

Table 5. Timewarp loss weighting factors

Dataset	$\mathcal{L}_{\text{lik}}(\theta)$	$\mathcal{L}_{\text{acc}}(\theta)$	$\mathcal{L}_{\text{ent}}(\theta)$
AD	0.99	0.01	0.1
2AA	0.9	0.1	0.1
4AA	1	0	0

Table 6. Timewarp training parameters

Dataset + training method	Batch size	No. of A-100s	Training time
AD — likelihood	256	1	1 week
AD — acceptance	64	1	2 days
2AA — likelihood	256	4	2 weeks
2AA — acceptance	256	4	4 days
4AA — likelihood	256	4	3 weeks

## F. Computing infrastructure

The training was performed on 4 NVIDIA A-100 GPUs for the 2AA and 4AA datasets and on a single NVIDIA A-100 GPU for the AD dataset. Inference with the model as well as all MD simulations were conducted on single NVIDIA V-100 GPUs for AD and 2AA, and on single NVIDIA A-100 GPUs for 4AA.