Dear referee and Editor George Jacoby,

Thank you to the referee for pointing out the mistakes, confusions in our manuscript. We agree with the points and made the following changes (in numerical order)

1. *[I had some trouble understanding the flow of the explanation as the equation numbers used in the text do not match the equation numbers behind the equations. …]*
Thanks the referee for pointing this out, we fixed the typo.

2. [*Figure 9 shows scatter plots of three different quantities against each other, presented in a matrix form. The plots on the diagonal show histograms and not scatter plots. The labeling implies that these are plots of the three quantities against themselves. This is not the case. I suspect that these are PDFs of the quantities but the labeling of the axes should make this clear as should the caption.]*
Thanks to the referee for pointing out this confusion. The diagonal plots are indeed histograms of the respective quantities and we have indicated as such in the caption and in the text.

3. *[Eq. 28 gives an expression for \hat{P}(k) but in this expression, there is an integral over all k. How can the result then be a function of k?]*
Thanks to the referee for pointing out this typo. The left hand side should be a function of k_{\tau, \bar{b}}, as defined in Eq. (14).

4. [*Fig.6 shows the window functions. For the higher l, the window functions spread out to both lower and higher k values. In the text, it is first written that tails are stronger for lower k modes where the foregrounds reside. However, then it goes on to state that foreground leakage to higher k modes is more pronounced. These two statements are in apparent conflict with each other, perhaps it would be better to formulate this in another way.*
There is also an "the our proposed .... estimator" some sentences along where either the or our should be removed.]
Thanks to the referee for pointing out the possible confusion. It turns out that there is no contradiction here: for window functions *centered* at higher \ell, the tails extend to lower k. This means that when one attempts to make measurements at high \ell, one is actually 'dragging up' power from low k (where the emission is foreground dominated). In other words, when attempting to measure power at some intermediate k, the tails of the window function may extend to lower k's, so that lower k modes masquerade as high k modes. Power has thus leaked upwards on the plane, causing the wedge. We have made minor edits to the text in the final paragraph of Section 4.2. to clear up our explanation. We have also fixed the typo pointed out by the referee.

5. [*Fig. 10 shows the benefit of the proposed scheme. However, this figure shows that if one takes only the equivalent baseline-pairs (dashed lines), one gets between 80 and 90% of the sensitivity. In other words, the nearly equivalent baselines only contribute at the 20 to 10% level. For HERA350 it's closer to 20% and one can argue whether this increase is really significant. What puzzles me is the statement that the PAPER results until now only use the theta_min~1 case. Looking at Fig. 10 this would imply it's operating at 20% of the achievable sensitivity AND it is not using all the equivalent baseline pairs. Surely PAPER is using all these equivalent baselines as there seems to be no reason not to do so. This would imply that PAPER128 uses theta_min~0.4 or so. Could the authors clarify this?]*

Thanks to the referee for pointing out this confusion. The point is correct that we find 80% to 90% sensitivity to be achievable with all the equivalent baselines. Currently the PAPER64 results have indeed only used 3 of the equivalent baselines due to safety precautions on the systematics at the time. As the referee correctly pointed out, if PAPER128 uses all the equivalent baselines, the sensitivity achieved would be roughly comparable to a cutoff at theta_min~0.4 or so.

6. [*The other question raised by Fig. 10 is where this increased sensitivity is contributing. Which k-modes benefit from this? Probably mostly the low k-modes with relatively little benefit for the high k-modes? It would be good to show this as well.]*
Thanks to the referee for this suggestion. The k modes benefitting from the sensitivity boost is determined by the length of the baselines involved. Intuitively the near-equivalent baselines of PAPER128 range over all lengths, while that of HERA are mostly longer baselines (the shortest baseline of HERA need to rotate 60 degrees to overlap with each other). We see this trend in Fig. 9, where for example the high \widetilde{\Theta} contributions of HERA-350 are peaked for baselines of length ~100 meters. We have added comments in the text regarding this distinction.

7. [*When the text mentions PAPER(128) it is described in the present tense suggesting it is still operational. I had understood that it has been decommissioned because of the HERA construction. If so, it would be better to use past tense when describing PAPER.]*

Thanks to the referee for this suggestion. We have changed the discussion of PAPER128 to past tense.

8. In addition, from the feedback we received from the community, we expanded the discussion of related works in the introduction.

Sincerely,
Yunfan Gerry Zhang, Adrian Liu, Aaron Parsons