We thank the reviewers for their positive feedback and insightful criticism. We first address some general comments appearing in more than one review and then proceed with addressing the additional comments of each reviewer.

Clarity of the technical parts of the paper. We agree with the reviewers that the proofs of our results in the submitted paper would greatly benefit by adding beforehand summaries of the high level ideas and proof sketches, at least for our main theorems. We commit on doing this for our revised version of our work. Specifically, upon acceptance, we plan on devoting part of the additional page for the camera-ready version of the paper to adding high-level proof summaries. We do agree with multiple reviewers that the arguments used are, for the most part, rather simple and indeed such an adjustment will hopefully make this (potentially appealing) aspect of this work more clear.

Applications of our results beyond sparse tensor PCA We thank the reviewers for asking whether our results for the Gaussian additive model can be applied to other models, beyond sparse tensor PCA. We first wanted to highlight that the reason we decided to focus solely on our application on sparse tensor PCA was simplicity; this is an extremely well-studied inference problem which is also easy to state, and it gives a clear example of the applicability of our main theorem to obtain tight results for all $k = o(p)$ and all $d \geq 2$.

As part of our ongoing work, we have also proven that our results in the submitted paper do imply the all-or-nothing phenomenon also for the k-Gaussian submatrix localization problem (a Gaussian version of the well-studied k-stochastic block model) [Banks et al. '18] when $\omega(1) = k = o(n^{\frac{1}{4}})$. We strongly believe that our method can be applied to establish the all-or-nothing phenomenon for various well-studied Gaussian additive models in the literature.

Techniques: simplicity and novelty Our second-moment-method techniques are indeed similar to ones existing in the literature (e.g. [Banks et al. '18, Perry et al. '20]); however, these works do not address the all-or-nothing nature of the recovery threshold. By contrast, our techniques are quite different from the statistical-physics-inspired tools of [Barbier et al '20]. We consider our ideas “simpler" as we do not attempt to fully characterize the limiting free energy of the model, but instead prove the all-or-nothing phenomenon directly via our characterization combined with the second-moment method. We commit to making this comparison clearer in our revision. We agree emphatically with Reviewer 5: our main aim in this paper is to present a clean and sharp condition equivalent with the all-or-nothing property. Once the statement of the theorem is guessed, the proofs follow the standard outline.

Reviewer 1.

Numerical simulations We thank the reviewer for suggesting numerical simulations to validate the theoretical findings. Unfortunately, such simulations are in our context prohibitive for a potentially fundamental reason, namely, that sparse tensor PCA is widely conjectured to be computationally hard at the information-theoretic threshold; that is, polynomial-time methods seem to require a much larger λ_N to work. As a corollary, simulating the performance of the Bayes-optimal estimator is potentially only possible with exponential-time methods around the critical $\lambda = 2 \log M_N$.

"The content of this work seems to be on the light side" As mentioned by Reviewer 5, the main contribution of this paper is to establish that the all-or-nothing phenomenon is equivalent to an explicit condition on the Kullback-Leibler divergence. In that sense, the goal of this paper is conceptual simplicity.

Reviewer 3.

Key messages demonstrated in prior work for sparsity $k = \Omega(p^{2/3}), k = o(p)$ We respectfully disagree with the reviewer that the all-or-nothing phenomenon for sparse tensor PCA for sparsity $k = \Omega(p^{2/3}), k = o(p)$ has been demonstrated in previous work. This has been done solely in the sparse *matrix* PCA case, i.e. when $d = 2$, in [Barbier, Macris '20]. In our result we establish it for all $k = o(p)$ and any $d \geq 2$.

More examples in Section 2 We wholeheartedly agree with the reviewer that adding more examples below our theorems would help the reader develop important intuition about our results. The reviewer also asked the interesting question of whether M_N uniform random points from the sphere would satisfy the overlap condition of Theorem 2. While we are not sure if this is the case, we can prove that as long as $M_N \to +\infty$, the all-or-nothing phenomenon nevertheless still holds holds for this example with high probability, by directly bounding the KL divergence and using Theorem 1. We thank the reviewer for raising this nice example, and we plan to add it to a revision to emphasize the applicability of Theorem 1.

Tightness of Theorem 2 We thank the reviewer for this question. Theorem 2 is not tight, as one can show directly via Theorem 1 that the all-or-nothing phenomenon holds for sparse tensor PCA when $d = 1$, yet the overlap condition fails. We plan to add this comment after the theorem in the revised version of our paper, to help the intuition of the reader.