HERA Memo 107
Received: 30 Nov. 2021

HERA Memo:
H1C IDR 3.2 Power Spectrum Analysis Updates and Choices

Joshua S. Dillon, UC Berkeley

November 30, 2021

1 The H1C IDR 3.2 Pre-Processing and Power Spectrum Pipeline

The H1C IDR 3.2 pipeline consists of three parts. Roughly speaking, they correspond to the three lines of pipeline schematic (Figure 3) in HERA Collaboration et al. (2021). The first is the analysis pipeline, which was described in HERA Memo #97 (Dillon, 2021). It includes redundant-baseline calibration (Dillon et al., 2020), absolute calibration (Kern et al., 2020a), RFI excision, calibration smoothing, and LST-binning. In H1C IDR 3.2, that LST-binning was done in four separate epochs, which can then be LST-binned together for the deepest data product.

The second step is somewhat confusingly known as "pre-processing." It is a set of systematics-mitigating operations performed on LST-binned visibilities whose output product is still visibilities. This has traditionally included a final round of flagging, in-painting to fill RFI gaps, cable reflection calibration, cross-talk subtraction via fringe rate-filtering (Kern et al., 2019; Kern et al., 2020b), further time averaging, and the combination of polarizations to produce pseudo-Stokes. It now also includes a second LST-binning step after cross-talk subtraction where the four epochs are combined.

The final step is power spectrum estimation, which includes estimating power spectra and noise for every pair of baselines in the data set. To simplify the statistics, power spectra are computed only from the cross of even and odd integrations (Tan et al., 2021). These power spectra are then averaged incoherently across all "cross-baseline pairs" (e.g. the power spectrum computed by crossing $\widetilde{V}_{12,13}$ with $\widetilde{V}^*_{25,26}$) that measure the same baseline separation, dropping "auto-baseline pairs."[1] Next they are averaged incoherently in time in 5 separate fields (i.e. ranges of LST). These are then spherically averaged to get the deepest upper limits on $\Delta^2(k)$.

For the most part, this pipeline is quite similar to that used for the H1C IDR 2.2 analysis and upper limits (HERA Collaboration et al., 2021). In this memo, I lay out the updates and analysis choices that differ between H1C IDR 2.2 and H1C IDR 3.2. Some of these are consequences of the changes further upstream in the analysis, like the more aggressive RFI flagging. Others are attributable to the challenges presented by a deeper data set covering a larger range of LSTs. Some reflect our new, more physical understanding of the cross-talk systematics (Dillon et al., 2021). Some are just minor tweaks. I will discuss the updates to pre-processing in section 2 and the updates to the power spectrum estimation in section 3.

2 Pre-Processing

To summarize, the pre-processing pipeline includes:

1. A final round of flagging times and frequencies.
2. In-painting of flagged channels using a wide-band delay CLEAN.
3. Calibrating out per-antenna cable reflections using the autocorrelations.

[1] This is maintains most of the sensitivity of a coherent average at the visibility level of all redundant baselines, followed by power spectrum estimation (Parsons et al., 2014).

4. Removing the cross-talk by using the PCA and GPR-based technique described in Kern et al. (2019) and Kern et al. (2020b), effectively eliminating the fringe-rates near zero.

5. LST-binning together the four epochs.

6. Time averaging from 21.4 second integrations to 214 second integrations.

7. Forming pseudo-Stokes I from xx and yy visibilities

The data products from each individual epoch's pre-processing can be found at NRAO in `/lustre/aoc/projects/hera/H1C_IDR3/IDR3_2/LSTBIN/epoch_*/preprocess`. The histories of the files include what processing has been done to them, but the file extensions are also a useful shorthand:

- `L`: LST-binned.
- `P`: In-painted.
- `X`: Cross-talk subtracted (and usually, implicitly reflection calibrated).
- `T`: Time-averaged.
- `K`: Formed pseudo-Stokes.

The data products after LST-binning of the four epochs togther can be found in `/lustre/aoc/projects/hera/H1C_IDR3/IDR3_2/LSTBIN/all_epochs_preprocessed/preprocess`.

The pre-processing pipeline is controlled by a script that allows some parallelization via the python `multiprocessing` module. That script is available here. It is parameterized by a set of `YAML` files, which are all available here. The second round LST-binner script is here and the YAML parameterizing it is here.

And now on to what's new!

2.1 By-Hand Flagging

In the H1C IDR 2.2 analysis, a final round of flagging was performed by looking at delay-CLEANed residuals in the final LST-binned visibilities and hand-flagging times and channels that were clear outliers. The additional flags were applied to all baselines as a single flagging waterfall. The process was documented in this notebook.

I took the four epochs without any pre-processing, LST-binned them together, and performed a high-pass delay filter at 2000 ns, which should remove both foregrounds and cross-talk. I then averaged the amplitude of all baselines together, weighting and then re-normalizing by the noise, so that frequencies and times with high delay structure—potentially RFI or some other artifact of mis-calibration and/or LST-binning—would stand out. That waterfall is shown in Figure 1.

Outlier channels and integrations were identified using time- and frequency-averages of the metric in the left panel of Figure 1. This process was performed iteratively until rough convergence. Finally, the residual metric was exported as an image and individual pixels were flagged by hand by painting them white in Photoshop and then exporting the bitmap and re-interpreting it as a boolean array. I made an effort to only flag coherent clumps of pixels—preferably rectangular regions—to avoid cherry-picking. This flagging preferentially flagged channels and integrations that were flagged on many nights but not all nights. In Figure 2 I show how the distribution of $N_{\rm samples}$ compares between the waterfall pixels that were not flagged after LST-binning and the subset of those that I hand-flagged. This makes sense: we might expect spectral discontinuities due to the combination of time-variable flagging patterns and calibration errors. The final flag waterfall, along with the bands chosen for power spectrum estimation based on these flags (see subsection 3.1), is shown in Figure 3.

Fundamentally, this step involved many small judgement calls about what to flag, rather than a precise algorithm. This may have led to overflagging, since some of the high delay structure was was likely CLEAN artifacts rather than real RFI. However, since all the flagging was done on the high delay components of baseline-averaged amplitudes, there's little risk of signal loss form this step.

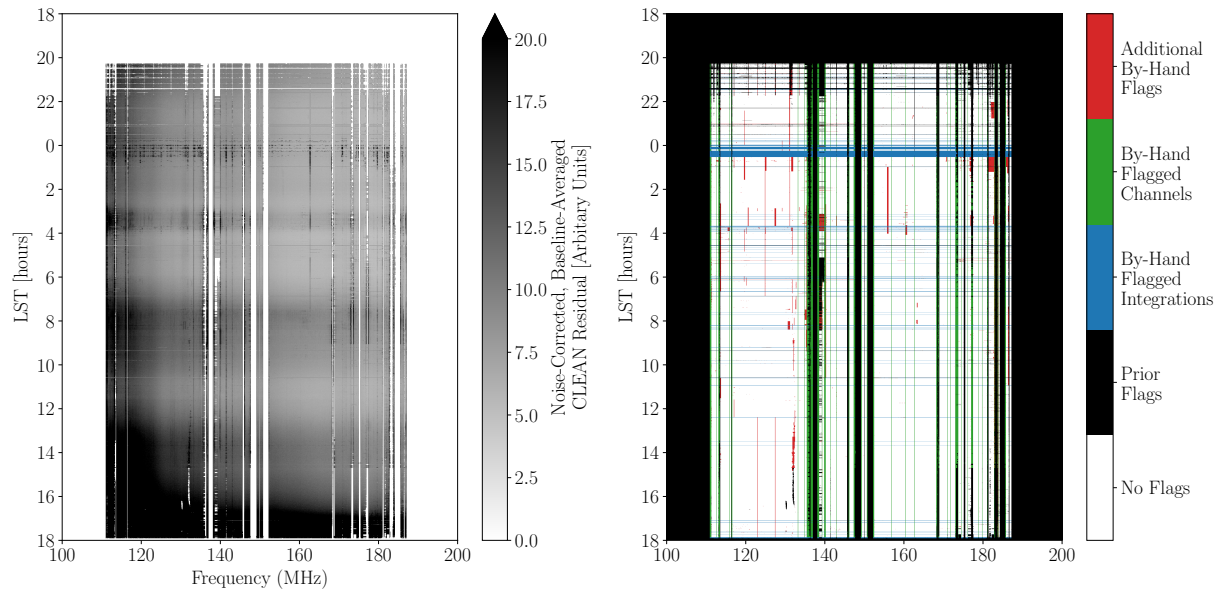

Figure 1: Here I illustrate how the by-hand flags were derived after LST-binned. *Left:* The baseline-average metric of residual spectral spectral structure after the 2000 ns high-pass delay filter shows clear discontinuities. Particularly telling are the channels that are partially flagged (white) but show substantial power at times they are not flagged, e.g. around ORBCOMM ($\nu \approx 137\,\mathrm{MHz}$) and around some of the digital TV stations ($\nu \gtrsim 170\,\mathrm{MHz}$). *Right:* Here is illustrate where the flags come from for each pixel in the final flagging waterfall (see Figure 3). Most of the by-hand flags are either entire channels or entire integrations, though a few do come from painting individual pixels in Photoshop.

2.2 Per-Epoch Systematics Removal

One of the biggest changes from H1C IDR 2.2 was necessitated by the split of the observing season into four epochs and the question of how and whem to combine them into one final data set. I had originally tried using the standard LST-binner in `hera_cal` to combine the four data sets before post-processing. However, I found evidence for substantial residual cross-talk even after applying the cross-talk subtraction procedure developed in Kern et al. (2019) and Kern et al. (2020b) and applied to HERA Collaboration et al. (2021). This was somewhat surprising since our physical model of the cross-talk indicates that it should be phase-stable with an amplitude structure following the autocorrelations (Dillon et al., 2021; Kern and Parsons, 2019). When we examine the four epochs combined without systematics subtraction in Figure 4, we can see that that is no longer the case.

The temporal discontinuity in the phase structure of the cross-talk is not entirely surprising in retrospect. If, as was speculated in Dillon et al. (2021), the delay structure beyond the first peak is attributable to multi-path over-the-air transmission from the receiverators, then it makes sense that that delay structure should change as new dishes were built and new feeds were raised.

To address this issue I performed the first few steps of the pre-processing pipeline before a second round of LST-binning. This includes RFI excision (with flags back-propagated from the first flagging mask developed with all epochs binned together), in-painting, reflection fitting, and cross-talk subtraction. I then developed a new second round LST-binner which, unlike the one in `hera_cal`, preserves in-painted channels appropriately (which are unflagged but have $N_{\mathrm{samples}} = 0$). Also, unlike the `hera_cal` LST-binner, which weights each sample in a given time and frequency bin by its number of samples, this code weights them by the frequency-averaged number of samples in that epoch to avoid introducing spectral structure. This code is also much simpler than the `hera_cal` version, since all epochs have been LST-binned to the same grid and therefore no rephasing is necessary. After LST-binning, the final time-averaging and conversion to pseudo-Stokes was performed.

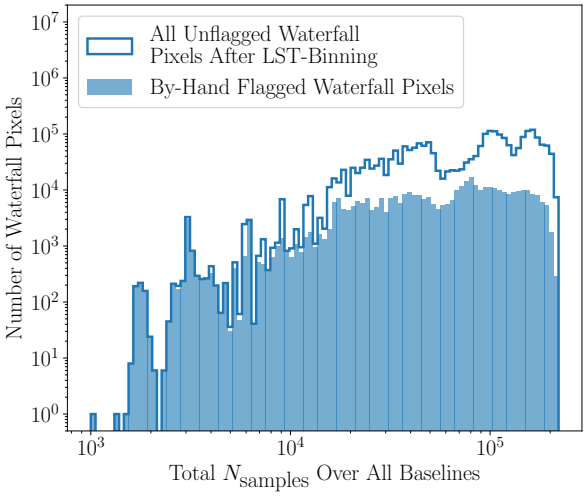

Figure 2: The multi-stage hand-flagging of the final flag waterfall was focused on flagging regions with high values in the noise-corrected, baseline averaged metric shown in Figure 1. As a result, many more partially-flagged (i.e. low-N_{samples} pixels were flagged. This preference is revealed in the histogram of N_{samples}, summed over all baselines, for both all unflagged pixels after LST-binning and those that were flagged by-hand (red, green, or blue in Figure 2.

This all seems to have worked petty well. Very individual few baselines show strong evidence for residual cross-talk after per-epoch systematics subtraction was performed (see subsection 3.3), though there is some evidence that the baselines with short ($< 15\,\text{m}$) projected East-West separations have residual cross-talk (see subsection 3.4).

2.3 Minor Changes

A number of other small analysis changes were made, including:

- I added an explicit LST branch cut for defining the order of files in pre-processing. It just so happens that the first integration on the first day of H1C IDR 3.2 is just after last integration on the last day of H1C IDR 3.2 in LST, so I use that to define the branch cut at 4.7 radians.
- I lowered the CLEAN tolerances on in-painting by a factor of 10 to account for increased sensitivity.
- While we believe the reflections should be time-stable within epochs, I noticed that our solutions were not. So I added a time-averaged reflection fit step which provides a starting point for the time-dependent (albeit smoothed) reflection fit algorithm used in H1C IDR 3.2.
- I increased the number of terms in the reflection fit. In H1C IDR 2.2, we fit 3 delay terms between 75 and 800 ns, then 5 terms between 100 and 800 ns, then 5 terms between 200 and 800 ns, then 5 terms between 250 and 800 ns, then 5 terms between 300 and 800 ns, then 5 terms between 700 and 1500 ns. I replaced the 3 terms 75 and 800 ns with 5 terms between 25 and 800 ns and then added another 5 terms between 1100 and 2500 ns. The former change was to try to fit the $\sim$20 m cables better. The latter change was to account for a few antennas with very long cables (see Dillon et al. (2021)).
- I increased the number of SVD to keep in modes cross-talk subtraction from 30 to 50, based on some mini-experiments on individual baselines where it seemed to get the residual cross-talk a bit closer to the noise. The effect was pretty small.
- In performing the SVD for cross-talk subtraction, I assigned zero weight to all times between 15.3 hours and 21 hours. This was meant primarily to isolate our most sensitive data from the effect of the

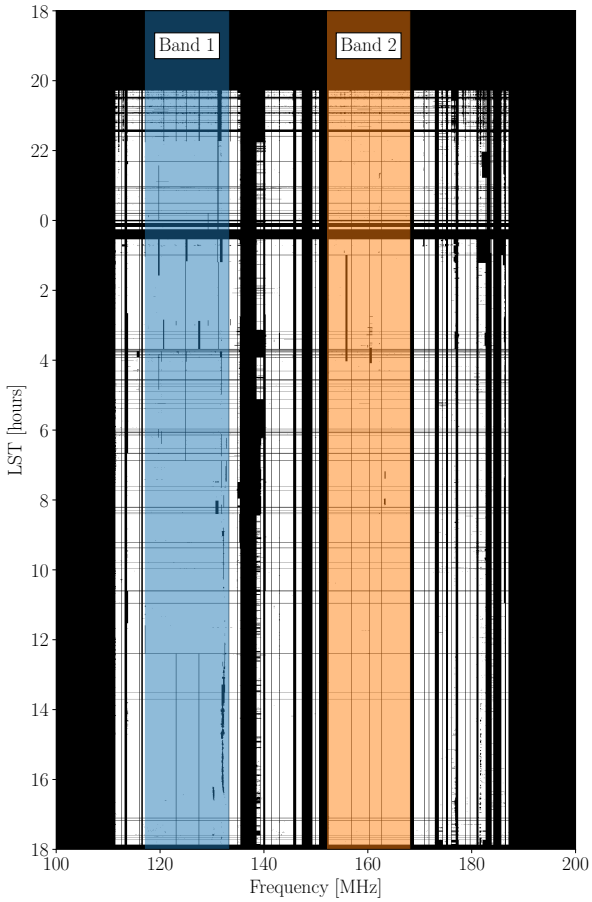

Figure 3: The final flagging pattern for H1C IDR 3.2 includes both flags from XRFI and by-hand flags as described in subsection 2.1. The edges of the two bands were then defined by picking ranges of frequencies with relatively minimal flagging. The additional flagging compared to H1C IDR 2.2 means that the definition of the bands is slightly different than in HERA Collaboration et al. (2021); see subsection 3.1 for details.

galaxy. Because the amplitude of the cross-talk scales with the autocorrelation and because cross-talk removal relies on temporal stability, having an order-of-magnitude brightness increase on the edge of the time range creates a ringing effect that makes subtraction worse over much of the prime LST range. The only downside of this is that the cross-talk model for LSTs near the galactic center is effectively extrapolated from earlier LSTs and thus we end up with a substantial under-subtraction of systematics at those LSTs.

3 Power Spectrum Analysis

Just like the pre-processing pipeline, the first part of the power spectrum pipeline is also controlled by a script that allows intra-node parallelization via the python `multiprocessing` module. That script is available here. It is parameterized by a set of `YAML` files, which are all available here. This stage produces power spectra for all baseline pairs within each redundant group. It also produces estimates of the noise power spectrum for all those baseline pairs based on the autocorrelations and the number of samples.

Power spectra for each epoch individually are available at NRAO in `/lustre/aoc/projects/hera/H1C_IDR3/IDR3_2/LSTBIN/epoch_*/pspec`. The full-season power spectrum container is available at `/lustre/aoc/projects/hera/H1C_IDR3/IDR3_2/LSTBIN/all_epochs_preprocessed/pspec/pspec.grp1.of1.LPXLTK.h5` (or without cross-talk subtraction at `pspec.grp1.of1.LPLTK.h5`).

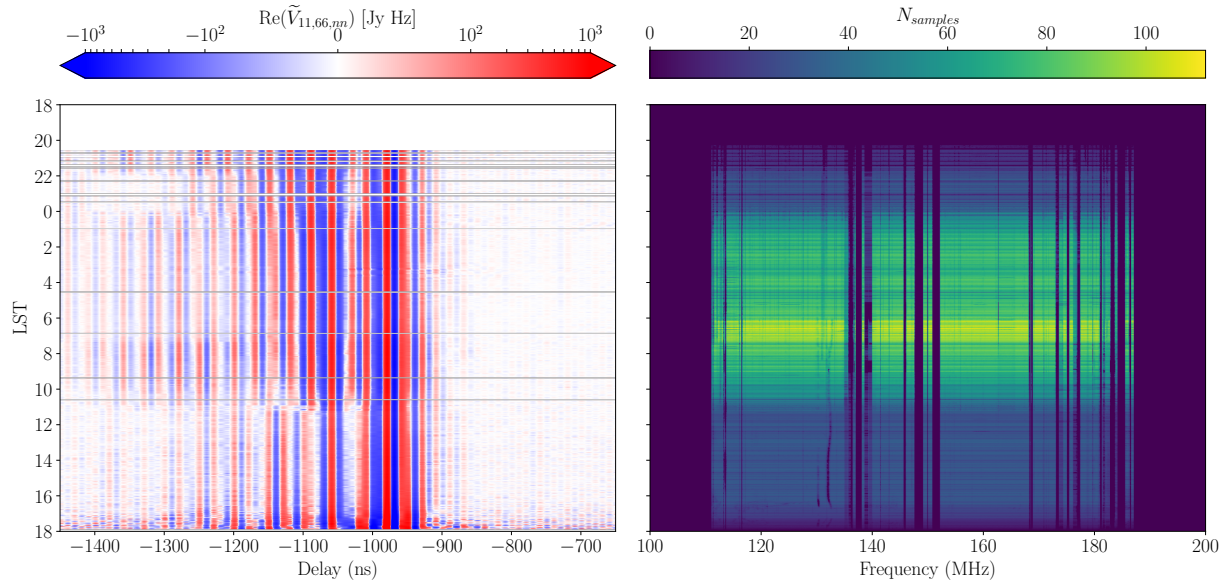

Figure 4: While the cross-talk systematics were phase stable in H1C IDR 2.2 (Kern et al., 2020b), there are clear temporal discontinuities in the cross-talk. Here I show the real part of a single visibility LST-binned across all four epochs without any pre-processing. We can see shifts in the cross-talk pattern, especially at larger magnitude delays, that correspond with the LST boundaries of epochs (as seen by the induced discontinuities in the N_{samples} waterfall). While main peak of the cross-talk around $\sim(-950\,\mathrm{ns})$ is phase stable, the clear temporal discontinuities break the cross-talk removal algorithm, which relies on temporal stability. This necessitates per-epoch cross-talk subtraction, since individual epochs show much less temporal structure in the cross-talk.

After the `pspec_pipe` script, the following steps take place in a Jupyter notebook:

1. Removal of baseline-pairs with high residual cross-talk (this is new; see subsection 3.3).
2. Averaging within redundant baseline groups.
3. Incoherent averaging within fields and formation of cylindrical power spectra with associated error bars.
4. Spherical averaging and formation of final spherical power spectra $P(k)$ and $\Delta^2(k)$ with associated error bars and window functions.

Once the cuts and analysis choices described here are finalized, that notebook can be released internally for review. We now turn again to what has changed between H1C IDR 2.2 and H1C IDR 3.2.

3.1 Tweaking Bands

Due to more aggressive RFI flagging (including hand flagging; see subsection 2.1 and Figure 3), I slightly modified the two bands in which power spectra were estimated. The result is summarized in Table 1. I have also reproduced Figure 12 of HERA Collaboration et al. (2021) for H1C IDR 3.2 in Figure 5 showing how these bands fall relative to the RFI occupancy when the Blackman-Harris tapering function is applied.

3.2 Redefining Fields

Given our wider range of observed LSTs and updated flagging algorithm (including a reduced overflagging of LSTs around the peak transit of Fornax A), I felt it appropriate to reevaluate our fields. This process

	H1C IDR 2.2	H1C IDR 3.2
Band 1 Channels (inclusive)	175–334	176–339
Band 1 Frequency Range	117.09–132.62 MHz	117.19–133.11 MHz
Band 1 Bandwidth	15.53 MHz	15.92 MHz
Band 1 Central Redshift	10.38	10.35
Band 2 Channels (inclusive)	515–694	535–696
Band 2 Frequency Range	150.29–167.77 MHz	152.25–167.97 MHz
Band 2 Bandwidth	17.48 MHz	15.72 MHz
Band 2 Central Redshift	7.93	7.87

Table 1: How the frequency bands changed from H1C IDR 2.2 to H1C IDR 3.2.

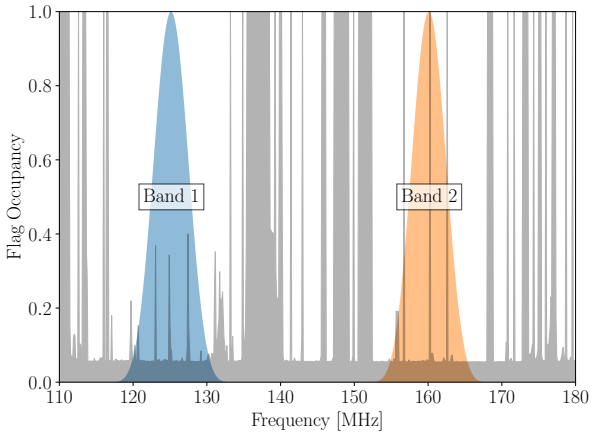

Figure 5: Here I reproduce Figure 12 of HERA Collaboration et al. (2021). The bands in H1C IDR 3.2 were defined to be similar to those that H1C IDR 2.2, but were moved slightly to avoid RFI (see Figure 3)

is vulnerable to fine tuning, so a few steps were taken to avoid that. First, field boundaries were restricted to quarter hour increments in LST. Second, the fields were defined largely by looking at two metrics that are relatively independent of the final power spectrum limits: baseline-averaged power spectra at $\tau = 0$ and baseline-averaged noise power spectra. Both of those are shown in Figure 6

My rationale for defining the fields is as follows. I wanted fields B, C, and D, to correspond decently well to fields 1, 2, and 3 in HERA Collaboration et al. (2021). So to cover the new LST ranges, I added fields A and E. Field A was easy to define, given the extensive flagging between 0 and 1 hours. The precise edges were defined to exclude first few stray and highly integrations and the last integration before the gap. Likewise, field B was defined to exclude the first integration after the gap and to get as close to Fornax A in the main lobe as possible while keeping its brightness comparable to its brightest point in the first sidelobe around 2 hours. Field C was defined to start at a roughly symmetrical place to where field 1 ended and to include the range from roughly 4–6 hrs of maximum sensitivity. Thus the upper field boundary was set by the sidelobe of the galactic anticenter at 6.25 hours. The boundary between fields D and E was set to be to include a roughly symmetrical sidelobe within field D, keeping the galactic anticenter as contained as possible to a single field. Field E ends a bit before where the cross-talk subtraction gets zero weight in the SVD.

The final definitions of the fields, as compared to the definitions in H1C IDR 2.2, are laid out in Table 2.

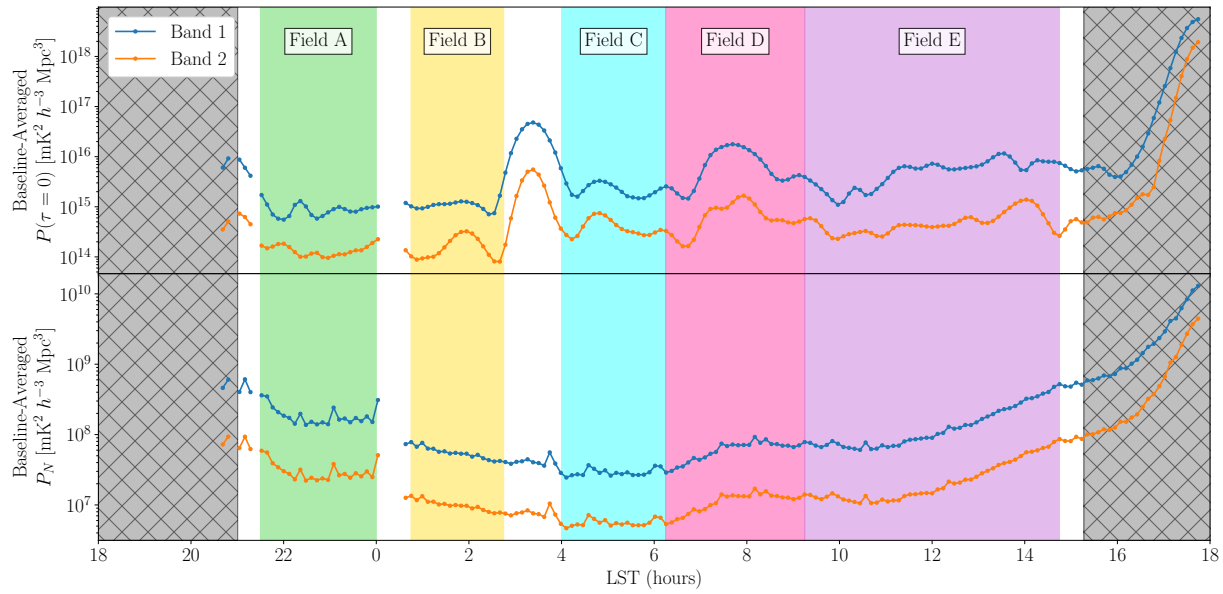

Figure 6: Here I show two metrics used to define the five fields in LST. The first is the baseline-averaged power spectrum at $\tau = 0$, which is a metric of the brightness of foregrounds at that LST. The second is the baseline-averaged noise power spectrum, which includes both foreground brightness and observing time. Both averages were performed using an inverse-variance weighting using the noise power spectrum. Grey hashed regions received zero weight during cross-talk subtraction.

Field	H1C IDR 3.2 LSTs	Closest H1C IDR 2.2 Field	H1C IDR 2.2 LSTs
A	21.5–0.0 hrs	-	-
B	0.75–2.75 hrs	1	1.25–2.7 hrs
C	4.0–6.25 hrs	2	4.5–6.5 hrs
D	6.25–9.25 hrs	3	8.5–10.75 hrs
E	9.25–14.75 hrs	-	-

Table 2: How the fields (i.e. the LST-ranges combined into single power spectra) changed from H1C IDR 2.2 to H1C IDR 3.2.

3.3 Cutting Baselines with Excess Cross-Talk

In the process of hunting down the excess residual cross-talk,[2] I developed a metric of per-baseline-pair residual cross-talk strength. Namely, I computed the average absolute value of the SNR (i.e. $|P(\tau,t)/P_N(\tau,t)|$) in the cross-talk delay ranges between ± 800 and ± 1500 ns and for the most sensitive fields (B, C, and D). Computing the expectation value of this a priori is challenging.[3] However, since the noise is white in τ, we can compare this statistics to a region of delay space that should be noise dominated, namely the range from ± 2800 and ± 3500 ns. In Figure 7 I show that quantity computed for all baseline-pairs in the 14.6 m E-W baseline group. I set the color scale to help highlight SNR in excess in the residual cross-talk range compared to the noise-dominated range.

Recall that the power spectra were formed by cross-multiplying adjacent 210 s integrations, so the power spectrum of $(11, 12) \times (142, 143)$ includes data from the opposite set of times as $(142, 143) \times (11, 12)$. Hence,

[2] A process that unveiled a statistical asymmetry between positive and negative delays that ultimately led us a better physical model for the cross-talk (Dillon et al., 2021).

[3] While the distribution of the real and imaginary parts of $P(\tau,t)/P_N(\tau,t)$ should be Laplacian (Tan et al., 2021), the distribution of the absolute value of that quantity, averaged over a range of potentially correlated times and highly correlated delays makes this problem analytically intractable. Another potential approach would be a noise-only simulation, though it is certainly easier to use noise-dominated delays.

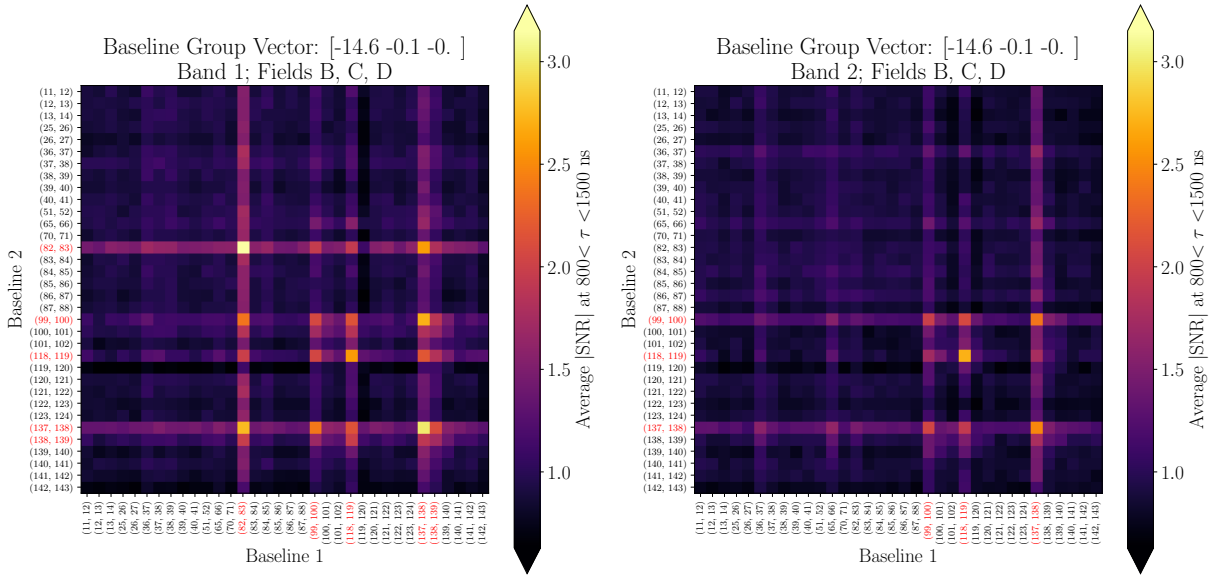

Figure 7: Here I show the average $|P(\tau,t)/P_N(\tau,t)|$ in the positive τ cross-talk delay ranges for all baseline-pairs in the 14.6 m E-W baseline group. If these baseline were consistent with noise, they would center on roughly 0.63 (black on this color scale, see Figure 8). Even after per-epoch systematics subtraction (see subsection 2.2), some baselines show substantially more residual cross-talk than others. This metric is used to find them on a per-baseline, band, and delay side (i.e. positive and/or negative delays) basis. Baselines are removed based on the value of the diagonal of this matrix, cutting entire rows and columns. The baselines flagged in this group at positive delays are labeled in red.

the lower and upper diagonals of Figure 7 are technically different, though they are highly correlated.

In order to avoid cutting individual baseline-pairs based on their power (albeit not in the delay range where we expect to set our best limits), I instead use the auto-baseline-pairs (the diagonal of Figure 7) to inform a cut. If any auto-baseline-pairs exceeded 1.5, I cut all baseline pairs that involve that baseline.[4] This cut was done separately for each frequency band. Similarly, the cut was done separately for positive and negative delays, since we now know that those are associated with different emitting antennas (Dillon et al., 2021). Baseline-pairs cut for excess positive delays in the cross-talk delay range were cut for all $\tau > 0$; likewise for negative delays.

The histogram of these auto-baseline-pair SNRs is shown in Figure 8. The vast majority of auto-baseline-pairs (between 96% and 99%) are kept. A slightly larger fraction of cross-baseline-pairs were removed because short baselines had more residual cross-talk and there were more baselines in the short baseline-groups. More baseline-pairs were cut at negative delays than positive delays because the antenna ordering means that negative delays were more often associated with antennas nearer the receiverator (again, see Dillon et al. (2021)).

This cut is the most surgical that I perform, in terms of removing individual power spectra before averaging. However, by only looking at high $|\tau|$ and using only the auto-baseline-pairs (which are excluded from the cylindrically and spherically averaged power spectra), we insulate ourselves from cherry-picking and signal loss.

3.4 Cutting on Minimum East-West Distance

Despite removing baseline-pairs with high residual cross-talk before performing the average within redundant baseline groups, some residual cross-talk apparently does not average down. If we compute this same

[4] I picked 1.5 based on the histograms in Figure 8 to remove values that appeared separate from the relatively continuous distribution. This is a judgement call.

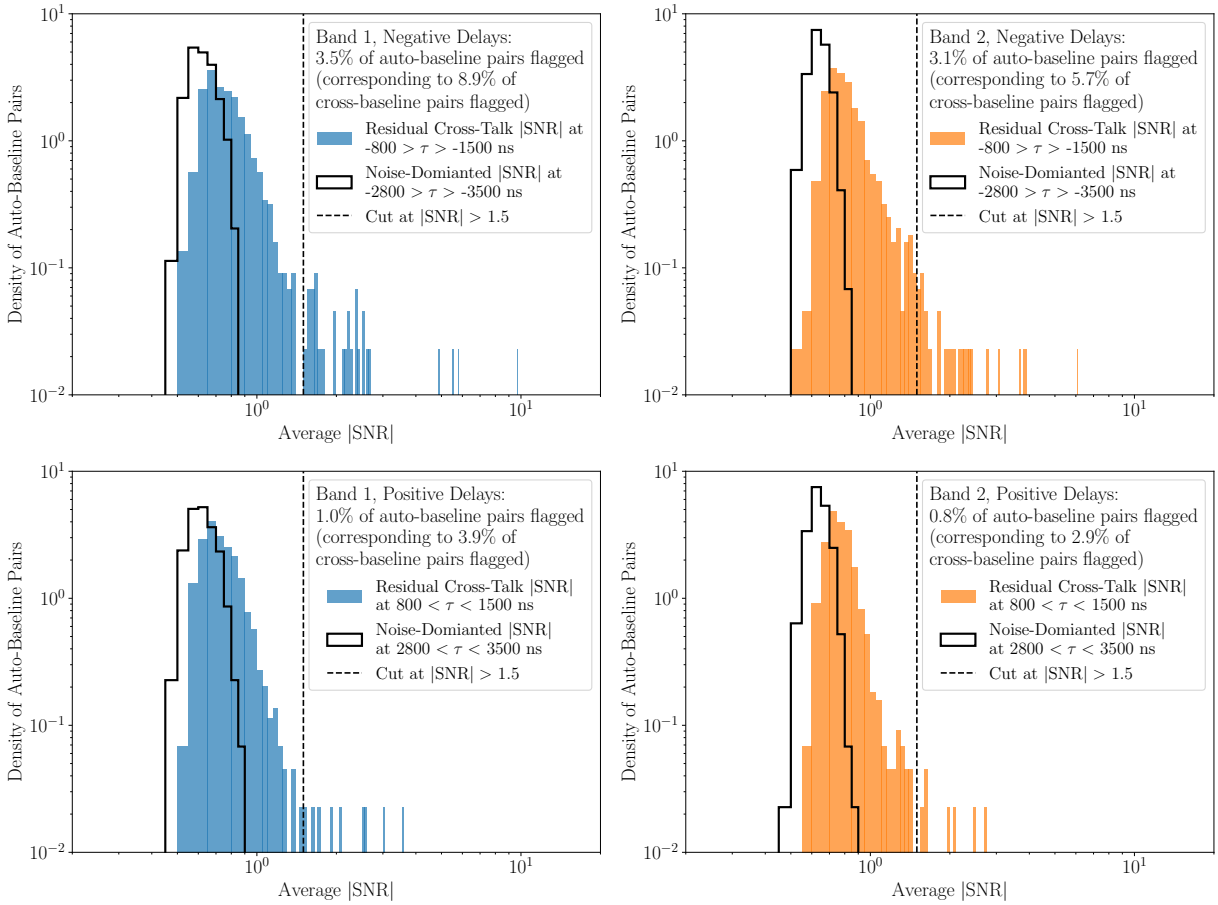

Figure 8: The distribution of magnitude of the SNR (i.e. $|P(\tau,t)/P_N(\tau,t)|$) averaged over the cross-talk delay ranges between ±800 and ±1500 ns and fields B, C, and D. This is the histogram of the diagonal of Figure 7, except for all baseline groups and now showing both positive and negative delays. Outlined in black is the equivalent histogram of the same quantity computed from power spectra between between ±2800 and ±3500 ns, a range of delays that should be noise-dominated. While it appears that there is consistently some residual cross-talk, a handful of baseline pairs are clear outliers, indicative of a poor cross-talk subtraction.

$|P(\tau,t)/P_N(\tau,t)|$ after redundant averaging, we find a clear pattern in the residual SNR in the cross-talk delay ranges (Figure 9). While we have already thrown out baselines with projected E-W distances less than 14.6 m (as was done in H1C IDR 2.2), we see that the remaining baselines with projected E-W distances less than 15 m show a pattern of cross-talk power in excess of noise.

This makes sense physically; these baselines have their main lobe closest to zero fringe-rate and so the cross-talk subtraction algorithm is the most conservative here for fear of signal loss. As was also the case in H1C IDR 2.2, the maximum fringe-rate $f_{\rm max}$ is given by

$$f_{\rm max} = 0.024\,{\rm mHz}\left(\frac{|b_{\rm E\text{-}W}|}{1\,{\rm m}}\right) - 0.28\,{\rm mHz} \quad (1)$$

where $b_{\rm E\text{-}W}$ is the East-West component of the baseline ($f_{\rm max}$ is also restricted to be less than 0.75 mHz. If the cross-talk is centered at 0 mHz but has some width in fringe-rate space, we expect that these baselines should be the first to show residual cross-talk as we integrate down, and that is precisely what we see. This was probably true in H1C IDR 2.2 as well, but the lower noise level in this data simply makes the systematics clearer.

I have therefore expanded the cut on projected E-W baseline from 14.6 m to 15 m based in the results in

Figure 9. This cuts out two columns of baselines and the last remaining 14.6 m baseline. This is unfortunate, because it is a big sensitivity hit. However, I would argue that it is the safest approach to this problem. It is a cut based on data from delays well-outside the region of interest for setting power spectrum limits. It is also a cut based on modifying an existing analysis choice in H1C IDR 2.2 rather than adding an entirely new cut, which insulates us somewhat from cherry picking.

I should also note that if there is going to be a future re-analysis of H1C IDR 3, perhaps importing back some of the analysis improvements from H4C, then figuring out how perform effective systematics mitigation on these short baselines is likely the lowest hanging fruit. Perhaps a more aggressive fringe rate filtering that is willing to incur more signal loss—and then calibrate it out via analytic projections and validation simulations—would be produce a more sensitive power spectrum upper limit.

3.5 Expanding the Wedge Buffer Cut

The final major change the power spectrum analysis is an increase of the "wedge buffer" parameter from 200 ns in H1C IDR 2.2 to 300 ns in H1C IDR 3.3. This additive buffer has a long history in the field, going back to Parsons et al. (2012), which suggested that a combination of foreground and beam chromaticity and the application of tapering functions in the delay power spectrum can extend power $\sim 0.15\,h\mathrm{Mpc}^{-1}$ beyond the horizon wedge.

The choice of 200 ns in HERA Collaboration et al. (2021) (equivalent to $0.11\,h\mathrm{Mpc}^{-1}$ at $z = 7.9$ and $0.10\,h\mathrm{Mpc}^{-1}$ at $z = 10.4$) was motivated by Figures 14 and 15 of that paper. As that paper states, "the exact buffer was chosen to mitigate foreground leakage beyond the horizon limit." In other words, the buffer value was picked to largely match the region in $k_\parallel$-$k_\perp$ space where the SNR was consistently much larger than 1. The idea was the balance allowing a small sliver of foreground-dominated data at low $k_\perp$ while not excluding too much noise-dominated data at high $k_\perp$.

I reproduce the equivalent of that Figure 15 (the cylindrically-averaged power spectrum and SNR for band 2) in Figure 10. Reducing the noise has somewhat increased the area where foregrounds and foreground leakage clearly dominates. This motivated me to increase the buffer as the simplest single-parameter change I could make to account for this. I wanted a new value of the buffer that roughly struck the same balance of low $k_\perp$ systematics getting through with high $k_\perp$ noise-dominated measurements getting excluded. As with previous parameter changes, I settled on 300 ns as a round number to avoid cherry picking precisely which bins get included and which do not. This produces at wedge buffer at $k = 0.15\,h\mathrm{Mpc}^{-1}$ at band 1 and $k = 0.17\,h\mathrm{Mpc}^{-1}$ at band 2, which is actually more in line with the value suggested by Parsons et al. (2012) and used in many HERA forecasts starting with Pober et al. (2014). This change effectively eliminates the first k bin measured in HERA Collaboration et al. (2021) and reduces the number of baselines going into two lowest remaining k bins.

3.6 Refining the Binning to Spherical k

I have adjusted how spherical k bins were defined. In H1C IDR 2.2, Δk was fixed to be $0.064\,h\mathrm{Mpc}^{-1}$. The goal of this value was to halve the native resolution of the power spectrum, binning two neighboring τ modes together into each spherical k. However, this value did not precisely match the the actual spacing in k (though it was picked to be quite close—within $0.0007\,h\mathrm{Mpc}^{-1}$). More importantly, centering the first bin at $k = 0$ leads to some values of fixed τ falling into multiple bins. I show how that binning sticks points at various $k_\parallel$ and $k_\perp$ into k bins in Figure 11.

In H1C IDR 3.2, the match for $\Delta k = 0.064\,h\mathrm{Mpc}^{-1}$ was worse, due to the new band definition. I have thus modified the notebook to pick the Δk independently for each band and to include precisely two τ bins in each spherical k for all k of interest. This yields $\Delta k = 0.0619\,h\mathrm{Mpc}^{-1}$ at band 1 and $\Delta k = 0.0709\,h\mathrm{Mpc}^{-1}$ at band 2. Then, to avoid the lowest k bin with data having only data from a single τ, I increased the bin center of the first k bin from 0 to $3\Delta k/4$. That change is downstream from the wedge buffer cut, so changing the buffer again might require revisiting the definition of the first k bin. The precise resulting binning is shown in Figure 12.

3.7 Minor Changes

Finally, there is one minor change to the spectrum analysis pipeline: an LST-branch cut was also added as a parameter for the power spectrum pipeline script.

4 What's next?

Once power spectra are formed for each baseline pair and the above analysis choices are made, the final cylindrically and spherically binned power spectra are produced within a Jupyter notebook.[5] For the purposes of reproducibility, it is my intention to release that notebook to the collaboration in this repo once the set of new analysis choices and cuts documents here achieves collaboration consensus.

References

Dillon, J. S. (2021). "HERA Memo #97: H1C Internal Data Release 3.2". reionization.org/memos.

Dillon, J. S., A. R. Parsons, and N. S. Kern (2021). "HERA Memo #104: A Physical Model for the H1C Cross-Talk Systematic". reionization.org/memos.

Dillon, Joshua S., Max Lee, Zaki S. Ali, et al. (Oct. 2020). "Redundant-Baseline Calibration of the Hydrogen Epoch of Reionization Array". In: MNRAS. DOI: 10.1093/mnras/staa3001. arXiv: 2003.08399 [astro-ph.IM].

HERA Collaboration, Zara Abdurashidova, James E. Aguirre, et al. (Aug. 2021). "First Results from HERA Phase I: Upper Limits on the Epoch of Reionization 21 cm Power Spectrum". In: *arXiv e-prints*, arXiv:2108.02263, arXiv:2108.02263. arXiv: 2108.02263 [astro-ph.CO].

Kern, N. S. and A. R. Parsons (2019). "HERA Memo #64: Refleciton Systematics in HERA H1C Data". reionization.org/memos.

Kern, Nicholas S., Joshua S. Dillon, Aaron R. Parsons, et al. (Feb. 2020a). "Absolute Calibration Strategies for the Hydrogen Epoch of Reionization Array and Their Impact on the 21 cm Power Spectrum". In: ApJ 890.2, 122, p. 122. DOI: 10.3847/1538-4357/ab67bc. arXiv: 1910.12943 [astro-ph.IM].

Kern, Nicholas S., Aaron R. Parsons, Joshua S. Dillon, et al. (2019). "Mitigating Internal Instrument Coupling for 21 cm Cosmology. I. Temporal and Spectral Modeling in Simulations". In: *The Astrophysical Journal* 884.2, p. 105. DOI: 10.3847/1538-4357/ab3e73. URL: https://doi.org/10.3847%2F1538-4357%2Fab3e73.

Kern, Nicholas S., Aaron R. Parsons, Joshua S. Dillon, et al. (Jan. 2020b). "Mitigating Internal Instrument Coupling for 21 cm Cosmology. II. A Method Demonstration with the Hydrogen Epoch of Reionization Array". In: ApJ 888.2, 70, p. 70. DOI: 10.3847/1538-4357/ab5e8a. arXiv: 1909.11733 [astro-ph.IM].

Parsons, A. R., A. Liu, J. E. Aguirre, et al. (June 2014). "New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at z = 7.7". In: ApJ 788, 106, p. 106. DOI: 10.1088/0004-637X/788/2/106. arXiv: 1304.4991.

Parsons, A. R., J. C. Pober, J. E. Aguirre, et al. (Sept. 2012). "A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature". In: ApJ 756, 165, p. 165. DOI: 10.1088/0004-637X/756/2/165. arXiv: 1204.4749 [astro-ph.IM].

Pober, J. C., A. Liu, J. S. Dillon, et al. (Feb. 2014). "What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization". In: ApJ 782, 66, p. 66. DOI: 10.1088/0004-637X/782/2/66. arXiv: 1310.7031.

Tan, Jianrong, Adrian Liu, Nicholas S. Kern, et al. (Aug. 2021). "Methods of Error Estimation for Delay Power Spectra in 21 cm Cosmology". In: ApJS 255.2, 26, p. 26. DOI: 10.3847/1538-4365/ac0533. arXiv: 2103.09941 [astro-ph.IM].

[5] Incidentally, all the figures here were also produced in that notebook, except Figure 1, Figure 2, and Figure 11.

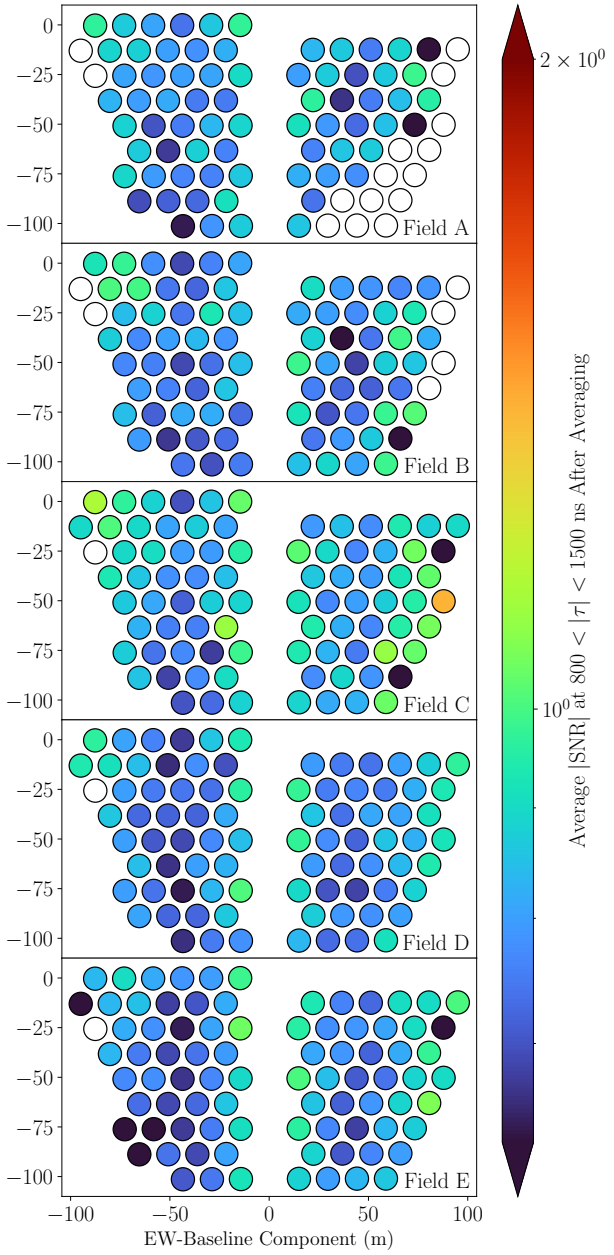

Figure 9: Baseline-group averaged residual SNR in the cross-talk delay ranges for band 2 data. Data consistent with noise should be roughly 0.63 (bottom of the color scale). The most important trend here is that the baselines with 14.6 m projected E-W distances show a clear pattern of residual cross-talk over the fields and over baseline groups with different projected N-S distances. Note that baselines with shorter projected E-W distances are excluded from the power spectrum calculation at an earlier stage because cross-talk subtraction cannot be performed without signal loss. Band 1 shows the same pattern. This motivates the cut on all baselines with projected E-W distances less than 15 m, which is more aggressive than in H1C IDR 2.2, which threw out baselines with E-W components less than 14.6 m. There are also some long baselines that show residual cross-talk, most of these have comparatively little observing time which might be affecting the cross-talk removal. For example, the brightest single baseline in this plot corresponds to only two unique baselines; it is not that unusual for both to have near-cutoff |SNR|. However, I feel that it is too dangerous to take this cut one step further and simply remove any redundantly-averaged baseline with high SNR.

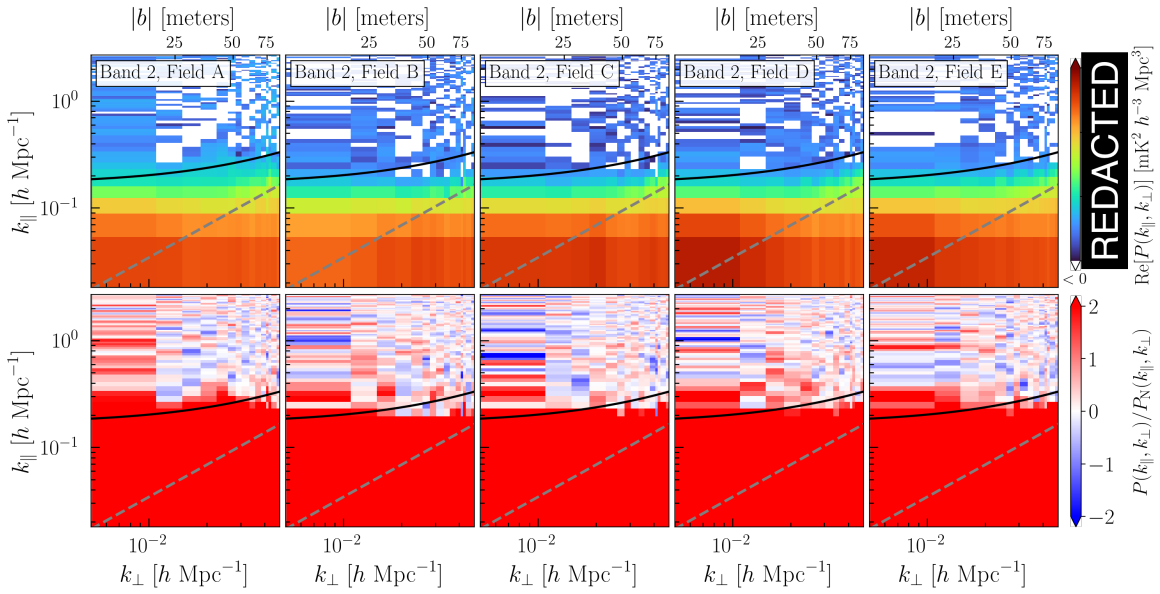

Figure 10: Here I show cylindrically-averaged $P(k)$ and SNR "wedge" plots for all five fields at Band 2. This plot is used to justify increasing the buffer (black line) beyond the horizon (dashed grey line) from 200 ns in H1C IDR 2.2 to 300 ns. I have redacted the color scale on the cylindrical power spectrum, since those results have not yet been released internally and since I do not want the specific $P(k)$ values measured to have any impact on the cuts.

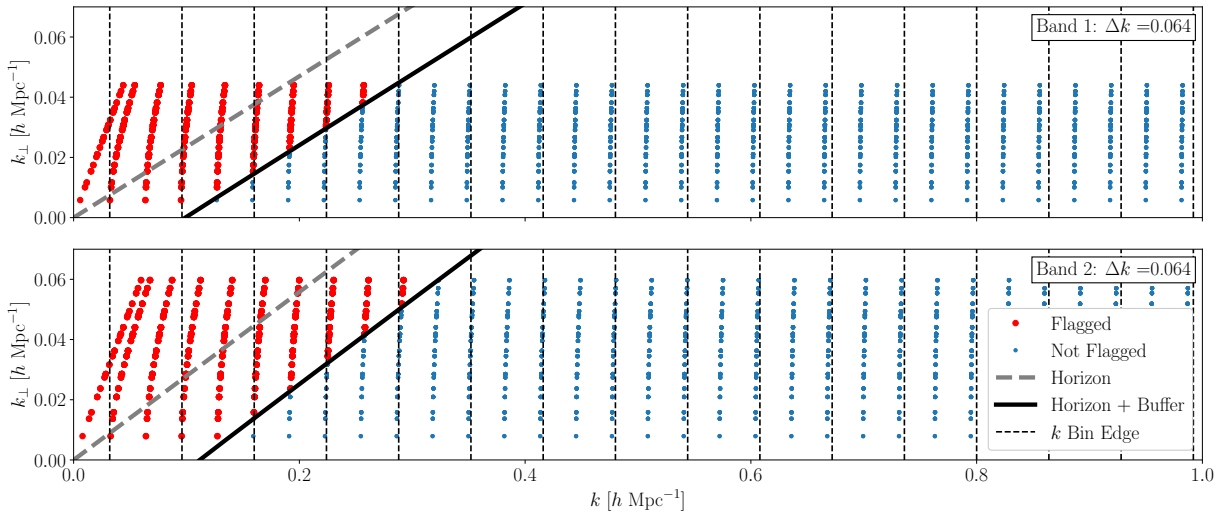

Figure 11: Here I show how spherical binning was performed in H1C IDR 2.2 for both bands. Nearly vertical stacks of points correspond to constant $k_\parallel$ or equivalently $|\tau|$. Points in red are flagged because they fall below the horizon wedge (plus the buffer). We can see that the choice of Δk is slightly mismatched to the natural spacing in τ, which is why I am proposing to change both Δk and the location of the first bin center (see Figure 12).

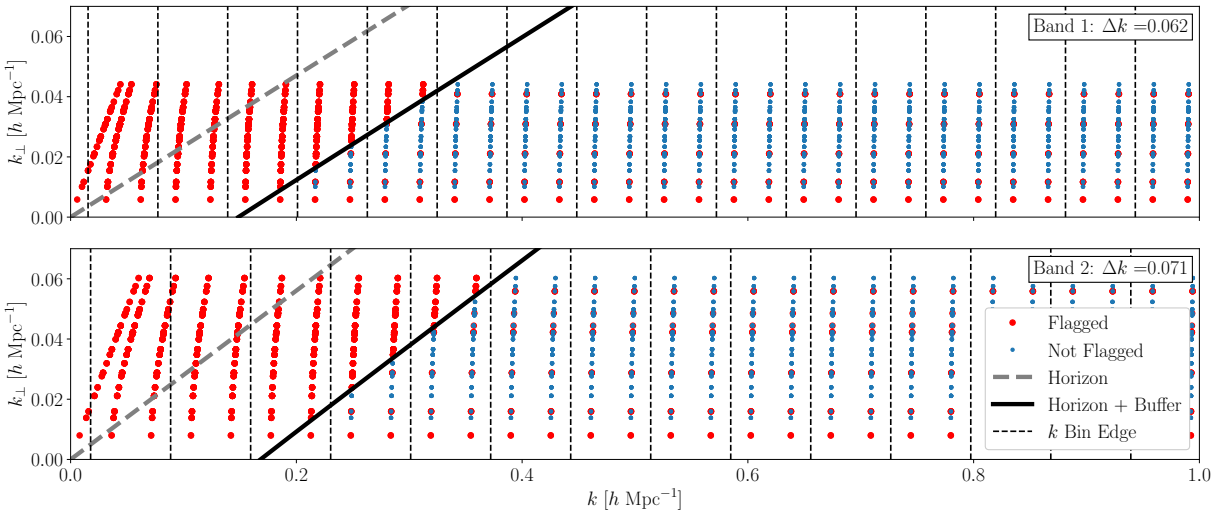

Figure 12: Here I show my modified spherical binning scheme for H1C IDR 3.3 (compare to H1C IDR 2.2 in Figure 11). Again, "vertical" stacks of points correspond to constant $k_\parallel$ or equivalently $|\tau|$. Points in red are flagged because they fall below the horizon wedge (plus the buffer) and/or because they come from baselines with a projected E-W distance of less than 15 m (see subsection 3.4). The increased wedge buffer is also reflected here (see subsection 3.6). The bin edges were picked to ensure that precisely two values of $|\tau|$ fall into each bin at the k values of interest. This required both making the Δk band-dependent and shifting the center of the first bin up to $3\Delta k/4$. While it is difficult to read how many data points go into the final k bins, given the overlap of points, if we count the number of unique baselines and $|k_\parallel|$ that go into each k bin, then for band 1 it's 0, 0, 0, 26, 94, 122, 122, 122, and so on. For band 2, it's 0, 0, 0, 26, 93, 120, 120, 120, and so on. The difference comes from precisely where the wedge and wedge buffer lie.