From Physics to Generative Models Ζ

Ziming Liu, IAIFI Journal Club, April 18 2023 **Advisor: Max Tegmark**







Q: Can physics help build generative models?



Our Universe/Physics is a generative model



Lin, Tegmark & Rolnick

Thank you for your listening!





Just kidding! **Need actually build generative models!**





Overview

1. Brief background: Diffusion Models (DM)

2. Poisson flow generative models (PFGM/PFGM++) <

3. GenPhys: From Physical Processes to Generative Models

Diffusion process

Electrostatics

Any physical process?



Generative Models

DALLE-2



ng of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck





esso machine that makes coffee from human souls, artstation



panda mad scientist mixing sparkling chemicals, artstation



a corgi's head depicted as an explosion of a nebula







a dolphin in an astronaut suit on saturn, artstation



a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese



a teddybear on a skateboard in times square

Figure 1: Selected 1024×1024 samples from a production version of our model.

Stable diffusion

Score-based models (diffusion models)



Physics interpretation

$$p(\mathbf{x}) = \exp(-E(\mathbf{x})) \quad \mathbf{s}_{\theta}(\mathbf{x})$$

 $\mathbf{s}_{\theta}(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p(\mathbf{x})$

 $\mathbf{x}) = -\nabla E(\mathbf{x})$





New samples $\dot{\mathbf{x}} = \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \sqrt{2} \dot{W}$

Courtesy of Yang Song https://yang-song.net/blog/2021/score/





Score-based models (diffusion models)



Courtesy of Yang Song https://yang-song.net/blog/2021/score/



Poisson Flow Generative Models Yilun Xu^{*}, Ziming Liu^{*}, Max Tegmark & Tommi Jaakkola Accepted by NeurIPS 2022





Basic idea



Generative modeling

data point/sample

data distribution

bijective map

electric lines

Electrostatics

Poisson Equation $\nabla^2 \varphi = -\rho(x), \ \nabla^2 \equiv$

Green function $\nabla^2 G(x, x') = -\delta(x - \delta(x - \delta(x$

$$\varphi(x) = \int G(x, x') \rho(x') dx' \qquad E(x) = -\nabla \varphi(x) \sim \int \frac{x - x'}{||x - x'||^N} \rho(x') dx'$$



$$\sum_{i=1}^{N} \partial_i^2$$

$$-x') \Longrightarrow G(x, x') \sim r^{-(N-2)}, \ r \equiv ||x - x'||$$
$$-\sum G(x) = \int x - x'$$

Poisson Flow: a New Flow Inspired by Electrostatics

- Interpret N-dim data distribution as charge density
- dimension **Z**
- distribution on the large hemisphere



Placing the charges on the z = 0 hyperplane in an N + 1-dim space augmented with

Electric Field Lines *define a bijection* between *data distribution* and a *uniform*

Generation 1. Uniformly sampling an initial sample (•) on the hemisphere.



Generation

Uniformly sampling an initial sample () on the hemisphere. Evolving the sample by following the corresponding electric field line.

Generation

Uniformly sampling an initial sample () on the hemisphere. Evolving the sample by following the corresponding electric field line.

Generation

Uniformly sampling an initial sample () on the hemisphere. Evolving the sample by following the corresponding electric field line.

2. Evolving the sample by following the corresponding electric field line.

Why bijection?

Why augmentation?

No augmentation (2D)

Demo (Heart)

Demo (PFGM word)

Demo (IAIFI word)

Experiments: Visualization of Backward ODE

						140				
					100			1.10	2022	
						$t \in I$				
			1996	200	200					
							200		3	
						2.9			Contraction of the	
						642				
			1.84							
						6.75				
					1.44					
										10 2 See
1		5.3								10000
									10 C	
			10.00					91 G		
					5. A.		100			

Experiments: Generation Quality and Speed (CIFAR-10)

		Qual	lity	Speed
Model	Invertible?	Inception Score (higher is better)	FID Score (lower is better)	NFE (lower is better
StyleGAN2-ADA	×	9.83	2.92	1
Diffusion (VP) - SDE	×	9.68	2.41	1000
Glow		3.92	48.9	1
Diffusion (VP) - ODE		9.47	2.86	134
PFGM (ours)*		9.68	2.35	110

Current SOTA results in normalizing flow family!

*: PFGM and Diffusion models use the same architecture, DDPM++ deep

Q: Relation between Diffusion Models and Poisson Flows?

Poisson Flow Generative Models ++

State of the Art Image Generation on CIFAR-10

Xu, Liu, Tong, Tian, Tegmark, and Jaakkola. PFGM++: Unlocking the Potential of Physics-Inspired Generative Models, arXiv: 2302.04265

Augmented dimensionality D

PFGM++ framework

Anchor the ODE by r = ||z|| due to symmetry

 $\mathbf{E}(\tilde{\mathbf{x}}) = \frac{1}{S_{N+D-1}(1)} \int \frac{\mathbf{x} - \tilde{\mathbf{y}}}{\|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|^{N+D}} p(\mathbf{y}) d\mathbf{y}$

$d\tilde{\mathbf{x}} = \mathbf{E}(\tilde{\mathbf{x}}) dt$

$\mathrm{d}z_i = \mathbf{E}(\tilde{\mathbf{x}})_{z_i} \mathrm{d}t$

 $\frac{\mathrm{d}r}{\mathrm{d}t} = \sum_{i=1}^{D} \frac{z_i}{r} \frac{\mathrm{d}z_i}{\mathrm{d}t} = \int \frac{\sum_{i=1}^{D} z_i^2}{S_{N+D-1}(1)r \|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|^{N+D}} p(\mathbf{y}) \mathrm{d}\mathbf{y}$ $= \frac{1}{S_{N+D-1}(1)} \int \frac{1}{\|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|^{N+D}} p(\mathbf{y}) d\mathbf{y}$

Balance Robustness and Rigidity by Controlling D

- Larger $D \rightarrow$ Better Accuracy/ Worse Robustness ullet
- Smaller $D \rightarrow$ Worse Accuracy/ Better Robustness •

Sweet spot D^* in the middle!

Experiments: Image Generation

Table 1. CIFAR-10 sample quality (FID) and number of function evaluations (NFE).

	Min FID \downarrow	Top-3 Avg FID \downarrow	NFE \downarrow
DDPM (Ho et al., 2020)	3.17	-	1000
DDIM (Song et al., 2021a)	4.67	-	50
VE-ODE (Song et al., 2021b)	5.29	-	194
VP-ODE (Song et al., 2021b)	2.86	-	134
PFGM (Xu et al., 2022)	2.48	-	104
PFGM++ (unconditional)			
D = 64	1.96	1.98	35
D = 128	1.92	1.94	35
D = 2048	1.91	1.93	35
D = 3072000	1.99	2.02	35
$D ightarrow \infty$ (Karras et al., 2022)	1.98	2.00	35
PFGM++ (class-conditional)			
D = 2048	1.74	-	35
$D ightarrow \infty$ (Karras et al., 2022)	1.79	-	35

Table 2. FFHQ sample quality (FID) with 79 NFE in unconditional setting

	$\operatorname{Min}\operatorname{FID}\downarrow$	Top-3 Avg FID \downarrow
D = 128	2.43	2.48
D = 2048	2.46	2.47
D = 3072000	2.49	2.52
$D \rightarrow \infty$ (Karras et al., 2022)	2.53	2.54

(with improved DDPM++/NCSN++ backbone in EDM)

Experiments: Image Generation

Table 1. CIFAR-10 sample quality (FID) and number of function evaluations (NFE).

	$\operatorname{Min}\operatorname{FID}\downarrow$	Top-3 Avg FID \downarrow	NFE \downarrow
DDPM (Ho et al., 2020)	3.17	-	1000
DDIM (Song et al., 2021a)	4.67	-	50
VE-ODE (Song et al., 2021b)	5.29	-	194
VP-ODE (Song et al., 2021b)	2.86	-	134
PFGM (Xu et al., 2022)	2.48	-	104
PFGM++ (unconditional)			
D = 64	1.96	1.98	35
D = 128	1.92	1.94	35
D = 2048	1.91	1.93	35
D = 3072000	1.99	2.02	35
$D \rightarrow \infty$ (Karras et al., 2022)	1.98	2.00	35
PFGM++ (class-conditional)			
D = 2048	1.74	-	35
$D ightarrow \infty$ (Karras et al., 2022)	1.79	-	35

[III] State of the Art Image Generation on CIFAR-10

Filter unta	GAN Score-based VAE	Flow-based	Transformer Diffusion	PFGM ResNet Optimal Transport	ЕВМ			Ed
Rank	Model	Clean- FID ↓ FID- 10k	Inception score bits/dimension	FID- 10k- Paper test	Code	Result	Year	,
1	PFGM++	1.74		PFGM++: Unlocking the Poten of Physics-Inspired Generative Models	tial O	Ð	2023	
2	EDM-G++ (unconditional)	1.77	2.55	Refining Generative Process w Discriminator Guidance in Sco based Diffusion Models	ith re- ()	÷	2022	s
3	StyleGAN-XL	1.85		StyleGAN-XL: Scaling StyleGA to Large Diverse Datasets	^{NN} O	Ð	2022	
4	STF (unconditional)	1.90		Stable Target Field for Reduce Variance Score Estimation in Diffusion Models	d O	Ð	2023	s
5	LSGM-G++ (FID)	1.94	3.42	Refining Generative Process w Discriminator Guidance in Sco based Diffusion Models	ith re- Ç	Ð	2022	s
6	LSGM (FID)	2.10	3.43	Score-based Generative Modeling in Latent Space	0	Ð	2021	s
7	Subspace Diffusion (NSCN++)	2.17	9.99	Subspace Diffusion Generative Models	0	Ð	2022	s
8	LSGM (balanced)	2.17	2.95	Score-based Generative Modeling in Latent Space	0	Ð	2021	s
9	NCSN++	2.20	9.73	Score-Based Generative Modeling through Stochastic Differential Equations	0	Ð	2020	s

https://paperswithcode.com/sota/image-generation-on-cifar-10

t Leaderboard
ags 🔊
PFGM
Diffusion ore-based
GAN
ore-based
ore-based
VAE
ore-based
VAE ore-based
ore-based

Experiments: Robustness

128

 $\alpha = 0.2$

 ∞

Experiments: Robustness in post-training quantization

Table 3. FID score versus quantization bit-widths on CIFAR-10.

Quantization bits:	9	8	7	6	5
D = 64	1.96	1.96	2.12	2.94	28.50
D = 128	1.93	1.97	2.15	3.68	34.26
D = 2048	1.91	1.97	2.12	5.67	47.02
$D \to \infty$	1.97	2.04	2.16	5.91	50.09

Q: Going beyond diffusion equation and **Poisson equation?**

GenPhys: From Physical Processes to Generative Models

Ziming Liu, Di Luo, Yilun Xu, Tommi Jaakkola, Max Tegmark. "GenPhys: From Physical Processes to Generative Models" arXiv: 2304.02637

From physics to generative models

Diffusion (Diffusion equation)

Electrostatics (Poisson equation)

Diffusion models

Poisson Flow Generative Models

Q: Is there a universal converter from physics to generative models? A: Yes, but...

Yes: A concrete protocol that converts physics to generative models

but: the converted generative models may not have desirable properties

Converter: partial differential equations (PDEs)

A physical process is described by a **PDE**

$$\hat{L}\phi \equiv F(\phi, \phi_t, \phi_{tt}, \nabla\phi, \nabla^2\phi, \dots$$

Physicists already know how to solve it, ...

A generative model is associated with a density flow (which is also a **PDE**)

$$\hat{M}(p, \mathbf{v}, R) \equiv \frac{\partial p(\mathbf{x}, t)}{\partial t} + \nabla \cdot [p(\mathbf{x}, t)\mathbf{v}(\mathbf{x}, t)] - R(\mathbf{x}, t) = p_{\text{data}}(\mathbf{x})\delta(t) \quad \text{(Density Flow)}$$
probability distribution velocity field birth/death rate

..., then we know how to solve this one, too. By "solve", we mean a design of (p, v, R).

..., if they are equivalent, ...

Converter

physical process

$$\hat{L}\phi \equiv F(\phi, \phi_t, \phi_{tt}, \nabla\phi, \nabla^2\phi, ...$$
generative model
$$\hat{M}(p, \mathbf{v}, R) \equiv \frac{\partial p(\mathbf{x}, t)}{\partial t} + \nabla \cdot [p(\mathbf{x}, t)\mathbf{v}(\mathbf{x}, t)]$$

Example: Diffusion equation

Source PDE

$$\phi_t - \nabla^2 \phi = 0$$

$$\phi_t - \nabla^2 \phi = \frac{\partial \phi}{\partial t} + \nabla \cdot (\phi(t))$$

Comparing the two sides gives:

Target PDE

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} + \nabla \cdot \left[p(\mathbf{x}, t) \mathbf{v}(\mathbf{x}, t) \right] - R(\mathbf{x}, t) = 0$$

 $(-\nabla \log \phi)) - 0 \quad \Leftrightarrow \quad \frac{\partial p}{\partial t} + \nabla \cdot (p\mathbf{v}) - R$

 $p = \phi, \quad \mathbf{v} = -\nabla \log \phi, \quad R = 0$

"score"

Example: Poisson equation

Source PDE

$$\phi_{tt} + \nabla^2 \phi = 0$$

$$-(\phi_{tt} + \nabla^2 \phi) = \frac{(\partial - \phi_t)}{\partial t} + \nabla \cdot \left[-\phi_t \left(\frac{\nabla \phi}{\phi_t} \right) \right] - 0 \quad \Leftrightarrow \quad \frac{\partial p}{\partial t} + \nabla \cdot \left(p \mathbf{v} \right) - R$$

Comparing the two sides gives:

p = - \mathcal{O}_t

Target PDE

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} + \nabla \cdot \left[p(\mathbf{x}, t) \mathbf{v}(\mathbf{x}, t) \right] - R(\mathbf{x}, t) = 0$$

$$\mathbf{v} = rac{
abla \phi}{\phi_t}, \quad R = 0$$

"Poisson field"

Example: Wave equation

Source PDE

$$\phi_{tt} - \nabla^2 \phi = 0$$

$$-(\phi_{tt} - \nabla^2 \phi) = \frac{(\partial - \phi_t)}{\partial t} + \nabla \cdot$$

Comparing the two sides gives:

$$p = -\phi_t,$$

Target PDE

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} + \nabla \cdot \left[p(\mathbf{x}, t) \mathbf{v}(\mathbf{x}, t) \right] - R(\mathbf{x}, t) = 0$$

$$\left[-\phi_t\left(-\frac{\nabla\phi}{\phi_t}\right)\right] - 0 \quad \Leftrightarrow \quad \frac{\partial p}{\partial t} + \nabla \cdot (p\mathbf{v}) - R$$

$$\mathbf{v} = -\frac{\nabla \phi}{\phi_t}, \quad R = 0$$

Caveat: prior distribution

Generative models should have data-independent priors: No matter what the initial distribution (data) is, the final distribution should be independent of it.

Which PDEs can give desirable generative models?

A PDE x is s-generative (s for smooth) if:

(C1) x can be converted to a density flow; (C2) The solutions of x become "smoother" over time.

- (C1) Case-dependent constructions are required, but are usually straightforward.
- (C2) turns out to be equivalent to a constraint on dispersion relations of PDEs. Intuition: Non-zero frequency modes should decay faster than zero modes.

Examples

				?	?		
equation	diffusion equation	Poisson equation	ideal wave equation	dissipative wave equation	Helmholtz equation	screened Poisson equation (Yukawa)	Schrödinger
PDE $\hat{L}\phi = 0$	$\phi_t - abla^2 \phi = 0$	$\phi_{tt}+ abla^2\phi=0$	$\phi_{tt} - abla^2 \phi = 0$	$\phi_{tt} + 2\epsilon\phi_t - abla^2\phi = 0$	$\phi_{tt}+ abla^2\phi+k_0^2\phi=0$	$\phi_{tt}+ abla^2\phi-m^2\phi=0$	$i\phi_t + abla^2$
Rewritten	$rac{\partial \phi}{\partial t} + abla \cdot (\phi(- abla \mathrm{log}\phi)) = 0$	$rac{\partial (-\phi_t)}{\partial t} + abla \cdot ((-\phi_t)(rac{ abla \phi}{\phi_t})) = 0$	$rac{\partial (-\phi_t)}{\partial t} + abla \cdot ((-\phi_t)(-rac{ abla \phi}{\phi_t})) = 0$	$rac{\partial (-\phi_t - 2\epsilon\phi)}{\partial t} + abla \cdot ((-\phi_t - 2\epsilon\phi)(rac{ abla \phi}{\phi_t + 2\epsilon\phi})) = 0$	$rac{\partial (-\phi_t)}{\partial t} + abla \cdot ((-\phi_t)(rac{ abla \phi}{\phi_t})) - k_0^2 \phi = 0$	$rac{\partial (-\phi_t)}{\partial t} + abla \cdot ((-\phi_t)(rac{ abla \phi}{\phi_t})) + m^2 \phi = 0$	$rac{\partial \phi ^2}{\partial t} + abla \cdot (\phi ^2)^2$
p	ϕ	$-\phi_t$	$-\phi_t$	$-(\phi_t+2\epsilon\phi)$	$-\phi_t$	$-\phi_t$	$ \phi $
\mathbf{v}	$- abla ext{log}\phi$	$\frac{\nabla \phi}{\phi_t}$	$-\frac{\nabla\phi}{\phi_t}$	$rac{ abla \phi}{\phi_t + 2\epsilon \phi}$	$-\phi_t$	$-\phi_t$	$ \phi ^2$
R	0	0	0	0	$k_0^2\phi$	$-m^2\phi$	0
G(r,t)	$rac{1}{(4\pi t)^{rac{N}{2}}} ext{exp}(-rac{r^2}{4t})$	$rac{1}{(t^2+r^2)^{rac{N-1}{2}}}$	$\frac{1}{\sqrt{t^2 - r^2}} \Theta(t - r) \ (\text{2D})$	$\frac{e^{-\epsilon t} \cosh(\epsilon \sqrt{t^2 - r^2})}{\sqrt{t^2 - r^2}} \Theta(t - r) $ (2D)	$(rac{k_0}{\sqrt{t^2+r^2}})^{rac{N-1}{2}}H^{(1)}_{rac{N-1}{2}}(k_0\sqrt{t^2+r^2})$	$\left(\frac{m}{\sqrt{t^2+r^2}}\right)^{\frac{N-1}{2}} K_{\frac{N-1}{2}}(m\sqrt{t^2+r^2})$	$rac{1}{\left(4\pi it ight)^{rac{N}{2}}}\mathrm{e}$
$\hat{G}(k,t)$	$\exp(-k^2t)$	$\exp(-kt)$	$\exp(\pm ikt)$	$ \exp(-\epsilon t + i\sqrt{k^2 - \epsilon^2}t) \ (k > \epsilon) \\ \exp(-(\epsilon + \sqrt{k^2 - \epsilon^2})t) \ (k \le \epsilon) \ \hat{G}(k, t) $	$\exp(-i\sqrt{k_0^2-k^2}t)~(k\leq k_0)\ \exp(-\sqrt{k^2-k_0^2}t)~(k>k_0)$	$\exp(-\sqrt{k^2+m^2}t)$	$\exp(i t)$
(C1)	Yes	Yes	No	Conditionally yes	Conditional yes	Yes	No
(C2)	Yes	Yes	No	Conditionally yes	Conditional Yes	Yes	No
Illustration ϕ							
s_generative?	Ves (Diffusion Models)	Ves (Poisson Flow)	No	Conditionally Vas (large c)	Conditionally yes (small k)	Vac	No
s-generative?	Tes (Diffusion Models)	ies (roissoil filow)	INU	Conditionally Tes (large e)	Conditionally yes (small k)	105	INC

Dispersion relations

(C2) The solutions of x become "smoother" over time.

 $\phi \sim e^{i(kx - \omega t)}$ Table 3: Dispersion relat PDE Physics $\phi_t - \nabla^2 \phi = 0$ Diffusion $\phi_{tt} + \nabla^2 \phi = 0$ Poisson $\phi_{tt} - \nabla^2 \phi = 0$ Ideal Wave $\phi_{tt} + 2\epsilon\phi_t - \nabla^2\phi = 0$ Dissipative wave $\phi_{tt} + \nabla^2 \phi + k_0^2 \phi = 0$ Helmholtz Screened Poisson $\phi_{tt} + \nabla^2 \phi - m^2 \phi = 0$ $i\phi_t + \nabla^2 \phi = 0$ Schrödinger

tion $\omega(k)$ of physical PDEs.	
Dispersion Relation	s-generative?
$\omega = -ik^2$	Yes
$\omega=\pm ik$	Yes
$\omega=\pm k$	No
$\omega = \begin{cases} i(-\epsilon \pm \sqrt{\epsilon^2 - k^2}) & k \le \epsilon \\ i\epsilon \pm \sqrt{k^2 - \epsilon^2} & k > \epsilon \end{cases}$	Conditionally Yes (large ϵ)
$\omega = \begin{cases} \pm \sqrt{k_0^2 - k^2} & k \le k_0 \\ \pm i\sqrt{k^2 - k_0^2} & k > k_0 \end{cases}$	Conditionally Yes (small k_0^2)
$\stackrel{\rarkow}{\omega}=\pm i\sqrt{k^2+m^2}$ $\omega=k^2$	Yes No

 $\operatorname{Im} \omega(k) < \operatorname{Im} \omega(0), \quad \text{for all } k > 0$

Other s-generative PDEs

Physics

Mixed diffusion Poisson Fractional diffusion Third-order "diffusion" Elasticity (Biharmonic)

Table 4: The dispersion relation suggests new PIGM

PDE	Dispersion Relation
$a\phi_{tt} - b\phi_t + \nabla^2 \phi = 0 \ (a > 0, b > 0)$	$\omega = \frac{i}{2a}(b \pm \sqrt{b^2 + 4ak^2})$
$\phi_t + (-\Delta)^{eta} \phi = 0 \ (eta > 0)$	$\omega = -ik^{2eta}$
$\phi_{ttt} - \Delta u = 0$	$\omega = (-i, e^{irac{\pi}{6}}, e^{irac{5\pi}{6}})k^{rac{2}{3}}$
$\phi_t + abla^2 abla^2 \phi = 0$	$\omega = -ik^4$

Open Questions

- and in practice)?
- Going beyond smooth and linear PDEs. E.g., Naiver-Stokes equation, reaction-diffusion equation

Navier-Stokes equation

Among s-generative PDEs, which one gives the best performance (in theory

Reaction-diffusion equation

Thank you for your listening!

