Appendix

A Results of Annealing Reflow with Different Fixed β

We present the sampled images and the corresponding FID over 5k images of Annealing Reflow training of different fixed β using the same data pairs in Fig. 7. We test the results with four different β values: $\{0, 0.1, 0.3, 0.5\}$ and found that the RK45 samples are similar in four different models but the quality of one-step generated samples is much worse when β is large. This observation validates the marginal preserving property of the Annealing Reflow training.

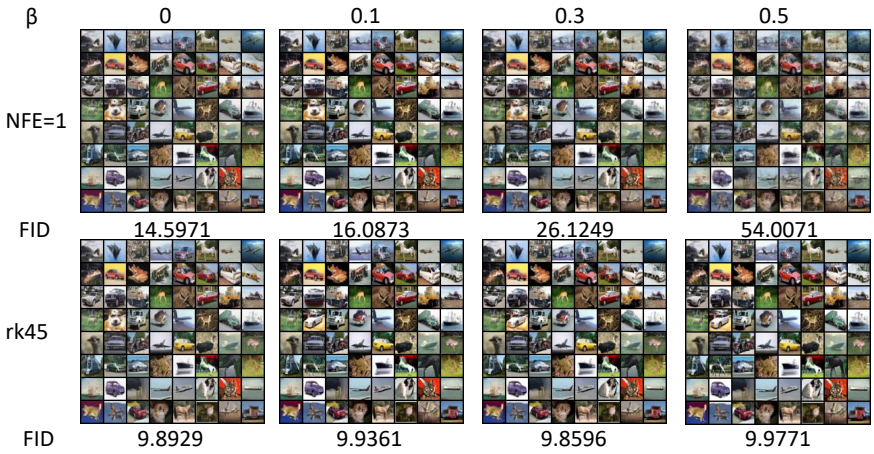

Fig. 7: Result on models with 15M parameters trained with different fixed β values. The FIDs are calculated with only 5k generated samples.

B More Experimental Details

In Tab. 6, we list all the architecture choices and related hyper-parameters in our experiments on CIFAR10 and FFHQ. In Tab. 7, we list all the architecture choices and related hyper-parameters in our experiments on ImageNet. The training of all the networks is smoothed by EMA with a ratio of 0.999999. Adam optimizer is adopted with a learning rate of $2e-4$ and the dropout rate is set to 0.15, following [29]. For the Annealing Reflow training, we use ℓ_2 loss with a uniform loss weight; for the distillation, we switch to the LPIPS loss. Most of the ablation experiments are conducted with configuration D using the data pairs from the 1-rectified flow teacher.

For the experiment on ImageNet, we also trained a model with configuration J in Tab. 7, which has almost half the parameters but similar MACs to configuration I. We get an FID of 8.86 from the 2-rectified flow with RK45 sampler

	A	B	C	D†	E	F	G
#Blocks	4	4	2	2	2	1	1
Base Channels	128	128	128	96	64	64	64
Channel Multiplier	(1, 2, 2, 2)	(2, 2, 2)	(1, 2, 2)	(1, 2, 2)	(1, 2, 2)	(1, 1, 2)	(1, 1)
Attention resolutions	(16,)	(16,)	(16,)	(16,)	(16,)	(16,)	(16,)
Batch Size	-	-	128	128	128	128	128
Batch Size (FFHQ distillation)	-	-	64	64	128	128	-
#Paras	61.8M	55.7M	27.9M	15.7M	7.0M	3.4M	1.2M
MACs	10.3G	20.7G	6.6G	3.7G	1.6G	0.7G	0.5G
FLOPs	22.0G	42.7G	13.9G	7.9G	3.6G	1.5G	1.2G

Table 6: Architecture configurations that are used in this work for CIFAR10 and FFHQ. † represents the default configuration for ablations. MACs and FLOPs are calculated with input shape $(1, 3, 32, 32)$.

		H	I	J
Architecture Configuration	#Blocks	3	2	2
	Base Channels	192	128	128
	Channel Multiplier	(1, 2, 3, 4)	(1, 2, 2, 4)	(1, 2, 2, 2)
	Attention resolutions	(32,16,8)	(32, 16)	(32, 16)
Training	Batch Size	-	96	96
	Batch Size (distillation)	-	64	64
Model Size	#Paras	259.9M	80.7M	44.7M
	MACs	103.4G	31.0G	28.1G
	FLOPs	219.4G	67.8G	61.9G

Table 7: Architecture configurations that are used in this work for the ImageNet dataset. MACs and FLOPs are calculated with input shape $(1, 3, 64, 64)$.

	Channel Multiplier	#Paras	MACs	FLOPs	FID (2-rectified flows)	FID (distilled flows)
I	(1, 2, 2, 4)	80.7M	31.0G	67.8G	8.89	12.34
J	(1, 2, 2, 2)	44.7M	28.1G	61.9G	8.86	12.52

Table 8: Comparison between two ImageNet experiments in this work.

(NFE≈40) and a final FID of 12.52 on the final one-step flow. The comparison is listed in Tab. 8.

For the Annealing Reflow training, we use 50k data pairs from the 1-rectified flow, 100k data pairs from EDM on CIFAR10 dataset, 200k data pairs from EDM on FFHQ dataset, and 400k data pairs from EDM on the ImageNet dataset. For distillation, we always simulate 500k data pairs from the 2-rectified flows.

For distillation, the loss after replacing the 2-rectified flow with the one-step model as mentioned in Sec. 4.1 is:

$$\mathcal{L}'_{2\text{-step}}(\phi') := \mathbb{E}_{\mathbf{x}_1 \sim \pi_1}\left[\int_0^1 \mathbb{D}(\mathbf{x}_1 - (1-t)\mathbf{v}_{\phi'}(\mathbf{x}_1, 1) - t\mathbf{v}_{\phi}(\mathbf{x}_t, t), \mathbf{x}_1 - \mathbf{v}_{\phi'}(\mathbf{x}_1, 1))\mathrm{d}t\right]. \tag{14}$$

where a stop-gradient operation is added to the first $\mathbf{v}_{\phi'}$. This will help the convergence of the training in practice and save one forward step of the 2-rectified flow.

All of the reference statistics for computing FID are from EDM [18]. All of the straightness is calculated using 100 Euler steps and averaging over 256 images, with the following Eq. (7).

C Additional Samples from SlimFlow

In this section, we provide some additional samples from our one-step models.

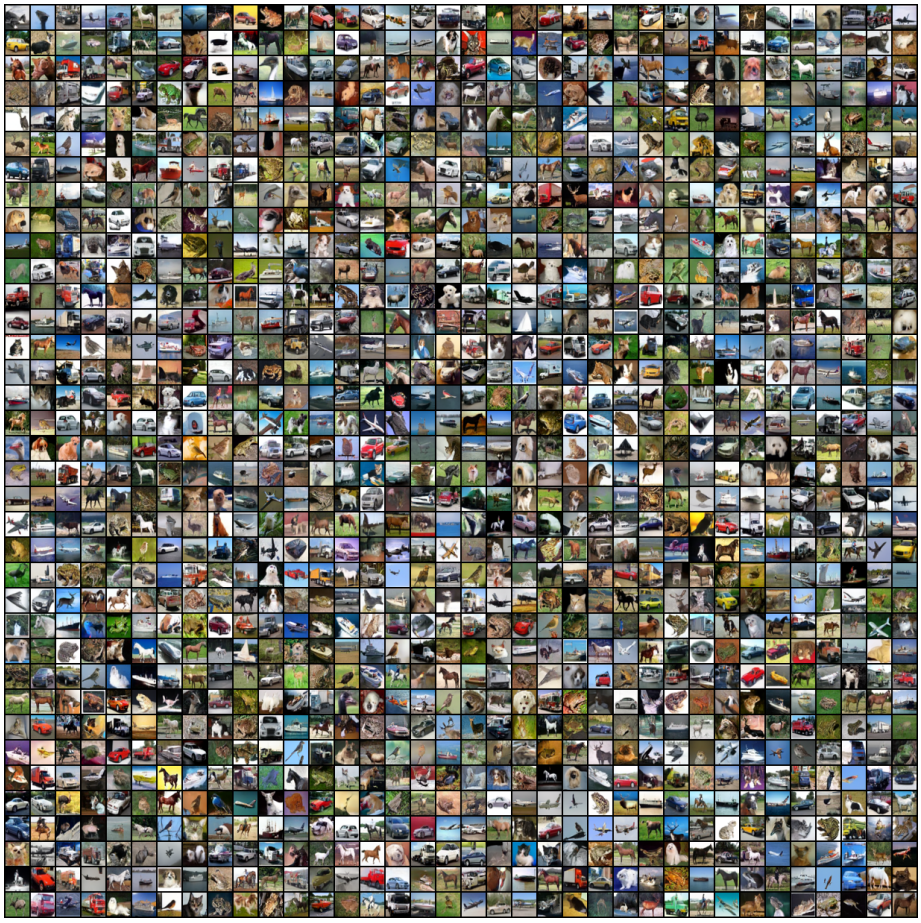

Fig. 8: Uncurated samples from unconditional CIFAR-10 32×32 using SlimFlow with single step generation (FID=4.53).

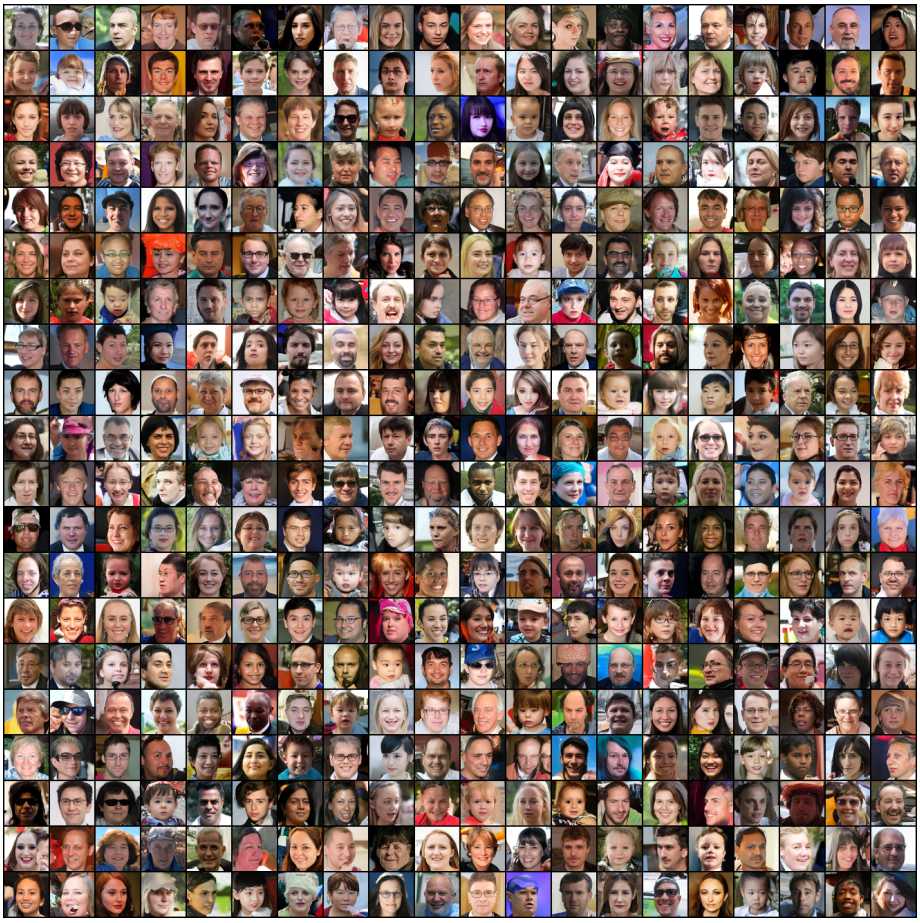

Fig. 9: Uncurated samples from unconditional FFHQ 64×64 using SlimFlow with single step generation (FID=7.21).

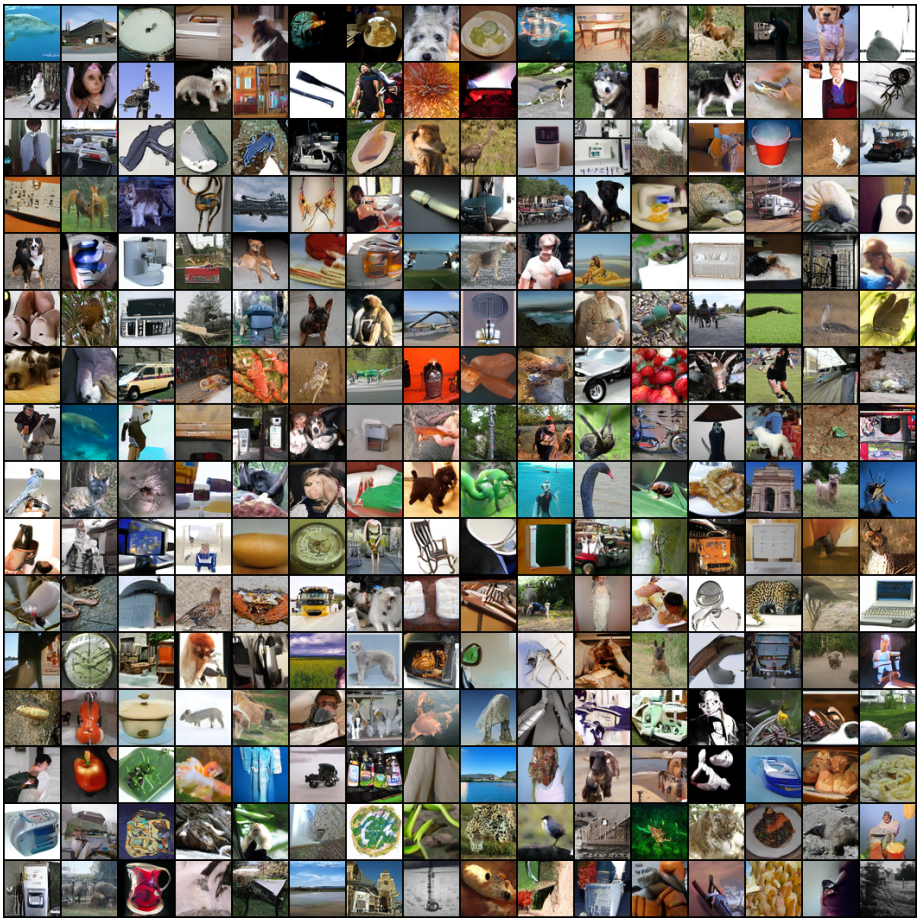

Fig. 10: Uncurated samples with random class labels from conditional ImageNet 64×64 using SlimFlow with single step generation (FID=12.34).

Fig. 11: Uncurated samples with three given classes from conditional ImageNet 64×64 using SlimFlow with single step generation (FID=12.34).