Neural QFT Transport States

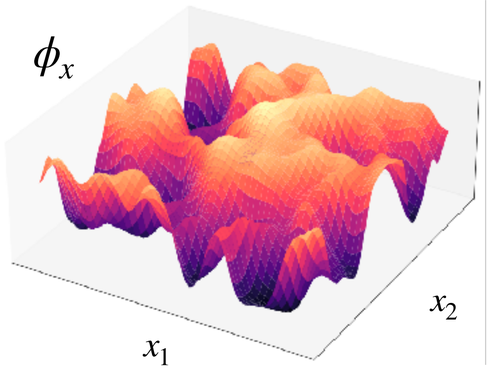
with Jordan Cotler & Michael Albergo

& optimizing diffusion model trajectories

with Zier Mensch, Miranda Cheng, Max Welling; GUD: [2410.02667]

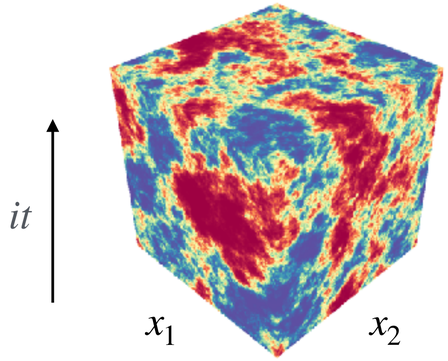
Mathis Gerdes — mgerdes@mit.edu — ECT*, 7 Sep 2026

QFT



$$\hat{H} = \sum_x \left[-\frac{1}{2}\partial^2_{\phi_x} + \frac{1}{2}(\nabla\phi)^2_x + \frac{1}{2}m^2\phi^2_x + \dots \right]$$

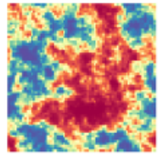
Imaginary time



$$\sim e^{-S[\phi]}$$

$$\langle \mathcal{O} \rangle_S \sim \sum_\phi \mathcal{O}(\phi)$$

Real time $|\Psi\rangle$

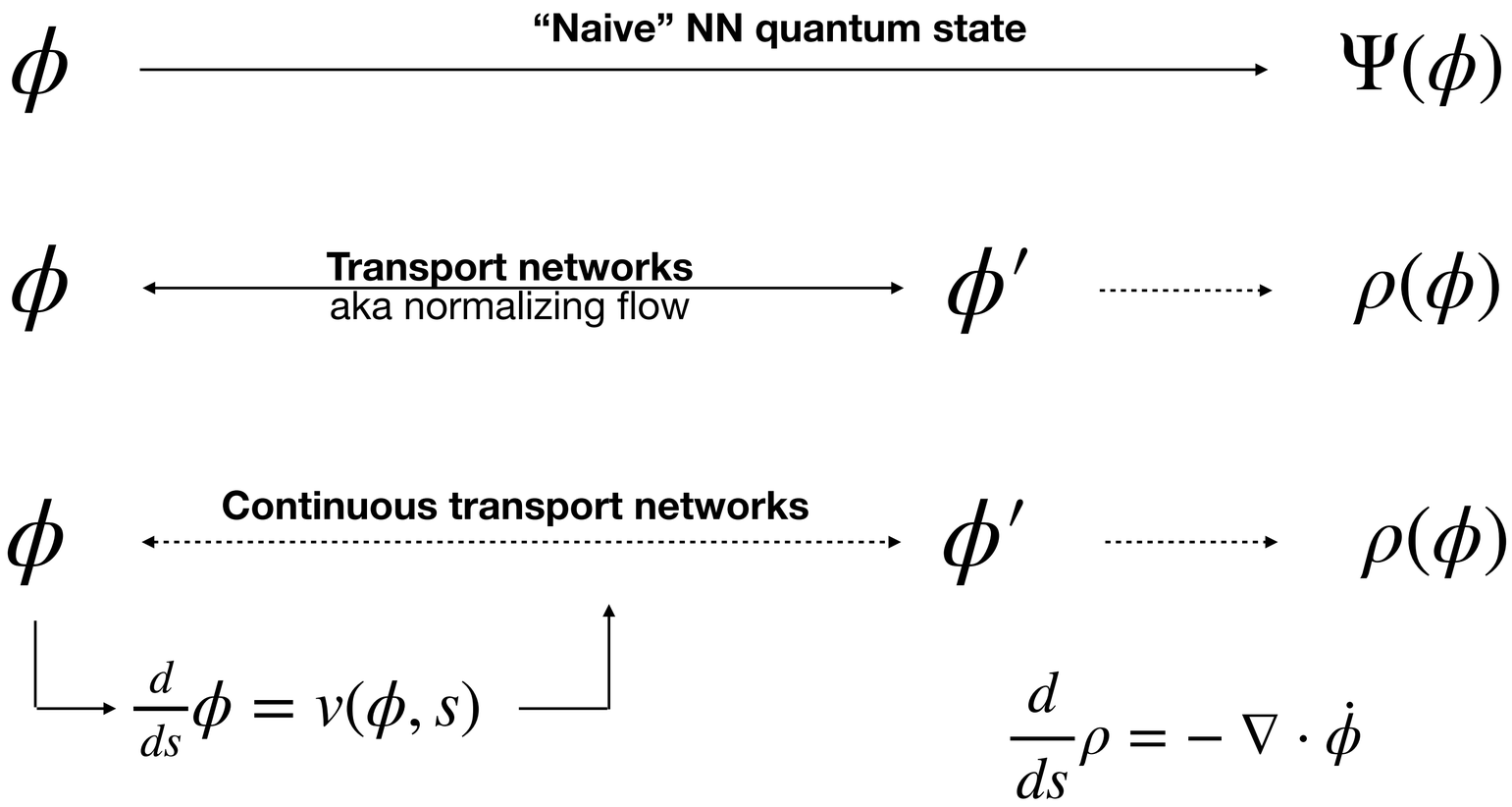


$$\phi \mapsto \psi(\phi) \in \mathbb{C}$$

$$\langle \Psi | \mathcal{O} | \Psi \rangle = \int d\phi \, \psi^* \hat{\mathcal{O}} \psi = \int d\phi \, |\psi|^2 \frac{\hat{\mathcal{O}}\psi}{\psi}$$

ML

$$\rho = |\Psi|^2$$


“Naive” NN quantum state
ϕ
$\Psi(\phi)$
Transport networks
aka normalizing flow
ϕ
ϕ'
$\rho(\phi)$
Continuous transport networks
ϕ
ϕ'
$\rho(\phi)$
$\frac{d}{ds}\phi = v(\phi, s)$
$\frac{d}{ds}\rho = -\nabla \cdot \dot{\phi}$

$$\langle \phi \,|\, \Psi \rangle = \psi(\phi) = \sqrt{\rho}$$

Ground state

$$\min_{\Psi} \langle \Psi \,|\, H \,|\, \Psi \rangle$$

$$\hat{H} \sim \frac{1}{2}\pi^2 + \frac{1}{2}(\nabla_x \phi)^2 + \frac{1}{2}m^2\phi^2 + \dots$$

$$\nabla_x \phi \;\to\; \phi_x - \phi_{x-\hat{\mu}}$$

$$\langle \Psi \,|\, \mathcal{O} \,|\, \Psi \rangle = \int d\phi \,|\psi|^2 \frac{\hat{\mathcal{O}}\psi}{\psi}$$

need $\dfrac{\hat{H}\Psi}{\Psi}\Big|_{\phi} = E_{loc}$

"Point operators"

$$\frac{V(\hat{\phi})\Psi}{\Psi}\Big|_{\phi} = V(\phi)$$

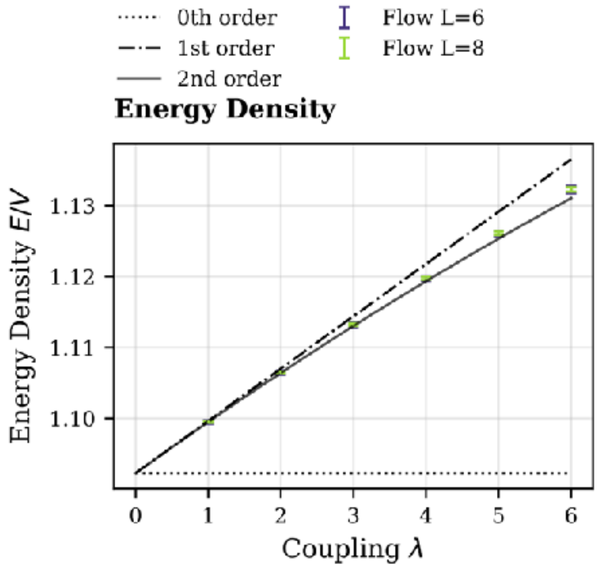


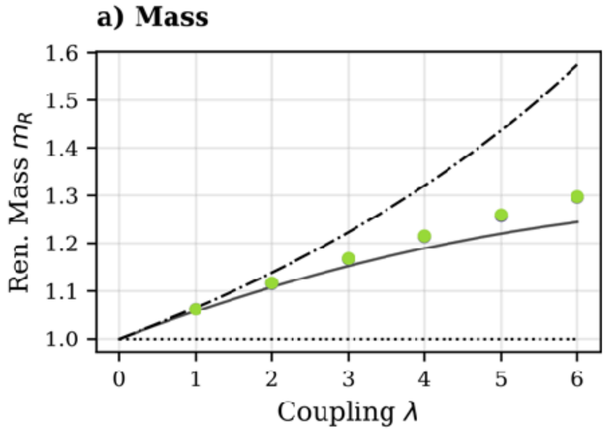
Kinetic term

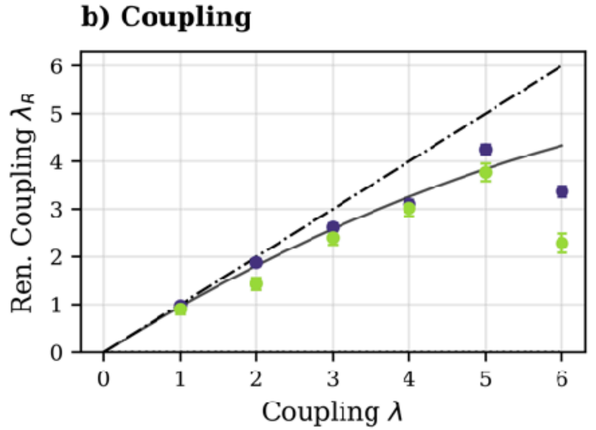
$$\hat{\pi}^2\Psi \propto \nabla^2\Psi$$



Attempt to use “discrete” normalizing flows



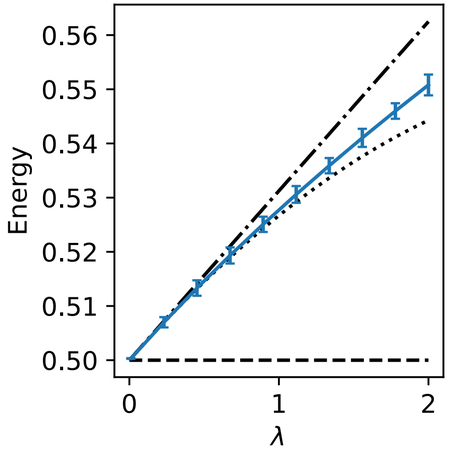




real NVP:
very bad

with analytic bijection:
works but still bad

Anharmonic oscillator



Analytic Bijections for Smooth and Interpretable Normalizing Flows

M. Gerdes, M. C.N. Cheng

ICML 2026, [2601.10774]

Co-propagate everything!

By taking derivatives of Fokker Planck equation, we find e.o.m for:

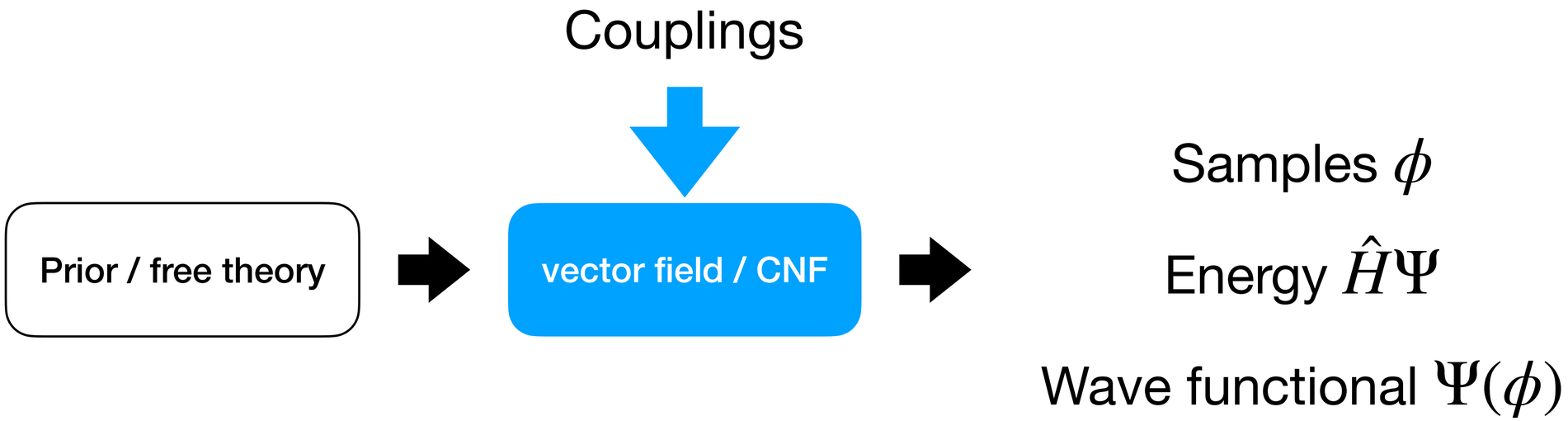
Density: $\log \rho$

Stein score: $s = \nabla \log \rho$

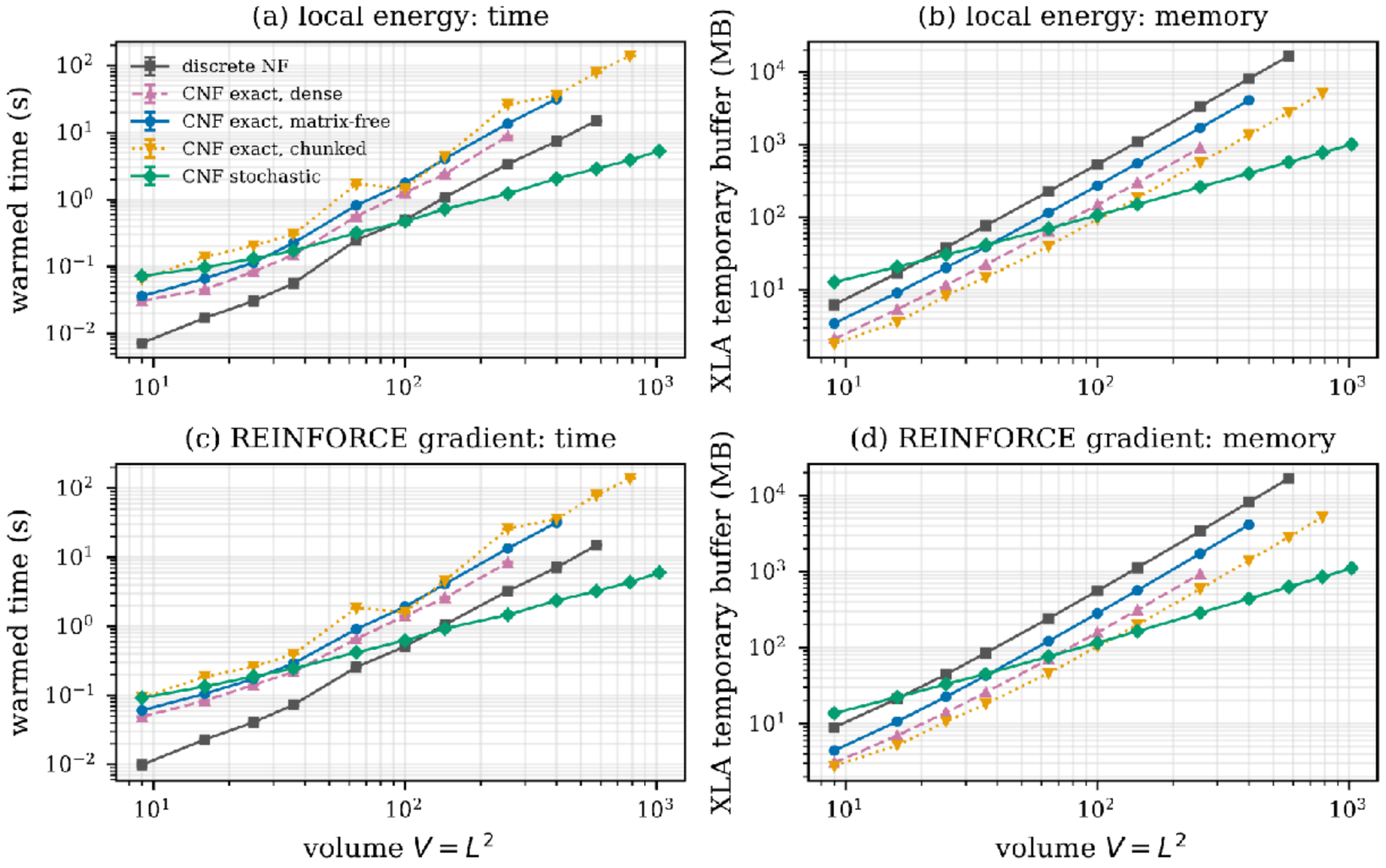
Hessian: $B = \nabla s$

} Kinetic term $\nabla^2 \psi = \nabla^2 e^{\frac{1}{2} \log \rho}$

Neural Transport States



Does it scale?



Exact:

both CNF & NF $\sim V^2$

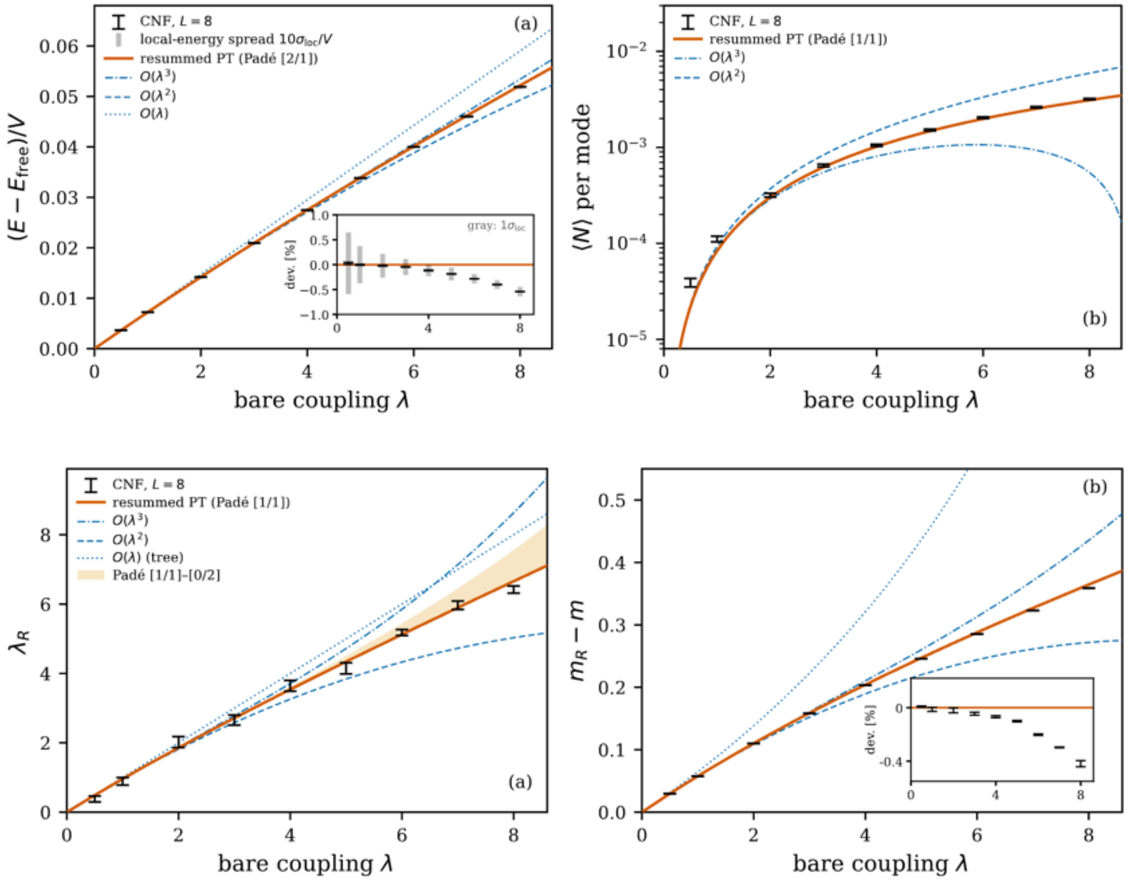
Stochastic laplacian:

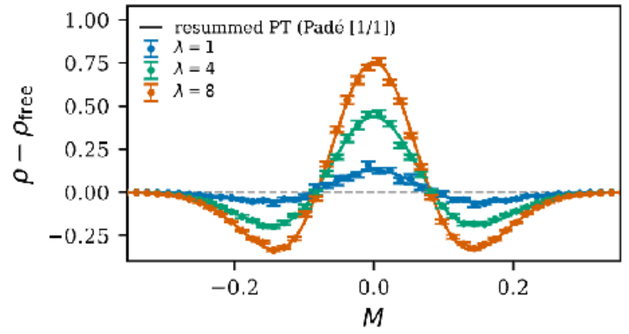
reduce to $\sim KV$

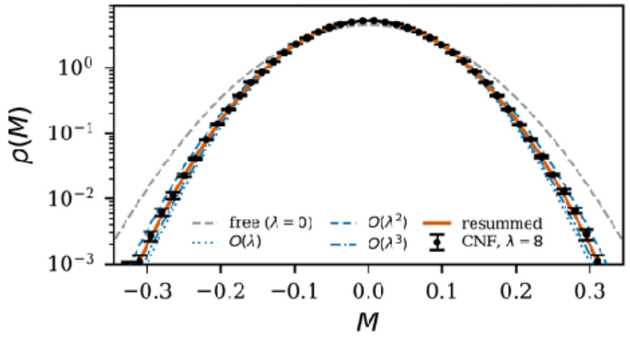
Perturbative validation

Here: 2+1d (1 and 3 + 1 similar)

Up to 1000 sites!







Critical regime validation

1+1d vs DMRG

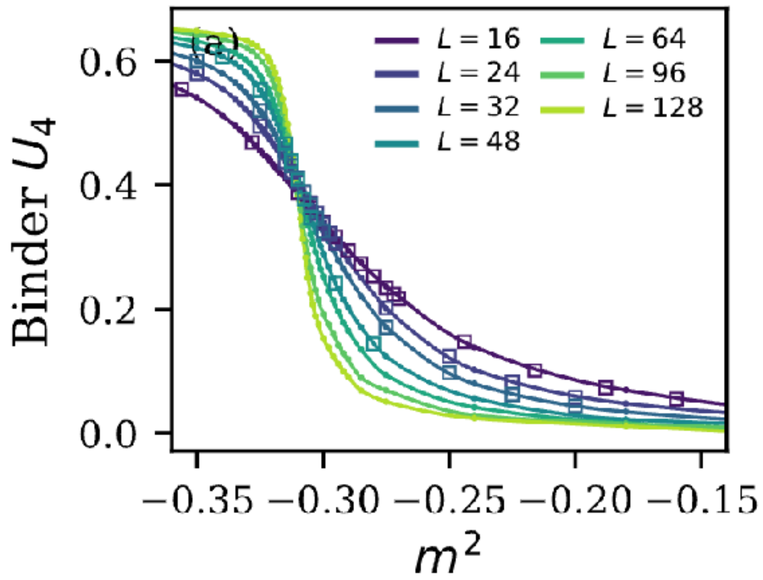
$$A_{\theta r} = |\langle \psi_\theta | \psi_r \rangle|$$

$A_{\theta r} \geq 0.999982$ **across the ladder**

L	$10^6(1 - A_{\theta r})$	seed offset / SEM
8	0.528 ± 0.063	
16	0.791 ± 0.178	
24	2.427 ± 0.441	
32	4.009 ± 0.176	
48	17.621 ± 0.583 †	

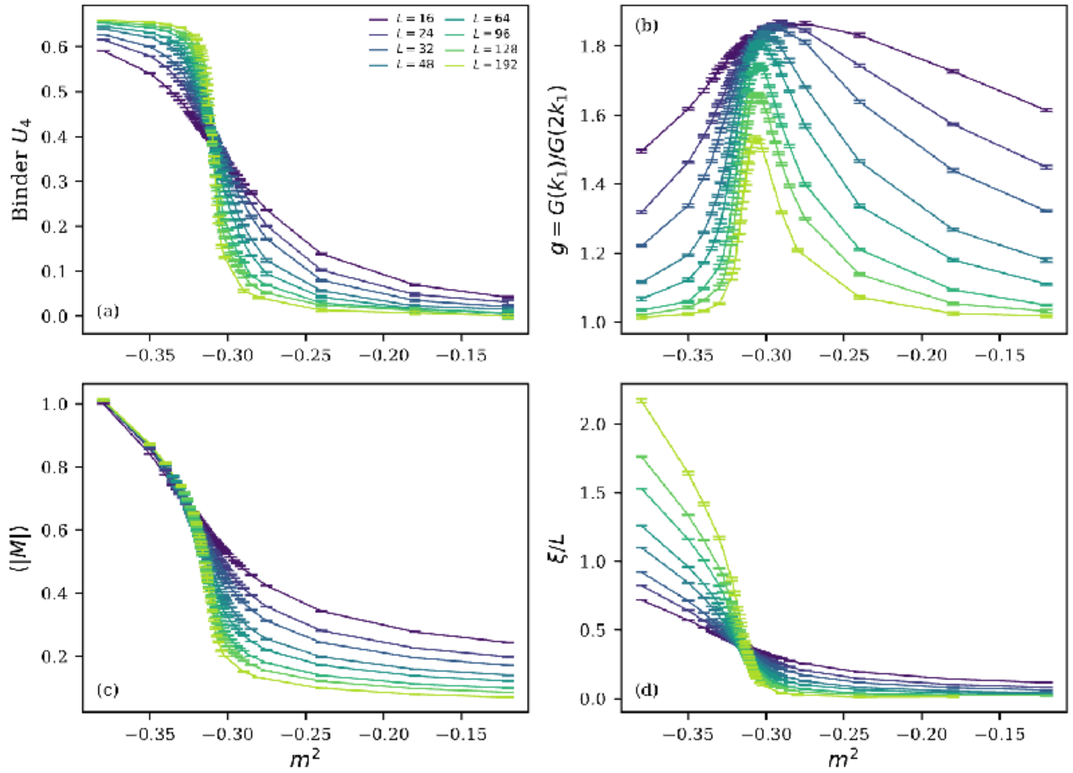
−2 0 +2

† finite reference not basis-converged

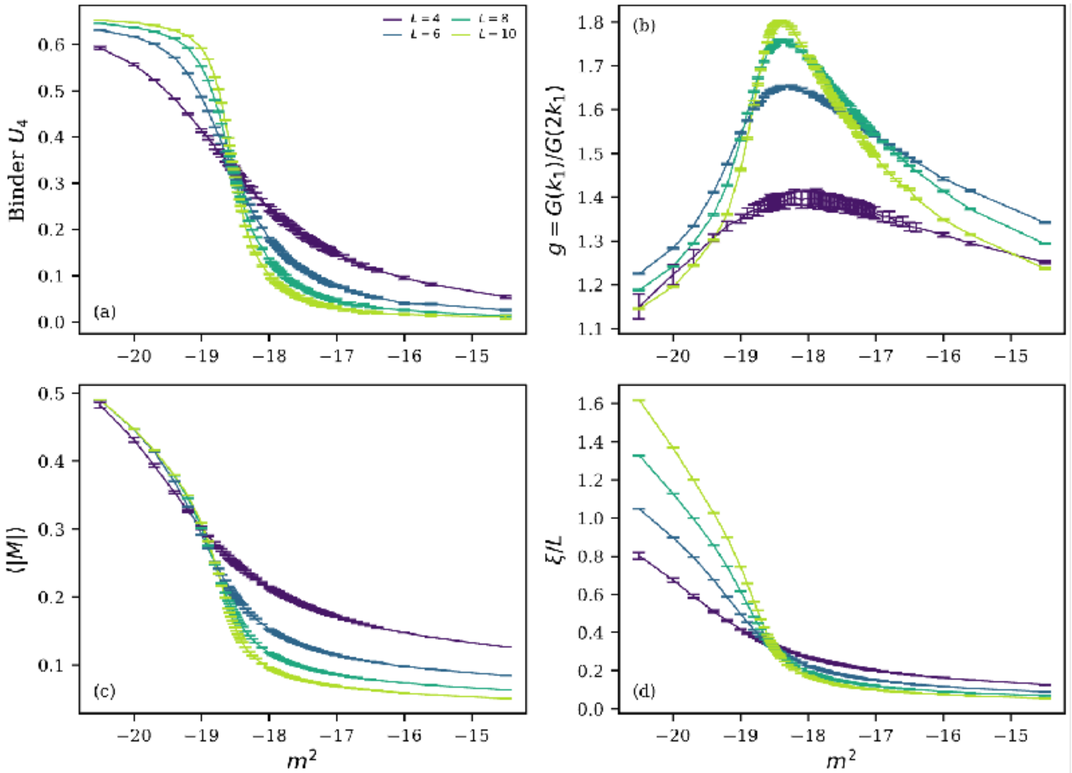


Critical regime validation

1+1d

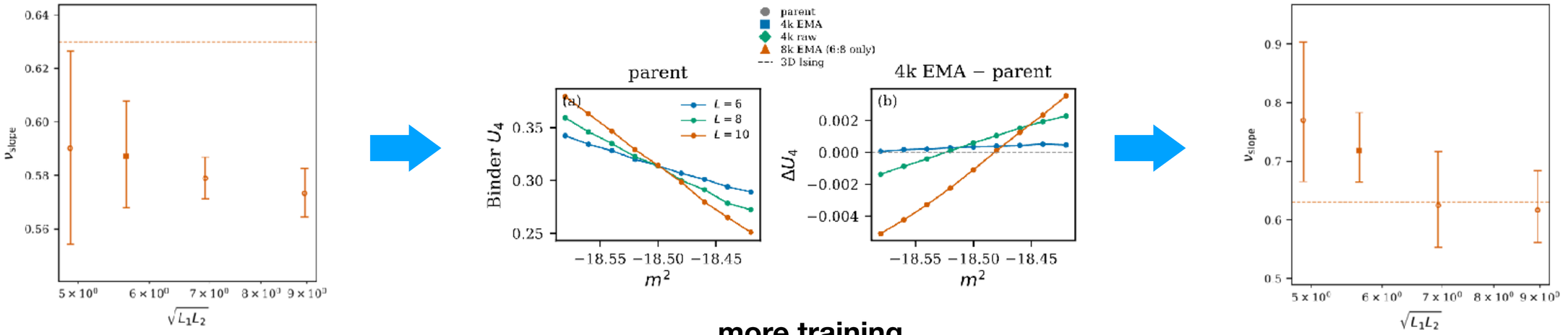


2+1d



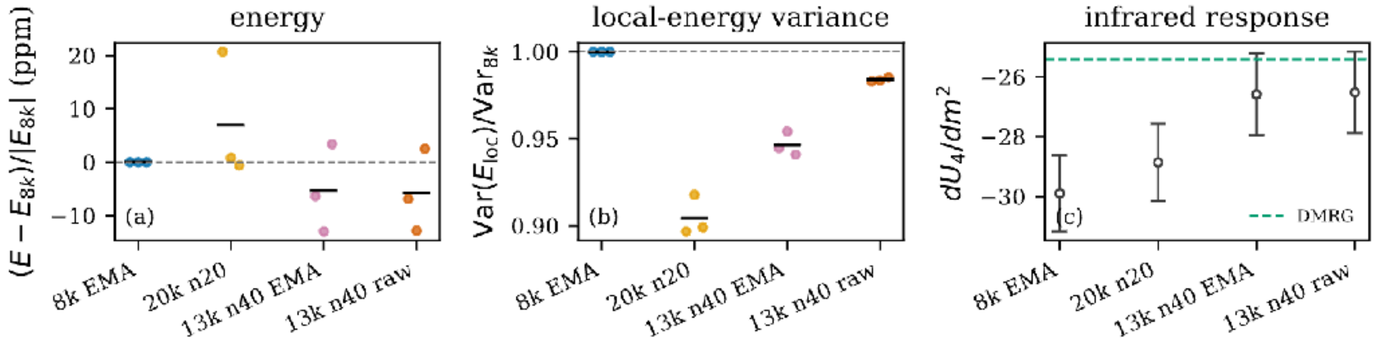
Critical regime validation

2+1d



more training

1+1d



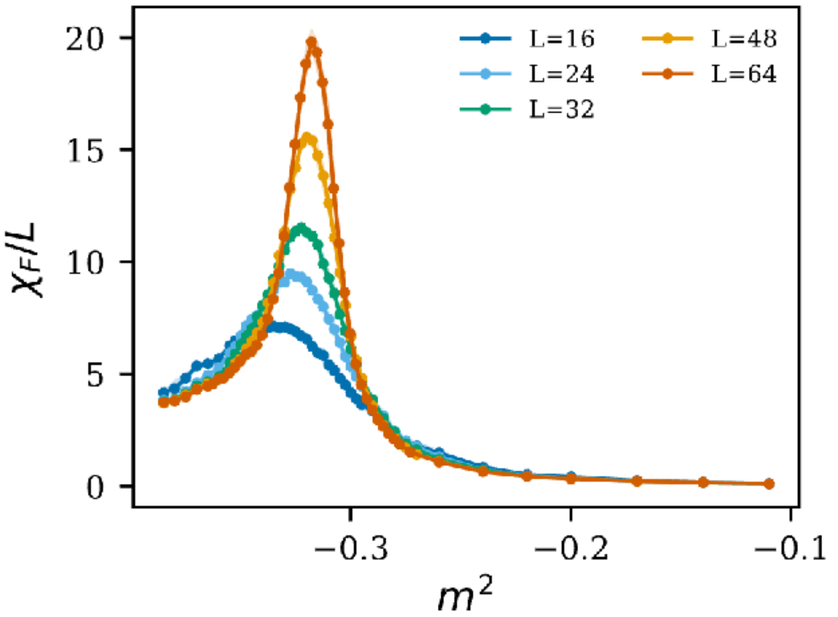
Quantum geometry

$$\chi_F = \frac{1}{4}\mathbb{E}_{\rho_\theta}\left[(\partial_{m^2}\log\rho_\theta)^2\right]$$

Fidelity susceptibility

$$F(m^2,\delta) = \left|\langle\psi(m^2-\delta/2)|\psi(m^2+\delta/2)\rangle\right|$$

$$\chi_F = \lim_{\delta\to 0}\frac{2[1-F(m^2,\delta)]}{\delta^2}$$



→ Locate phase transition without order parameter!

Neural Transport States

- A normalized continuous-field transport state gives reusable access to: samples, amplitudes, wavefunctional derivatives at large lattice volume

- Easy to fool oneself: strong validation highlights almost energy-blind critical optimization error that further refinement can ***still*** repair

Difficulties in Applying the Variational Principle to Quantum Field Theories[1]

Richard P. Feynman

California Institute of Technology
Pasadena, California 91125, U.S.A.

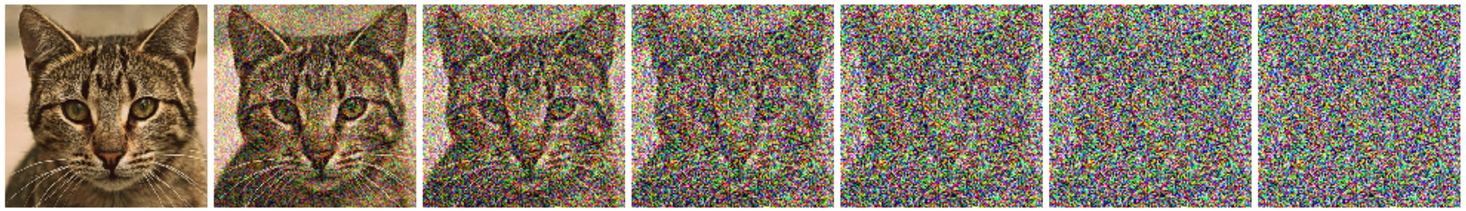
Introduction

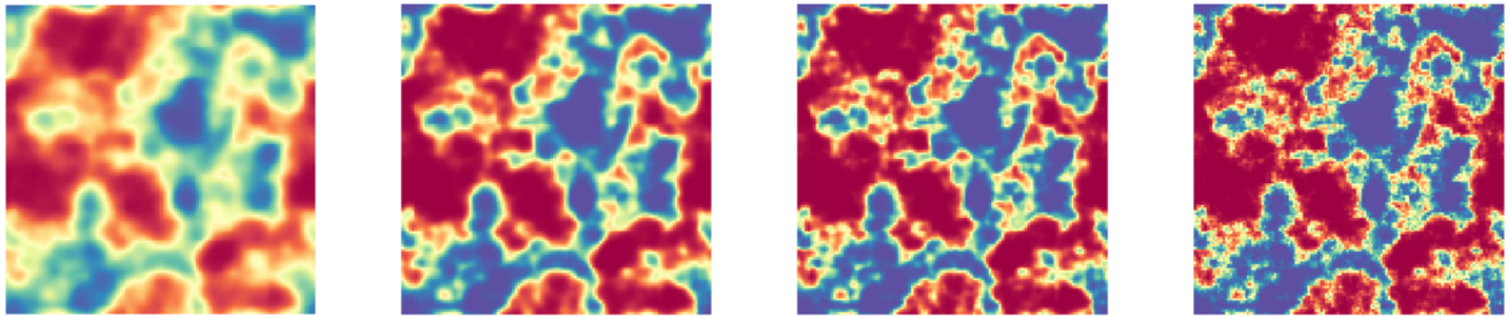
I'd like to talk on some work I did on the variational principle in field theory. At one time I thought that the brute force method of doing arithmetic on the machines will never get anywhere and we will probably end with something more old-fashioned, i.e. some analysis plus the machines to help us with the analytic equations, and so I tried to do something along these lines with quantum chromodynamics. So I'm talking on the subject of the application of the variational principle to field theoretic problems, but in particular to quantum chromodynamics.

I'm going to give away what I want to say, which is that I didn't get anywhere! I got very discouraged and I think I can see why the variational principle is not very useful. So I want to take, for the sake of argument, a very strong view – which is stronger than I really believe – and argue that it is no damn good at all!

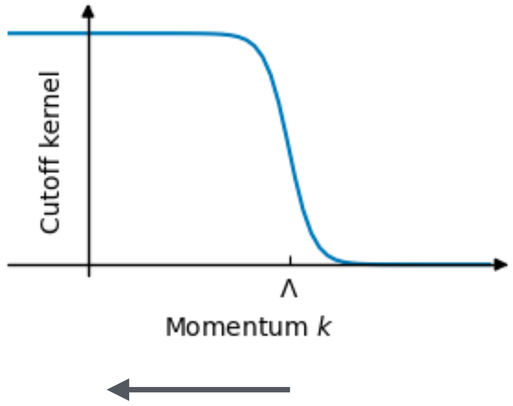
1988

RG — Diffusion models





Momentum space RG



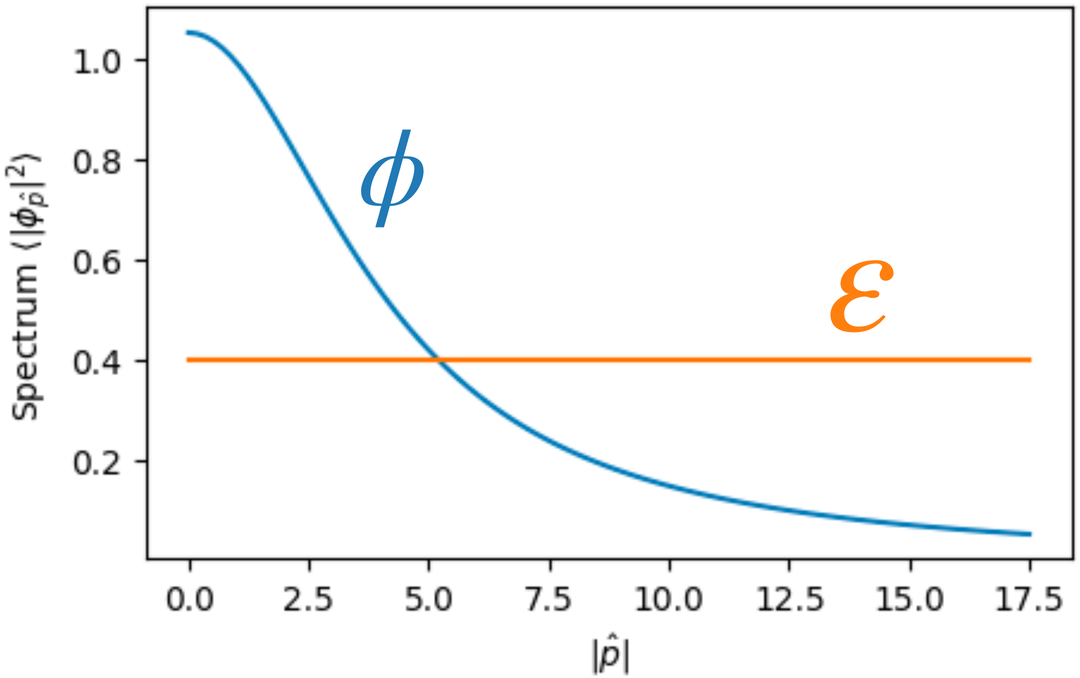
Set high-frequency modes to 0

GUD: [2410.02667]

“Cat” by Adlum Ramadani, Wikimedia Commons, CC BY-SA 3.0; noised

RG — Diffusion models

Forward OU: $\phi(t) = \alpha_t \phi(0) + \sigma_t \varepsilon$

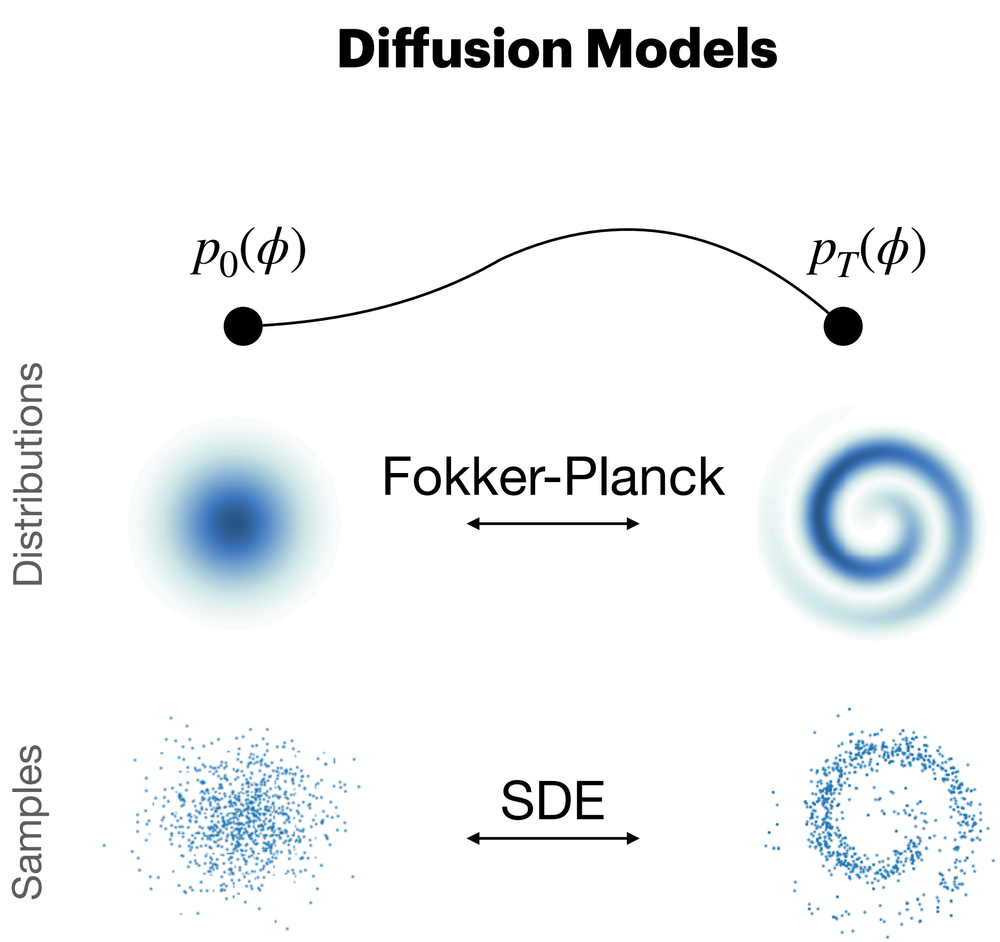




Diffusion models already implicitly destroy information by-scale!

- Multi-scale information erasure implicit, depending on data magnitude.
- No explicit control over this.

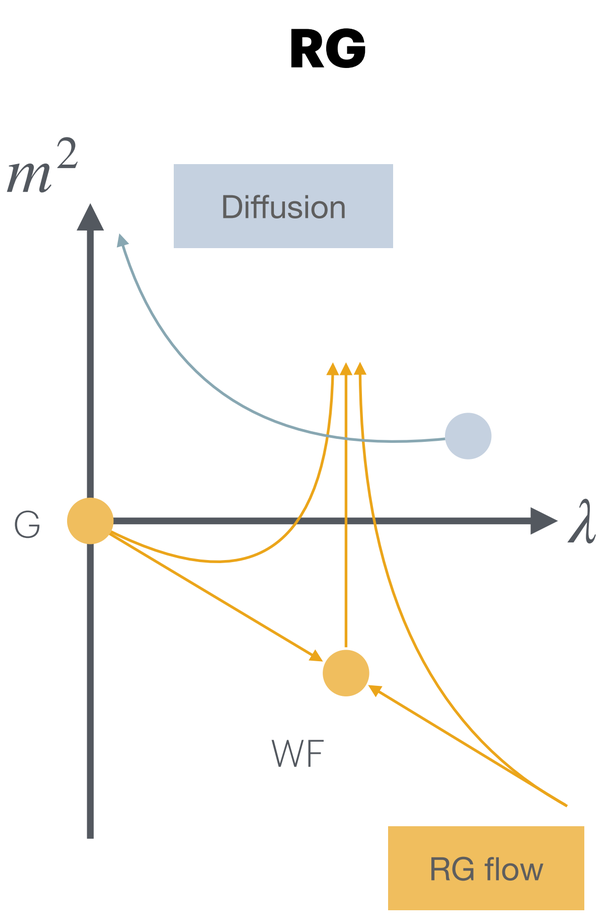
Flows in Distribution Space



Prior
RG fixed point

Basis
Fourier/Real space

Schedule
Cutoff kernel



Component-wise diffusion

Need simple, solvable process to get $p_t\left(\phi(t) \mid \phi(0)\right)$

General SDE: $d\phi = F(\phi, t)\, dt + G(\phi, t)\, dw$

Ornstein-Uhlenbeck process ✅

Simultaneously diagonalizable: $d\phi = UAU^\dagger \phi\, dt + UBU^\dagger\, dw$

Variance preserving: $d\phi = \frac{1}{2} U\boldsymbol{\beta}U^\dagger \phi\, dt + U\sqrt{\boldsymbol{\beta}}SU^\dagger\, dw$

Components $\chi = U\phi$

$\boldsymbol{U}$	$\boldsymbol{S}$	$\boldsymbol{\beta}(t)$
Component basis, "scheme"	Noise "color" & fixed point	Multi-scale erasure schedule

GUD: [2410.02667]

Optimal Path

- Large under-explored design degree of freedom
- Naive isotropic Gaussian “has” to be suboptimal (depending on dataset)
- Generalizes common, simpler “optimal time parametrization” goal

