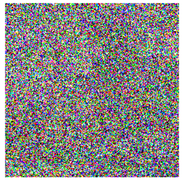
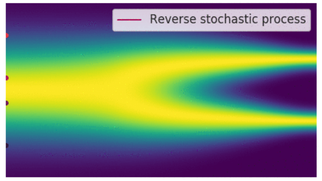
Parallel Sampling of Diffusion Models

Andy Shih, Suneel Belkhale, Stefano Ermon, Dorsa Sadigh, Nima Anari

andyshih@cs.stanford.edu





[Song 2021]

$$\mathrm{d}\boldsymbol{x}_t = [f(t)\boldsymbol{x}_t - g^2(t)\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t$$

can reverse the SDE

[Anderson 1982]

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = f(t)\boldsymbol{x}_t - \frac{1}{2}g^2(t)\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)$$

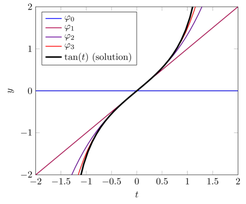
can also write as ODE

[Maoutsa 2020]

	DDPM [Ho 2020]	DDIM [Song 2021]	DPMSolver [Lu 2022]	ParaDiGMS [ours]
Sample Method	SDE (euler maruyama)	ODE (euler)	ODE (heun)	ODE (picard+ euler/heun)
Speed	Slow 1000 steps	Fast 50 steps	Fast 50 steps	Fast 1000 steps
Quality	Best	Good	Good	Best
		trade quality for speed	trade quality for speed	trade compute for speed

Picard Iterations

Solve discretized ODE by iterating until convergence

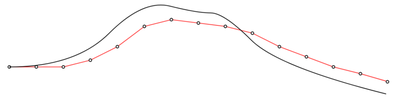


$$\varphi_{k+1}(t) = y_0 + \int_{t_0}^{t} f(s, \varphi_k(s))ds$$

continuous

$$\varphi_{k+1}(\frac{j}{N}) = y_0 + \frac{1}{N}\sum_{i=0}^{j-1} f(\frac{i}{N}, \varphi_k(\frac{i}{N}))$$

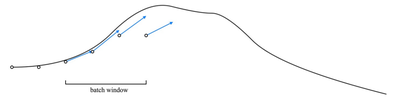
discrete

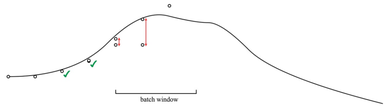


Practical Issues

out of memory?

Batching: process k timesteps at a time with sliding window





approximate?

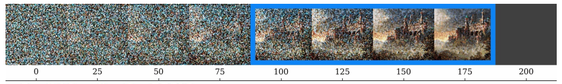
No measurable degradation on standard benchmarks

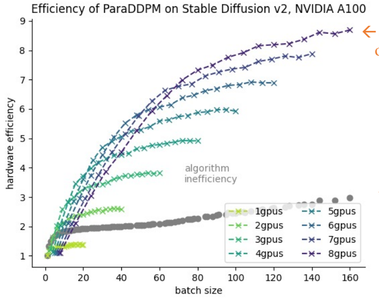
If $\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^{K-1}\|^2 \le 4\epsilon^2\sigma_t^2/T^2$

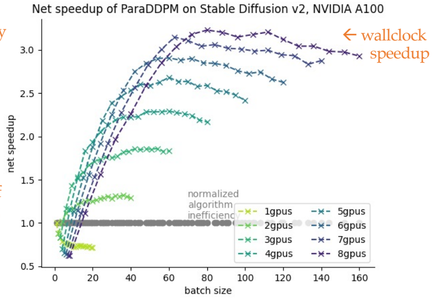
Then $D_{\mathrm{TV}}(\mathcal{N}(\boldsymbol{x}_t^K, \sigma_t^2\boldsymbol{I}) \,||\, \mathcal{N}(\boldsymbol{x}_t^\star, \sigma_t^2\boldsymbol{I})) \le \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x}_t^K, \sigma_t^2\boldsymbol{I}) \,||\, \mathcal{N}(\boldsymbol{x}_t^\star, \sigma_t^2\boldsymbol{I}))}$

$$= \sqrt{\frac{\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^\star\|^2}{4\sigma_t^2}} \le \sqrt{\frac{\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^{K-1}\|^2}{4\sigma_t^2}} \le \frac{\epsilon}{T}$$

We can bound total variation distance







3x speedup!

Robotics

Franka Kitchen	Sequential Model Evals	Sequential Reward	Sequential Time per Episode	ParaDiGMS Model Evals	ParaDiGMS Parallel Iters	ParaDiGMS Reward	ParaDiGMS Time per Episode	Speedup
DDPM	100	0.85 ± 0.03	112s	390	25	0.84 ± 0.03	33.3s	3.4x
DDIM	15	0.80 ± 0.03	16.9s	47	7	0.80 ± 0.03	9.45s	1.8x
DPMSolver	15	0.79 ± 0.03	17.4s	41	6	0.80 ± 0.03	8.89s	2.0x

Latent Image

StableDiffusion-v2	Sequential Model Evals	Sequential CLIP Score	Sequential Time per Sample	ParaDiGMS Model Evals	ParaDiGMS Parallel Iters	ParaDiGMS CLIP Score	ParaDiGMS Time per Sample	Speedup
DDPM	1000	32.1	50.0s	2040	44	32.1	16.2s	3.1x
DDIM	200	31.9	10.3s	425	16	31.9	5.8s	1.8x
DPMSolver	200	31.7	10.3s	411	16	31.7	5.8s	1.8x

Pixel Image

LSUN Church	Sequential Model Evals	Sequential FID Score	Sequential Time per Sample	ParaDiGMS Model Evals	ParaDiGMS Parallel Iters	ParaDiGMS FID Score	ParaDiGMS Time per Sample	Speedup
DDPM	1000	12.8	24.0s	2556	42	12.9	8.2s	2.9x
DDIM	500	15.7	12.2s	1502	42	15.7	6.3s	1.9x