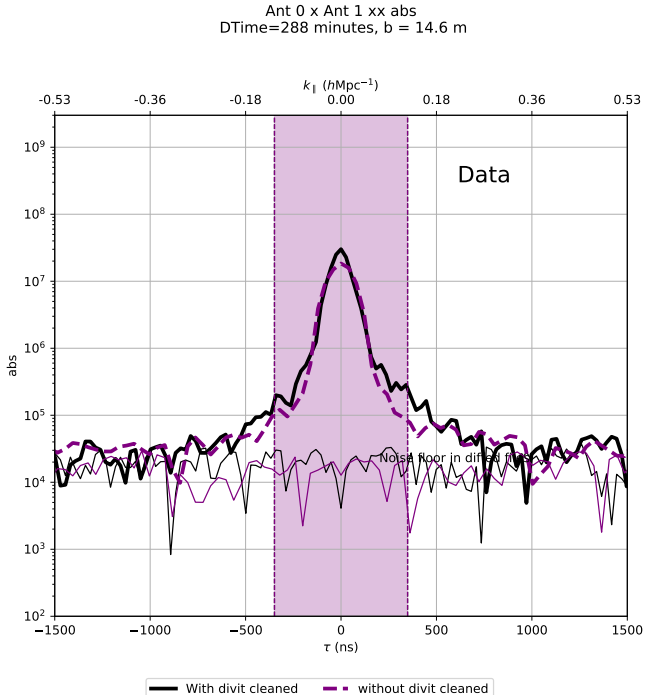

Figure 11. Same as Fig. 8 but now with a cross correlation. Super-horizon emission remains even when the divot is excluded. We note that this power is lower than the level of the charge-pump spikes in the auto-correlations and so may be present in those as well.

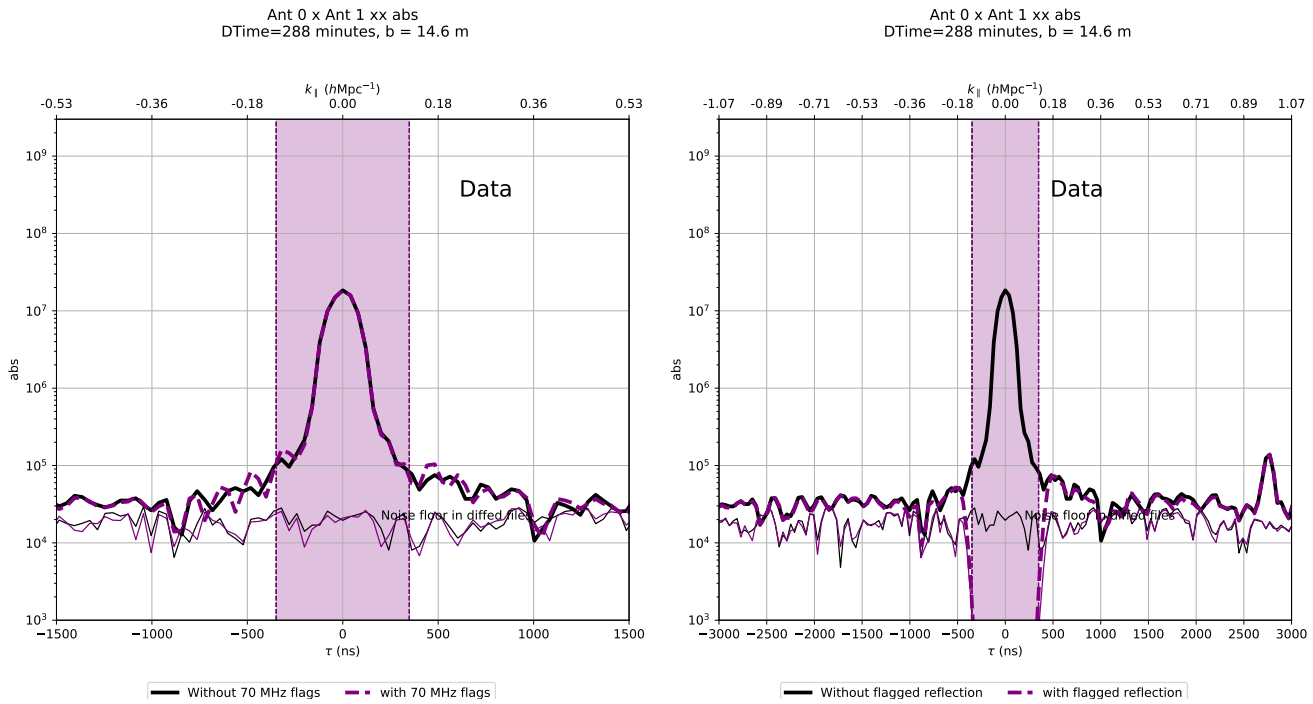

Figure 12. Left: delay domain views of filtering a visibility to exclude the 2600 ns reflection probably sourced by the optical fiber. Right: The same comparison but now examining the effect of flagging 70 MHz to see if residual power is being caused by an RFI/charge-pump spike. Neither attempt to correct super-horizon power has much of an effect.

Feature	Frequencies	Delays	Crosses?	Autos?	Amplitude	Integrates down?
Charge Pump	All	All	Maybe	Yes	-40 to -10 dB	
RFoF Reflection	All	2600 ns	Yes	Probably[a].	-40dB	
divot	50-60 MHz	$\lesssim 1\mu$s	Maybe	Yes	-30 dB	
"Cross Wings"	75 MHz and $\sim$ 70 MHz	$\lesssim 1\mu$s.	Yes	Maybe[b]	-30 dB	

[a](hidden by switching noise)
[b](hidden by switching noise)

Table 1. A table of super-horizon power sources. All features can be compared directly in Fig. **??**.

6. WHAT NEXT?

Before we pull the trigger on 350 feeds, we need to understand and mitigate the spectral structures listed in Table 1.

7. ACKNOWLEDGEMENTS

I thank Bryna Hazelton for substantial feedback on this analysis and suggesting the terminology "delay wings". I also credit Miguel Morales, Danny Jacobs, Jacqueline Hewitt, Juan-Mena Para, David DeBoer, Aaron Parsons, and Nima Razavi-Ghods, Joe Lazio, and Tzu-Ching Chang for useful discussions on the analysis in this memo.

REFERENCES

Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219

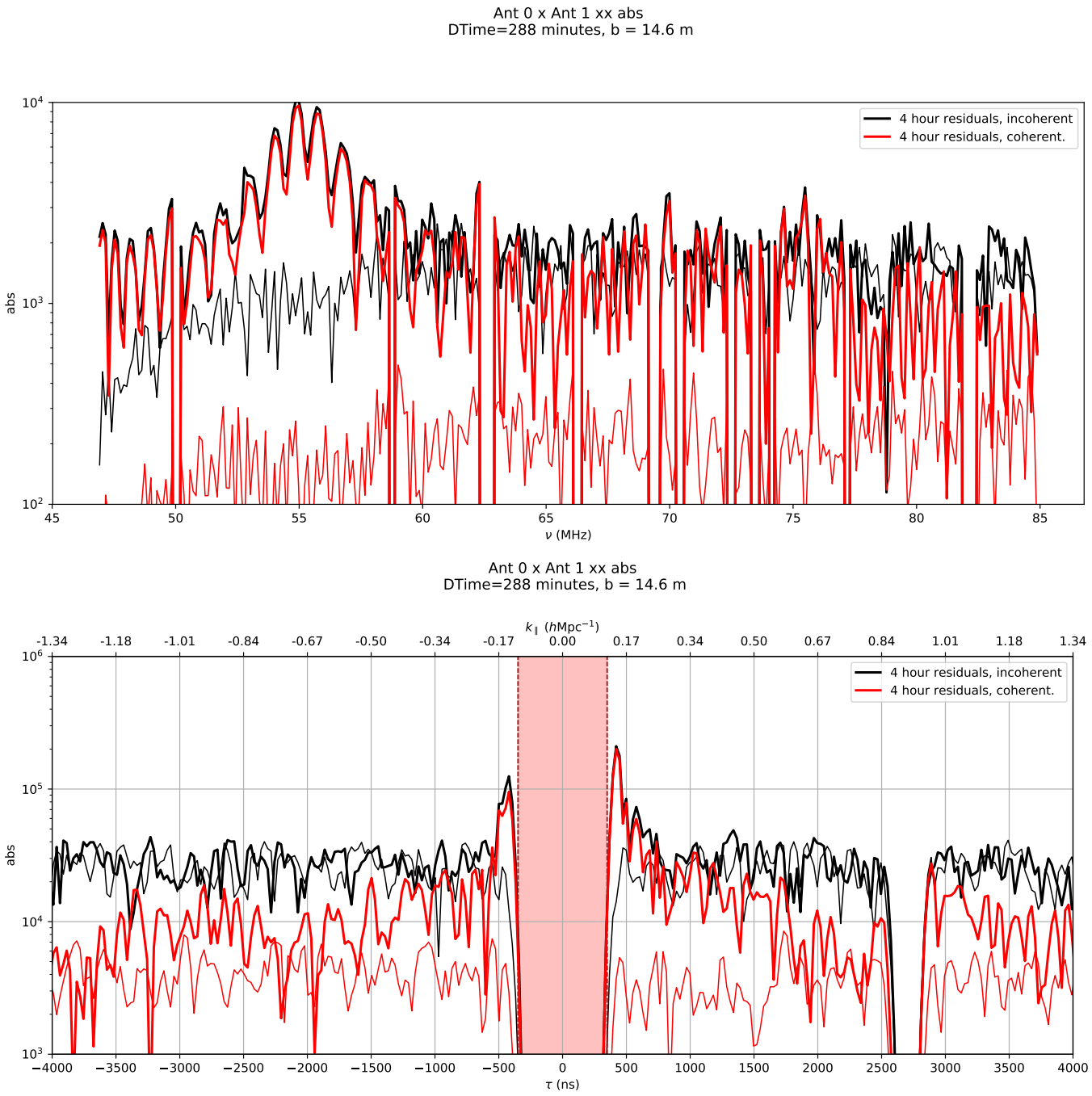

Figure 13. Top: Frequency domain residuals for a coherently (red) and incoherently (black) baseline. Bottom: The same in the delay domain.

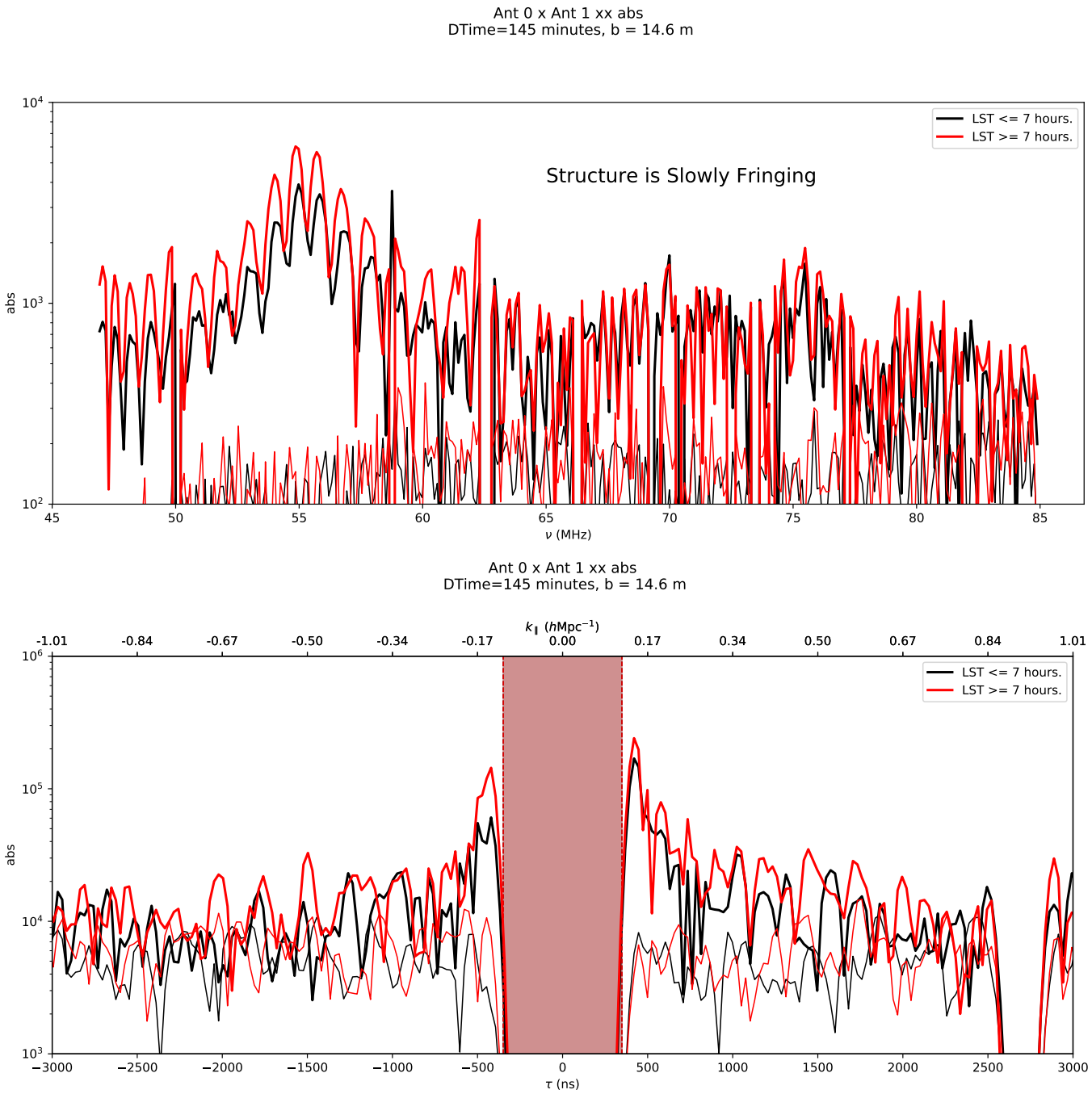

Figure 14. Top: A comparison of residuals from two coherent integrations, each over ~ 2 non-overlapping hours of LST. The high-delay residuals agree very well, suggesting further that they arise primarily from low delay-rate sources (instrumental or otherwise).

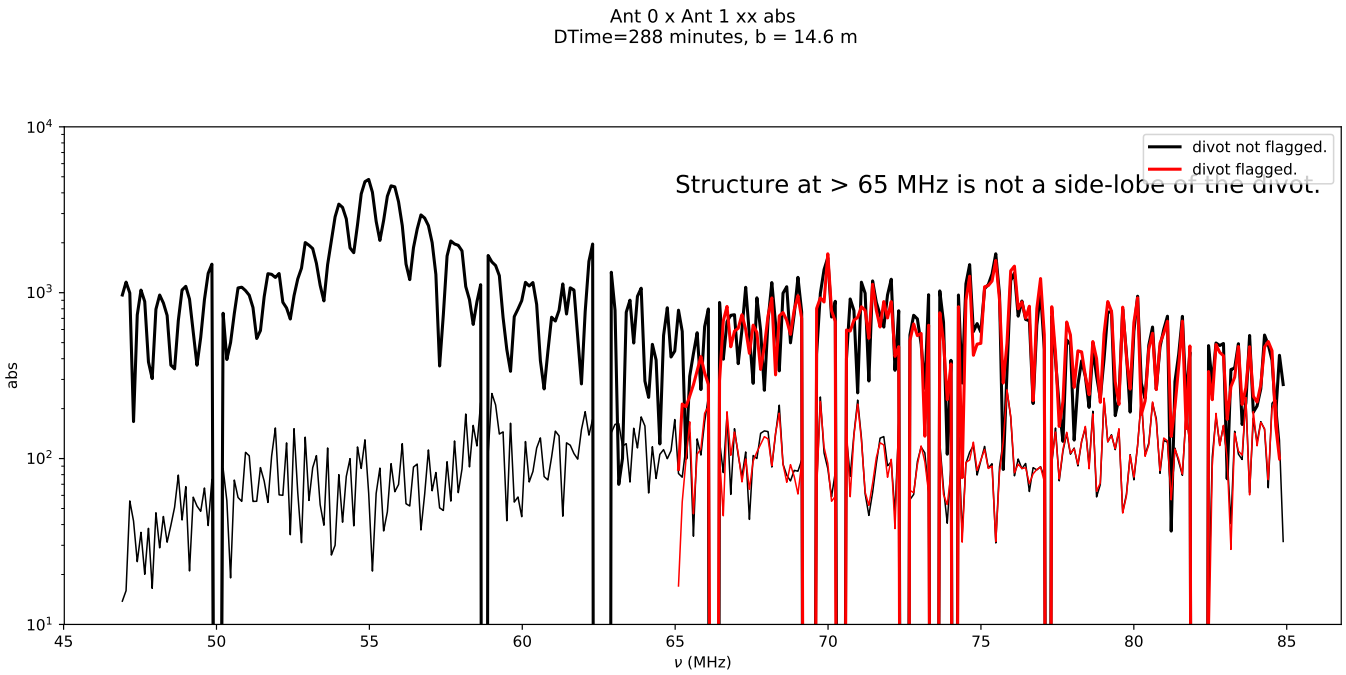

Figure 15. Comparing the residuals with and without the divot masked.

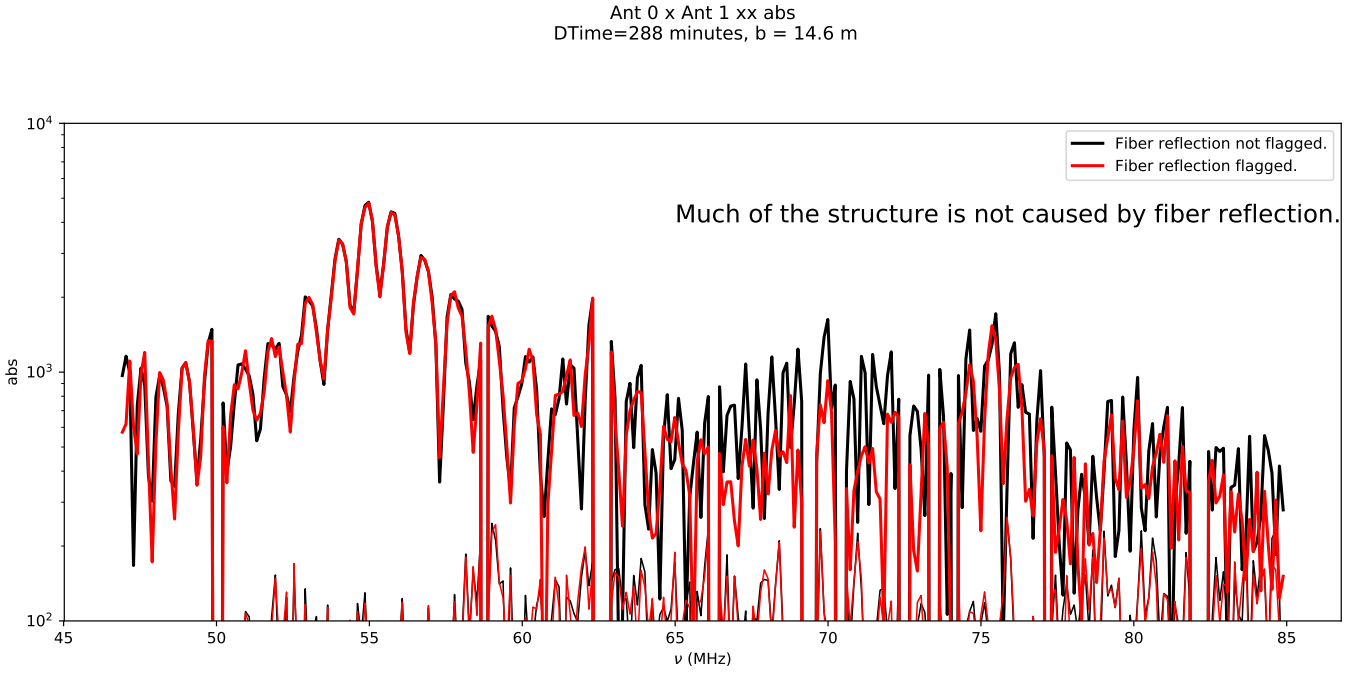

Figure 16. Testing the impact of masking fiber delays.

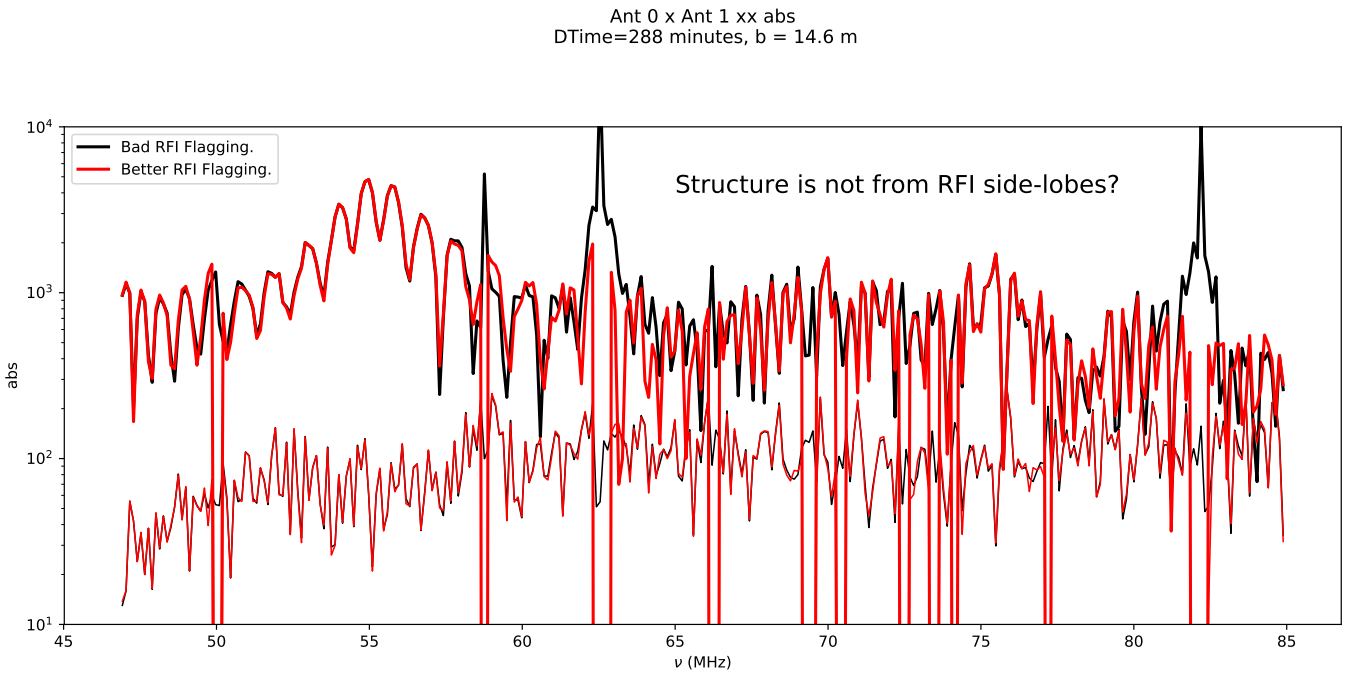

Figure 17. A jack-knife of whether the high delay structures are coming from RFI side-lobes.