We thank the reviewers for their valuable feedback. We are pleased that they recognize the novelty of our theoretical results (R1), find the universality of the lower bound interesting (R3), and believe that prior work is described well (R1).

Clarity (R2): Thank you for pointing out the need for more emphasis on our setting, such as the consideration of only single node interventions. We will emphasize this in the abstract and more clearly throughout the paper.

We will extend our comparison to previous bounds and emphasize our bound has a unique flavor, since it is *instance-wise* rather than minimax or an average over a MEC. As mentioned on line 292, Hauser and Buhlmann (2014) and Shanmugam et al. (2015) prove minimax bounds, which both involve the size of largest maximal clique. For single-node interventions, the bound of Shanmugam et al. (2015) specializes to exactly the size of the largest maximal clique - a trivial result, since the MEC contains a DAG with the largest maximal clique upstream of all other nodes. Meanwhile, our bound is much stronger: regardless of the position of the largest maximal clique, it is a fundamental barrier to structure learning; moreover, the bound over *all* cases is only a factor of two smaller than the bound on the *worst* case.

R2 points out the paper might appear dense and not accessible to non-expert audiences. We have strived to make the paper as clear and self-contained as possible, e.g., by providing several examples and an overview of all required concepts, including standard graphical models and graph theory definitions. On the other hand, we do appreciate the recommendations that will make it even more accessible and streamlined. For example, we will clarify which edges are learned by a single-node intervention: these are the edges incident on the intervened node, as well as any logical implications of these edges (via the application of Meek's rules, reviewed in Appendix A). Regarding the dense notation, in the main paper we have tried to introduce only notation and definitions necessary to state the main results and give an intuition of the proofs. In fact, any apparent simplicity in the proofs is partially the result of careful definitions.

Novelty/Applicability (R1, R2): We disagree with the claims that this work is incremental or narrowly interesting. As pointed out by R1, this work contains several novel theoretical results. Most importantly, our characterization of verifying intervention sets is a clear step toward a better understanding of identification of causal structures, which has numerous applications. Many of our results are potentially useful in more realistic settings, e.g. with larger interventions. Furthermore, our results, especially the residual characterization, may serve as a template for similar results in the presence of latent variables or cycles.

Experiments (R2, R3): As suggested, we will report the running time of the policies on large graphs that are not tree-like. As a preliminary result, we find that for 80-node graphs generated in the same manner as our smaller graphs, DCT is roughly 30% faster than Coloring. However, both policies take more than an order of magnitude longer on these denser graphs. The comparison between DCT and Coloring in Fig. 6c,d is direct, we ask R2 for further clarification.

(R1.1) Directed cycle definition: Thank you for pointing this out; the definition of a directed cycle should not include asterisks on the left-hand sides of each edge. The new definition will correspond to the standard one for chain graphs.

(R1.2) Intervention policy definition: We will clarify that a policy is a map from *interventional* essential graphs.

(R1.3) Definition 2: See Line 62: * is a wildcard; we will recall this usage in the definition to bolster clarity.

(R1.4) ic-ratio: The ic-ratio of a policy on a DAG measures how many interventions the policy uses to orient that DAG, relative to the minimum number of interventions that can orient it. A policy is optimal on an instance if the ic-ratio is 1.

(R2.1) $m(D)$ computation: We compute the actual $m(D)$ in our experiments. An efficient method is described in Algorithm 4 of the Appendix; we will add a reference to the algorithm in Section 5.

(R2.2) Appendix 591-593: We will add an explanation. Briefly, in a clique, each pair of nodes must have at least one member intervened to establish the orientation of the edge between them. The smallest set satisfying this criterion has size $\lfloor \frac{n}{2} \rfloor$. This is formally proven in Shanmugam et al. (2015), Theorem 4, where n is assumed even for simplicity.

(R2.3) Appendix 613: Thank you for pointing this out; the reference should be 2014.

(R3.2) $C_{\max}$ definition: Thanks for the correction. $C_{\max}$ should be a number, as in Thm 3, not a set. $C_{\max}$ is maximum number of maximal cliques in any chain component of the essential graph. As mentioned in lines 228-231, $\log n$ scaling is optimal for trees (Greenewald et al., 2019). On the other extreme, if the essential graph is a single clique, the log factor disappears. To the best of our knowledge, no work has established approximation factors before.

(R3.2) intersection-incomparability: We will add a discussion of this condition. First, intersection-incomparable chordal graphs have been introduced under the name "uniquely representable chordal graphs" (Kumar and Madhavan 02), which we will mention more clearly, and include familiar classes of graphs such as proper interval graphs. Second, while the assumption is necessary for our analysis of the DCT policy, the policy still performs well on graphs that do not satisfy it, as demonstrated by our experiments, suggesting that the assumption is not too restrictive in practice.