We thank all the reviewers for the positive feedback and thoughtful comments. Below we address all the comments and questions in order.

On the concern of reviewer 2 and reviewer 3 about the novelty and efficiency of DDLS:

0. We ran additional experiments to include the FID score progression during the MH/Langevin dynamics, with direct comparasion to MH-GAN baseline. We adopt the same WGAN architecture in our own implementation and report FID score for each 20 iterations on CIFAR-10, up to 640 iterations following the same setting in MH-GAN. We see that DDLS in latent space enjoys much faster mixing in the first 200 iterations.

1. Although the stationary distribution which we are sampling from in the vanilla GAN case is the same as DRS/MH-GAN, this stationary distribution was intractable before our work. Both DRS and MH-GAN had to use a proposal rejection/acceptance scheme. Our work proves that the distribution was indeed tractable in the latent space, with a clear and insightful enerygy-based model (EBM) formulation. This is our main technical contribution. The Langevin sampling scheme is only made possible thanks to tractable EBM formulation. So we think it's unfair to say that our contribution is just to replace MH sampling in MH-GAN with Langevin dynamics. We also extended the formulation to other GANs such as WGANs and shows its efficiency in our experiments.

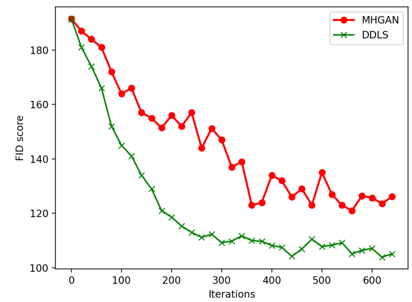

Figure 1: Progression of FID Score with MH/Langevin dynamics sampling steps, averaged by multiple runs.

2. In theory, DRS and MH-GAN samples from the same stationary distribution as ours, however in practice, independent sampling schemes such as DRS or MH-GAN can be very inefficient, and the number of steps required to move the sampler distribution to the stationary distribution can be too large for any reasonable computing resources. For example, consider training a generator on the MNIST dataset, where the generator produces 0.1% of number "0", when real data contains 10% of number "0" (this is common for GANs which are not very good at balancing different modes). Then if we use MHGAN to generate 100 numbers to simulate the real data distribution, in which 10 numbers should be "0", you have to generate 10,000 and reject 9,000 of them, even if these samples are perceptually good. In realistic cases where only a finite number of samples can be used, this inefficiency will seriously hurt the resulting distribution sampled by MHGAN. However, with our method, the gradient of the discriminator can guide the Langevin dynamics to move towards 0, which can be much more efficient. As shown in Figure. 1, we compare the FID of different sampling steps of the three methods on CIFAR-10. The FID score of our method goes down quickly with just a few MCMC steps, while for MH-GAN, the score goes down slowly, and stabilized at a much higher value. This observation confirms our claim that although in theory MH-GAN can achieve the same distribution of our method, it may take unbounded sampling steps to achieve that. So in practice the sampling distribution is much worse than in our work.

Other concerns:

Reviewer 1: Thanks for the valuable comments! Our method does need a generator pass to get better z. We will improve the writing of the paper to make it clearer. We will also include more discussions about the related works mentioned.

Reviewer 2: On the WGAN assumption, p_t and p_g can be close if p_g is close enough to p_d, this can be achieved after the training. Note our definition of $p_t = p_g(x)e^{-E(x)}/Z$ is a product distribution which is different from $p_t(x) = e^{-E(x)}/Z$ in [1,2]. In their definition, p_t and p_g may not be very close. Also, in LOGAN, the training explicitly takes gradient steps in latent space, which makes their distribution of negative samples more close to p_t. Please also refer to R.3 below for the FID score issue.

Reviewer3: Images with large resolution requires extra computational resources which we didn't have at the moment of submission, but we will add them in the next version. We have discussed why our method is much more efficient than MH-GAN in detail above. We didn't include the FID score on CelebA since the lack of FID scores in baseline methods, which makes direct comparaions impossible.

We didn't claim doing MCMC in the pixel space is not possible at all, but it is inefficient and hard to tune. Pixel-level EBMs are very sensitive to hyperparameters, needs a bunch of training tricks, very slow to train, and their performance is not comparable with GANs and our model.

Reviewer4: The main motivation of considering $p_t = p_g(x)e^{D(x)}/Z$ is to find an approximate energy-based model which is close enough to the actual WGAN model. In this way we can do latent space MCMC and get a distribution which is much closer to the data distribution than the original generator distribution. We will revise the paper to make it more clear, thanks for your comment.