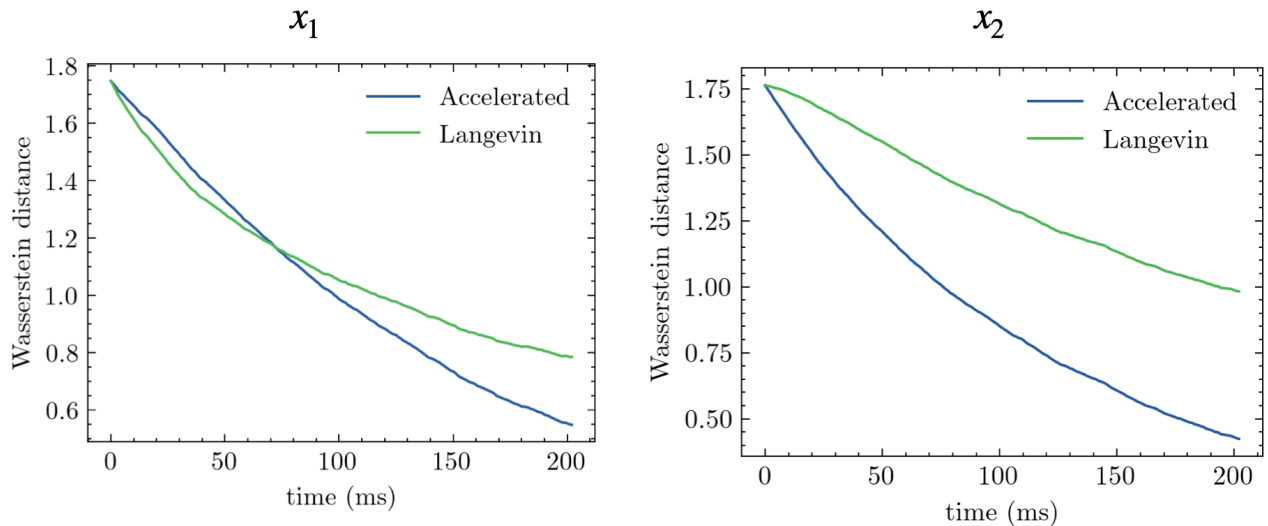
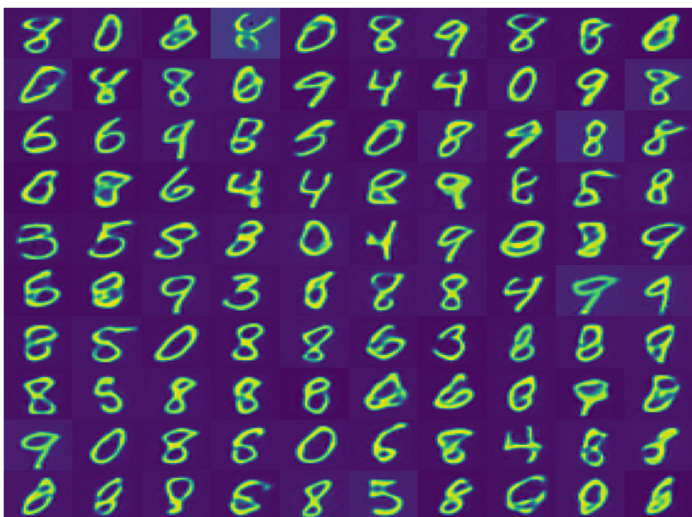


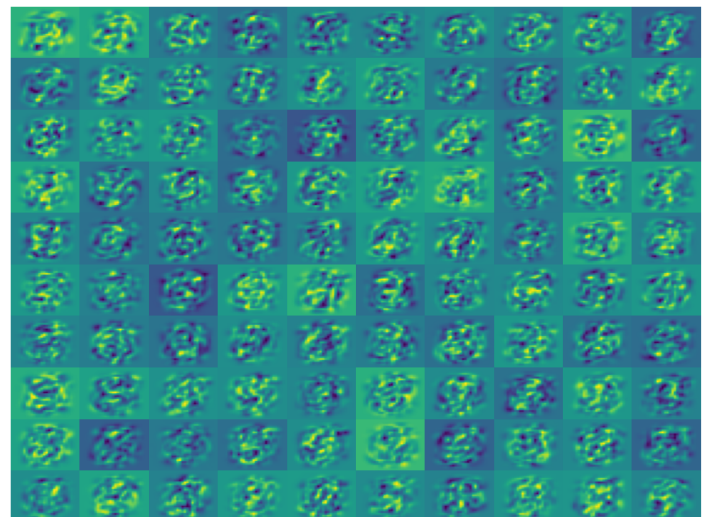
(S1) same experiment as in figure 1 of the paper but now neurons are using tanh activation function. It is clear that SO-FR and SO-SC learn score functions that are symmetric w.r.t. some point, but RSN is not constrained in such way and learns a better score function, leading to better sampling performance.



(S2) See our response to reviewer pX3M weakness 2, where we propose a new score matching loss to accelerate the sampling speed of RSN. We test this idea by training the RSN to sample from the same mixture distribution in Figure 2. The figures shows the Wasserstein distance between the marginalized distributions of generated samples and true samples for both dimensions



(S3) Samples produced by our reservoir sampler when trained to fit the latent distribution produced by a non-linear convolution autoencoder from MNIST. The distribution was 32-dimensional. The reservoir sampler had 500 hidden neurons and was trained for 400 epochs.



(S4) Samples produced by our reservoir sampler when trained to fit the latent distribution produced by a sparse coding model from MNIST. The distribution was 3136-dimensional (four-time overcomplete basis learned by sparse coding). The reservoir sampler had 30000 hidden neurons and was trained for 500 epochs.