We thank the reviewers for their kind comments, and for their consensus view that our theoretical results on TV modulus of continuity (Theorem 1) and statistically optimal robust learning of degree bounded Ising Models (Theorem 4) are novel, non-trivial and interesting. We are also thankful for the reviewers' concrete suggestions on improving the draft, which we will incorporate in the final version of our work. We begin by addressing two common concerns raised by reviewers before addressing specific questions raised by each reviewer.

Computational Efficiency. We agree with the reviewers that our proposed estimators are not computationally efficient. In particular, our estimator proposed in Equation 10 runs in exponential time due to the sizes of the coverings that the optimization problem is defined over. However, note that in stark contrast to many other problems studied in robust statistics, even ignoring considerations of computational efficiency, basic statistical questions of the fundamental limits of robust estimation for Ising models are not well-understood. Our main contribution is in giving statistically optimal estimators, thereby, establishing tight information theoretic rates for estimation of Ising models under contamination. Designing polynomial-time estimators for robust learning of Ising models that are also statistically optimal is certainly an interesting open problem, and we leave that for future work.

Dobrushin's condition. We work in the high-temperature regime i.e., $\max_{u \in V} \sum_{v \in V} |\theta_{uv}| < 1$. Note that while this may seem restrictive, this assumption is widely popular for studying Ising Models, for example, see related works in statistical physics [**2**, **5**], mixing times of Glauber Dynamics [**3**, **1**], correlation decay [**4**], and more recently in estimation and testing problems [Daskalakis et al. (2019), Dagan et al. (2020)]. Note that in Appendix A, we also provide preliminary results for Ising models which don't satisfy the Dobrushin condition.

AR1: *Related Work and TV-Cover.* We thank the reviewer for pointing us to a recent work by Devroye et al, and in particular, for providing an alternate exponential time estimator for $G_{p,k}$ based on TV-covers. We will update our manuscript add the relevant citation and also add discussion of this alternate exponential time estimator.

AR2: *Inconsistent Estimator.* We agree with the reviewer that our estimator doesn't recover the true model even in the infinite sample limit. However, note that this is not a limitation of our work, and is true in general (even for the most basic problem of mean estimation) in Huber's model. In particular, as shown in our lower bound (Lemma 1) , there exists two Ising models whose TV distance is $O(\epsilon)$, but the parameters are $\Omega(\epsilon)$ far apart. This shows that in the contaminated setting, one cannot hope to recover the exact model parameters. The interesting question then becomes whether we could recover the parameters up to this unavoidable bias due to Huber contamination: which we show our estimators are able to do.

AR3: *Model Width as relaxed parameter.* We thank the reviewer for raising this subtle issue. Note that under the high-temperature regime, the model width is upper bounded by 1, thereby, making it a constant. However, in Appendix A, we present additional preliminary results for Ising models with unrestricted model width ω, and seek optimal dependence on ω as well.

Applications of Modulus of Continuity. We believe that our modulus of continuity analysis can improve several existing results. For example, in the uncontaminated setting, our result directly improves the sample complexity analysis of the maximum likelihood graph decoder in Santhanam and Wainwright (2012). Moreover, we believe that our result can also improve testing of Ising Models, either by improving the sample complexity of existing symmetric KL-divergence based algorithm of Daskalakis et al (2018), or by using the contrapositive of our Theorem, to potentially design new tests based on ℓ_2 error of neighborhood vectors.

AR4: *Comparison to Lindgren et al.* Lindgren et.al. work in a more general setting, but do not distinguish between high and low-temperature regimes. In particular, they provide an estimator which achieves an upper bound of $O(\sqrt{\epsilon})$ and a lower bound of $\Omega(\epsilon)$. In contrast, our results show improved error rates in the high-temperature regime, and match their results in the low-temperature one (See Appendix A).

We thank the reviewer for their suggestions and will certainly include a more detailed exposition of known information theoretic limits of learning Ising models in the uncontaminated setting, highlighting the dependencies on λ, ω, k/d.

Additional references:

(1) De Sa, Olukotun and Ré. Ensuring rapid mixing and low bias for asynchronous Gibbs sampling, 2016.

(2) Dobrushin and Shlosman. Completely analytical interactions: constructive description, 1987.

(3) Kulske. Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures, 2003.

(4) Kunsch. Decay of correlations under Dobrushin's uniqueness condition and its applications, 1982.

(5) Stroock and Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice, 1992.