From Entropy to Artistry: on Thermodynamics and Generative AI

Stephan Mandt
Department of Computer Science
University of California, Irvine

A physicist's perspective on diffusion models

Part 1:

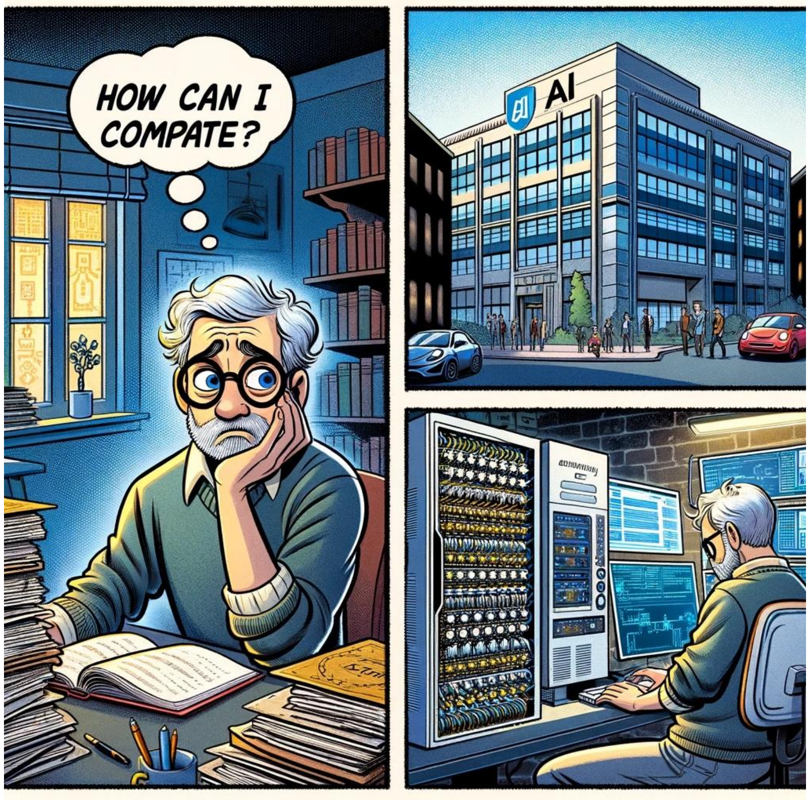
- Understanding the design space of diffusion models
- Efficient samplers for accelerating diffusion models
- Data communication with diffusion models

Part 2:

- Super-resolving atmospheric convection with diffusion models

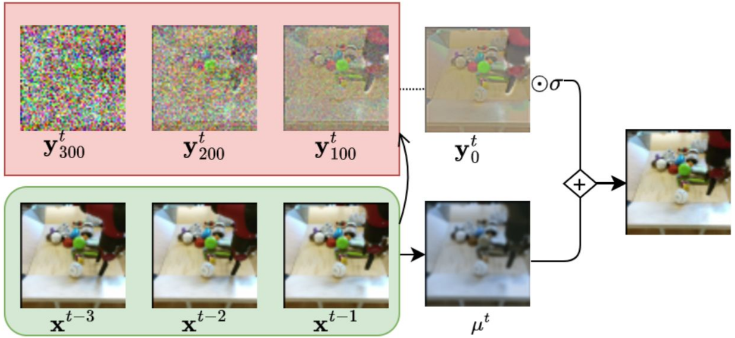
Diffusion Models for Image Generation

Prompt:
"I'm an AI professor desperately worried to compete with industry research on computing resources. Depict my situation in a comic."



ChatGPT / OpenAI April 2024

Diffusion Models for Video Generation



R. Yang, P. Srivastava, S. Mandt. arXiv 2022
J. Ho et al., NeurIPS 2022
Sora, OpenAI, 2024

Diffusion Models for Precipitation Super-Resolution

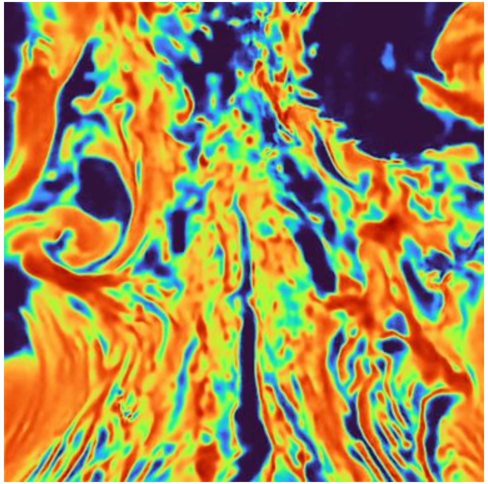
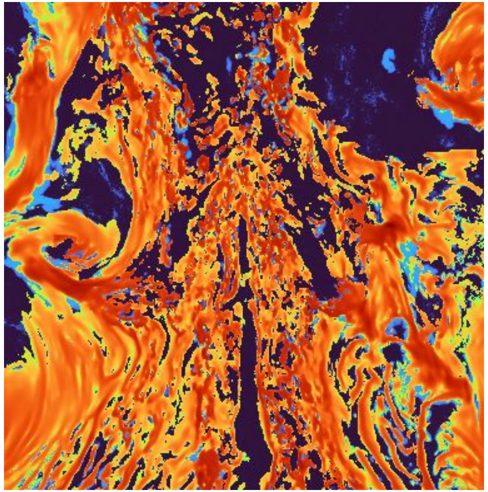
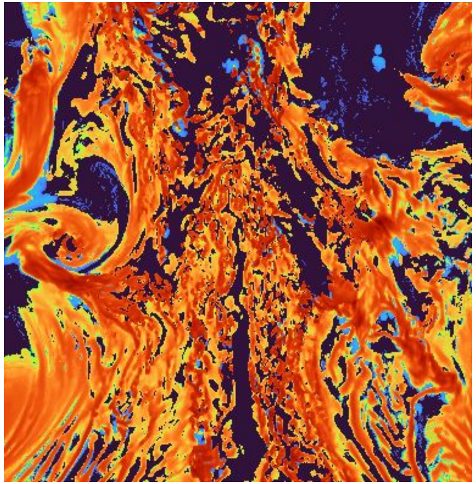


Image SR



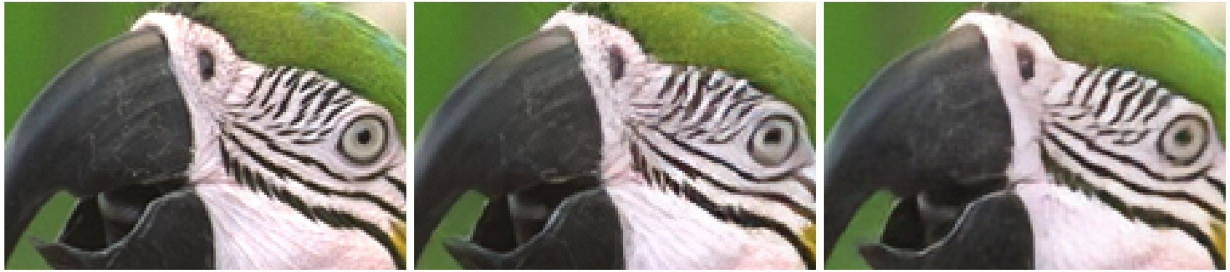
Video SR
(ours)



Truth

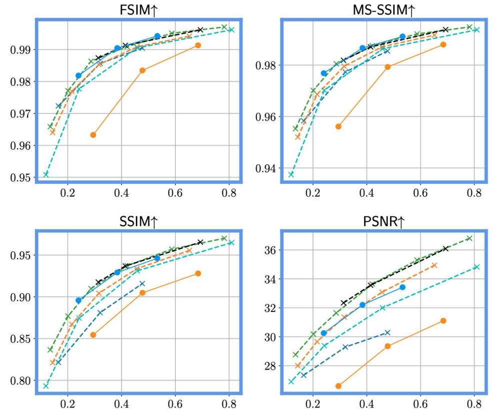
P. Srivastava, R. Yang, G. Kerrigan, G. Dresdner, J. McGibbon, C. Bretherton, S. Mandt. arXiv:2312.06071

Diffusion Models for Data Compression



(a) Ground Truth (b) CDC($\rho = 0.9$) (bpp=0.205) (c) HiFiC (bpp=0.207)

(d) Ground Truth (e) CDC($\rho = 0.9$) **(bpp=0.398)** (f) HiFiC (bpp=0.456)



R. Yang and S. Mandt. Lossy Image Compression with Conditional Diffusion Models. NeurIPS 2023.

This talk: physics, information, and generative modeling

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EWEISS@BERKELEY.EDU
University of California, Berkeley

Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation

these models are unable to aptly describe structure in rich datasets. On the other hand, models that are *flexible* can be molded to fit structure in arbitrary data. For example, we can define models in terms of any (non-negative) function $\phi(\mathbf{x})$ yielding the flexible distribution $p(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$, where Z is a normalization constant. However, computing this

Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems

Gavin E. Crooks[1]

Received October 20, 1997; final December 16, 1997

An equality has recently been shown relating the free energy difference between two equilibrium ensembles of a system and an ensemble average of the work required to switch between these two configurations. In the present paper it is shown that this result can be derived under the assumption that the system's dynamics is Markovian and microscopically reversible.

KEY WORDS: Nonequilibrium statistical mechanics; free energy; work; thermodynamic integration; thermodynamic perturbation.

1. INTRODUCTION

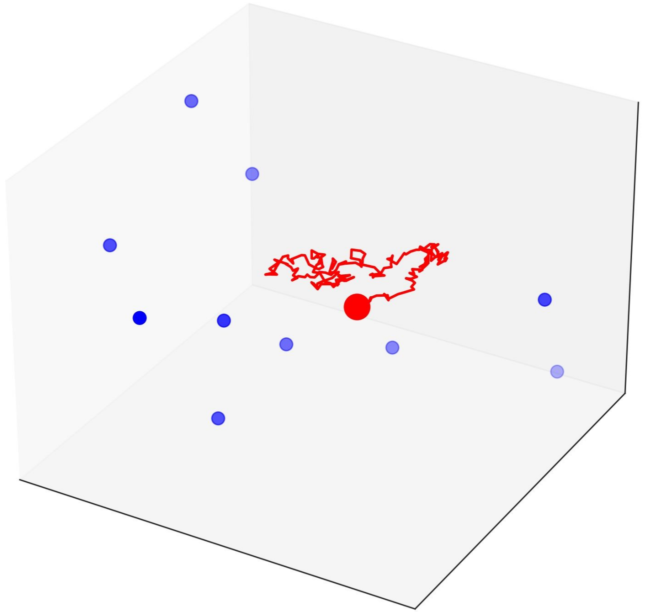
Consider a classical system in contact with a constant temperature heat bath where some degree of freedom of the system can be controlled.

- Diffusion models are rooted non-equilibrium thermodynamics
  - Theory of irreversible processes; entropy production
  - Also connects to information theory and efficient data communication

Diffusion Models

Background: Brownian Motion

- Heavy particle (red) in a “bath” of particles (blue), frequent collisions
- Whole system is Newtonian/deterministic, but subsystem *appears* stochastic
- Stochastic process perspective:
  - deterministic “drift” *f* (external forces)
  - stochastic “diffusion” *dW* (due to collisions)



$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

drift

diffusion

Diffusion in Generative AI

"**Creating noise from data is easy**… "

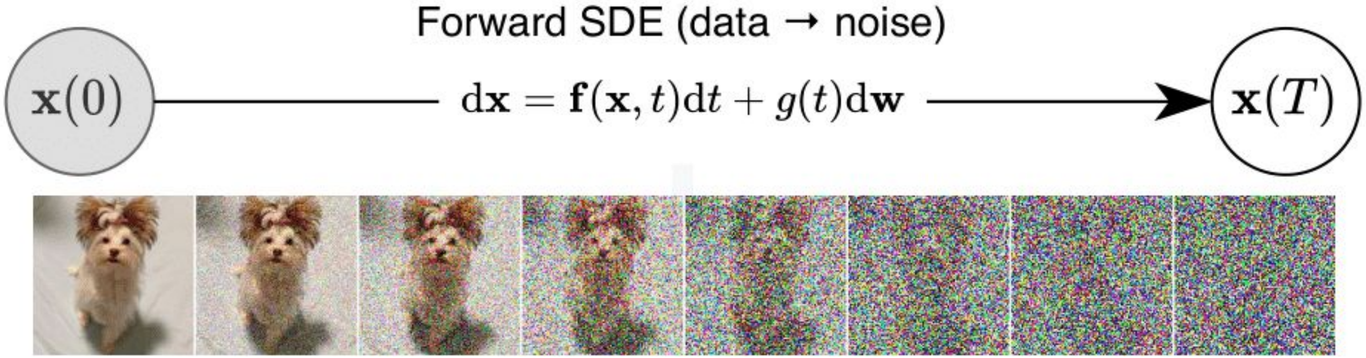


Image Credits: https://arxiv.org/abs/2011.13456

Diffusion in Generative AI

"Creating noise from data is easy… **creating data from noise is generative modeling**"

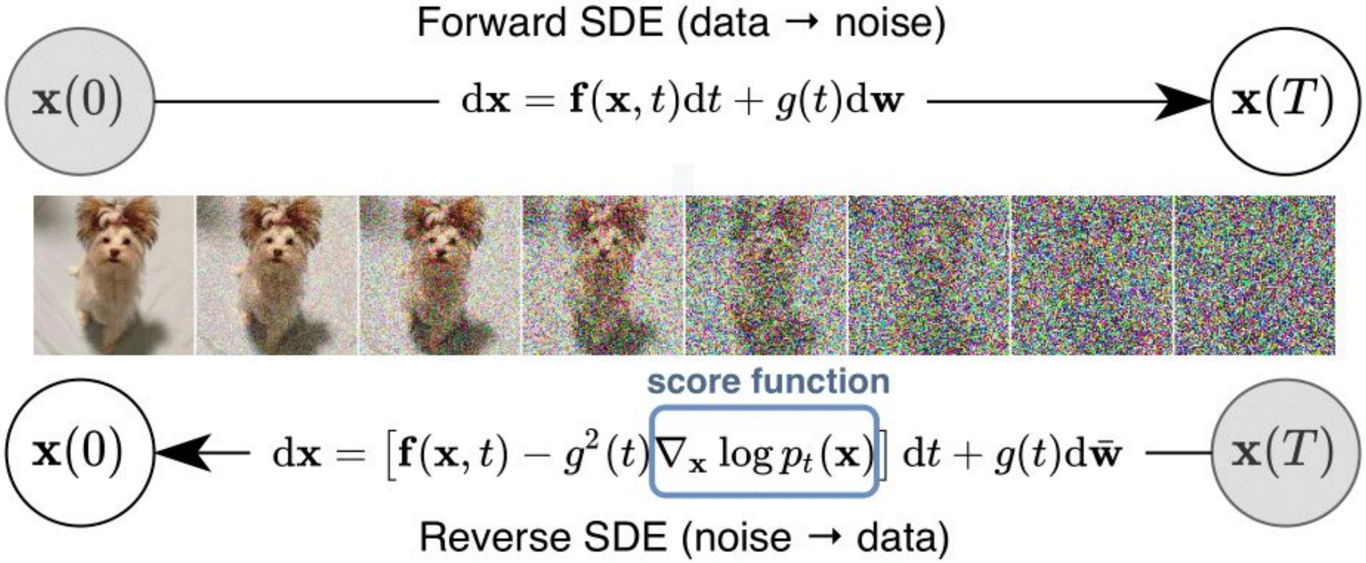


Image Credits: https://arxiv.org/abs/2011.13456

Forward Process

$$dz_t = F_t z_t dt + G_t dw_t, \quad t \in [0, T]$$

- Let z = (x,m) be a more general coordinate (e.g., pixel space + “momenta”)
- Increases entropy by transforming data to noise
- Usually no trainable parameters, no neural networks
- Desired: convergence to a Gaussian (so that we can sample from the inverse process)

Reverse Process

$$dz_t = [F_t z_t - G_t G_t^\top \underbrace{\boxed{\nabla \log p(z_t)}}_{\text{Score}}]dt + G_t dw_t$$

- *Reduces* entropy by transforming noise to data
- Notably, the **score** $\nabla \log p(z_t)$ enters the process (flow to high density regions)
- Unfortunately, we don’t know the score!

Reverse Process

$$dz_t = [F_t z_t - G_t G_t^\top \boxed{s_\theta(z_t, t)}]dt + G_t dw_t$$

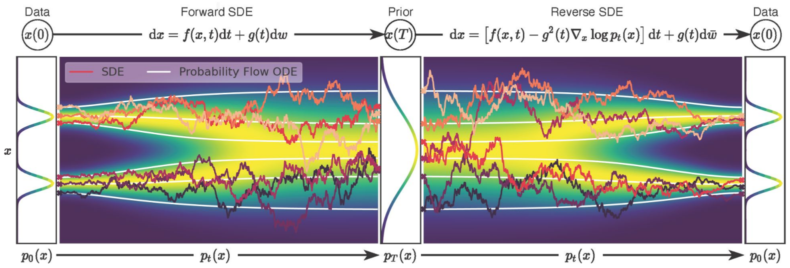
Approximated score

- *Reduces* entropy by transforming noise to data
- The score $\nabla \log p(z_t)$ is approximated using a neural network: $s_\theta(z_t, t)$
- Once the score is learned, we can sample from the model by solving the stochastic differential equation numerically
- The score is learned using **score matching** (regression); skip details.

Part 1:
Augmented Diffusions
and Efficient Integrators

Current diffusion models are inspired by physics

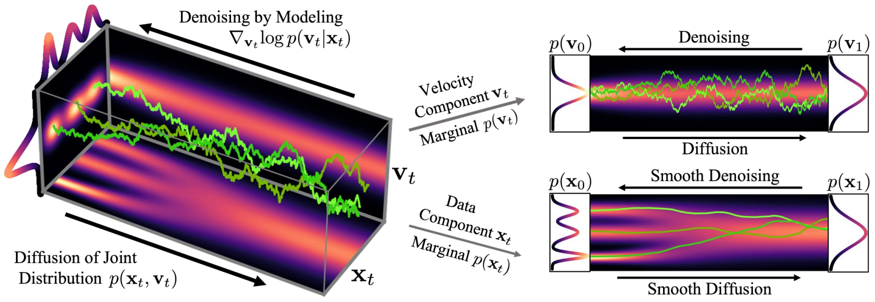
Diffusion in "position" space



- Perform diffusion only in the data space, z = x
- Follow an Ornstein-Uhlenbeck process

$$F_t = -\frac{1}{2}\beta_t I_d \qquad G_t = \sqrt{\beta_t} I_d$$

Diffusion in "phase space"



- Perform diffusion in an augmented space i.e. $z_t = [x_t, m_t]$
- Inspired from Molecular Dynamics

$$F_t = \left(\beta \begin{pmatrix} 0 & M^{-1} \\ -1 & -\Gamma M^{-1} \end{pmatrix} \otimes I_d\right), \quad G_t = \left(\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2\Gamma\beta} \end{pmatrix} \otimes I_d\right)$$

Image Credits: Song et al., Dockhorn, Vahdat, Kreis. ICLR 2022

How to design new diffusion models beyond physics intuition?

- **Augmented dimensions** that aren't necessarily momenta
  - Current models require dim(x) = dim(m) to match
- **Noise sources** that aren't necessarily thermal noise
  - Current models couple thermal noise with momentum, if available
- **Drift forces** that aren't necessarily conservative
  - Forces do not necessarily have to be gradients of scalar potentials

Ma, Chen, and Fox. A Complete Recipe for Stochastic Gradient MCMC. NeurIPS 2015
Singhal, Goldstein, Ranganath. Where to diffuse, how to diffuse, and how to get back. ICLR 2023
Pandey and Mandt. A Complete Recipe for Diffusion Generative Models. ICCV 2023.

A Complete Recipe for Diffusion Models

- **A1:** Consider position and auxiliary variables: $z = [x, m]^\top$
- **A2**: Consider continuous-time, first-order Markov process: $dz = f(z)dt + \sqrt{2D(z)}dw_t$
- **A 3**: Demand converge to a **simple**, pre-specified prior: $p_s(z) \propto \exp(-H(z))$

  e.g., $H(z) = |x|^2 + |m|^2$

- **Result**: The following parameterization is **complete (always exists & unique):**

$$f(z) = -(D(z) + Q(z))\nabla H + \tau(z)$$

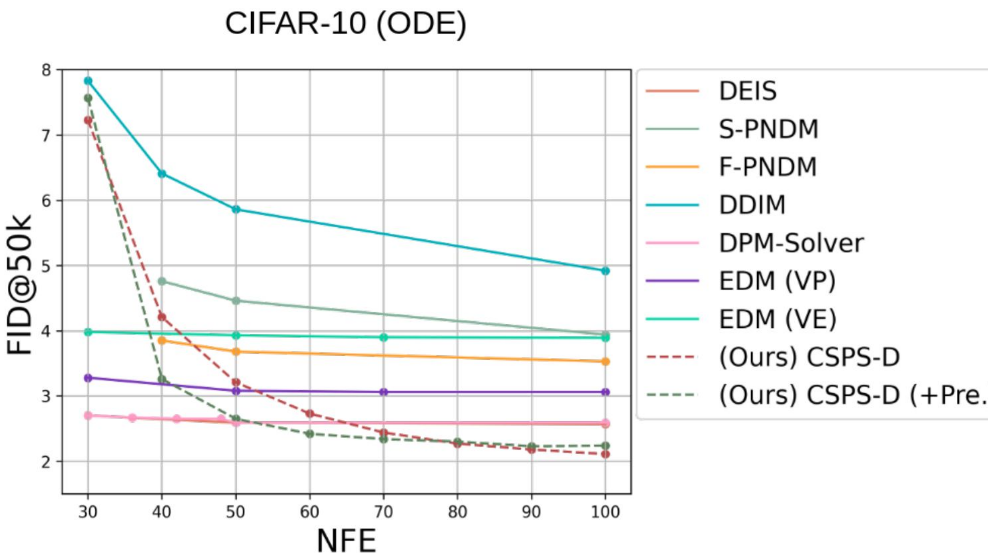
$$\tau_i(z) = \sum_{j=1}^{d} \frac{\partial}{\partial z_j}(D_{ij}(z) + Q_{ij}(z))$$

Pandey and Mandt. A Complete Recipe for Diffusion Generative Models. ICCV 2023.

Phase-Space Langevin Diffusion: A New Sampler

$$d\begin{pmatrix} x \\ m \end{pmatrix} = \frac{\beta}{2}\begin{pmatrix} -\Gamma & M^{-1} \\ -1 & -\nu \end{pmatrix}\begin{pmatrix} x \\ m \end{pmatrix} dt + \begin{pmatrix} \sqrt{\Gamma\beta} & 0 \\ 0 & \sqrt{M\nu\beta} \end{pmatrix} dw_t,$$

- The noise parameter Γ is "unphysical", but improves convergence
- Noise sources in both position and momentum
- Works even better with splitting integrators



Pandey, and Mandt, ICCV 2023; Pandey, Rudolph, and Mandt, ICLR 2024.

Phase-Space Langevin Diffusion: A New Sampler

Pandey, and Mandt, ICCV 2023; Pandey, Rudolph, and Mandt, ICLR 2024.

Diffusion models may live in poorly-conditioned geometries

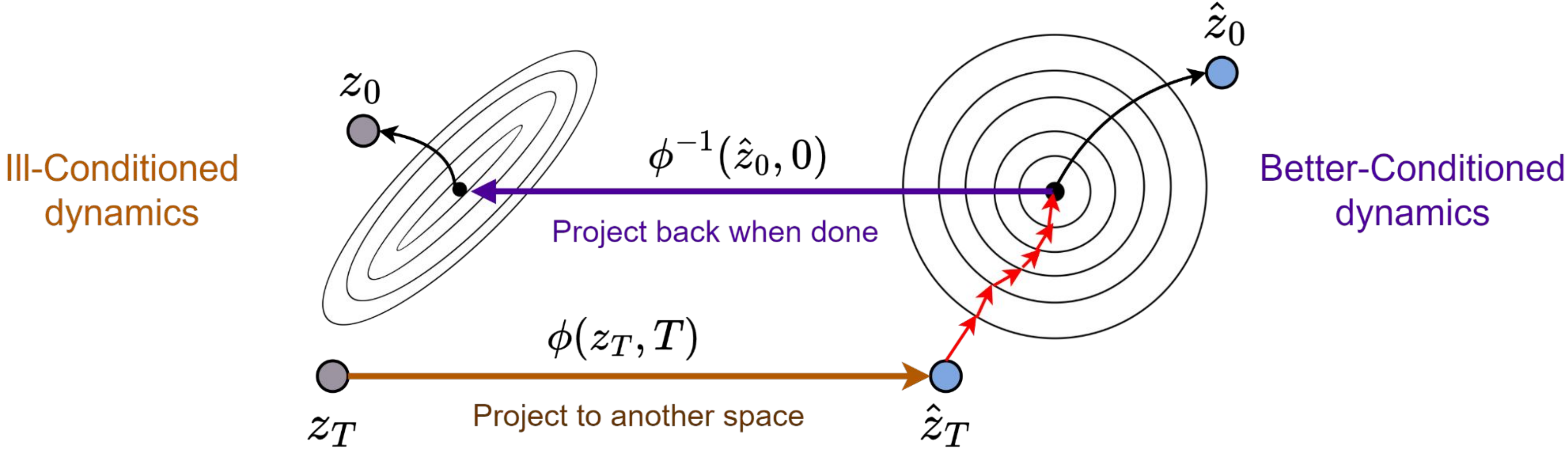
$$dz_t = \left[F_t z_t - \frac{1}{2} G_t G_t^\top s_\theta(z_t, t)\right] dt$$

Linear drift; could be solved analytically in isolation

Diffusion coefficient matrix; can complicate sampling geometry

- Claim: the geometry is "ill-conditioned" (as in optimization)
- Intuition: change of coordinates before simulating the equations
  - Simpler equations can be solved in fewer discretization steps
  - Eliminate the state-independent drift

Conjugate Integrators



Works in already trained models! E.g., OpenAI diffusion models trained on ImageNet

Pandey, Rudolph, Mandt. Efficient Integrators for Diffusion Generative Models. ICLR 2024.

Conjugate Integrators

Consider linear transformation, $\hat{z}_t = A_t z_t$

Score parameterization: $s_\theta(z_t, t) = C_{\text{skip}}(t) z_t + C_{\text{out}}(t) \epsilon(z_t, t)$

Define

$$A_t = \exp\left(\int_0^t B_s - F_s + \frac{1}{2} G_s G_s^\top C_{\text{skip}}(s) ds\right),$$

$$\Phi_t = -\int_0^t \frac{1}{2} A_s G_s G_s^\top C_{\text{out}}(s) ds$$

After some straightforward math:

$$d\hat{z}_t = A_t B_t A_t^{-1} \hat{z}_t dt + d\Phi_t \epsilon_\theta \left(A_t^{-1} \hat{z}_t, t\right)$$

Optional damping term;
can set B=0 or B =-λI

Linear preconditioner, can be precomputed

Conjugate Integrators - Theory

- **Connections to ODE stability criteria**
  - Let $\mathcal{G}$ be the flow map of an ODE integrator. *Stability* implies that $\forall \Delta \; \exists \epsilon$:

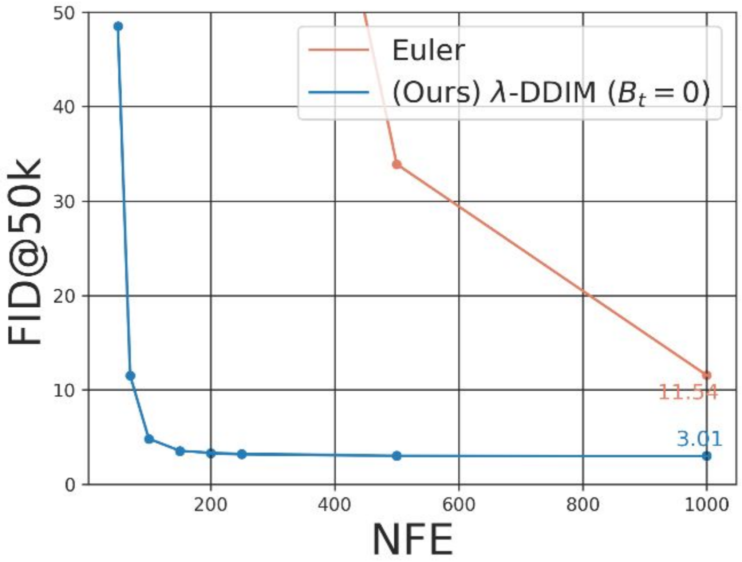
$$\|\mathcal{G}_h(\hat{z}(t)) - \mathcal{G}_h(\hat{z}_t)\| \leq \Delta, \qquad \text{s.t. } \|\hat{z}(t) - \hat{z}_t\| < \epsilon,\ \epsilon > 0, \Delta > 0$$

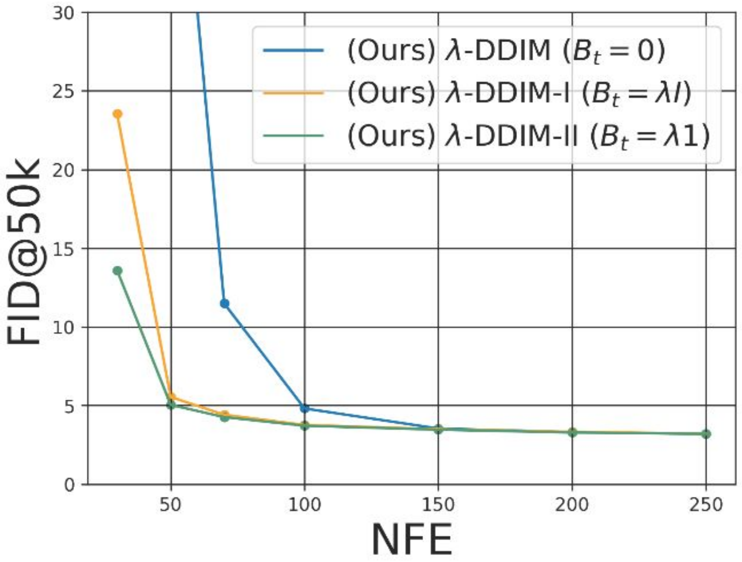
Stability analysis considers the Eigenvalues of the linearized flow operator
Different choices of B can enhance stability, improving over existing samplers

- **Connections to prior methods:**
  - Interestingly, $B_t = 0$ corresponds to DDIM for diffusion models.
  - Also connections with fast samplers based on exponential integrators

Pandey, Rudolph, Mandt. Efficient Integrators for Diffusion Generative Models. ICLR 2024.

Speed vs. Sample Quality Tradeoffs



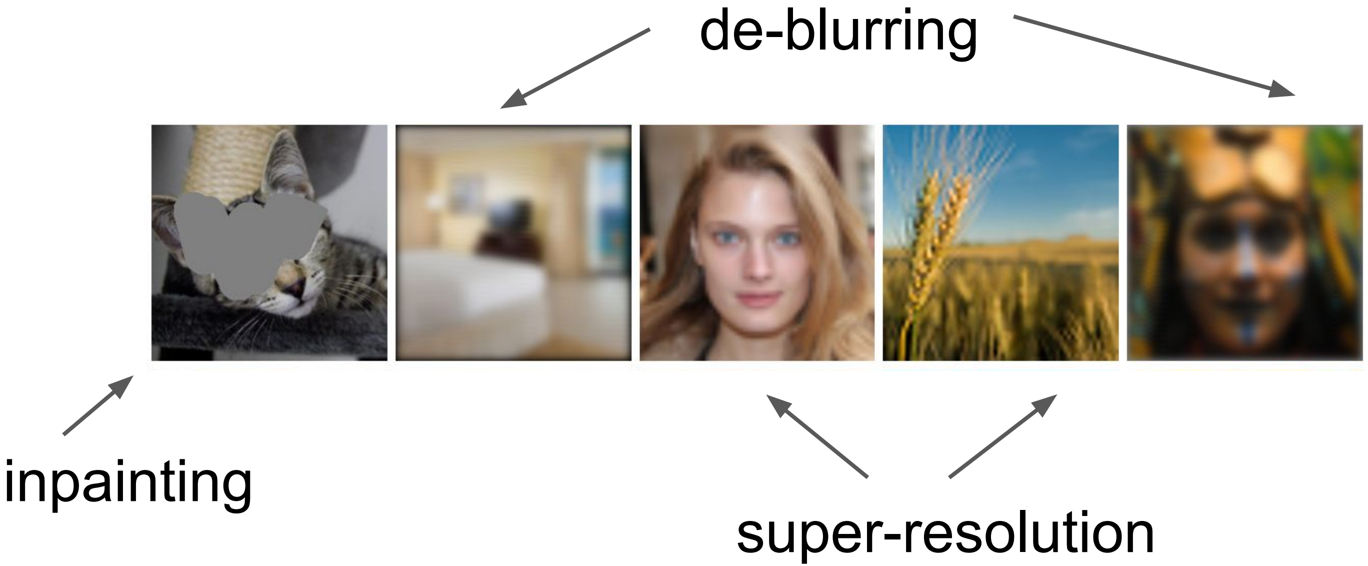


- Dramatic speed improvements; in particular for phase-space diffusions
- Non-zero damping term **λ** can further improve performance

Now, consider inverse problems

Degradation operator H:

$$\mathbf{y} = \boldsymbol{H}\mathbf{x}_0 + \sigma_y \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \boldsymbol{I}),\ \mathbf{x}_0 \sim p_{\text{data}}$$



Now, consider inverse problems

Degradation operator H:

$$\mathbf{y} = \boldsymbol{H}\mathbf{x}_0 + \sigma_y \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \boldsymbol{I}),\ \mathbf{x}_0 \sim p_{\text{data}}$$

We will consider two types of iterative refinement models:

- Diffusion models
- Flow matching models

$$\text{Diffusion:} \qquad d\mathbf{x}_t = \left[\boldsymbol{F}_t \mathbf{x}_t - \frac{1}{2}\boldsymbol{G}_t \boldsymbol{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \boldsymbol{y})\right] dt, \tag{3}$$

$$\text{Flows:} \qquad d\mathbf{x}_t = \boldsymbol{b}(\mathbf{x}_t, \boldsymbol{y}, t)\, dt,$$

Pandey, Yang, Mandt. Fast Samplers for Inverse Problems in Iterative Refinement Models. https://arxiv.org/pdf/2405.17673

Fast Samplers for Inverse Problems in Iterative Refinement Models

Degradation operator H:

$$\mathbf{y} = \boldsymbol{H}\mathbf{x}_0 + \sigma_y \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \boldsymbol{I}),\ \mathbf{x}_0 \sim p_{\text{data}}$$

Proposition 1. *For the conditional diffusion dynamics defined in Eqn. 3, introducing a diffeomorphism, $\bar{\mathbf{x}}_t = \boldsymbol{A}_t \mathbf{x}_t$, where,*

$$\boldsymbol{A}_t = \mathbf{exp}\left(\int_0^t \boldsymbol{B}_s - \boldsymbol{F}_s ds\right), \qquad \boldsymbol{\Phi}_t = -\int_0^t \frac{1}{2}\boldsymbol{A}_s \boldsymbol{G}_s \boldsymbol{G}_s^\top \boldsymbol{C}_{out}(s) ds, \tag{6}$$

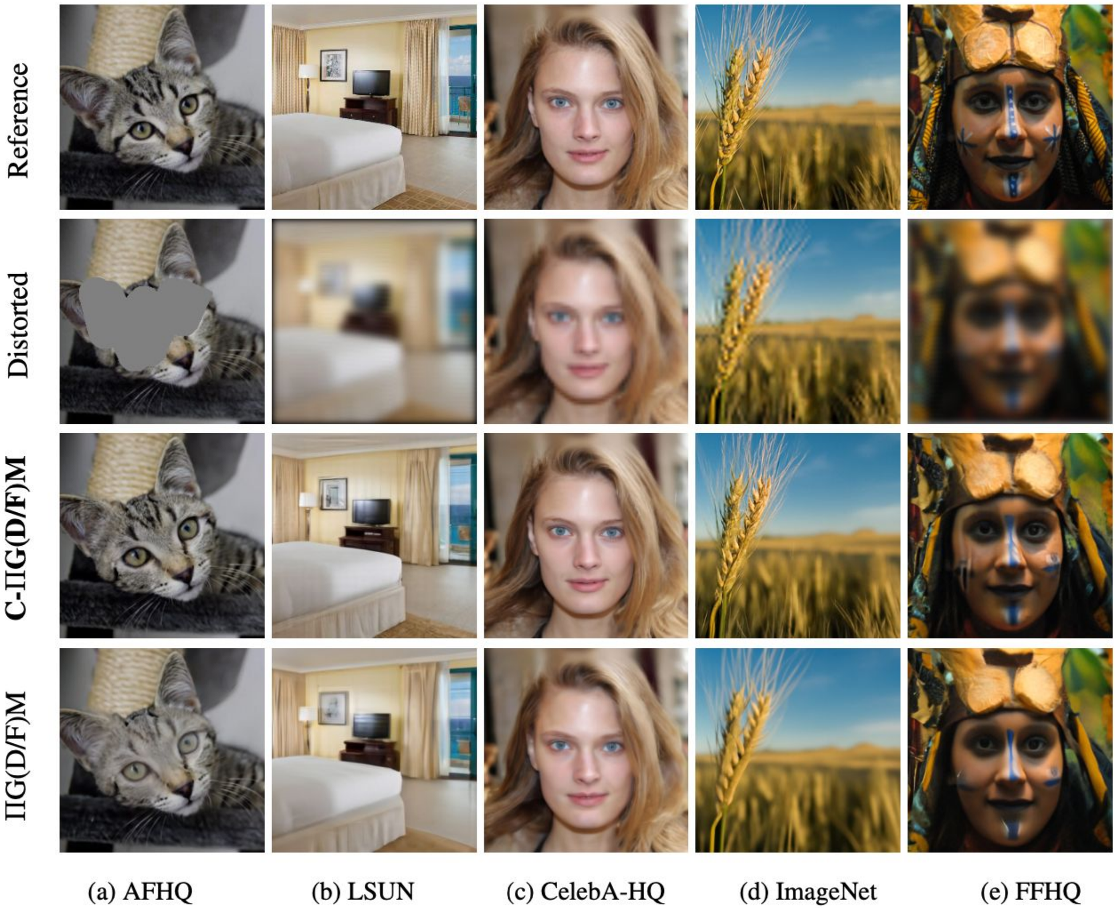
induces the following projected diffusion dynamics,

$$d\hat{\mathbf{x}}_t = \boldsymbol{A}_t \boldsymbol{B}_t \boldsymbol{A}_t^{-1} \hat{\mathbf{x}}_t dt + d\boldsymbol{\Phi}_t \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \frac{w_t r_t^{-2}}{2} \boldsymbol{G}_t \boldsymbol{G}_t^\top \frac{\partial \hat{\mathbf{x}}_0}{\partial \mathbf{x}_t}^\top (\boldsymbol{H}^\dagger \boldsymbol{y} - \boldsymbol{P}\hat{\mathbf{x}}_0) dt, \tag{7}$$

where $\boldsymbol{H}^\dagger = \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}$ and $\boldsymbol{P} = \boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{H}^\top)^{-1}\boldsymbol{H}$ represent the pseudoinverse and the orthogonal projector operators for the degradation operator $\boldsymbol{H}$. (Proof in Appendix A.2)

Main idea: assign different dynamics to degradation operator’s null space and its orthogonal complement.

Conjugate Integrators: high-quality generation in only five iterations



Pandey, Yang, Mandt. https://arxiv.org/pdf/2405.17673

Conjugate Integrators: high-quality generation in only five iterations

Ground Truth

Regular integrator (NFE=5)

Conjugate Integrator (NFE=5)

Pandey, Yang, Mandt. Fast Samplers for Inverse Problems in Iterative Refinement Models. https://arxiv.org/pdf/2405.17673

Fast Samplers for Inverse Problems in Iterative Refinement Models

Flow Results	NFE	LPIPS↓		KID$\times 10^{-3}$ ↓		FID↓	
		C-ΠGFM	ΠGFM	**C-ΠGFM**	ΠGFM	**C-ΠGFM**	ΠGFM
Inpainting	5	**0.125**	0.240	**17.6**	167.0	**26.95**	161.49
	10	**0.074**	0.188	**8.0**	86.6	**14.64**	94.91
	20	**0.065**	0.144	**4.6**	54.4	**10.93**	65.39
Super-Resolution	5	**0.063**	0.091	**5.5**	17.5	**13.08**	21.84
	10	**0.058**	0.076	**3.6**	12.2	**10.65**	16.73
	20	**0.064**	0.069	3.9	**3.5**	11.07	**10.23**
Deblurring	5	**0.083**	0.114	**3.7**	10.9	**12.86**	18.97
	10	**0.077**	0.088	**5.0**	7.0	**14.41**	15.09
	20	0.080	**0.073**	7.9	**3.1**	17.10	**11.35**

Diffusion Results	NFE	C-ΠGDM	ΠGDM	DPS	DDRM	C-ΠGDM	ΠGDM	DPS	DDRM	C-ΠGDM	ΠGDM	DPS	DDRM
Super-Resolution	5	**0.220**	0.306			**2.7**	6.3			**37.31**	49.06		
	10	**0.206**	0.252	0.252	0.318	**1.6**	4.8	5.8	14.1	**34.22**	44.30	38.18	51.64
	20	**0.207**	0.222			**1.7**	2.5			**34.28**	37.36		
Deblurring	5	**0.272**	0.349			**3.89**	14.1			**44.42**	63.94		
	10	**0.272**	0.294	0.619	0.336	**3.6**	5.3	59.5	12.3	**43.37**	47.80	139.58	62.53
	20	0.268	**0.259**			**3.5**	4.2			**43.70**	44.20		

Pandey, Yang, Mandt. Fast Samplers for Inverse Problems in Iterative Refinement Models. https://arxiv.org/pdf/2405.17673

Reverse diffusion as progressive decompression

A diffusion model can be understood as:

- Denoising autoencoder at multiple noise levels [Vincent 2011, Song & Ermon, 2019]
- Learning to reverse an SDE (mostly this talk) [Song et al., 2021]
- **Deep hierarchical VAE** [Sohl-Dickstein et al., 2015, Ho et al., 2020, Kingma et al., 2021]

-VLB(**x**) = $\mathbb{E}\left[D_{\mathrm{KL}}[q(\mathbf{z}_T \,|\, \mathbf{X}) \,\|\, p_T(\mathbf{z}_T)]\right] + \sum_{s=1}^{T-1} \mathbb{E}\left[D_{\mathrm{KL}}[q(\mathbf{z}_s \,|\, \mathbf{Z}_{s+1}, \mathbf{X}) \,\|\, p(\mathbf{z}_s \,|\, \mathbf{Z}_{s+1})]\right]$

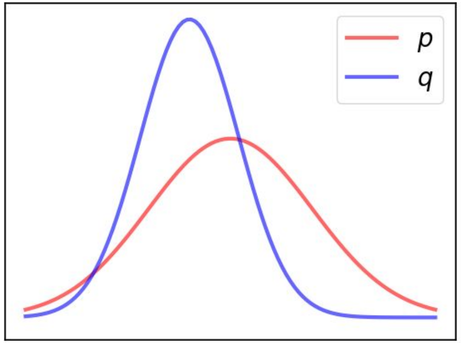
Entropy reduced in every reverse diffusion step

- Can this information be efficiently transmitted between a sender and receiver?

Background: Relative entropy coding

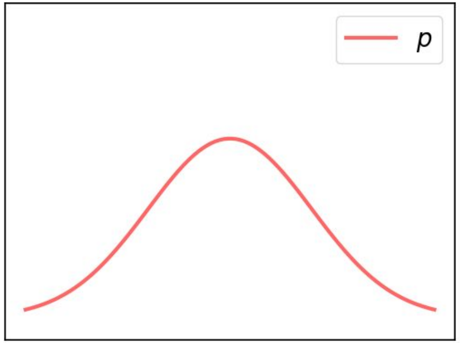
Problem: Alice wants to transmit a sample from **q** to Bob, under shared prior **p**, using KL(**q**||**p**) bits.

Alice



`01011010111...`

Bob



Background: Relative entropy coding

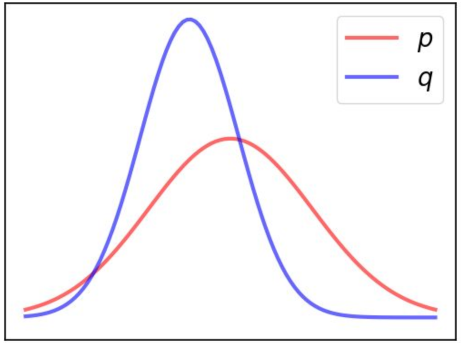
Idea (sketch):

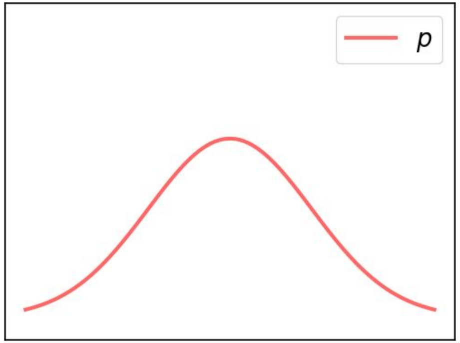
- Let Alice and Bob share a random seed.
- Sample from p many times until we hit a “good” (high likelihood under q) sample
- Transmit the index K in binary.

Alice

01011010111…

Bob





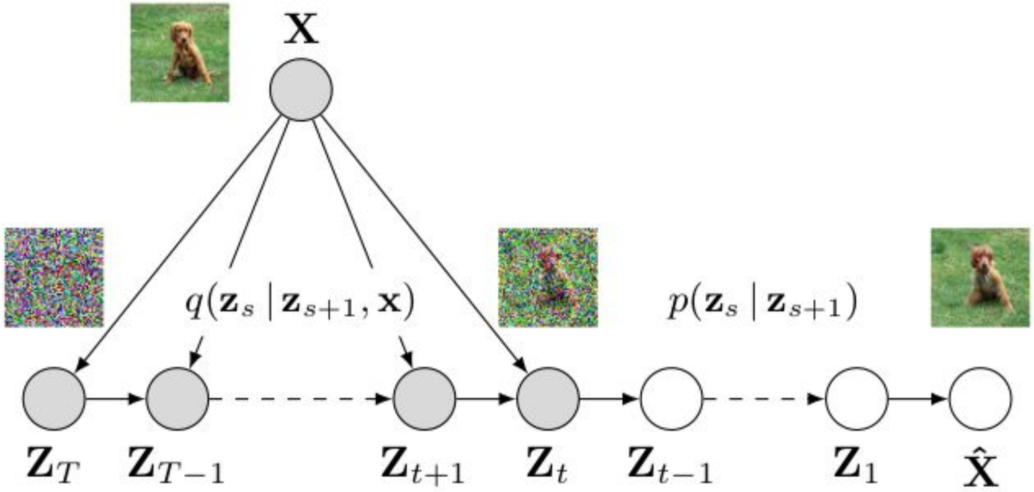
Diffusion models for transmitting information

Problem setup: Alice wants to transmit data **x** to Bob using -VLB(**x**) many bits.

$$\mathbb{E}\left[D_{\mathrm{KL}}[q(\mathbf{z}_T \,|\, \mathbf{X}) \,\|\, p_T(\mathbf{z}_T)]\right] + \textstyle\sum_{s=1}^{T-1} \mathbb{E}\left[D_{\mathrm{KL}}[q(\mathbf{z}_s \,|\, \mathbf{Z}_{s+1}, \mathbf{X}) \,\|\, p(\mathbf{z}_s \,|\, \mathbf{Z}_{s+1})]\right]$$

posterior

shared prior

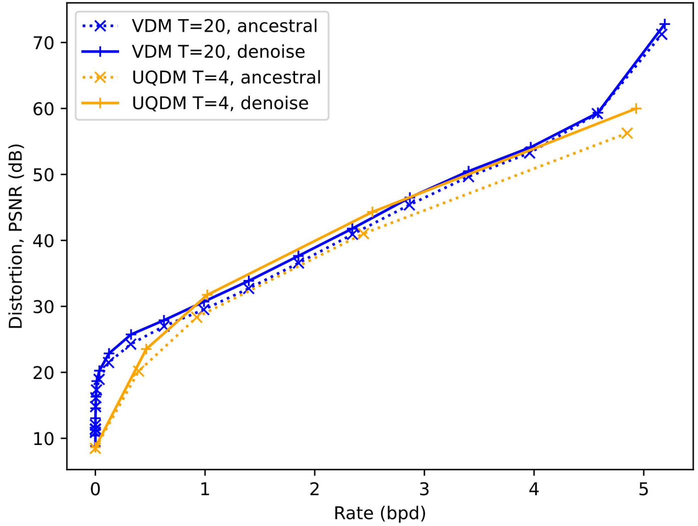


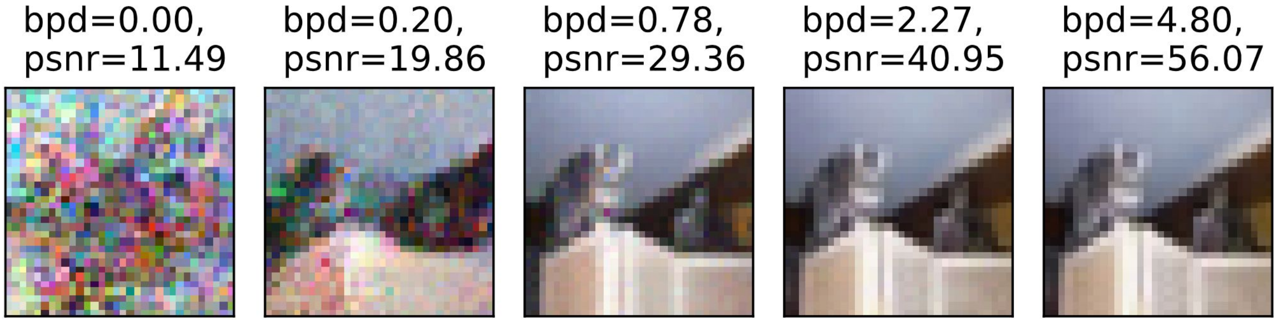
- Works in theory; doesn't scale with dimension
- Practical solution: *universal quantization*
- Requires re-design of the diffusion process, Gaussian→ Uniform

Preliminary results

CIFAR data

- Promising results when compared to Gaussian diffusion at T=20 discretization steps
- Still work in progress; not competitive at larger T & compared to SOTA models

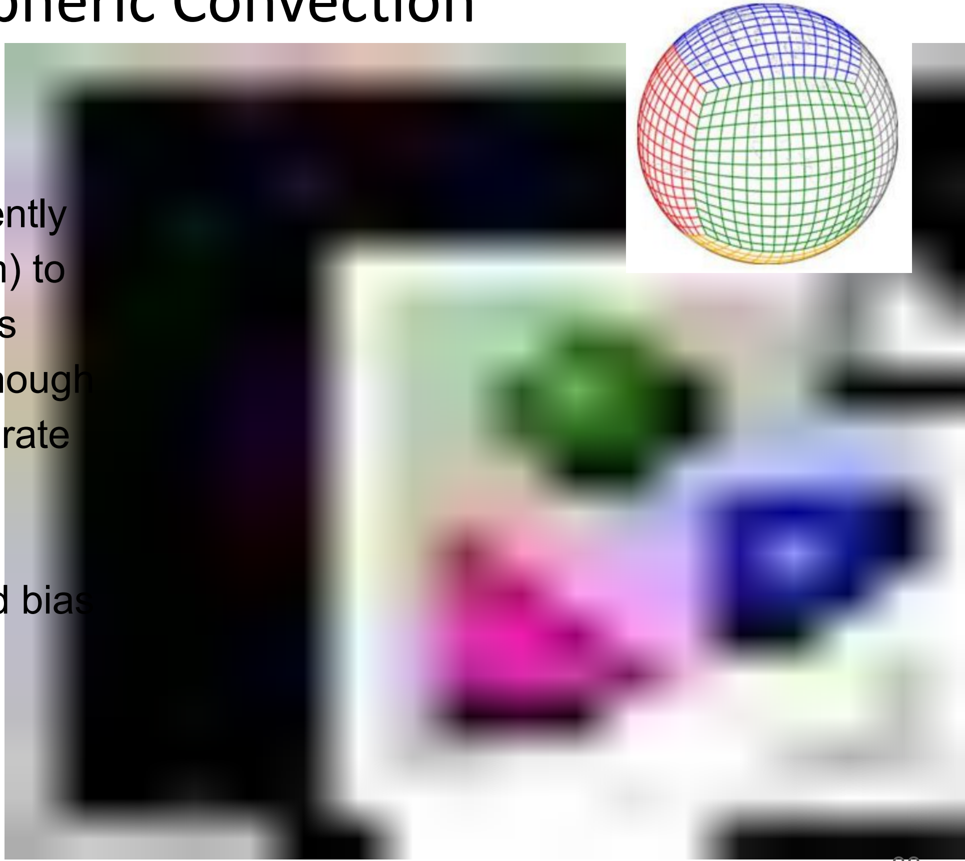




Yang, Mandt, and Theis. An introduction to neural data compression. Foundations and Trends in Computer Vision, 2023

Part 2: Emulating Thermodynamic Processes with Diffusion Models

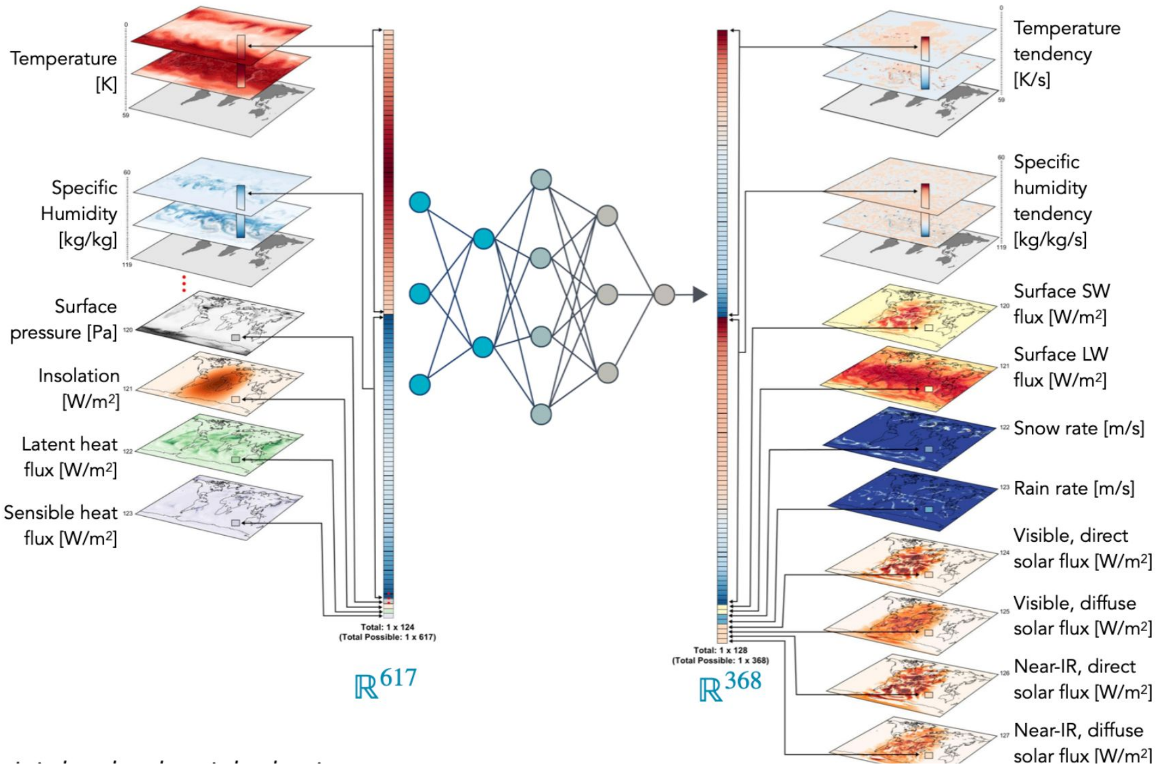
Generative Modeling for Atmospheric Convection



- The climate modeling dilemma:
  - **Either** simulate the climate at sufficiently high spatial resolution (e.g., a few km) to capture, e.g., cloud-related processes
  - **Or**, simulate the climate for a long-enough time (several decades) to make accurate predictions on global warming
- Unresolved processes are huge drivers of uncertainty and introduce randomness and bias
- Can we use generative modeling to stochastically downscale a low-resolution simulation?

Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

Aside: physics-ML hybrid models may solve the resolution dilemma

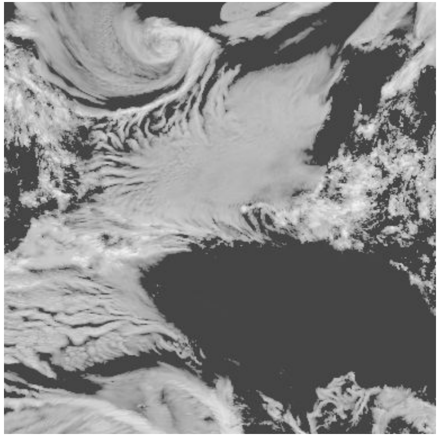


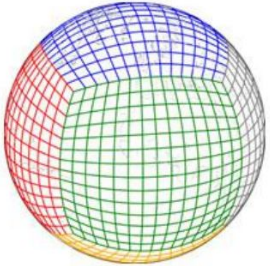
Yu, Sungduk, et al. "ClimSim: A large multi-scale dataset for hybrid physics-ML climate emulation." *NeurIPS 2023*

The data: high-resolution atmospheric simulation

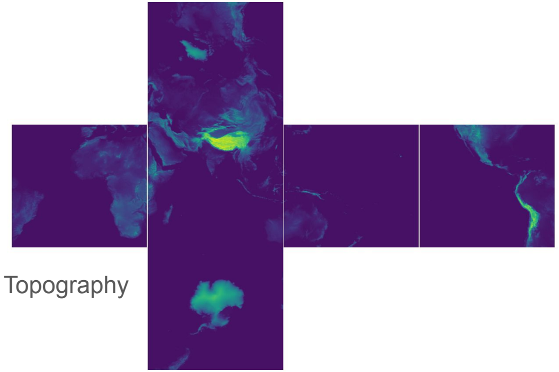
Focus on precipitation channel

- FV3GFS global atmosphere simulation dataset (Allen Institute for AI)
- Captures all relevant physical fields, including precipitation (rain, snow)
- 25 km resolution, 3-hourly





Cube Sphere Gridding



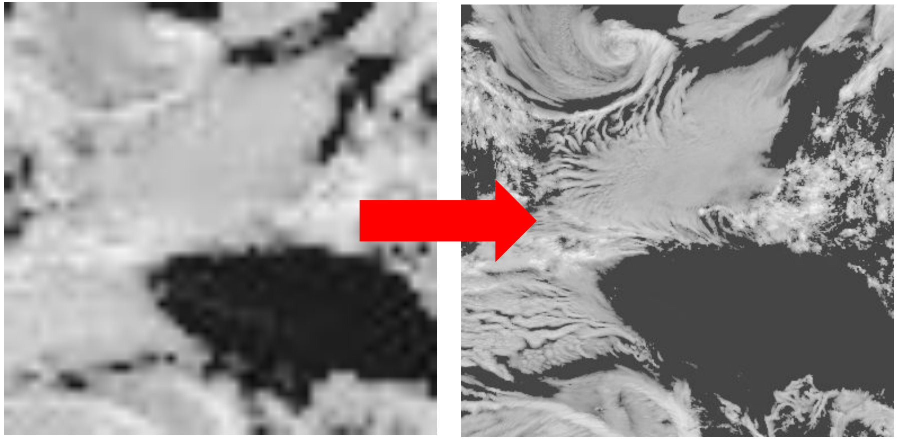
Topography

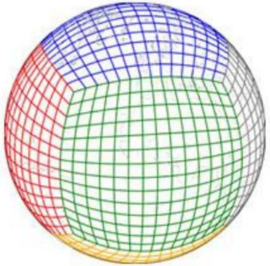
Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

The data: high-resolution atmospheric simulation

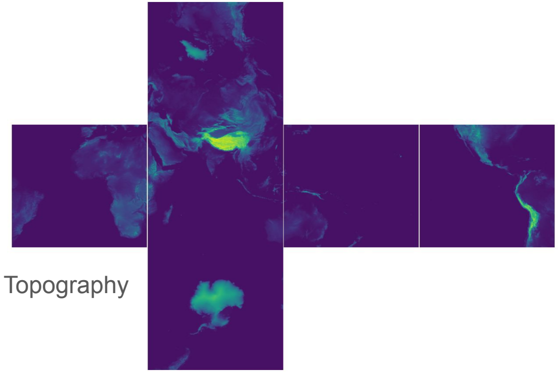
Focus on precipitation channel

- FV3GFS global atmosphere simulation dataset (Allen Institute for AI)
- Captures all relevant physical fields, including precipitation (rain, snow)
- 25 km resolution, 3-hourly





Cube Sphere Gridding



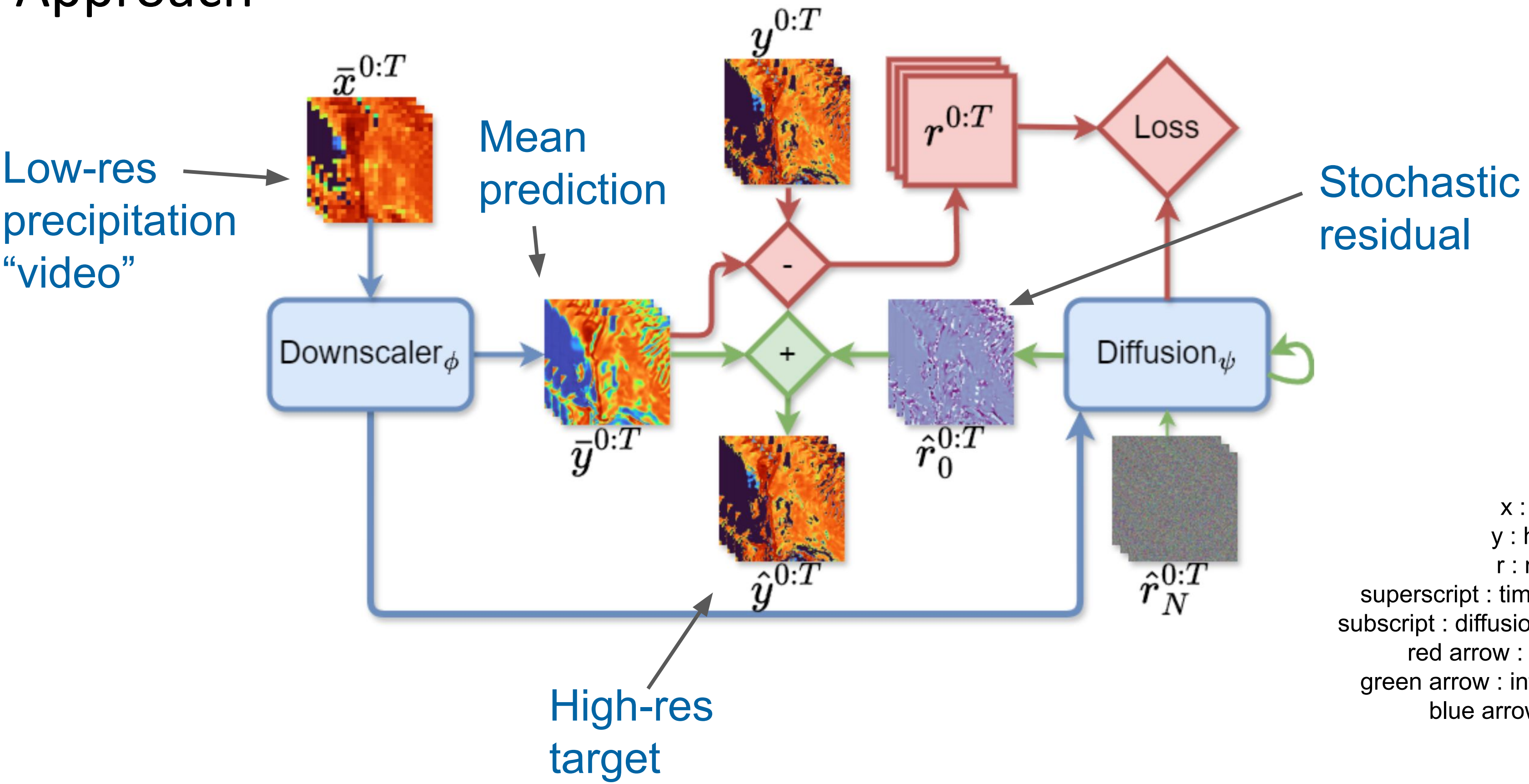
Topography

Goal: super-resolve the precipitation channel

- Allows to run a cheap model to simulate many years
- Use a super-resolution model to convert the data to high spatial resolution

Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

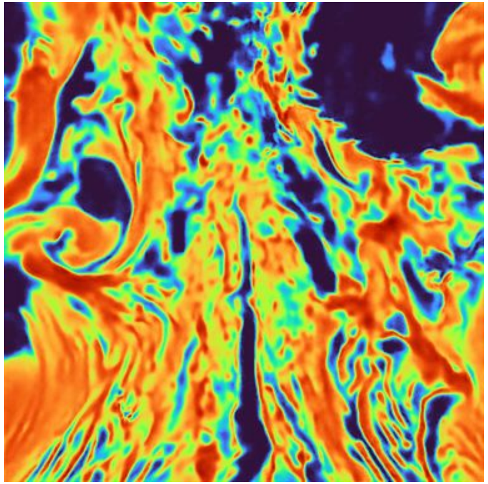
Approach



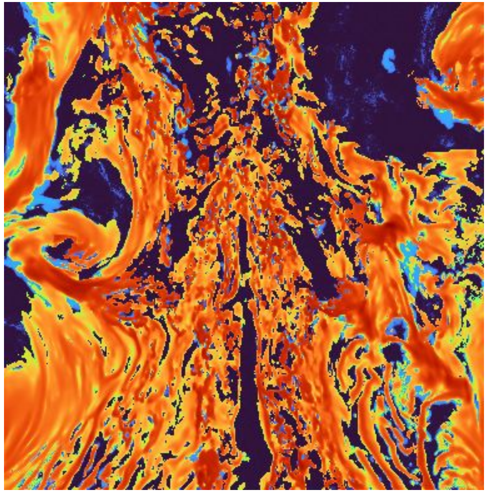
x : low res
y : high res
r : residual
superscript : time index
subscript : diffusion index
red arrow : training
green arrow : inference
blue arrow : both

Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

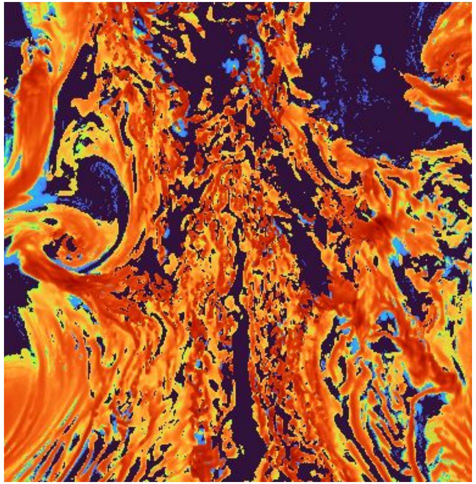
Video Diffusion accurately captures temporal information



single-frame SR



multi-frame SR

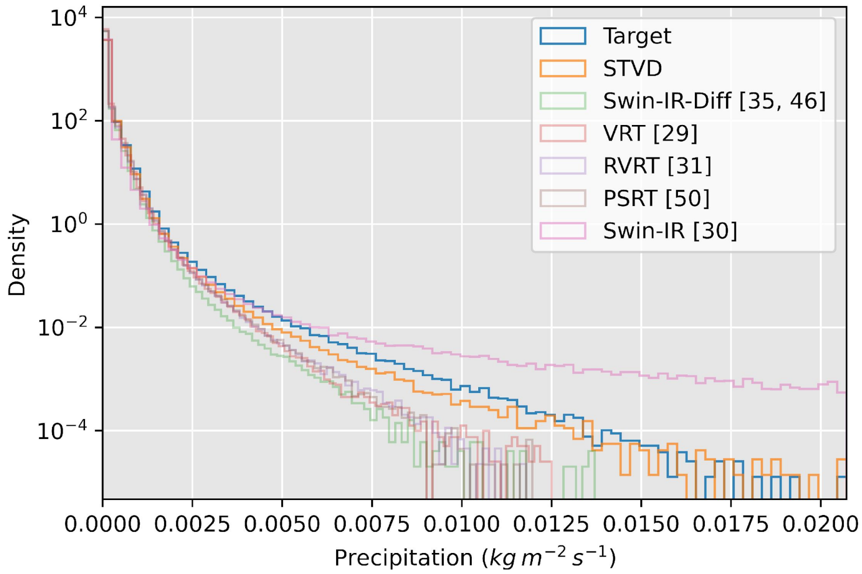


Truth

Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

Deterministic models underestimate extreme precipitation

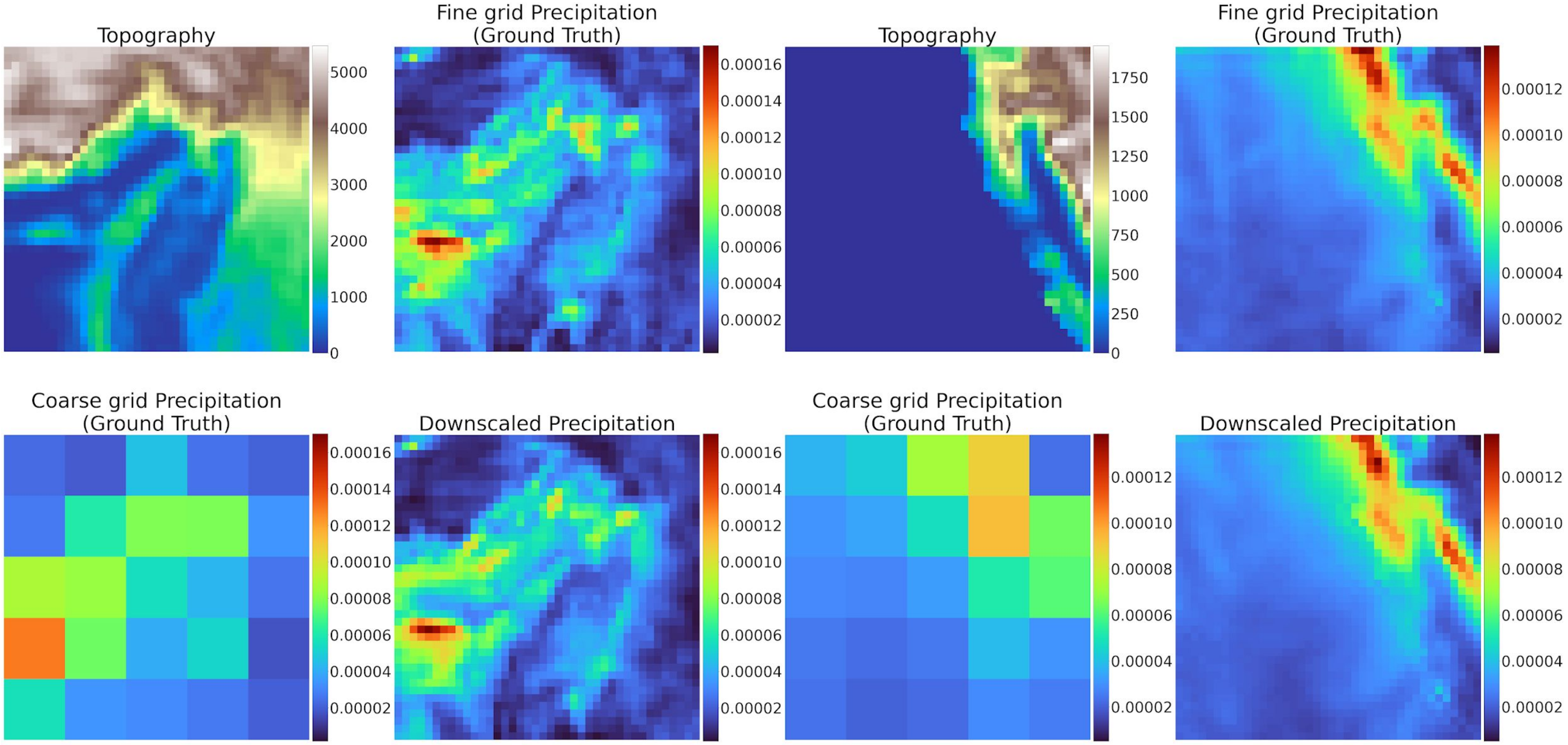
- Downscaling especially challenging for precipitation because of rare extreme events
    - Few geographical regions with extreme rainfall
    - Cyclones/extreme weather events
- While most models can capture low precipitation regions quite well, they heavily downplay the extremes



Precipitation distribution over three-hour windows on all grid points around the globe

Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

Precipitation: Annual Averages

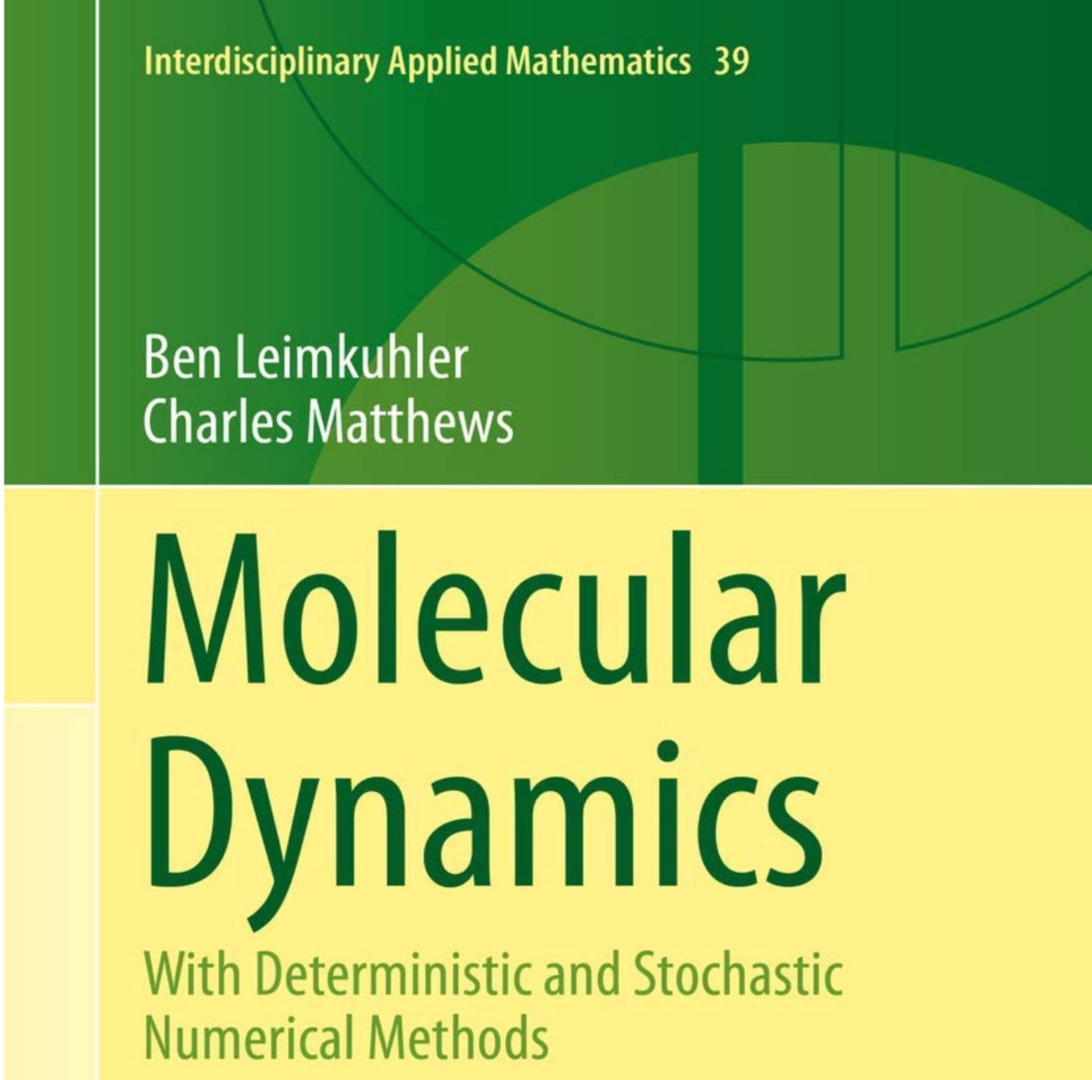


Srivastava, Yang, Kerrigan, Dresdner, McGibbon, Bretherton, Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. arXiv 2023

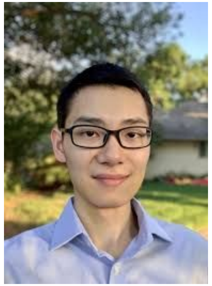
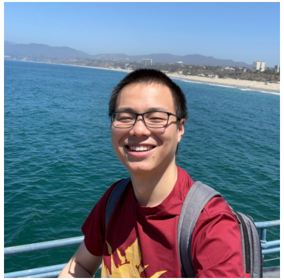
Diffusion Modeling in Molecular Dynamics

A treasure trove for diffusion generative modeling

- Phase space methods
- Conjugate Integrators
- Splitting Integrators



Summary


Kushagra Pandey, Yibo Yang, Ruihan Yang, Prakhar Srivastava

- The thermodynamics of diffusion
  - Origins of diffusion models in thermodynamics
  - Can adopt ideas such as physics-inspired generative processes (and beyond)
  - Efficient sampler design
  - Diffusion models as efficient progressive coders
- Diffusion for thermodynamics/climate
  - Climate data as a playground for generative modeling
  - Requires stochasticity to capture distribution-level properties (e.g., annual precipitation)
- Open questions / future research:
  - How to incorporate physical constraints into modeling convection
  - Unpaired distribution-to-distribution translation between climate models
  - Theoretical analysis of diffusion models: dynamical phase transitions, critical slowing down