This paper provides a useful analytic framework to compute the contribution of common-mode LNA noise to visibilities in Hydrogen cosmology experiments. As a convenient analytical treatment to one of the major sources of contamination in 21cm experiments, I believe that this work deserves publication after several issues are addressed. I believe that most of the issues in the paper arise from lack of explanation and follow up on various systematics in the simulations and measurements. These issues fall under the following major themes which I will list here before diving into specific examples and requests.

- There is very scant discussion around how the measurements are performed with both the VNA and the telescope correlator including essential details like the experimental setup (diagrams of the signal chains), how calibration is performed, and how the data are processed. Without these details, it is very difficult for any reader to judge the trustworthiness of the presented measurements and the conclusions that the paper draws from them. I give specific examples /suggestions below.
- Measurements are presented without any indication of thermal noise or systematics error levels. Plots that allow the reader to judge the level of systematics and noise in these measurements need to be included. I give specific suggestions below.
- There does not appear to be any attempt within the paper to check the trustworthiness of the numerical simulations by altering solver / meshing properties aside from the S21/S12 symmetry check which appears to be inconclusive. If convergence cannot realistically be checked with the large full-array simulation, the authors should be able to check it with a smaller simulation of two dishes. I think the authors should seriously consider how important multi-dish reflections are by analyzing a simulation that they are confident does not have numerical artifacts with 3 vs. 2 dishes and if the 3rd dish only has a small effect they should just use two dishes instead of the full array rather than having an untrustworthy simulation.
- There are numerous instances where plots scales are not matched, making comparisons between different panels difficult (specific instances below).

After these issues are addressed I will be happy to recommend this paper for publication.

Abstract

- *Optional:* This is a stylistic suggestion so the authors are welcome to ignore it – I think it's helpful to summarize the paper's conclusions in the abstract. Right now the abstract describes what the authors did but it doesn't include any take-aways.

Section 1

- Second-to-last paragraph. I think the authors should be more precise about how cross-coupling ideas were developed in the various works cited here. Kern studied the reflections of sky signals between the two antennas in a single baseline while Josaitis and Fagnoni extended the model to include reflections off of all antennas in the array (not just the two antennas in a particular baseline). Fagnoni studied this using electromagnetic simulations while Josaitis studied this with a semi-analytic model.

- It might be nice to mention why we worry about over-the-air LNA cross-talk more than noise from other parts of the signal chain (I presume high reverse isolation?).

Section 2

- Second paragraph – please include a detailed drawing or photograph of the feed and dish design.

Section 4

- Please include detailed renderings of the 3d dish/feed models used in the simulations.
- I worry that a lot of the problems associated with the simulation reliability have to do with the IES solver accuracy setting (medium 1e-3) which might not be sufficient to faithfully reproduce -80 dB features along with the meshing. In the literature, I have not seen such structures in cross-coupling simulations before, even for large arrays. Even in this paper, the features are not present in Fig. 14 with the upgraded dish with even lower coupling. Have the authors checked whether the fine-scale frequency structures change when they change IES solver to higher accuracy?
- I think that the authors should be able to reproduce the IES results with one other technique (such as TDS) in a scaled down simulation (with 2-3 dishes). If they do not agree with the current settings, then they should increase the accuracy until there is reasonable agreement.
- If the IES simulation cannot be made to be trustworthy with the full-array simulation, the authors should use a 2-to-3 dish simulation with settings that are sufficient for trustworthy results. The settings should be chosen so that when the accuracy is increased, there is not a large difference in the new results and they should also be in reasonable agreement with another solver. I believe this is possible if the contributions from secondary reflections off of neighboring dishes is sufficiently small. This can be checked by comparing a 3 dish simulation to a 2 dish simulation (both with solver settings that have been checked to be trustworthy).
- Fig. 4 – use consistent y-scales between all the panels.
- Mention somewhere what the off-zenith angle is for NCP observations.
- Show 3d-renderings of the NCP and zenith observation configuration for a subset of the dishes so we can visualize what the light-travel paths are between the dishes taking into consideration things like how the lines-of-sight between feeds intercept with dish edges and other components.
- Section 4.3. Scattering matrices aren't always symmetric. I believe this is the case here but the authors need to spell out clearly why this is true in this situation such as citing a theorem or paper that explains that S-params are symmetric in this situation.
- In discussion of Fig 6. thhe authors note that the delay-domain S-parameters make "physical sense" even though they are not equal. They note that the delays of reflection peaks line up with inter-antenna spacings but they do not comment on whether the amplitudes of the reflection peaks make sense given the geometry of the array and the light propagation paths. For example, the amplitudes of the two peaks in 2V x 8V are switched between S21 and S12. Does this make sense? Why is the S21 in all simulations consistently larger than S12?

- Fig 4 (and 5?): There is a lot of fine-scale structure in these simulated S-parameters which may be unphysical owing to the time scales involved in the reflections. These fluctuations are comparable to the differences between S21 and S12. The authors should comment on the physicalities of these structures. Are they at physical delays? Do they change significantly with the solver settings?
- Consider making Fig 6 have a log y-scale so that we can better see how much the features decay with time. On a linear scale this is hard to read after 0.2 microseconds.
- Make a similar figure to Fig. 6 but for the zenith pointing.
- End of 4.3 – I think that the paper needs to get to the bottom of why the disagreement exists and the source of the fine-scale spectral structure. I think this is possible with the current available resources if they explore simulations with just 2 or 3 antennas. I'm happy to concede this if they explicitly show that having all 32 antennas in the simulation is more important to the accuracy then improving the current solver settings.
- End of section 4 – Without increasing resources, the authors can check the accuracy of the simulation of the full array by slightly lowering the accuracy / mesh resolution and seeing if the results change much.

Section 5

- In section 5, there is no citation for equation (23) and very little information on how the measurements of receiver temperature are obtained, just a very unclear short description of the measurement which does not have enough information to verify as legit and with no citation to a better explanation. Either expand this discussion or add citation to a paper with a good description of the measurement procedure.
- In equation 23, P is not defined.
- There is no discussion about the validity of these measurements. The authors describe a fitting procedure, they should have plots comparing the fitted models to data at the very least.
- Why are their negative values for the measurements of T_b in Fig. 7? From equation 15 it seems like T_b should be positive definite.
- Please include diagrams for the measurement setups described in section 5.
- In section 5, why do the authors think that S21 was smaller pointing at NCP?
- In section 5.2 – how were the measurements calibrated? Show a diagram please.

Section 6.

- Fig 8: the fine scale structures in the S21 parameters look like numerical noise, related to previous discussion, the authors should track down whether this is the case and try to mitigate it. Can the authors confirm that the ripples in th measurement panel are not sourced by cable reflections and uncalibrated spectral structure? Show a measurement where the measurements are terminated by 50 Ohms with nothing else changed so we can see where the systematics floors are.
- Please include some discussion on the data-reduction involved with producing the nightly mean plots (you can cite Wu 2021 but i think it'd be helpful for the reader to at least know the basic parameters of the calibration and observations).

- The authors should also include information on the signal chain such as the various stages of amplification, filtering, digitization, and correlation.
- Fig. 11 – Need to show thermal noise level so reader can understand the measurement uncertainties. It would also be helpful to compare the means from subsets of nights so the reader can understand how time variable these residuals are.
- Explain how you converted visibilities to temperature units (or cite work on this).
- Fig 11 y-axis label is pixelated / does not show up clearly.
- Can the authors confirm that in Fig 11. The orange and blue lines are the predicted visibility based on VNA measurements and CST simulations? If they could clarifying this in the legend of Fig 11 I think this would make things clearer. Also clarify whether or not there is any time averaging in Fig. 11 (I assume there doesn't need to be).
- In Fig. 12 legend, the authors should clarify that the blue lines are the averaged visibility data (from the correlator) and the orange lines are VNA measurements.
- Please provide details about the sky map such as the specific source catalog, the model of diffuse emission, and the software used to perform the simulation.
- Please provide details on the beam model used beyond "crude beam model" is it analytical? If so, give the formula. If it's a simulation, please say how it was simulated.
- There is a lot of fine scale spectral structure in the averaged visibilities that is not present in the averaged sky model. Can the authors comment about what the sources of these structures might be? For example, could they be uncalibrated cable reflections? Are they errors introduced by the calibration algorithm / processing?
- It looks like both the averaged visibilities and the VNA noise measurements have significant 5 MHz ripples which correspond to roughly to round-trip travel times in 30 meter coax cables. Does the signal chain contain coaxial cables? What attempts were made to calibrate these structures out in both the VNA and visibility measurements?
- How well are averaged visibilities reproduced from night to night?

Section 7

- In Fig. 13 it looks like the cross-coupling in the new dish design is much smoother than the simulated cross-coupling in the current TDPA design. The authors should ideally make sure the settings for the solvers are as close as possible. If this isn't possible, give more information on the solver used for the update vs the current TDPA design.
- In Fig. 14, is the red line for the cross-talk derived from the VNA measurements or simulations? Both of these seemed to have a lot of spurious spectral structure either from numerical simulation errors or calibration artifacts / systematics. The authors should discuss whether these need to be taken into consideration when using this work to predict future performance.
- It might be helpful to show the updated design in Fig. 14 but I also think that it makes sense to leave this out since it could be a spoiler for the upcoming work focused on the improved design. Maybe mention that the performance of the updated dish design will similarly be considered in detail in Podczerwinski in prep.

Section 8

- The authors conclude that because the cross-coupling does not dominate mean nightly visibilities that we don't need to consider it at this stage but Fig 14 suggests that it is a major obstacle to an HI detection. It also seems to be several orders of magnitude greater rather than “same order of magnitude” that authors mention. I think this needs to be clarified in the text.

- A conclusion the authors reach is that the cross-talk has high chromaticity but we don’t know how much we trust whether this chromaticity is real since we don’t know whether it is numerical artifacts (in simulations) or calibration artifacts (in measurements). The authors should discuss whether they think we can trust the predictions of fine-scale structure – rule out the possibilities of calibration artifacts or numerical noise or say that there needs to be further investigation to check what the true chromaticty of the cross coupling is.