We thank the reviewers for their insightful comments. We first respond to the common concerns across the four reviews before addressing specific reviewer feedback.

Strength of results, hyperparameters, additional experiments: All reviewers observe we achieve only modest gains compared to the log-uniform spacing in terms of final log likelihood. We emphasize that while this is true, the main benefit of our approach is in avoiding the costly grid required by the log/linear-uniform spacing schedule. To respond to R4's request, as measured in wall-clock time, our GP-bandit schedule takes approx. 12 hrs to train a VAE on MNIST compared to the approx 160hrs training time for the grid searched log schedule (8 hrs/run x 20 runs). We will update the main text to make these considerations clear, and add an additional appendix section benchmarking our schedule against baselines in terms of total wall-clock time.

To respond to R4, all results were obtained from a single seed. We will update the figures in the final draft to average results across five seeds, and have included preliminary results on MNIST in Figure 1. As per R1's request, we will also include the training loss in the final version, which closely mirrors the test loss curves.

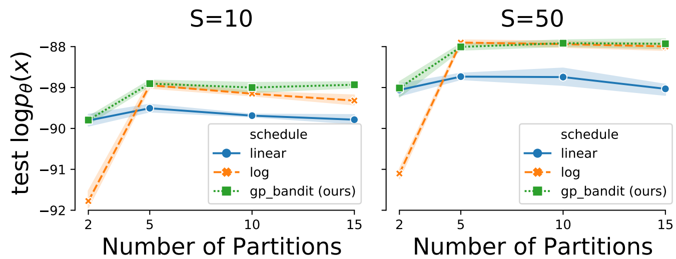

Figure 1: VAE MNIST, 6k epochs, 5 seeds

R2 raises a concern about the number of hyperparameters in our method. The GP-bandit introduces three additional hyperparameters which can be learned directly from data by maximizing the GP marginal likelihood, and therefore require no additional hand-tuning by practitioners as discussed in Appendix B. The exploration-exploitation trade off parameter κ_t is set using Theorem 1.

Figure 2 clarity: We agree with R2 and R4 that a more thorough description of both Fig. 2 and Fig. 3 are needed. In Fig. 2, we investigate the bandit exploration / exploitation behavior at early, middle, and late stages of training by showing where the bandit positions a single β_1. Color encodes timestep, so clusters of similar colors indicate the bandit is "exploiting" a particular region, particularly in regions where the blue line, i.e. GP mean, is high. This indicates that our reward model expects the TVO objective will improve with this choice of β_1. We also show the variance of our GP surrogate model across training phases and see both the mean and variance of predicted reward decreases as optimization converges.

The key takeaway from this figure is that we see the bandit exploits early on in training, and that the surrogate reward function correctly learns that the reward for choosing the correct location decreases as optimization proceeds and the curve flattens.

Figure 3 clarity: R2 correctly observes that it is difficult to know the shape of the integrand. We agree, and this is in fact one of our primary motivations for using a bandits schedule, as bandit optimization is uniquely suited to scenarios where one has little knowledge of the form of the reward function. We show possible example shapes for the integrand in the subpanel of Fig. 3, reflecting bandit choices of β in the middle (orange) and late (green) stages of training. These are based on the intuition that β choices should be concentrated in regions where the integrand is changing quickly, allowing the left Riemann approximation to capture the most area with a fixed budget of d partitions. We also assume that a perfectly flat curve will result in uniform β choices. We will update the caption of Fig. 3 to make this clear.

(R4) Comparison with Bogunovic et. al [5]: We agree that the positioning of our contribution with respect to [5] requires further clarification, and regret this oversight. Theorem 1 improves the bound on the maximum mutual information gain $\tilde{\gamma}$ from [5] by using Cauchy Schwarz and Jensen's inequality rather than an analysis of the optimality conditions (cf. eq. 61 in [5]). We have updated the main text in 4.3 to make it clear that Theorem 1 follows by using our tighter bound on $\tilde{\gamma}$ in Theorem 4 of [5], and simplified the derivation in the appendix to refer to [5] directly where possible.

(R4) Discussion on spatial covariance: R4 asks for clarification on covariance functions which can be used alongside our projection operation. Since the projection preserves the input space, any PSD covariance function will maintain this property after projection. We recommend choosing a kernel for which bounds on the maximum information gain are known, such as exponentiated quadratic, Mattern, or linear from Srinivas et al 2010 [35].

Typos: R1, R3, R4 helpfully point out a number of typos and suggestions to improve the writing which we will incorporate into the final draft.

(R4) Ref [43] in the supplement: Martin J Wainwright. "Basic concentration bounds." In *High-dimensional Statistics: A non-asymptotic viewpoint*. Chapter 2, pages 21–57. 2019