[Cotler, Rezchikov 2202.11737]

[Cotler, Rezchikov 2308.12355]

Renormalizing Diffusion Models

SEMON REZCHIKOV

PRINCETON UNIVERSITY / IAS

[Wilson, Kogut '74] [Polchinski '84]

First tool: Field Theory

Euclidean scalar field theory on $\mathbb{R}^d$, $\phi : \mathbb{R}^d \to \mathbb{R}$

Probability functional $P_\Lambda[\phi(x)] \propto e^{-S_\Lambda[\phi]}$

RG scale $\Lambda \sim 1/\ell$

Exact renormalization group (ERG) flow equation:

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \mathcal{F}\left[P_\Lambda[\phi], \frac{\delta P_\Lambda[\phi]}{\delta \phi}, \frac{\delta^2 P_\Lambda[\phi]}{\delta \phi \delta \phi}, \ldots\right]$$

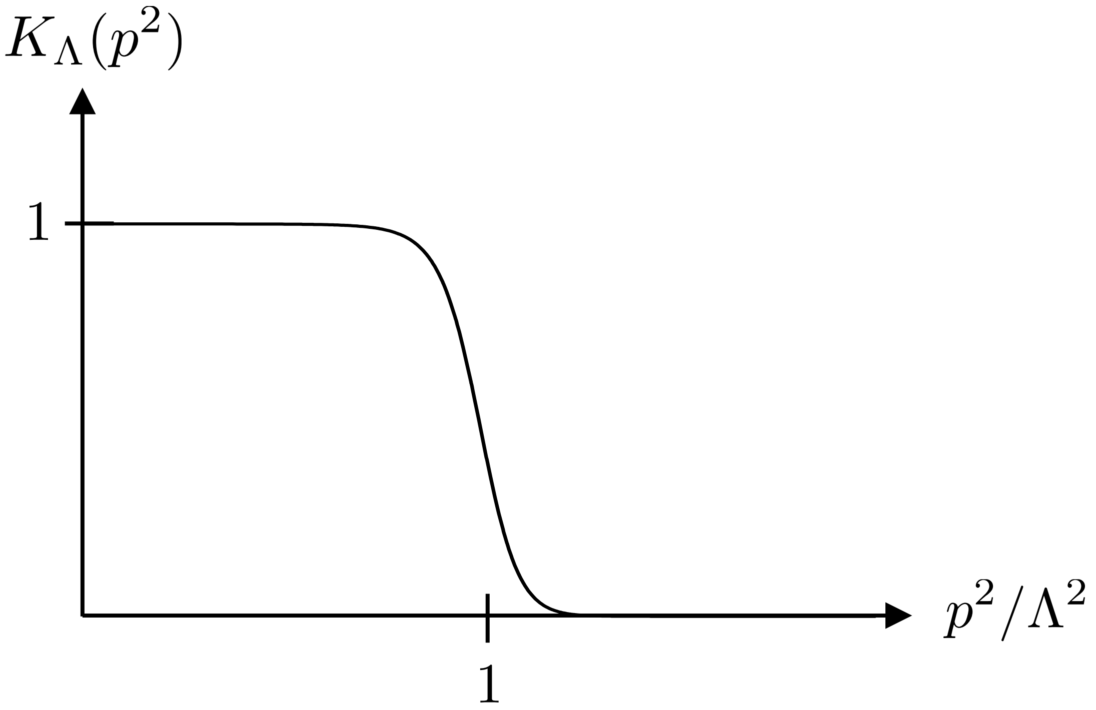
[Polchinski '84]

Exact Renormalization Group

Assume the source vanishes above the cutoff scale

Euclidean scalar field with a source:

$$Z_\Lambda[J] := \int [d\phi]\, e^{-\frac{1}{2}\int \frac{d^dp}{(2\pi)^d}\left(\phi(p)\phi(-p)(p^2+m^2)K_\Lambda^{-1}(p^2)+J(p)\phi(-p)\right)-S_{\mathrm{int},\Lambda}[\phi]}$$



Physics below the cutoff scale is preserved under RG flow:

★ $$-\Lambda\frac{d}{d\Lambda}Z_\Lambda[J] = C_\Lambda\, Z_\Lambda[J]$$

Renormalization Group Flow as an SDE

Wilson and Polchinski observed that in the continuum, certain natural renormalization group schemes can be viewed as an equation of Fokker-Planck type

Polchinski's equation was originally written as

$$\frac{\partial S_{\mathrm{int},t}[\phi]}{\partial t} = \frac{1}{2}\int d^d p\,(2\pi)^d (p^2+m^2)^{-1}\,\frac{\partial_t K_t(p^2)}{\partial t}\left\{\frac{\delta S_{\mathrm{int},t}}{\delta\phi(p)}\frac{\delta S_{\mathrm{int},t}}{\delta\phi(-p)} - \frac{\delta^2 S_{\mathrm{int},t}}{\delta\phi(p)\delta\phi(-p)}\right\}$$

Renormalization Group Flow as an SDE

This equation is not completely obviously in Fokker-Planck form, and it is an equation for $S_{\mathrm{int},t}[\phi]$ rather than an equation for

$$P_t[\phi] = \frac{1}{Z_t}\, e^{-S_t[\phi]}$$

But the resulting equation for the latter also turns out to be of convection-diffusion type

As such, it corresponds to stochastic dynamics for the fields

Polchinski equation as a Fokker-Planck equation

Polchinski equation

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \int d^d p \, \Lambda \frac{\partial C_\Lambda(p^2)}{\partial \Lambda} \frac{\delta^2}{\delta\phi(p)\delta\phi(-p)} P_\Lambda[\phi] + \int d^d p \, \Lambda \frac{\partial C_\Lambda(p^2)}{\partial \Lambda} \frac{\delta}{\delta\phi(p)} \left(C_\Lambda^{-1}(p^2)\, \phi(p)\, P_\Lambda[\phi] \right)$$

Fokker-Planck equation

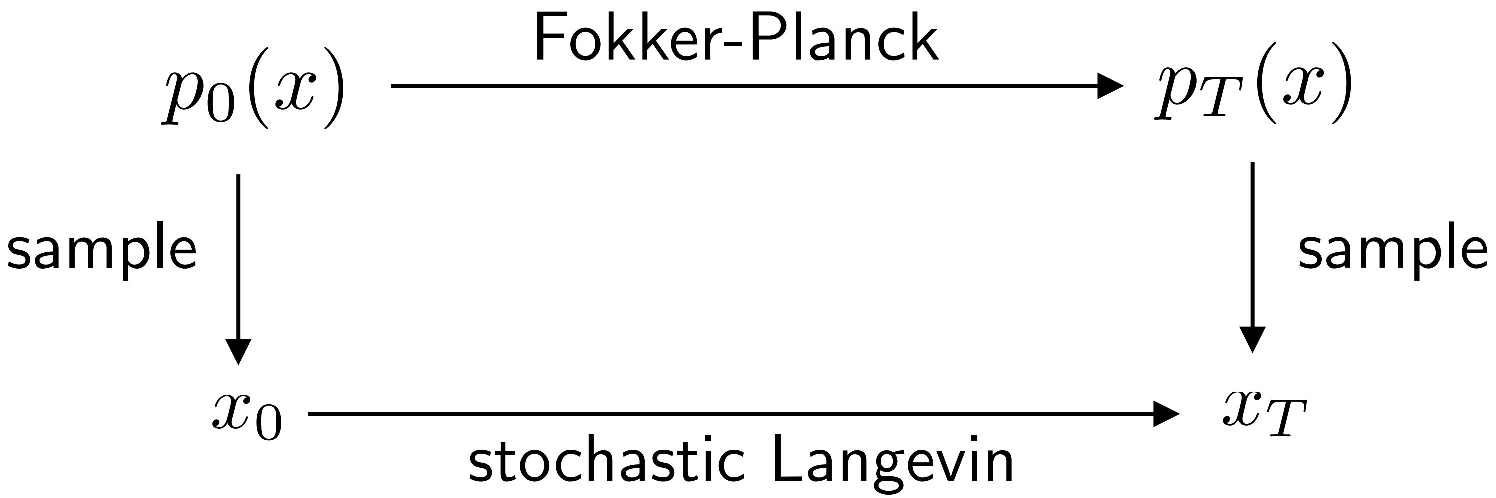
$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i (\partial^i V(x)\, p_t(x))$$

Second tool: Fokker-Planck versus Langevin

Fokker-Planck: $\frac{\partial}{\partial t} p_t(x) = \partial_i(\partial^i V(x)\, p_t(x)) + \partial_i \partial^i p_t(x)$ **PDE**

Stochastic Langevin: $dx_t = -\partial_i V(x_t)\, dt + \sqrt{2}\, dW_t$ **Stochastic ODE**

Question: Given some $p_0(x)$, how do we sample from $p_T(x)$?



[Cotler, Rezchikov '23]

Polchinski SDE

$$\partial_t \widetilde{\phi}_t(\mathbf{p}) = \partial_t \log(K_t(|\widehat{\mathbf{p}}|))\, \widetilde{\phi}_t(\mathbf{p}) + \sqrt{-\frac{1}{L^d}\, \frac{1}{|\widehat{\mathbf{p}}|^2 + m^2}\, \partial_t K_t(|\widehat{\mathbf{p}}|)\; \widetilde{\eta}_t(\mathbf{p})}$$

$$\mathbb{E}[\widetilde{\eta}_t(\mathbf{p})] = 0\,, \quad \mathbb{E}[\widetilde{\eta}_t(\mathbf{p})\widetilde{\eta}_s(\mathbf{k})] = \delta(t-s)\,\delta_{\mathbf{p},-\mathbf{k}}$$

$$K_t(|\widehat{\mathbf{p}}|) = e^{-|\widehat{\mathbf{p}}|^2/(e^{-t}\Lambda_0)^2}$$

Exact Renormalization Group (ERG) Schemes

One can look for all possible flows of $P_t[\phi]$ which are:

- Of convection-diffusion type
- Erase the information of frequencies of ϕ above some $f(t)$
- Preserve correlators of functions of the fields which only depend on frequencies below $f(t)$

Such schemes are studied and essentially classified using the Wegner-Morris-Wetterich equation. Any one of them is valid, but they can have *very different* numerical properties. They also have an entropic gradient flow formulation. **[Cotler, Rezchikov ’22]**

Wegner-Morris equation as a Fokker-Planck equation

Wegner-Morris equation [Wegner '74] [Morris '95]

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = \frac{1}{2} \int d^d p \, B_\Lambda(|p|) \left(\frac{\delta^2 P_\Lambda[\phi]}{\delta\phi(p)\delta\phi(-p)} + 2 \frac{\delta}{\delta\phi(p)} \left(\frac{\delta \widehat{S}_\Lambda[\phi]}{\delta\phi(-p)} P_\Lambda[\phi] \right) \right)$$

Fokker-Planck equation

$$\frac{\partial}{\partial t} p_t(x) = \partial_i \partial^i p_t(x) + \partial_i (\partial^i V(x) \, p_t(x))$$

$\Longrightarrow$ $t = -\log(\Lambda/\Lambda_0)$

[Cotler, Rezchikov '22]

Functional generalization of optimal transport

Initial and final distributions: $P_1[\phi], P_2[\phi]$

Transport kernel: $\Pi[\phi_1, \phi_2]$

$$\int [d\phi_2]\, \Pi[\phi_1, \phi_2] = P_1[\phi_1]\,, \quad \int [d\phi_1]\, \Pi[\phi_1, \phi_2] = P_2[\phi_2]$$

Cost: $\mathcal{C} : \mathscr{F} \times \mathscr{F} \to \mathbb{R}$

Minimize

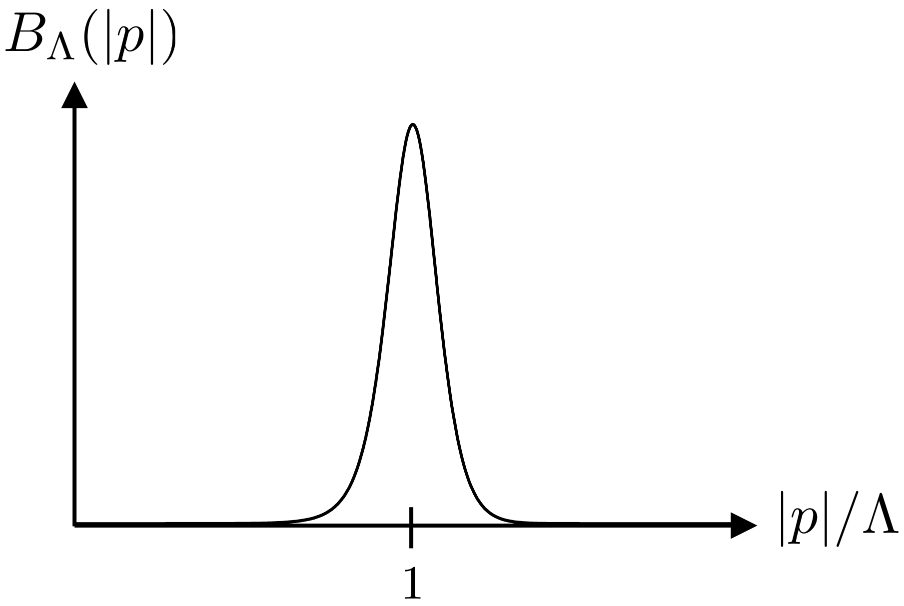
$$\mathsf{K}[\Pi] = \int [d\phi_1]\, [d\phi_2]\, \Pi[\phi_1, \phi_2]\, \mathcal{C}[\phi_1, \phi_2]$$

Wasserstein-2 distance

$$\mathcal{W}_2(P_1, P_2) := \left(\inf_{\Pi \in \Gamma(P_1, P_2)} \int [d\phi_1]\, [d\phi_2]\, \Pi[\phi_1, \phi_2] \int d^d x\, d^d y\, B_\Lambda^{-1}(x-y)\, (\phi_1(x) - \phi_2(x))\, (\phi_1(y) - \phi_2(y)) \right)^{1/2}$$

[Cotler, Rezchikov '22]

Functional generalization of optimal transport



Wasserstein-2 distance

ERG kernel

$$\mathcal{W}_2(P_1, P_2) := \left(\inf_{\Pi \in \Gamma(P_1, P_2)} \int [d\phi_1]\, [d\phi_2]\, \Pi[\phi_1, \phi_2] \int d^d x\, d^d y\, B_\Lambda^{-1}(x-y)\, (\phi_1(x) - \phi_2(x))\, (\phi_1(y) - \phi_2(y)) \right)^{1/2}$$

[Cotler, Rezchikov '22]

RG flow equation

Distribution of interest: $P_\Lambda[\phi] = \frac{1}{Z_{P,\Lambda}} e^{-S_\Lambda[\phi]}$

"Background" distribution: $Q_\Lambda[\phi] = \frac{1}{Z_{Q,\Lambda}} e^{-2\hat{S}_\Lambda[\phi]}$

Relative entropy: $S(P\|Q) = \int [d\phi]\, P[\phi] \log(P[\phi]/Q[\phi])$

$$-\Lambda \frac{d}{d\Lambda} P_\Lambda[\phi] = -\nabla_{\mathcal{W}_2} S(P_\Lambda[\phi] \,\|\, Q_\Lambda[\phi])$$

Wegner-Morris equation

RG flow is a gradient flow with respect to the relative entropy!

Carosso RG

Possibly the simplest SDE which turns out of this type may be

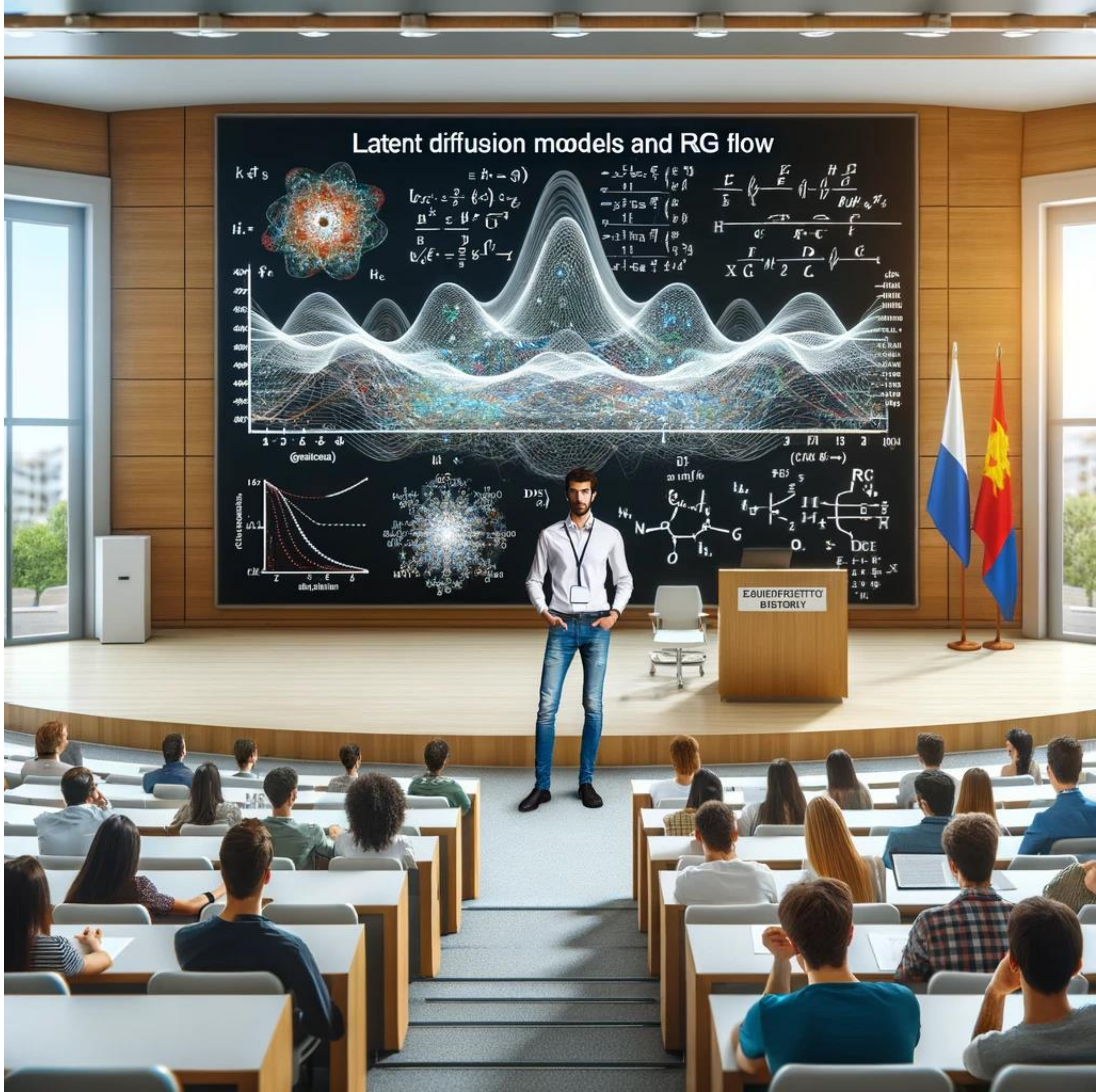
$$\partial_t \widetilde{\phi}_t(\mathbf{p}) = -|\widehat{\mathbf{p}}|^2\, \widetilde{\phi}_t(\mathbf{p}) + \widetilde{\eta}_t(\mathbf{p})$$

This SDE was studied by Carosso **['20]** in order to make the connection between the **gradient flow**, used in the lattice field theory community for scale-setting and inspired by Lüscher's ideas on the **Wilson flow** **['09]**, and the **renormalization group** precise

It turns out to be of Wegner-Morris-Wetterich type.

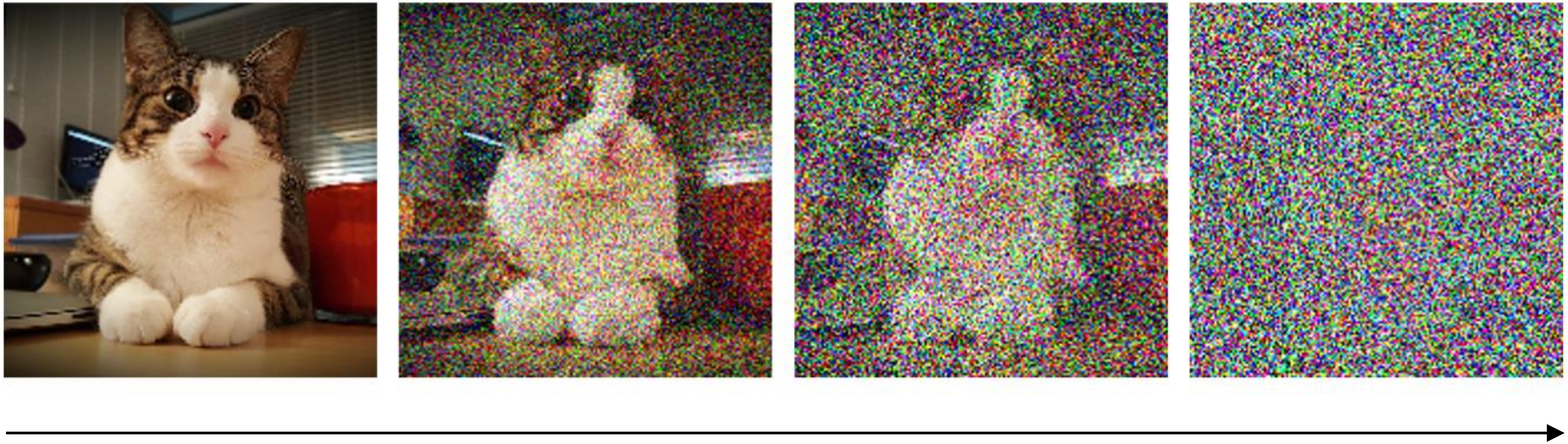
Q: Can you make an image of a scientist giving a talk about latent diffusion models and RG flow?

GPT-4 + DALL-E 2:



Score-based diffusion models

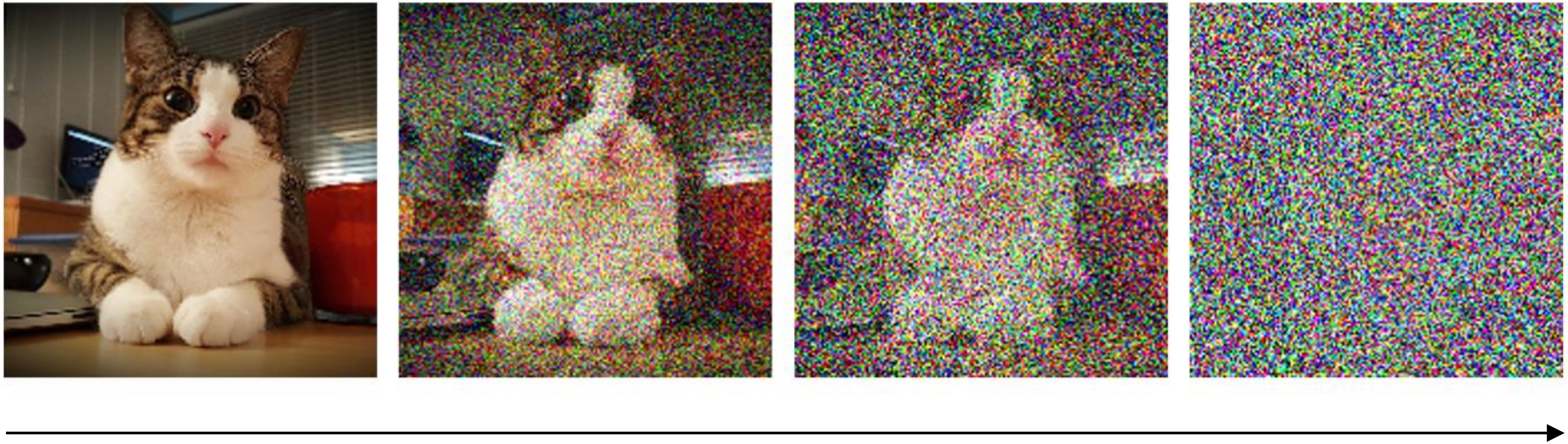
[Song et al. ‘19]
[Ho, Jain, Abbeel ‘20]
[Song et al. ‘21]



Source: NVIDIA

Score-based diffusion models (for fields!)

[Song et al. ‘19]
[Ho, Jain, Abbeel ‘20]
[Song et al. ‘21]



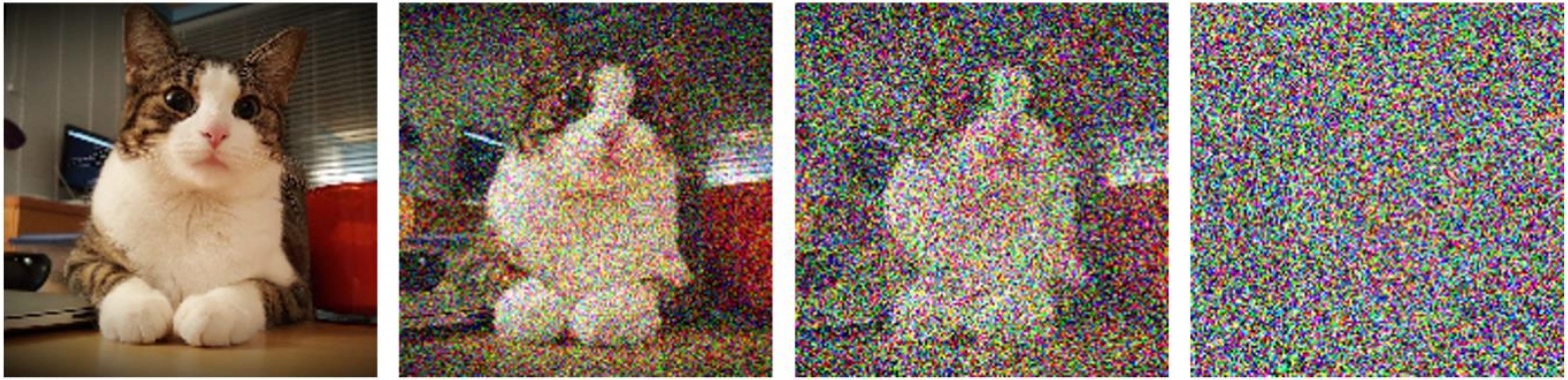
flow of true score $s_t^*(x)$

(using an ERG scheme!)

Source: NVIDIA

Score-based diffusion models (for fields!)

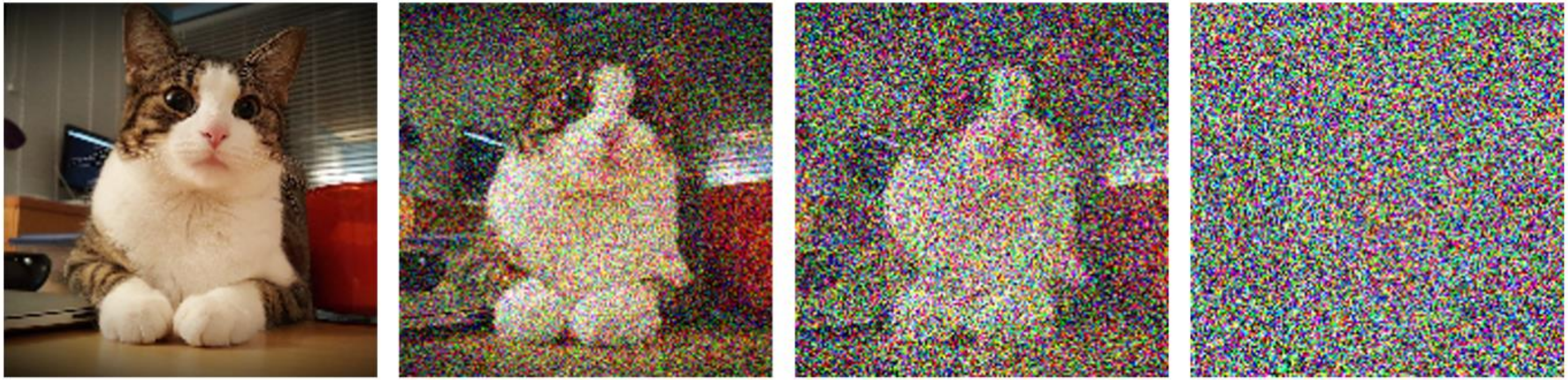
[Song et al. '19]
[Ho, Jain, Abbeel '20]
[Song et al. '21]



Source: NVIDIA

Score-based diffusion models (for fields!)

[Song et al. ‘19]
[Ho, Jain, Abbeel ‘20]
[Song et al. ‘21]



reverse flow of learned score $s_t^\theta(x)$

Score is gradient of action; effective field theory!

Source: NVIDIA

Diffusion Models

Score function: $s_t(x) = \nabla \log p_t(x)$

Forwards Equation

$$dx = f(x,t)\,dt + g(t)\,dB_t$$

Backwards equation [Anderson '82]

$$dx = \big(-f(x,\tilde{t}) + g(\tilde{t})^2\,\nabla \log p_{\tilde{t}}(x)\big)\,d\tilde{t} + g(\tilde{t})\,dB_{\tilde{t}}$$

Limiting distribution

$$\lim_{t\to\infty} p_t(x) = p_\infty(x)$$

Backwards equation: ODE

$$\frac{dx}{d\tilde{t}} = -f(x,\tilde{t}) + \frac{1}{2}\,g(\tilde{t})^2\,\nabla \log p_{\tilde{t}}(x)$$

Diffusion Models

Parameterize the score and optimize one of these functionals:

1. KL Divergence: $\mathrm{KL}(p_t|q_t) = \int dx\, p_t(x)\, \log\left(\frac{p_t(x)}{q_t(x)}\right)$

2. Fischer Divergence: $D_F(p_t|q_t) = \int dx\, p_t(x)\, \left|\nabla \log p_t(x) - \nabla \log q_t(x)\right|^2$

Can (and may) integrate either over t

3. Alternate form: $\mathbb{E}_{t\sim[0,T]}\left[\lambda(t)\, \mathbb{E}_{x\sim p_0(x)} \mathbb{E}_{x'\sim p(x',t;x,0)}[\nabla_{x'} \log p(x', t; x, 0) - s_\theta(x', t)]\right]$

Variational Approximation

Importantly these different objectives are connected:

An integral of Fischer divergences is the ELBO for the forwards KL.

$$\mathrm{KL}(p_{t=0} \,|\, p_{\theta,t=0}) \leq \mathbb{E}_{t\sim[0,T]}\left[\lambda(t)\,\mathbb{E}_{x\sim p_t^{\mathrm{data}}(x)}\left[|\nabla \log p_t(x) - s_\theta(x,t)|^2\right]\right] + \mathrm{KL}(p_{t=T}|p_{\theta,t=T})$$

Note also that these are **different** from those objectives usually used in normalizing flows, which are based on the **reverse KL**. These require **samples**, which we produce (in the context of field theory) by **running an (adaptive) MCMC chain during training**.

Multiscale Diffusion Models

It is natural to generate images by first generating a coarse-grained approximation and then filling in the fine-grained details. This strategy is often used by ML practitioners.

1. Same SDE as Carosso investigated as a noising process in **[Hoogeboom-Salimans]**
2. Terms like $-a(t)\phi$ correspond to field renormalizations; carefully tuned!
3. In practice people use conditional diffusion models. Or more complex multiscale schemes **[2209.14125, 2205.01490, 2208.05003, 2106.15282, 2104.07636]**. `Latent Diffusion’ **[2112.10752]** does diffusion in a learned latent space which is implicitly multiscale.

``The same equations have the same solutions.’’ –Richard Feynman.

Two Potential Directions

Thus we see that certain *multiscale diffusion processes* can be interpreted as *physical RG schemes* when applied to distributions coming from statistical field theory.

- One can use this connection to suggest new methods in ML
- One can also use this connection to try to build interpretable ML models for studying field theories.

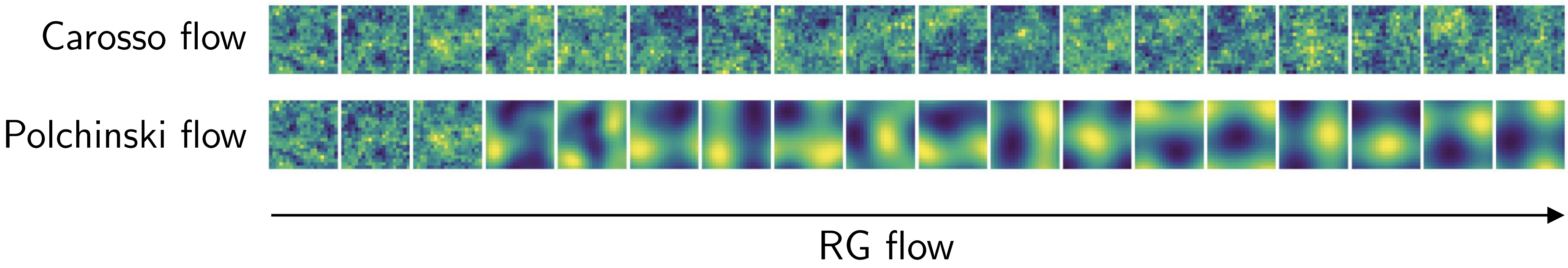
Two natural applications to field theory are to *sampling and effective field theory*, and to *variational ansatzes for ground states of QFTs.*

[Cotler, Rezchikov ‘23]

Numerics

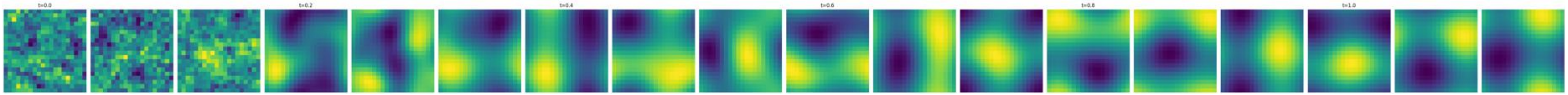
In the paper we developed algorithms based on insights from RG to build tailored flow-based and diffusion model algorithms for field theories

Here let us show some results from our learned models in the context of Euclidean scalar ϕ^4 theory in 2 dimensions (20 x 20 grid)

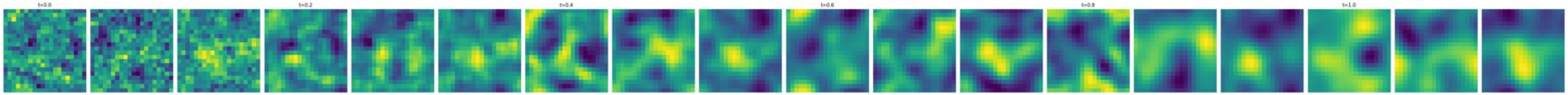


Numerics: Polchinski/Carosso on ϕ^4

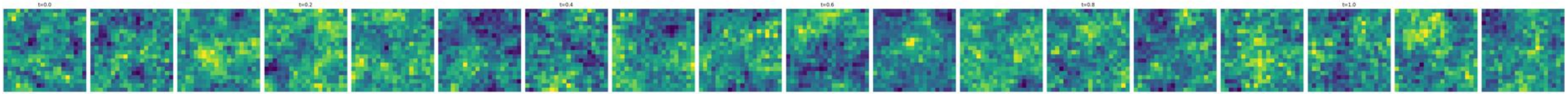
Samples from Polchinski RG



Zoom in to initial segment:

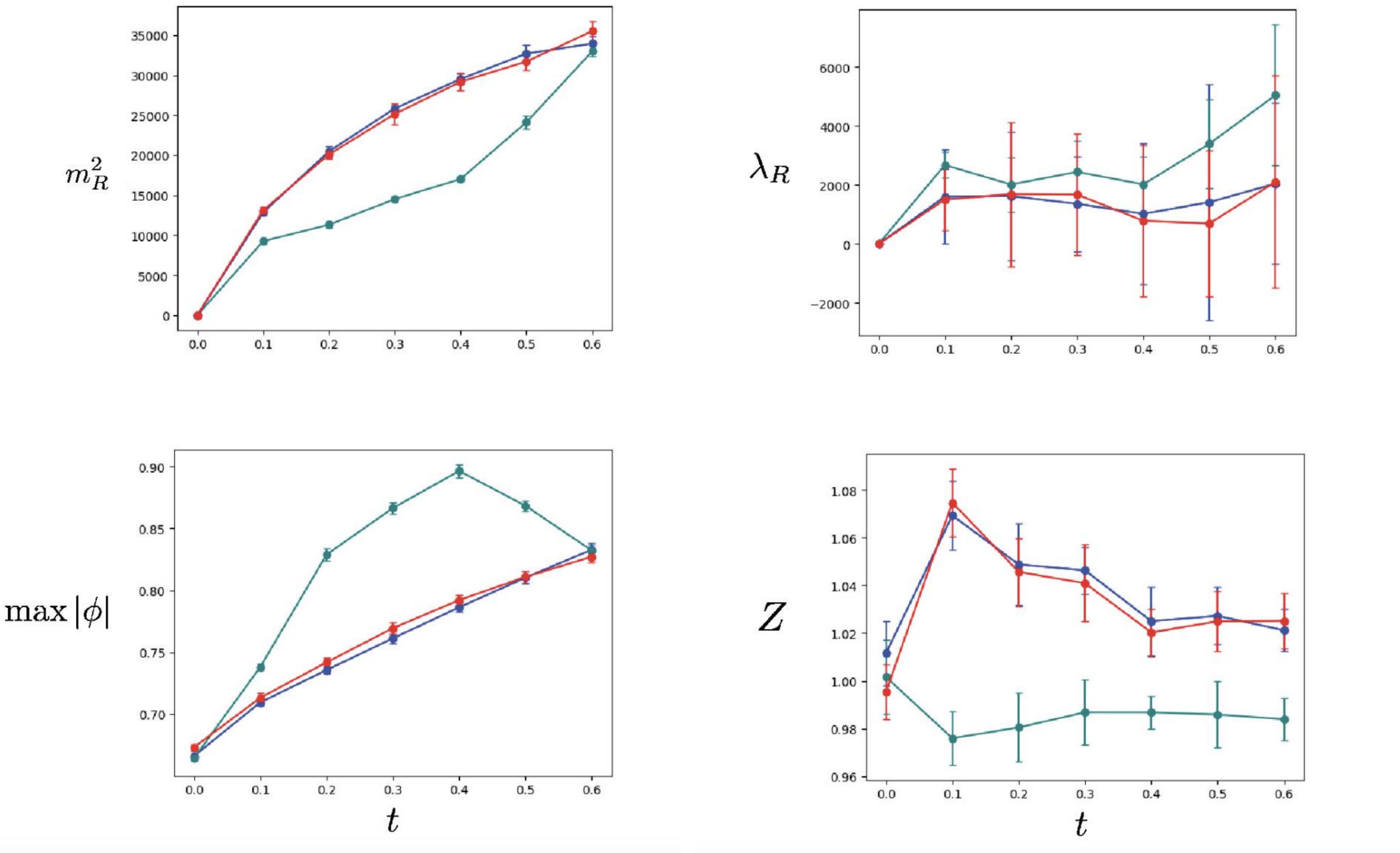


Samples from Carosso RG:



[Cotler, Rezchikov '23]

Numerics: Learning the RG flow of ϕ^4.



Exact RG flow

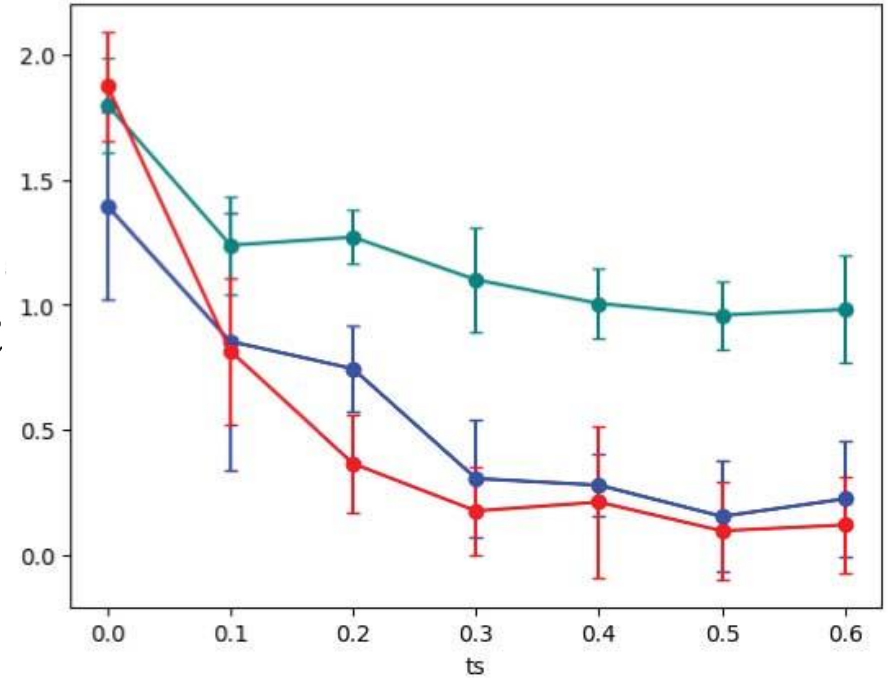
Learned RG flow

Flow based model learned using [Gerdes et al. '22]

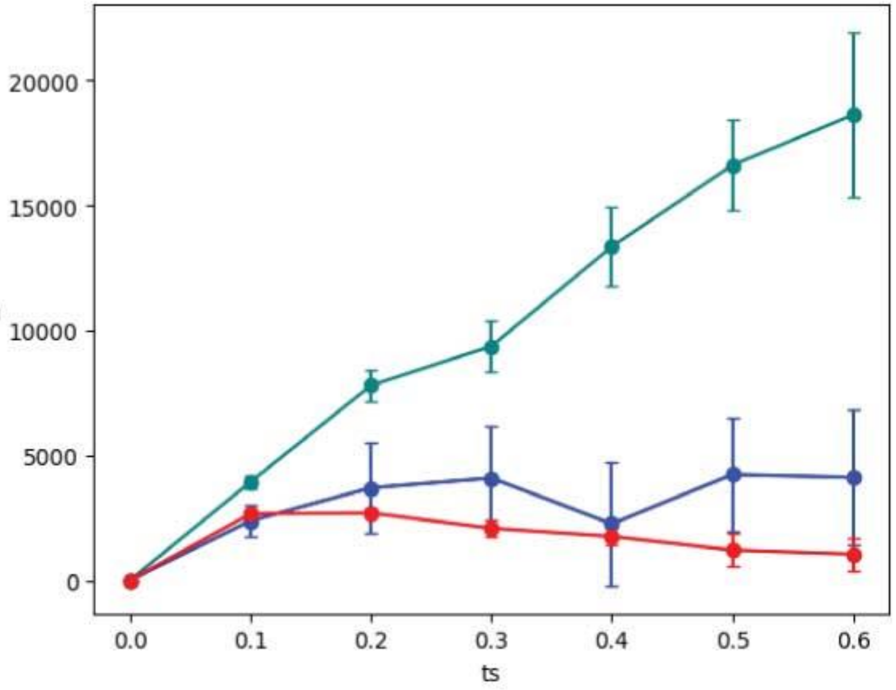
Numerics: Learning the RG flow of ϕ^4.

Carosso RG

$r_t^2 \lambda_R$



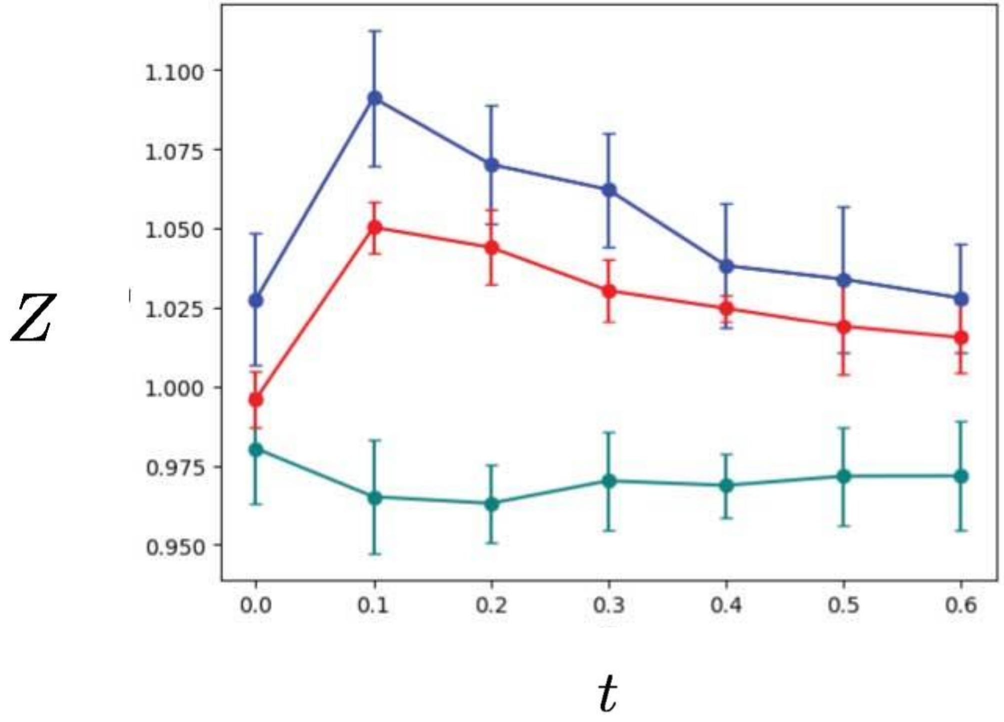
λ_R

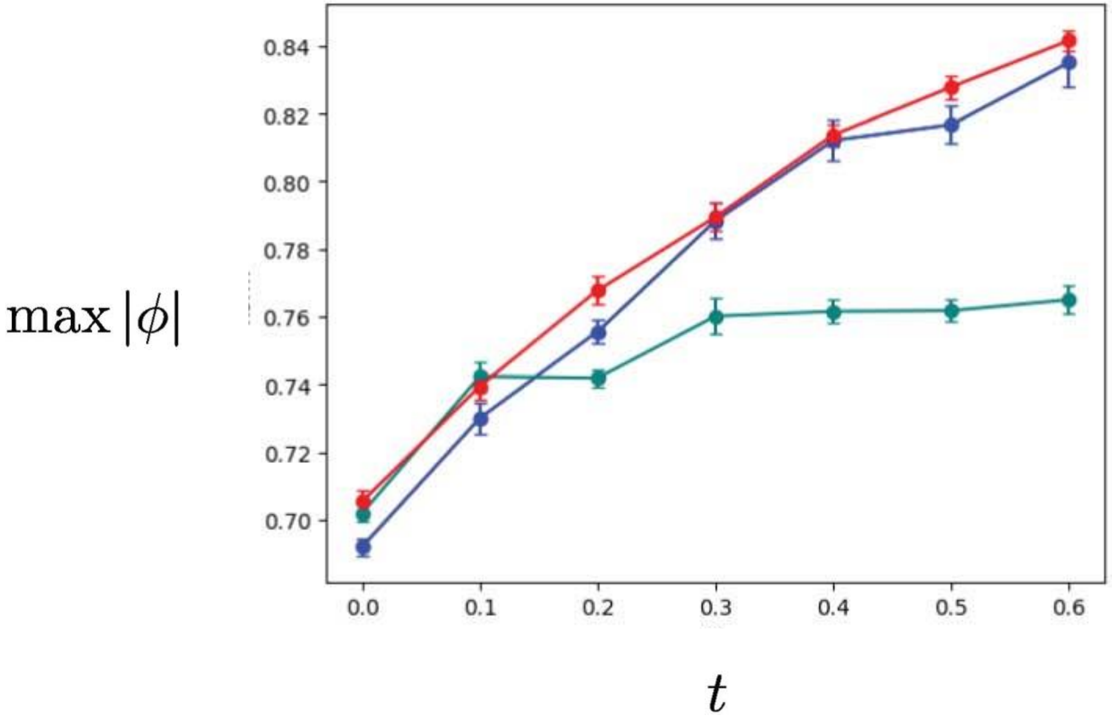


$$\lambda_R \approx -(m_R L)^2 \, \frac{\widetilde{G}_{4,\mathrm{conn}}(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0})}{\widetilde{G}_2(\mathbf{0},\mathbf{0})^2}$$

Numerics: Learning the RG flow of ϕ^4.

Carosso RG





Connections to other stories

Mathematical physics (**[Bauerschmidt-Bodineau], [Bauerschmidt '23]**)

Mixing of Markov chains in high-temperature ϕ^4, sine-Gordon

Optimal transport (**above, also [Cotler, Rezchikov '22]**)

Bakry-Emery, log-Sobolev inequalities, connection to correlation decay

Bayesian inference **[Berman et al.]**

Stochastic localization (**[Montanari '23], [Eldan-Koehler-Zeitouni '21], many others**)

Many powerful results about spin glasses, convex geometry, mixing of Markov chains

Wilson/gradient flow (**[Lüscher], [Carosso], many others**)

Wilson flow is an interesting smoothing process (Yang-Mills gradient flow)

Motivation for work on normalizing flows

Summary

Many rich connections between field theory and latent diffusion models

Can leverage these to invent physically motivated sampling algorithms for field theories

Multiple possible problems, e.g. sampling along RG flows, discovering phase transitions, sampling ground states of quantum field theories

Insights from RG flows may also help to improve latent diffusion models for image generation

Much more to understand and explore! **THANK YOU!**

Further Directions

- Models that *infer the effective action* **[Cotler, Rezchikov WIP]**

$$s_t^{\theta}(\phi) = \left(\sum_{i=1}^{r} f_i(\theta^i, t)\, g_i(\phi) \right) + \tilde{s}_t^{\tilde{\theta}}(\phi)$$

- Searching for *phase transitions? Goldstone modes?* using neural networks

- A *lot of design*

Designing Renormalizing Diffusion Models

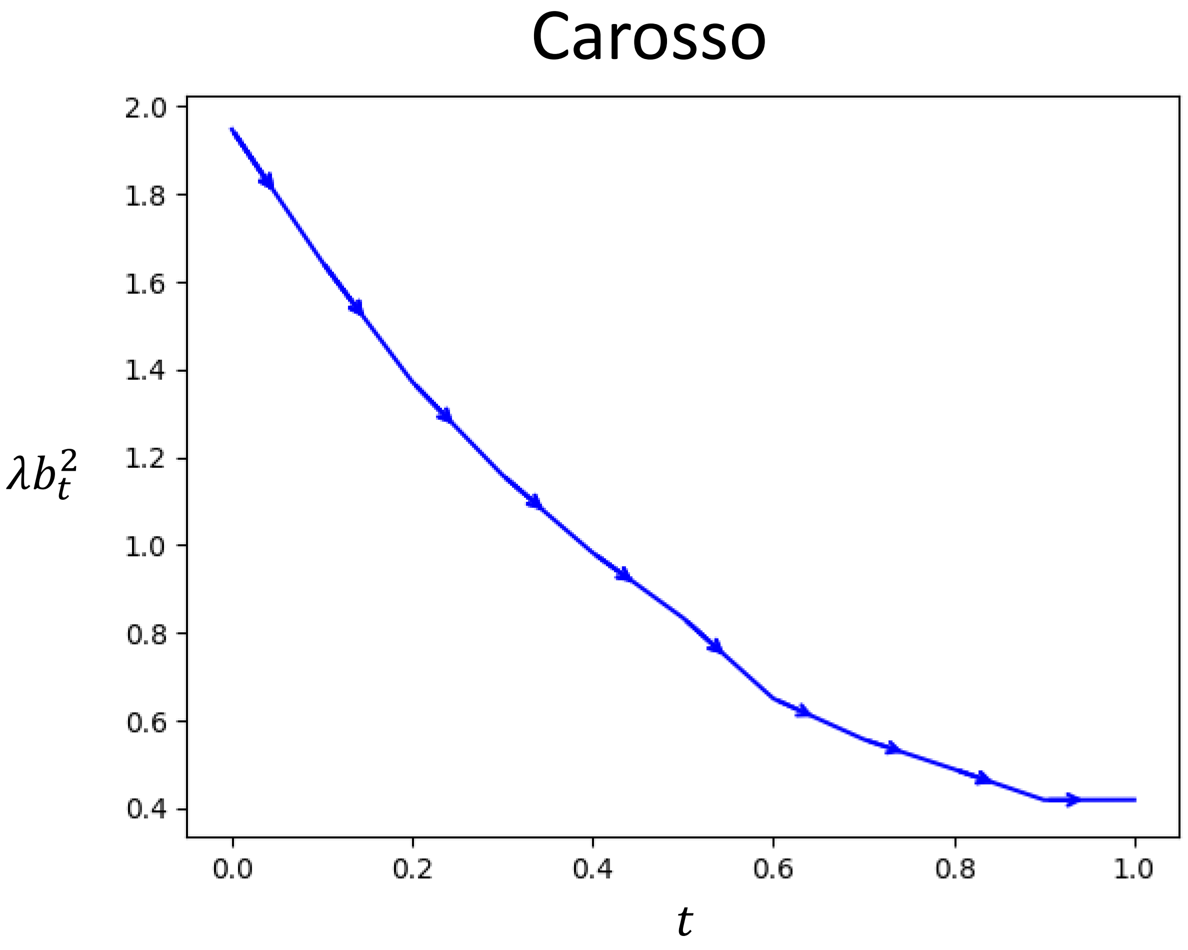
- One can incorporate field and *space* rescalings into the flow or estimators:

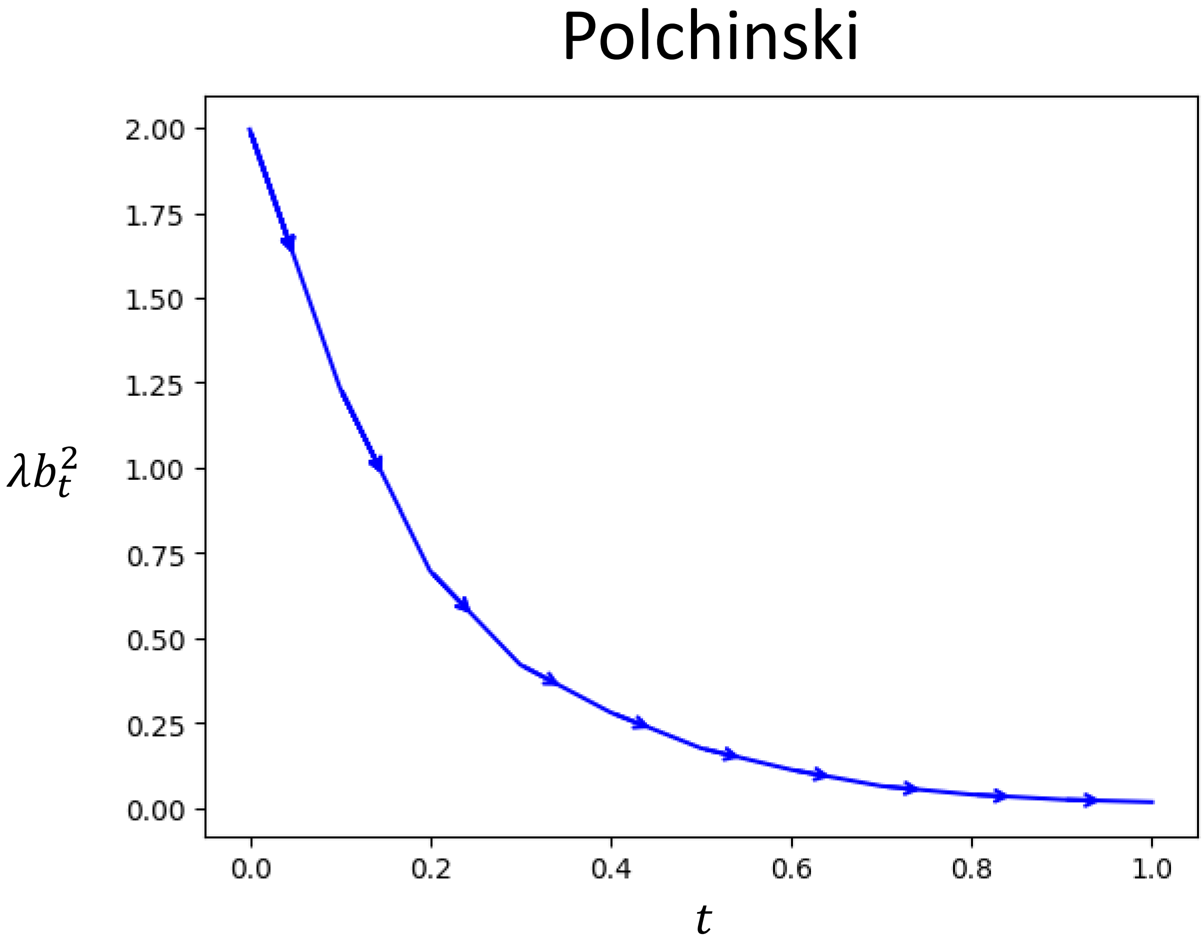
$$\mathbf{p} \to b_t \mathbf{p}\,, \quad \widetilde{\phi}(\mathbf{p}) \to b_t^{-\frac{d+2}{2}}\, \widetilde{\phi}(\mathbf{p})$$

- A *multiscale adaptive sampling algorithm* ***on a lattice of fixed size!***

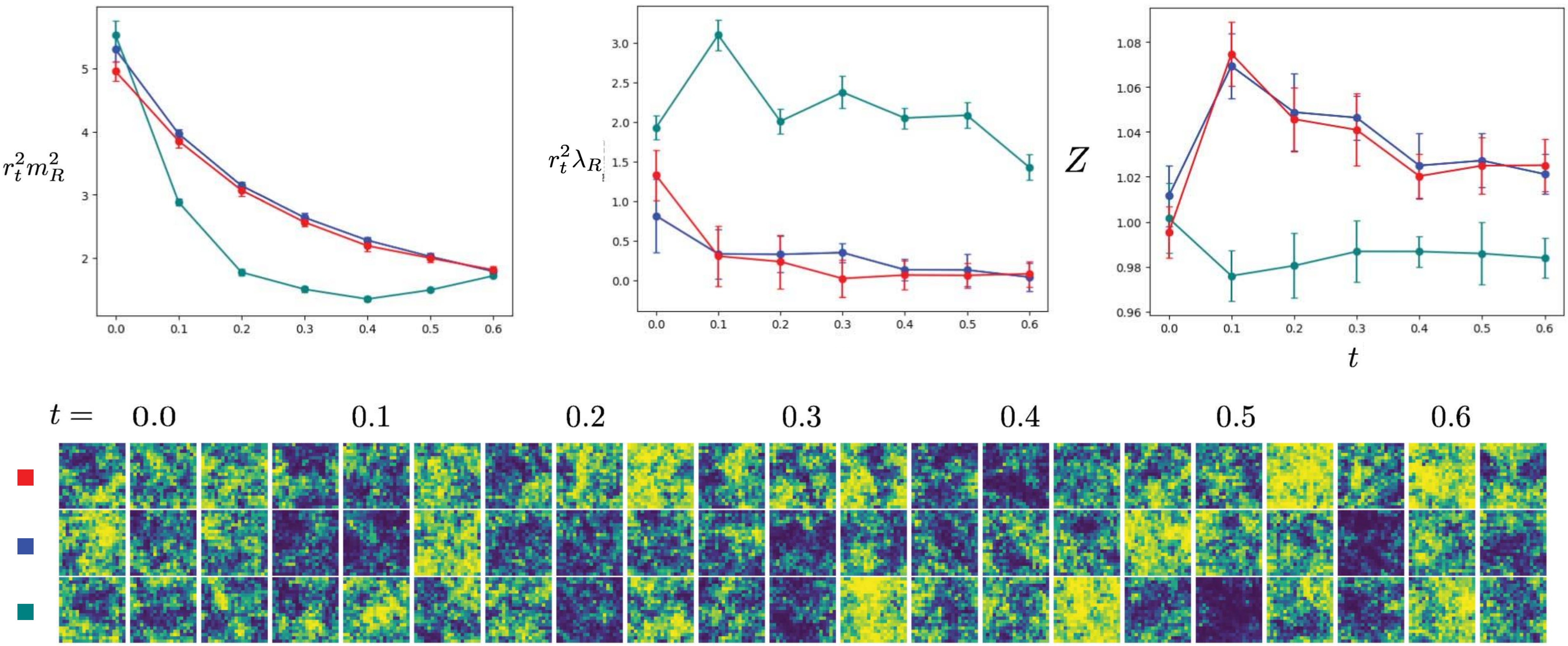
  (compare with Ryan Abbott's talk Lattice 2023 on `block RG')
- This can radically change the numerical behavior *as in ML*:
  - Polchinski flow tends to a delta-function (variance *collapsing)*
  - Carosso flow tends to a free field theory *(*variance *preserving)*
  - `Renormalized' Polchinski tends to white noise (variance *preserving* but in a different sense!)
- We found (as in ML) that tuning the flow is crucial for good numerical behavior
- The score function is the gradient of the action…

Numerics: Polchinski/Carosso on ϕ^4



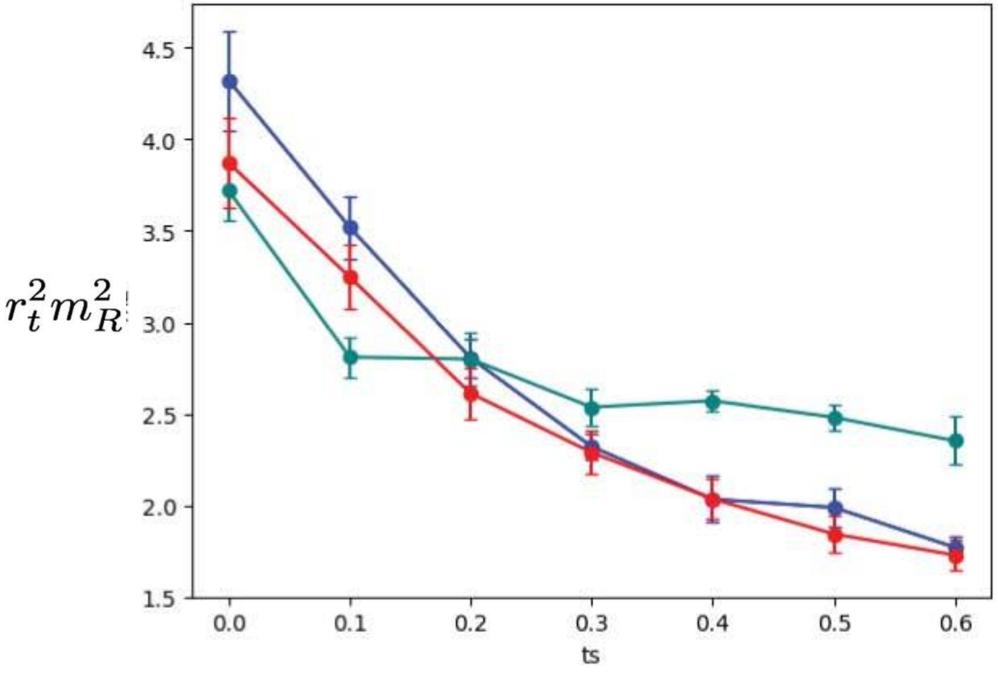


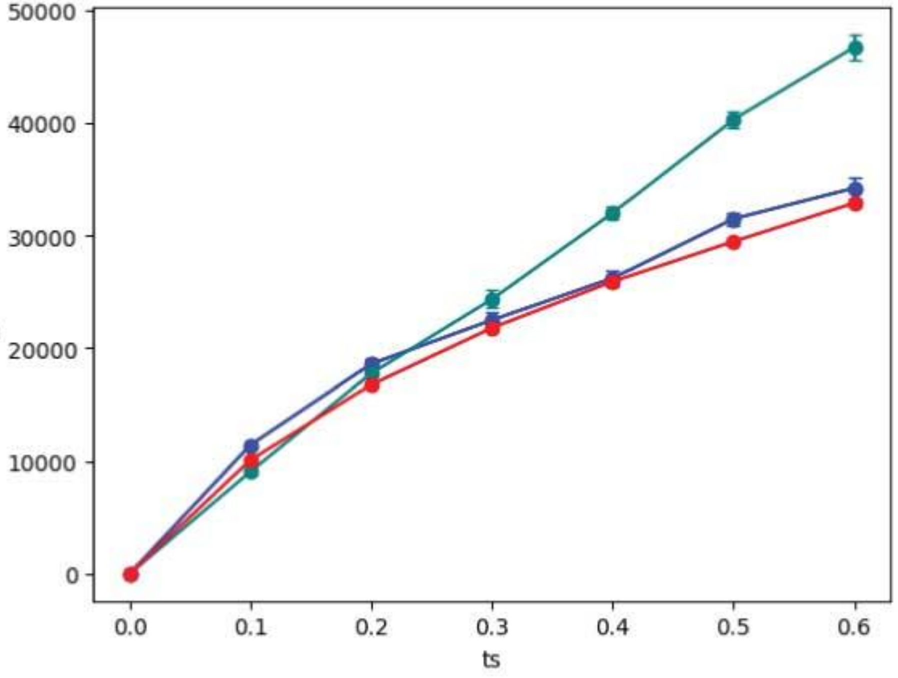
Alternate Parameters (close to critical point)



Numerics: Learning the RG flow of ϕ^4.

Carosso RG





$$\frac{1}{m_R^2} = \xi^2 \approx \frac{1}{4L^2} \sum_{\mathbf{p}\in\{(1,0),(0,1),(-1,0),(0,-1)\}} \frac{1}{\frac{4N^2}{L^2}\sin^2(\pi/N)} \left(\frac{\widetilde{G}_2(\mathbf{0},\mathbf{0})}{\widetilde{G}_2(\mathbf{p},-\mathbf{p})} - 1 \right)$$

Variational methods for ground states?

Instead let us Wick-rotate and consider *QFT*.

If the ground state wave functional is nondegenerate then (up to a phase) it is the square root of a probability density.

This probability density *also flows under RG* and thus one can use the same methods to design a variational ansatz for a lattice ground-state wave functional.

Turning this into a good numerical ansatz will involve **careful design of the RG SDE**.

Renormalization Group Flow as an SDE

This may seem confusing because we think of RG as a ***deterministic*** procedure which ***averages and rescales*** fields

However, the resulting evolution of the ***probability density*** can also be generated by an equivalent *stochastic* dynamics

(This is essentially an instance of the continuity equation)

In our work, we explore the connection between these stochastic RG schemes both theoretically and numerically

Connections to Other Stories

- Mathematical physics (**[Bauerschmidt-Bodineau]**, **[Bauerschmidt ‘23]**):
  - Mixing of Markov chains in high temperature ϕ^4, sine-Gordon
- Optimal Transport (above, also **[Cotler, Rezchikov]**)
  - Bakry-Emery, log-Sobolev inequalities, connection to correlation decay
  - Bayesian inference **[Berman et al.]**
- Stochastic Localization (**[Montanari ‘23]**, **[Eldan-Koehler-Zeitouni ‘21]**, **many others!**)
  - Many powerful results about spin glasses, convex geometry, and mixing of Markov chains
- Wilson/Gradient Flow (**[Lüscher]**, **[Carosso]**, **many others!**)
  - Wilson flow is an interesting smoothing process (Yang-Mills gradient flow/Heat equation in continuum); Carosso connected this to ERG
  - Motivation for normalizing flow work