Carnegie Mellon University

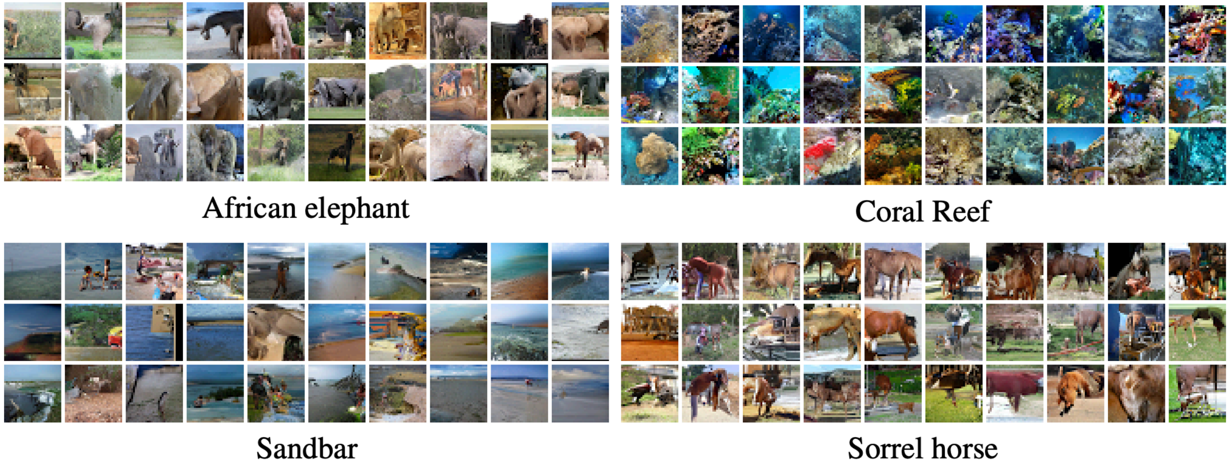
Text2PDE: Latent Diffusion Models for Accessible Physics Simulation

Anthony Zhou, Zijie Li, Amir Barati Farimani
Department of Mechanical Engineering
Carnegie Mellon University

Michael Schneier, John R. Buchanan Jr.
Naval Nuclear Laboratory

How does this work fit into the field?

Neural surrogates have come far: can we envision **new ways** of using physics simulators?

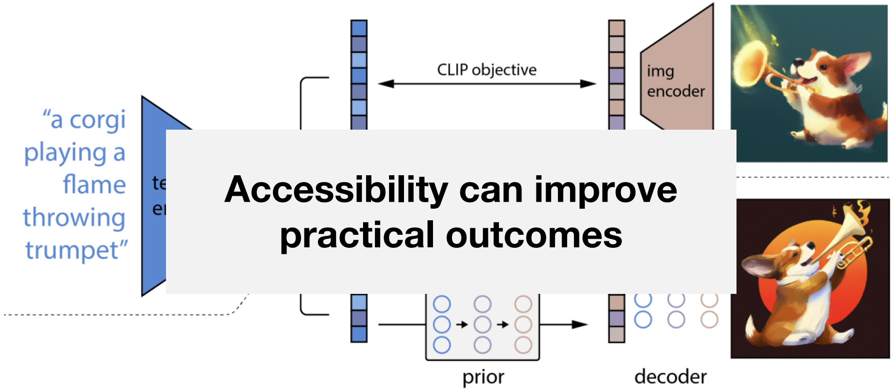


Class-conditioned image generation (PixelCNN, 2017)

A. van den Oord, et. al. Conditional Image Generation with PixelCNN Decoders.

How does this work fit into the field?

Neural surrogates have come far: can we envision **new ways** of using physics simulators?



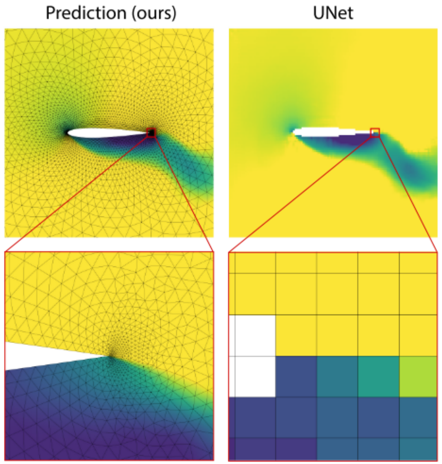
Accessibility can improve practical outcomes

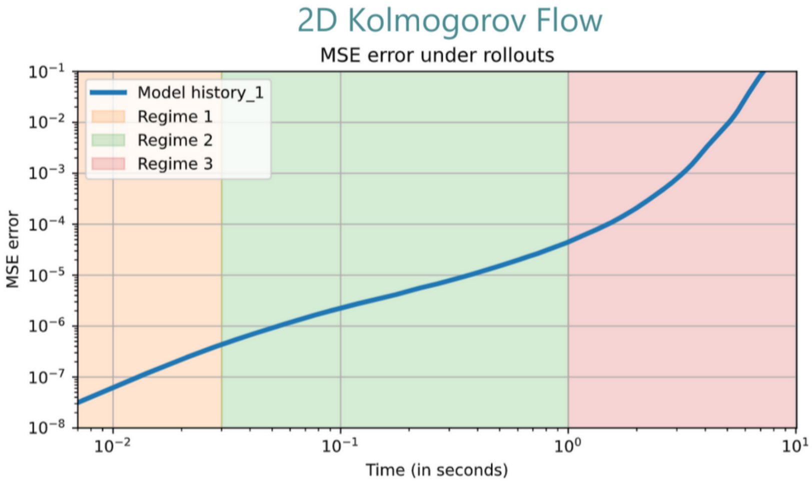
Text-conditioned image generation (DALL-E 2, 2022)

Impact of image generation models is largely due to the **accessibility** of text prompting.

A. Ramesh, et. al. Hierarchical Text-Conditional Image Generation with CLIP Latents.

What are the criteria for text-based physics simulators?





Main differences w/ image generation:

- **Accuracy**: Predicted physics must be physically consistent/stable
- **Unstructured data**: Physics naturally benefits from areas of refinement

T. Pfaff, et. al. Learning mesh-based simulation with Graph Networks. Phillip Lippe, Microsoft AI4Science, https://drive.google.com/file/d/1Qk7hX1InUmFwr9wk0HuWmB-OJ9llbAMG/view

How can generative modeling reduce error propagation?

$$p(\mathbf{u}|\mathbf{u}^0, B[\mathbf{u}]) \approx p_\theta(\mathbf{u}|\mathbf{u}^0, B[\mathbf{u}]) \qquad \mathbf{u} \in \mathbb{R}^{T\times M\times d_p}$$

Approximate the conditional probability distribution of **all timesteps** at once.

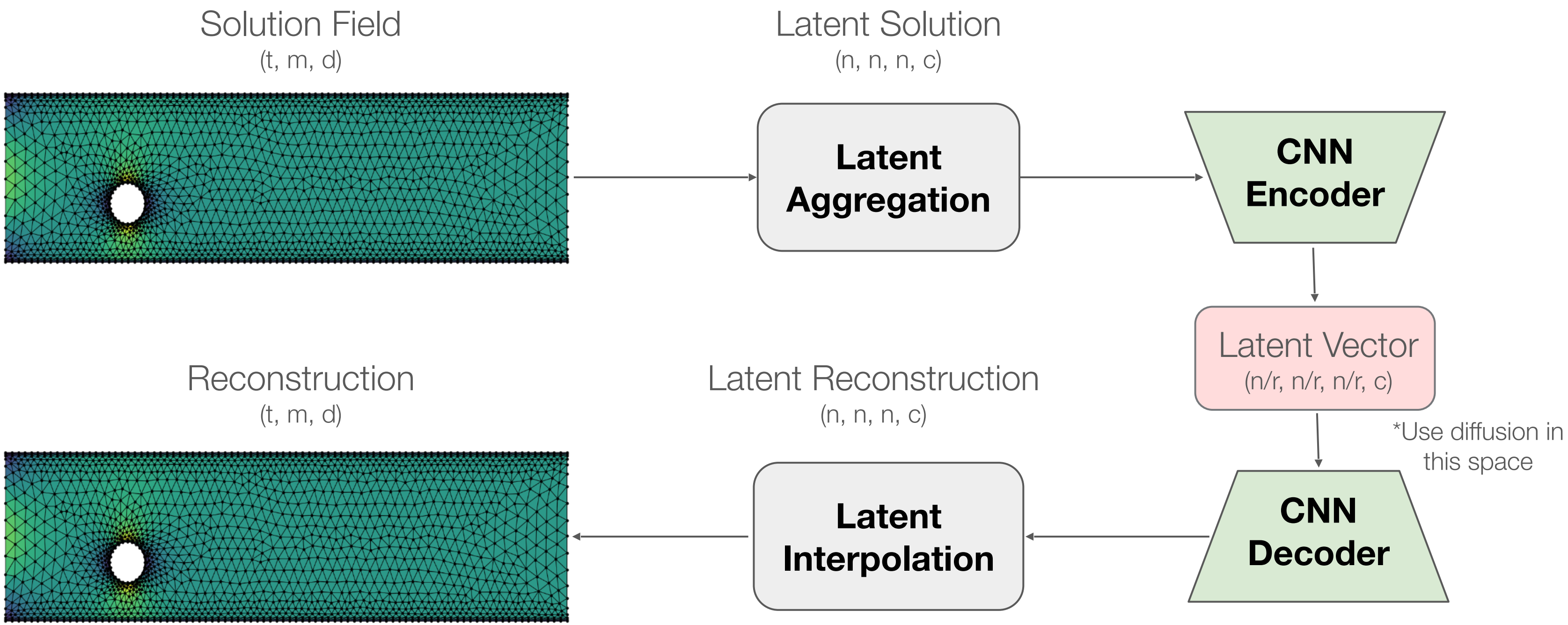
Challenge: Distribution of all possible solutions is extremely high dimensional (128x128 grid, 48 timesteps, 3 channels = ~2 million DoFs)

- Diffusion models have high capacity to model complex distributions
- Use a latent space to make training/inference tractable

Latent Diffusion can be an effective model choice to address temporal accuracy

How do we generate solutions at arbitrary discretization?

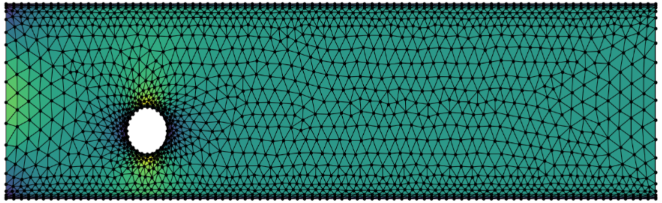
Perform diffusion on a regular latent space, use autoencoder to map from mesh and grid spaces.



Carnegie Mellon University

Can we use different conditioning modalities?

Physics



Latent Aggregation

CNN Encoder

*Mesh encoder, trained from scratch

Embedding

Text

“Fluid passes over a cylinder with a radius of 4.97 and position: 0.35, 0.14. Fluid enters with a velocity of 0.20. The Reynolds number is 190.”

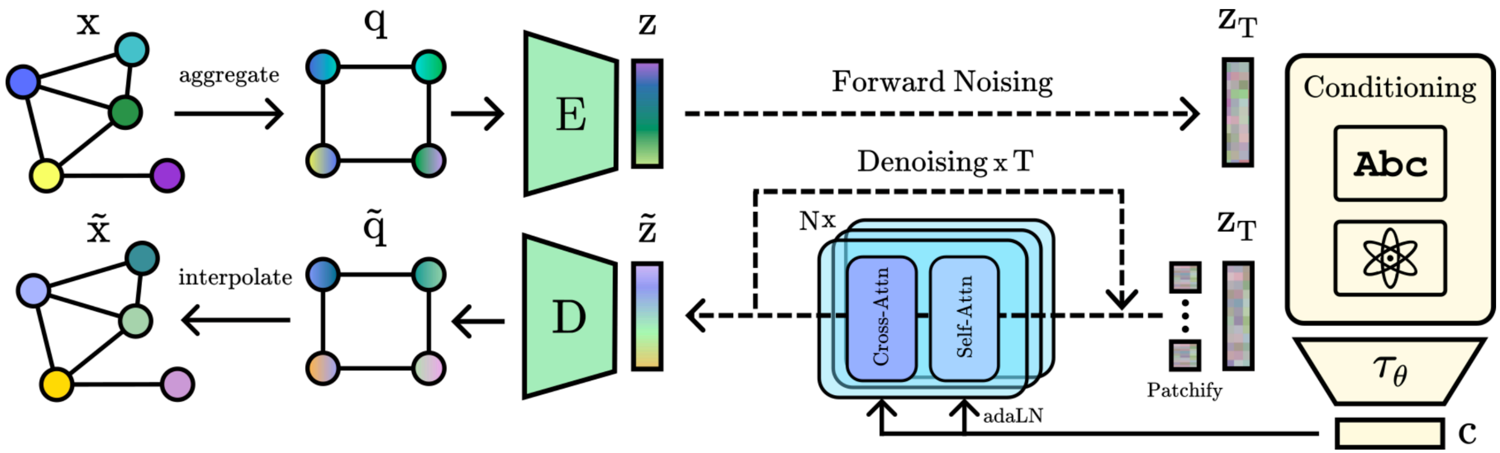
RoBERTa

*Pretrained and fine-tuned

Embedding

*Need to make a **captioned** dataset*

Bringing it All Together



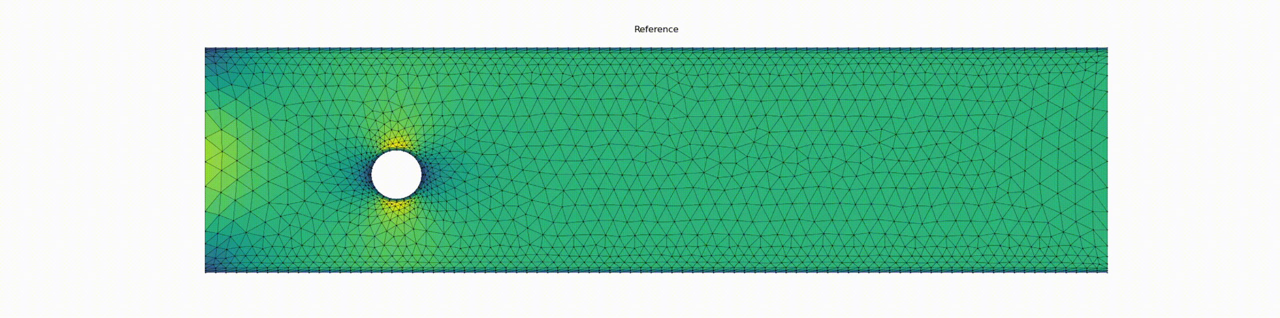
Recap of main features:

- Diffusing all timesteps at once in a latent space
- Using a mesh autoencoder to adapt to arbitrary discretizations
- Allow conditioning on either text or physics modalities

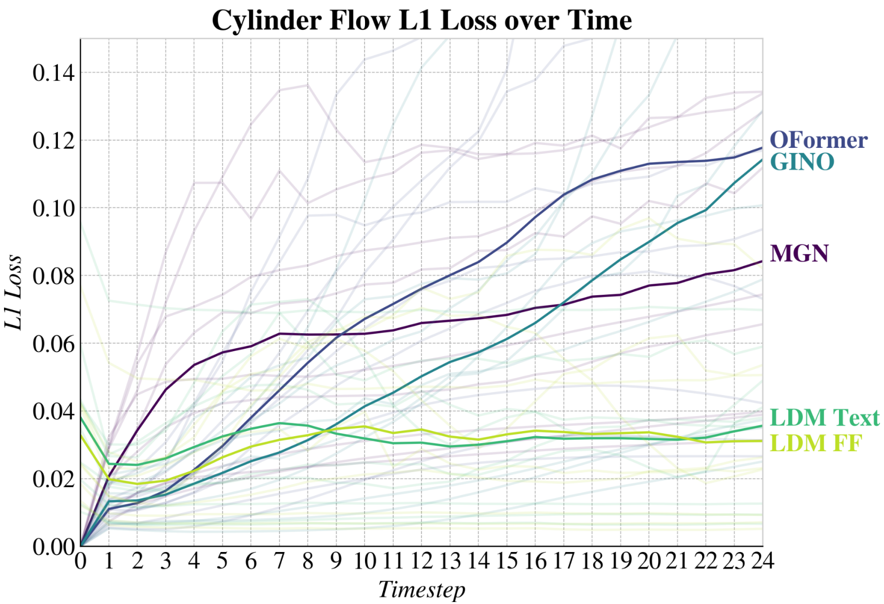
Cylinder Flow Results

Model	Params	Tflops	L2 Loss
GINO	72M	0.73	0.2445
MGN	101M	32.16	0.2617
OFormer	131M	17.34	0.3386
LDM_S-FF	198M	0.81	0.1522
LDM_M-FF	667M	1.16	**0.1309**
LDM_S-Text	313M	0.83	0.1796
LDM_M-Text	804M	1.20	0.1476

- LDMs can be very **efficient** from using a latent space
- LDMs can be **accurate** from using spatio-temporal prediction
- Text-conditioning can be accurate in constrained setups



Generative vs. Autoregressive Rollout Accuracy



LDMs are able to have around **constant** temporal error whereas autoregressive models **accumulate** error

3D Turbulence Results

Model	Params	Tflops	L2 Loss	D_{TKE}
FNO	1.02B	1.62	0.862	6.524
FactFormer	41.4M	53.8	0.795	6.022
Dil-Resnet	24.8M	249	0.707	6.153
LDM-FF	2.72B	9.78	**0.602**	**5.630**
LDM-Text	2.73B	9.43	0.693	5.653

- LDMs have some scaling ability to more complex systems
- Many details are lost, however, statistics can be captured

Thank you for listening!

Paper: https://arxiv.org/abs/2410.01153

Code: https://github.com/anthonyzhou-1/ldm_pdes

Carnegie Mellon University

Mechanical and AI Lab (MAIL)

NAVAL NUCLEAR LABORATORY

NERSC