Title: TBA - Machine Learning Initiative Seminar

Speakers: Kim Nicoli

Collection/Series: Machine Learning Initiative

Subject: Other

Date: October 11, 2024 - 2:30 PM

URL: https://pirsa.org/24100108

UNIVERSITÄT BONN
RHEINISCHE
FRIEDRICH-WILHELMS-
UNIVERSITÄT BONN

TRA
MATTER
TRANSDISCIPLINARY
RESEARCH AREA

To the Latent Space and Beyond:
New Frontiers in Generative Modelling in Physics

Kim A. Nicoli

University of Bonn, HISKP (Helmholtz Institute for Radiation and Nuclear Physics)

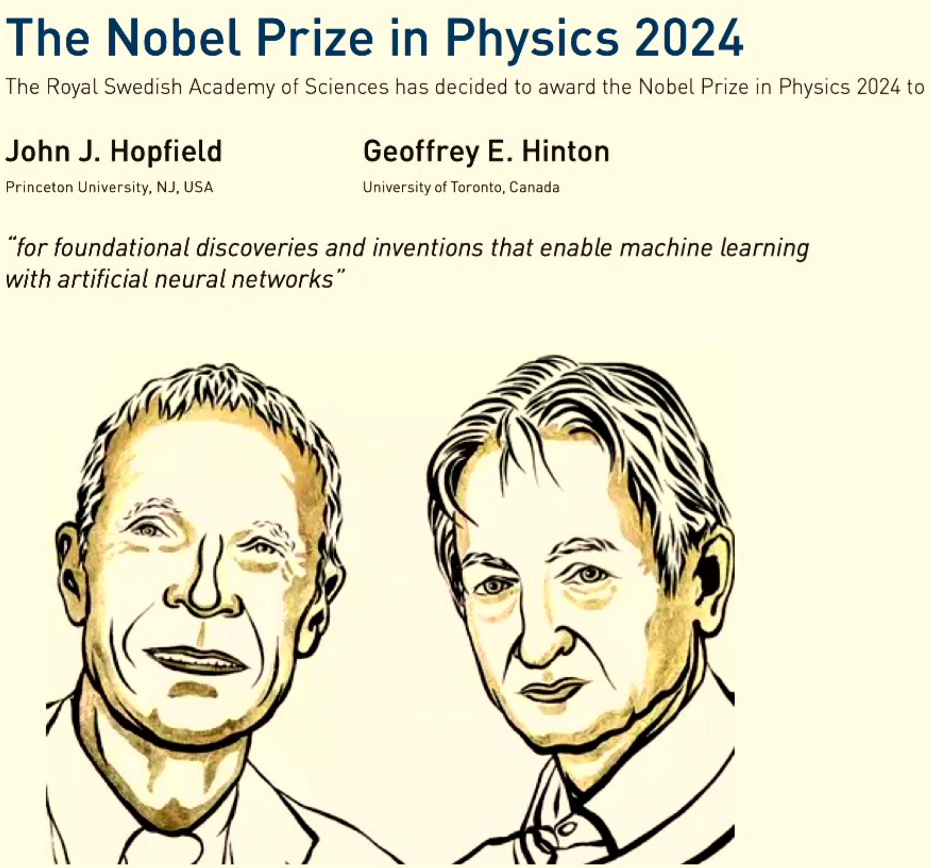
Talk based on PRE 101, 023304 (2020), PRL 126, 032001 (2021), PRD 108, 114501 (2023) and unpublished work.

Collaborators: C. Anders, E. Berkowitz, A. Bulgarelli, E. Cellini, L. Funcke, T. Hartung, K. Jansen, P. Kessel, J. Kreit, T. Luu, K-R. Müller, A. Nada, S. Nakajima, M. Rodekamp, D. Schuh, P. Stornati

Friday, 11th Oct. 2024 - Perimeter Institute ML Initiative Seminar Series

Image credits: generated with leonardo.ai

An exciting week for ML in the sciences



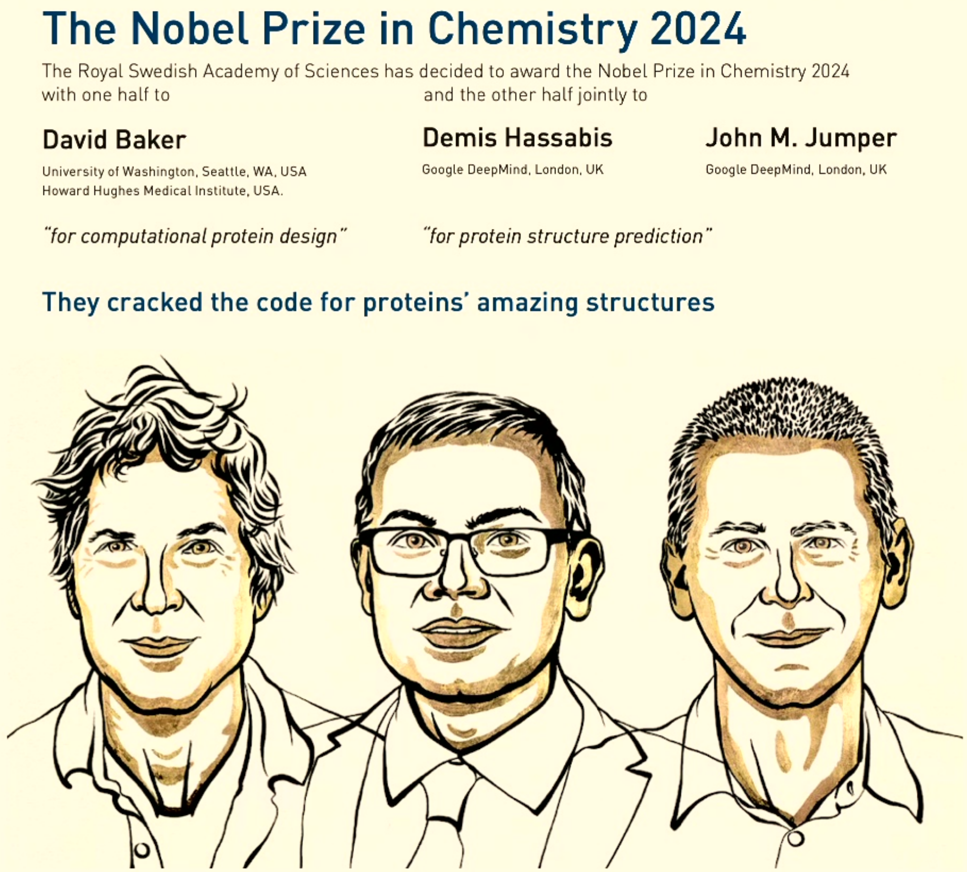
The Nobel Prize in Physics 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield
Princeton University, NJ, USA

Geoffrey E. Hinton
University of Toronto, Canada

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"



The Nobel Prize in Chemistry 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry 2024 with one half to

David Baker
University of Washington, Seattle, WA, USA
Howard Hughes Medical Institute, USA.

"for computational protein design"

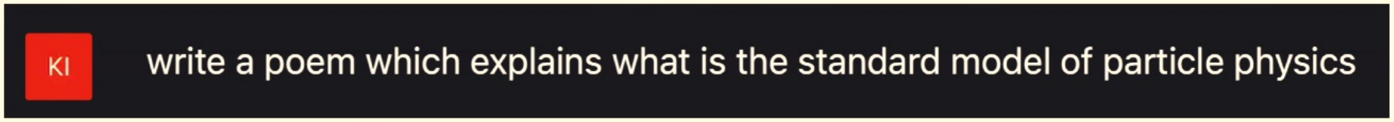
and the other half jointly to

Demis Hassabis
Google DeepMind, London, UK

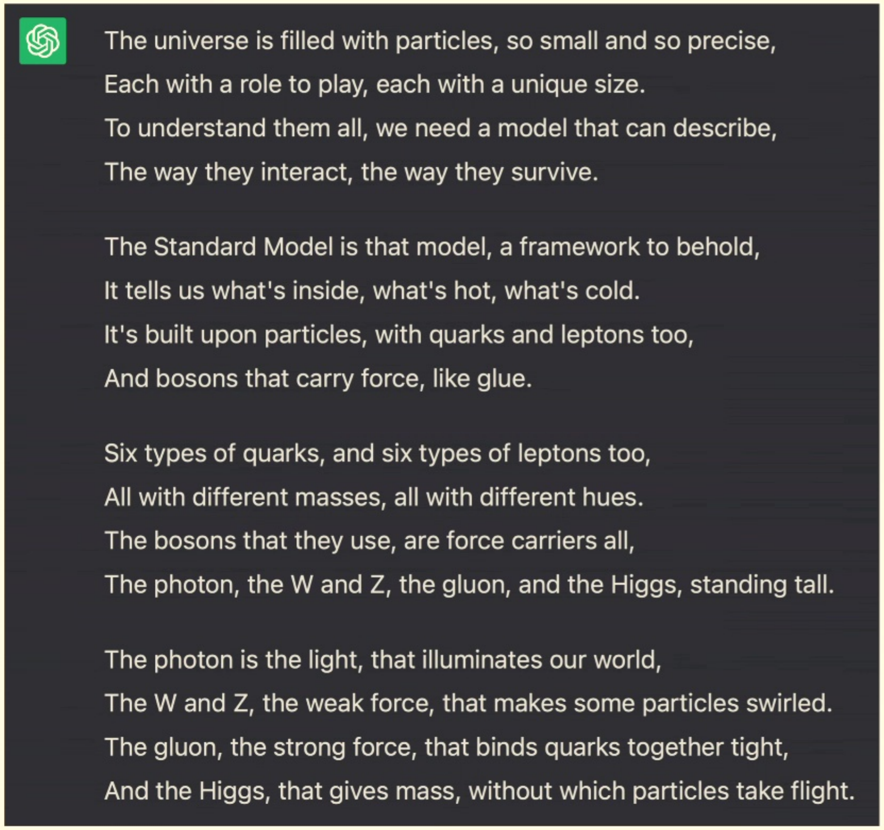
John M. Jumper
Google DeepMind, London, UK

"for protein structure prediction"

They cracked the code for proteins' amazing structures



write a poem which explains what is the standard model of particle physics

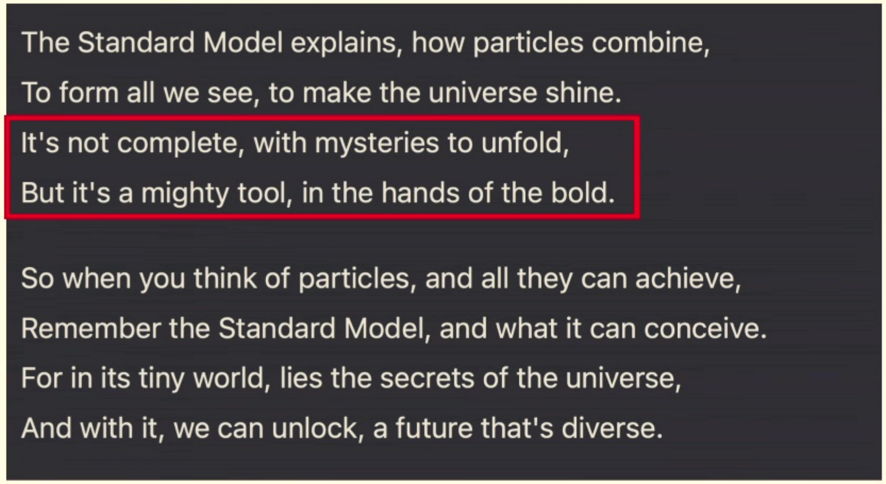


The universe is filled with particles, so small and so precise,
Each with a role to play, each with a unique size.
To understand them all, we need a model that can describe,
The way they interact, the way they survive.

The Standard Model is that model, a framework to behold,
It tells us what's inside, what's hot, what's cold.
It's built upon particles, with quarks and leptons too,
And bosons that carry force, like glue.

Six types of quarks, and six types of leptons too,
All with different masses, all with different hues.
The bosons that they use, are force carriers all,
The photon, the W and Z, the gluon, and the Higgs, standing tall.

The photon is the light, that illuminates our world,
The W and Z, the weak force, that makes some particles swirled.
The gluon, the strong force, that binds quarks together tight,
And the Higgs, that gives mass, without which particles take flight.



The Standard Model explains, how particles combine,
To form all we see, to make the universe shine.
It's not complete, with mysteries to unfold,
But it's a mighty tool, in the hands of the bold.

So when you think of particles, and all they can achieve,
Remember the Standard Model, and what it can conceive.
For in its tiny world, lies the secrets of the universe,
And with it, we can unlock, a future that's diverse.

Credits: ChatGPT (v3.5)

Generative AI for Physics : Past, Present, and Future

F. Noé, et al., Science, eaaw1147 (2019)

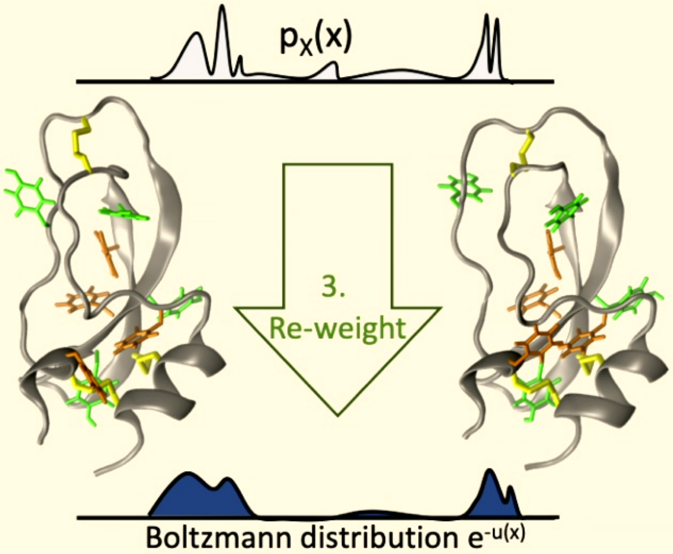


Image credits: F. Noé, et al., Science, eaaw1147 (2019)

Generative AI for Physics : Past, Present, and Future

Lattice Scalar Field Theories (ϕ^4)

M. S. Albergo, et al., Phys. Rev. D 100, 034515 (2019)

K. A. Nicoli, et al., Phys. Rev. Lett. 126, 032001 (2021)

P. deHaan, et al., arXiv: 2110.02673 @ ML4Pys workshop (2021)

L. Vaitl et al., arXiv: 2206.09016 @ ICML (2022)

A. Matthews et al., arXiv:2201.13117 @ ICML (2022)

M. Caselle, et al., J. High Energ. Phys. 2022, 15 (2022)

M. Gerdes, et al., SciPost Phys. 15, 238 (2023)

A. Singha, et al., Phys. Rev. D 107, 014512 (2023)

…

Lattice Gauge Theories ($\mathrm{U}(1), \mathrm{SU}(N)$)

G. Kanwar, et al., Phys. Rev. Lett. 125, 121601 (2020)

S. Bacchio, et al., Phys. Rev. D 107, L051504 (2023)

R. Abbott et al., Phys. Rev. D 106, 074506 (2022)

M. S. Albergo, et al., Phys. Rev. D 106, 014514 (2022)

R. Abbott, et al., arXiv:2305.02402 (2023)

J. Finkenrath, arXiv: 2201.02216 (2022)

…

Sampling Multimodal Densities in QFT

D.C. Hackett et al., arXiv:2107.00734 (2021)

K. A. Nicoli, et al., arXiv: 2111.11303 @ LATTICE21 (2021)

K. A. Nicoli, et al., Phys. Rev. D 108, 114501 (2023)

B. Maté et al., TMLR 2835-8856 (2023)

V. Kanaujia et al., arXiv:2401.15948 (2024) …

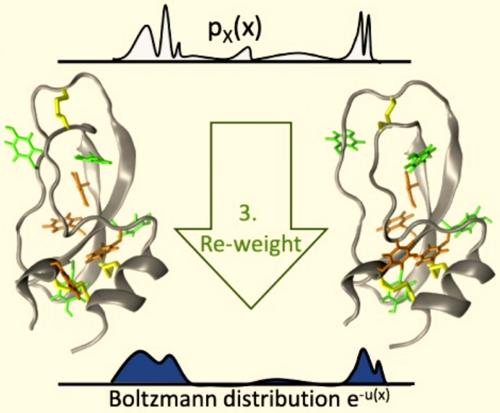


Image credits: F. Noé, et al., Science, eaaw1147 (2019)

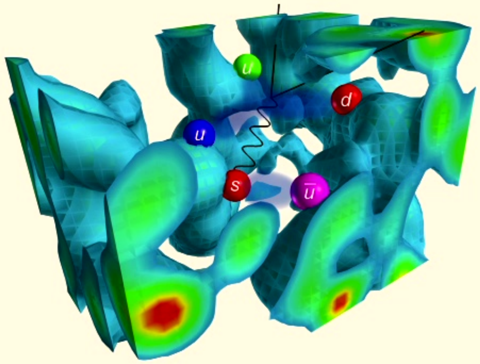


Image credits: Lattice QCD © Derek Leinweber/CSSM/University of Adelaide

Scaling to Larger Lattices

L. Del Debbio, et al., Phys. Rev. D 104, 094507 (2021)

R. Abbott et al., Eur. Phys. J. A 59, 257 (2023)

A. Faraz et al., arXiv:2308.08615 (2022)

B. Maté et al., arXiv: 2401.00828 (2024)

R. Abbott et al., arXiv: 2401.10874v1 (2024)

J. Finkenrath, arXiv: 2402.12176 (2024)

…

Autoregressive Models in Stat. Mech.

D. Wu et al., Phys. Rev. Lett. 122 (8), 080602 (2019)

K. A. Nicoli, et al., Phys. Rev. E 101 (2), 023304 (2019)

P. Bialas et al., Computer Physics Communications 281 (2022)

P. Bialas et al., Phys. Rev. E 107 (1), 015303 (2023)

…

Related Papers and Reviews

F. Noé, et al., Science, eaaw1147 (2019)

S. Chen, et al., Phys. Rev. D 107, 056001 (2022)

B. Maté et al., arXiv: 2210.13772 (2022)

L. Vaitl et al., MLST 3 (4), 045006 (2022)

Caselle, M., et al., J. High Energ. Phys. 2024, 48 (2024)

Cranmer K. et al., Nature Reviews Physics 5 (2023)

…

And many many more…

Path-Integral Formulation

Path integral is the basic tool for **quantising** fields and **computing** expectation values of physical observables $\mathcal{O}$:

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int D[\phi]\, \mathcal{O}(\phi)\, \exp\{-S(\phi)\}$$

Path-Integral Formulation

Path integral is the basic tool for **quantising** fields and **computing** expectation values of physical observables $\mathcal{O}$:

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int D[\phi]\, \mathcal{O}(\phi) \ \exp\{-S(\phi)\}$$

computed over a Boltzmann-like probability density:

$$p(\phi) = \frac{e^{-S(\phi)}}{Z} \longrightarrow$$

known in **closed form** up to a **numerically intractable** normalisation

$$Z = \int D[\phi] e^{-S(\phi)}$$

Lattice Quantum Field Theory

- The **path integral** reduces from a **functional integral** to a **high-dimensional** ordinary integral

$$\langle \mathcal{O} \rangle = \int \prod_{x \in \Lambda} \mathrm{d}[\phi(x)]\, \mathcal{O}(\phi)\, p(\phi)$$

- The field configuration $\phi(x)$ is now a **random variable** of size Λ (lattice volume).
- With the Markov-Chain Monte Carlo (MCMC) algorithm, we can **sample** and estimate observables

$$\langle \mathcal{O} \rangle_p = \int D[\phi]\, \mathcal{O}(\phi)\, p(\phi) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(\phi_i)$$

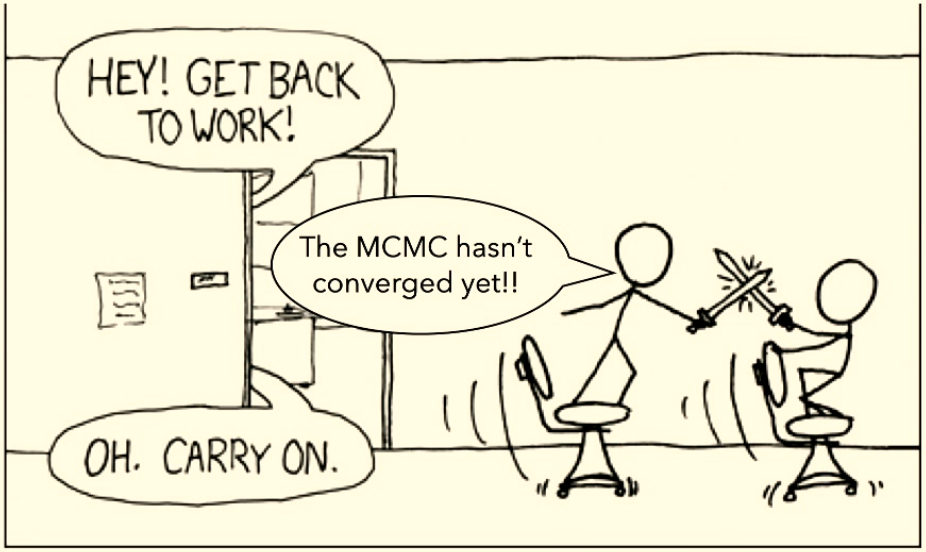
where $\phi_i \sim p$ (Boltzmann-like density)

Markov-Chain Monte Carlo (MCMC)

MCMC: sequentially proposes the next sample and guarantees to eventually converge to a target density.

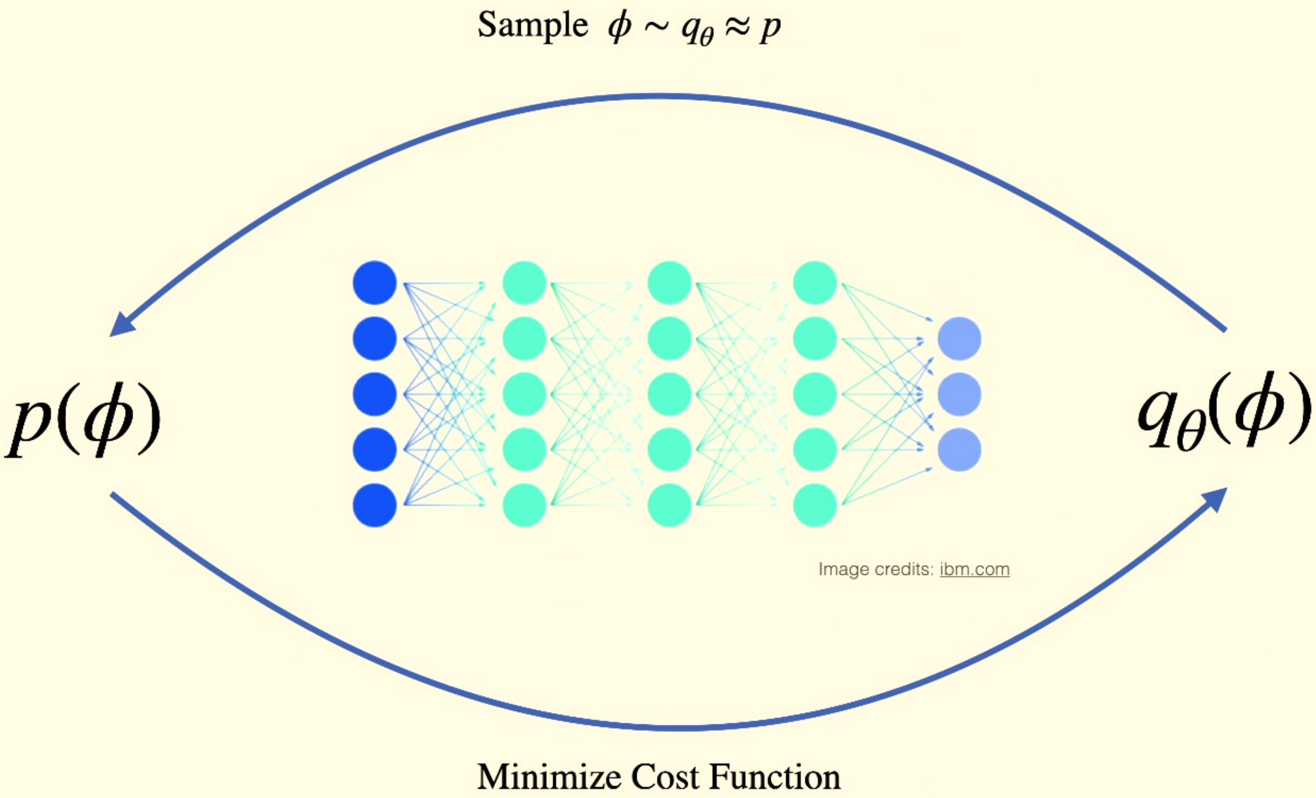
However, MCMC algorithms come at a **cost**:

- 👎 **Sequential** ⟹ MCMC chains cannot be parallelized.
- 👎 **Critical slowing down** ⟹ Struggles around phase transitions.
- 👎 Long range autocorrelations ⟹ **large statistical errors.**
- 👎 The partition function Z **is unknown**.
- 👎 No direct estimation of **thermodynamic observables.**



Adapted from: xkcd/303

Density Estimation with Deep Generative Models



Sample $\phi \sim q_\theta \approx p$

$p(\phi)$

$q_\theta(\phi)$

Image credits: ibm.com

Minimize Cost Function

Taxonomy of Generative Models

DGM	sampling probability	Normalization
GAN	none	X
VAE	approximate	✓
ARNN	exact	✓
NF	exact	✓
DDPM	approximate*	✓

Adapted from Table 2.1, **NKA**, PhD diss., Technische Universität Berlin, (2023)

* can be exact under some circumstances, e.g., Sec 4.3 from Yang et al., ACM (2023)

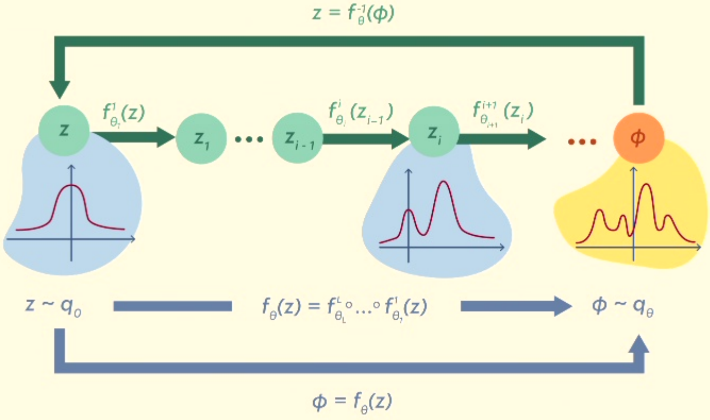
Sampling with Normalizing Flows

- NFs learn parametric maps f_θ (**<u>diffeomorphism</u>**)
- Transform samples from a prior $z \sim q_0$ into configurations $\phi \sim q_\theta$

$$f_\theta : z \sim q_0 \rightarrow \phi = f_\theta(z) \sim q_\theta$$

The parametric function needs to fulfill certain criteria:

- Bijective transformation $\phi = f_\theta(z)$
- Invertible and differentiable
- Tractable Jacobian

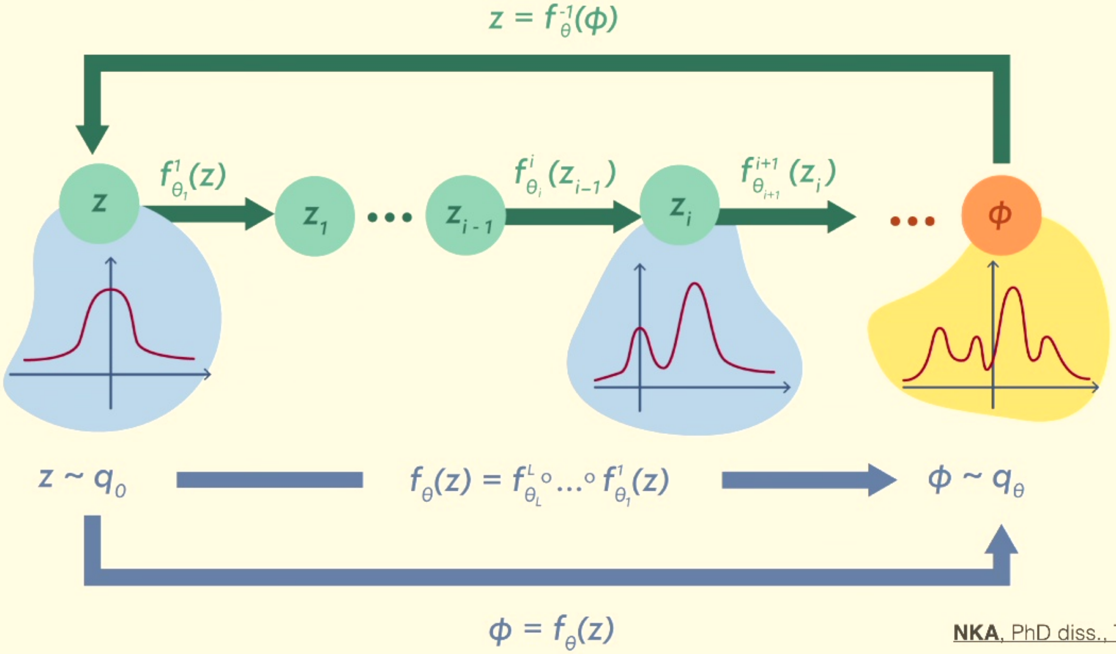


<u>**NKA**, PhD diss., Technische Universität Berlin, (2023)</u>

Sampling with Normalizing Flows

The probability density q_θ can be computed:

$$q_\theta(\phi) = q_0\left(f_\theta^{-1}(\phi)\right)\left|\det\left(\frac{\partial f_\theta}{\partial z}\right)\right|^{-1}$$



NKA, PhD diss., Technische Universität Berlin, (2023)

How do we train a generative model?

Often the variational density q_θ is trained by minimizing the **Reverse-KL** divergence:

$$KL(q_\theta||p) = \int D[\phi] q_\theta(\phi) \ln \frac{q_\theta(\phi)}{p(\phi)} \equiv \mathbb{E}_{q_\theta}\left[\ln \frac{q_\theta(\phi)}{p(\phi)}\right] .$$

How do we train a generative model?

Often the variational density q_θ is trained by minimizing the **Reverse-KL** divergence:

$$KL(q_\theta||p) = \int D[\phi] q_\theta(\phi) \ln \frac{q_\theta(\phi)}{p(\phi)} \equiv \mathbb{E}_{q_\theta}\left[\ln \frac{q_\theta(\phi)}{p(\phi)}\right] .$$

Since we know the target $p(\phi)$ is a Boltzmann distribution $p(\phi) = Z^{-1} \exp\{-S(\phi)\}$

$$KL(q_\theta||p) = \mathbb{E}_{q_\theta}\left[\ln \frac{q_\theta(\phi)}{p(\phi)}\right] = \mathbb{E}_{q_\theta}\left[\ln q_\theta(\phi) + S(\phi) + \ln Z\right]$$

How do we train a generative model?

Often the variational density q_θ is trained by minimizing the **Reverse-KL** divergence:

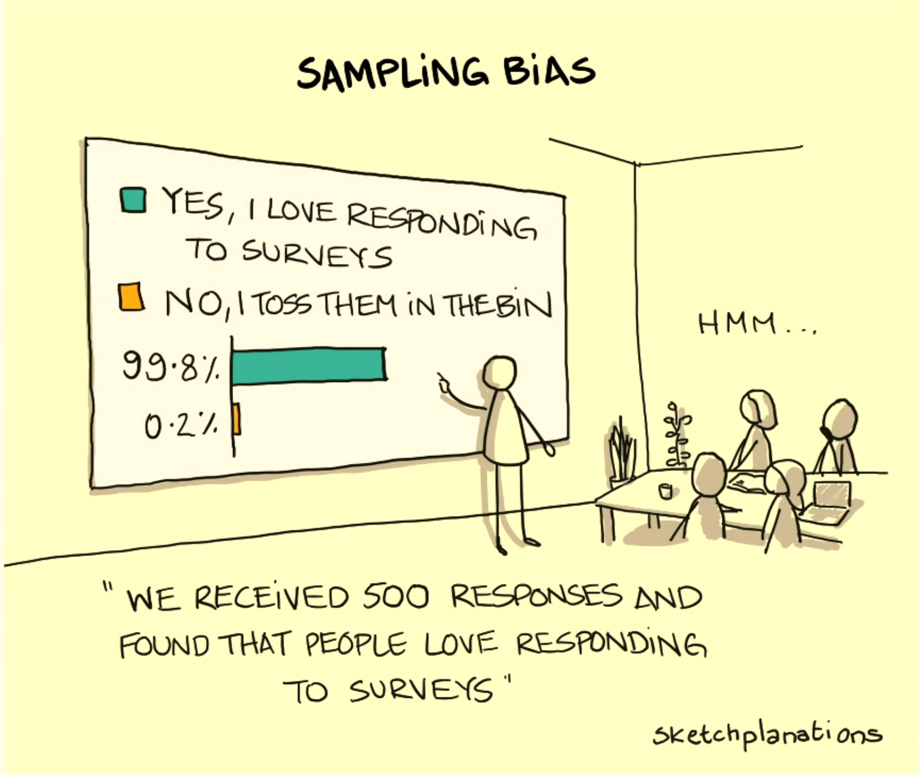
$$KL(q_\theta||p) = \int D[\phi] q_\theta(\phi) \, \ln \frac{q_\theta(\phi)}{p(\phi)} \equiv \mathbb{E}_{q_\theta}\left[\ln \frac{q_\theta(\phi)}{p(\phi)}\right] .$$

Since we know the target $p(\phi)$ is a Boltzmann distribution $p(\phi) = Z^{-1} \exp\{-S(\phi)\}$

$$\nabla_\theta KL(q_\theta||p) = \mathbb{E}_{q_\theta}\left[\nabla_\theta \ln q_\theta(\phi) + \nabla_\theta S(\phi) + \cancel{\ln Z}\right]$$

Training can be performed **by self-sampling** from the model we are training**!**

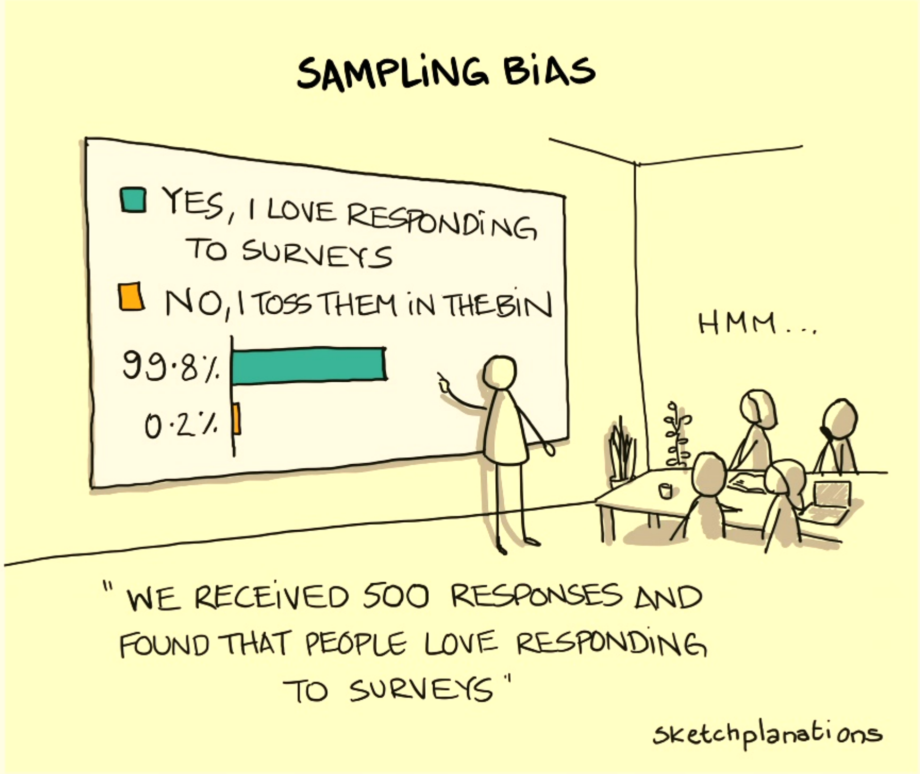
Are these samplers unbiased?



$$p(\phi) = q_\theta(\phi)$$ 🤔

Image credits: sketchplanations

Are these samplers unbiased?



$$\cancel{p(\phi) = q_\theta(\phi)}$$ 😭

$$p(\phi) \approx q_\theta(\phi)$$ 😕

Image credits: sketchplanations

Neural Importance Sampling (NIS)

$$p(\phi) \approx q_\theta \sim \phi_i$$ where $$p(\phi) = \frac{\exp\{-S(\phi)\}}{Z}$$

Recall: The partition function Z *can not* be directly estimated by MCMC.

$$Z = \int D[\phi]\, \exp\{-S(\phi)\} = \int D[\phi]\, q_\theta(\phi)\, \tilde{w}(\phi)$$ where $$\tilde{w}(\phi) = \frac{\exp\{-S(\phi)\}}{q_\theta(\phi)}$$

$$Z \overset{\mathrm{MC}}{\approx} \hat{Z} = \frac{1}{N}\sum_{i=1}^{N} \tilde{w}(\phi_i) \qquad \phi_i \sim q_\theta$$

Asymptotically Unbiased Estimation of Physical Observables

$$\langle \mathcal{O} \rangle_p = \langle w\mathcal{O} \rangle_{q_\theta} \overset{\text{MC}}{\approx} \frac{1}{N} \sum_{i=1}^{N} w(\phi_i)\mathcal{O}(\phi_i) \quad \phi_i \sim q_\theta$$

Asymptotically Unbiased Estimation of Physical Observables

$$w(\phi) = \frac{p(\phi)}{q_\theta(\phi)} = \frac{1}{\hat{Z}} \frac{e^{-S(\phi)}}{q_\theta}$$

$$\langle \mathcal{O} \rangle_p = \langle w \mathcal{O} \rangle_{q_\theta} \overset{\text{MC}}{\approx} \frac{1}{N} \sum_{i=1}^{N} w(\phi_i) \mathcal{O}(\phi_i) \qquad \phi_i \sim q_\theta$$

Asymptotically Unbiased Estimation of Physical Observables

$$\hat{F} = -T \ln \hat{Z}$$ ⚠

$$\langle \mathcal{O} \rangle_p = \langle w \mathcal{O} \rangle_{q_\theta} \overset{\text{MC}}{\approx} \frac{1}{N} \sum_{i=1}^{N} w(\phi_i) \mathcal{O}(\phi_i) \qquad \phi_i \sim q_\theta$$

$$\hat{P}, \hat{H}$$

NKA, Nakajima, Strodthoff, Samek, Müller, and Kessel, Phys. Rev. E (2020)

NKA, Anders, Funcke, Hartung, Jansen, Kessel, Nakajima, Stornati, Phys. Rev. Lett. (2021)

Summary and Conclusions (Part I)

- **Asymptotically unbiased samplers** can be constructed from trained DGMs (NIS or NMCMC).
- The **partition function** and **thermodynamic** observables can be **directly estimated**.
- Use **inductive biases**, e.g. symmetries, bootstrapping, annealing, etc., for enhance training.
- Sampling from DGMs is **embarrassingly parallelizable** **(i.i.d)** $\neq$ MCMC (**sequential**).

Summary and Conclusions (Part I)

TL;DR

Deep Generative Models (DGMs) are promising candidates for the next generation of sampling algorithms

The End… ?

We are certainly not done...

- Roadmap to full Lattice QCD →

Deploy state-of-the-art normalizing flow models on **exascale computing** to reach HMC comparable scale.

Cranmer K. et al., Nat. Rev. Phys. 5 (9) (2023)

Nice talk by Gurtej Kanwar at the Lattice Conference in 2023

- Scaling to larger lattices →

Learning **(local) defects** allows to scale to large lattices and gives accurate estimates of entanglement entropies.

Bulgarelli, Cellini, Kühn, Jansen, Nada, Nakajima, Panero, **NKA,** arXiv:XXXXXX (2024)

- Inductive bias (symmetries) →

Unsupervised learning of probability distributions for **condensed matter theories** with complicated topologies.

Schuh, Kreit, Berkowitz, Funcke, Luu, **NKA**, Rodekamp, arXiv:XXXXXX (2024)

- Software framework →

Developing a unified, accessible, modular framework for testing new theories and implementing new techniques.

NKA, Anders, Funcke, Jansen, Nakajima, Kessel Lattice Conf. 2023 PoS 286 (2024)

Direct Estimation of Entanglement Entropy with Normalizing Flows

In collaboration with: A. **Bulgarelli**, E. Cellini, K. Jansen, S. Kühn, A. Nada, S. Nakajima, M. Panero

Why Are we Interested in Entanglement Entropy?

The study of **entanglement** in many body systems finds applications in:

- Quantum phases of matter and quantum phase transitions [1].
  - Study of the **effective number** of degrees of freedom [2].
- High-energy physics.
  - Confinement, see [3].
- Quantum gravity and AdS/CFT [4].
- Resource for quantum computing, quantum simulations and technologies.
  - Enhance the engineering of quantum simulators through the study of entanglement [5, 6].
- Dirac Medallists 2024: Casini, Huerta, Ryu, Takanagi
  - For *"their insights on quantum entropy in quantum gravity and quantum field theories"*.

Image credits: https://iqim.caltech.edu

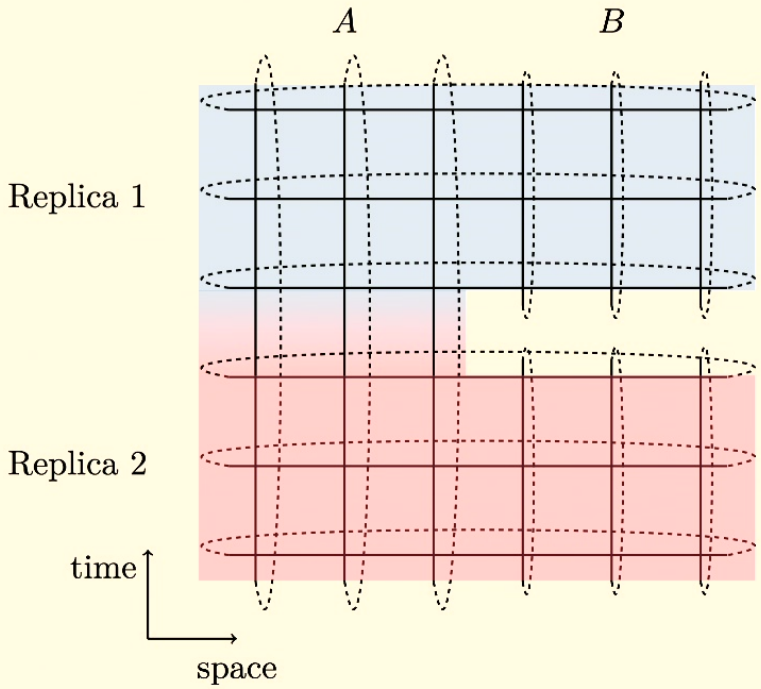
[1] Vidal et al., Phys. Rev. Lett. 90, (2003)
[2] Casini, Huerta, Phys. Lett. B (2004)
[3] Klebanov et al., Nuclear Phy. B, (2008)
[4] Ryu, Takanagi, Phys. Rev. Lett. 96 (2006)
[5] Daley et al., Phys. Rev. Lett. 109 (2012)
[6] Abanin et al., Phys. Rev. Lett. 109 (2012)

Entanglement Entropy

- **Entanglement entropy:** ideally computed the von Neumann entropy, but is hard to calculate numerically

(**Rényi entropy**: UV divergent quantity) Calabrese and Cardy, J. Stat. Mech. (2004)

$$S_n = \frac{1}{1-n} \log \mathrm{Tr}\rho_A^n \underset{\text{replica trick}}{\Longrightarrow} S_n = \frac{1}{1-n} \log \frac{Z_n}{Z^n}$$



See also: Calabrese and Cardy, J. Phys. A (2009), Bulgarelli and Panero, JHEP (2023), Bulgarelli and Panero, JHEP (2024)

Entanglement Entropy

- **Entanglement entropy:** ideally computed the von Neumann entropy, but is hard to calculate numerically

(**Rényi entropy**: UV divergent quantity) Calabrese and Cardy, J. Stat. Mech. (2004)

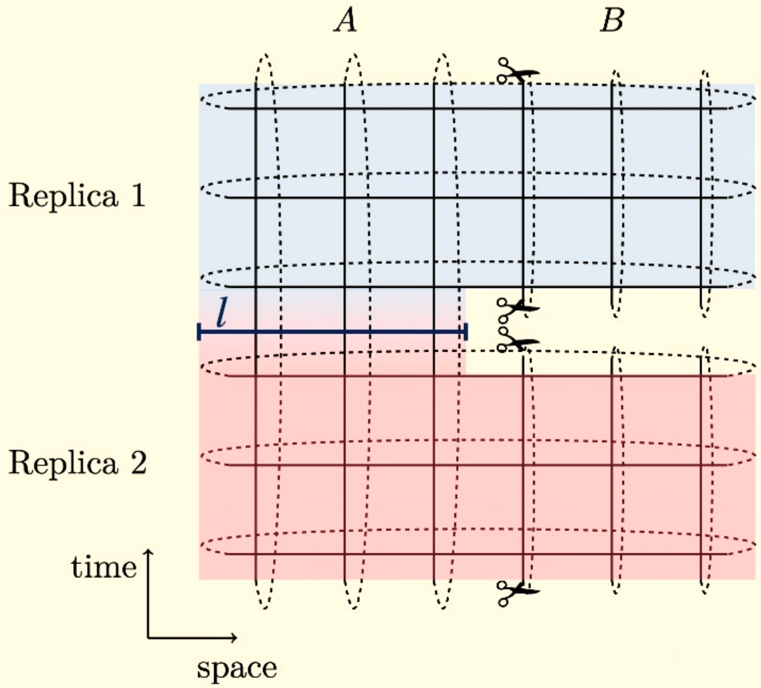
$$S_n = \frac{1}{1-n} \log \mathrm{Tr}\rho_A^n \xrightarrow[\text{replica trick}]{} S_n = \frac{1}{1-n} \log \frac{Z_n}{Z^n}$$

(**Entropic C-function**: derivatives of Rényi entropy)

$$C_n = \frac{l^{D-1}}{|\partial A|} \frac{\partial S_n}{\partial l} = \frac{l^{D-1}}{|\partial A|} \frac{1}{n-1} \log \frac{Z_n(l)}{Z_n(l+1)}$$

Where l is the **lenght of the *cut*** linking the two replicas

N.B. We assume unitary lattice spacing, i.e., $a = 1$ such that $l/a = l$



See also: Calabrese and Cardy, J. Phys. A (2009), Bulgarelli and Panero, JHEP (2023), Bulgarelli and Panero, JHEP (2024)

Entropic C-function

- The Renyi entropies/entropic c-functions can be computed on the lattice using different approaches:
  - **Quantum Monte Carlo** Hastings et al., Phys. Rev. Lett. 104 (2010)
  - **Non-Equilibrium MCMC** Bulgarelli and Panero, JHEP (2023), JHEP (2024)
  - **Tensor Networks** Feldman et al., PRX Quantum (2022), Okunishi et al., JPSJ (2022)
  - **Autoregressive Neural Networks** Bialas et al., arXiv:2406.06193 (2024)

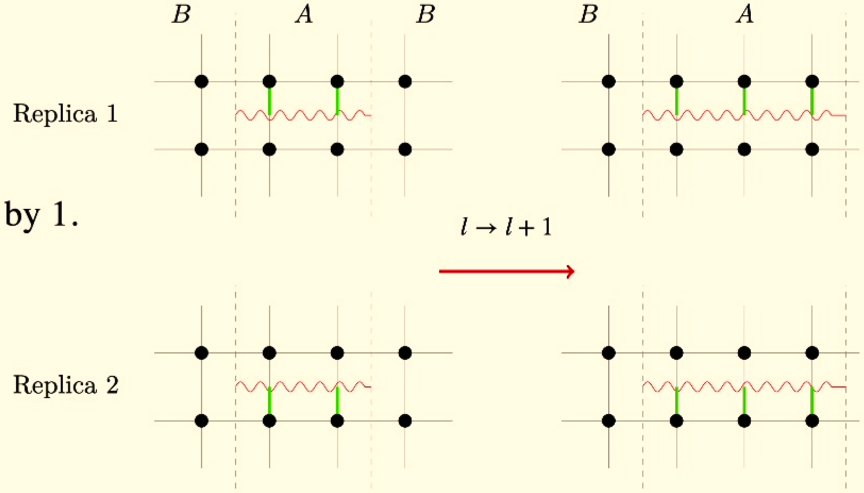
⚠ **We** train on a **reduced number of degrees of freedom** while **they** train to sample the **entire lattice**.

- We want to **learn** the effect of increasing the "*cut*" between the replicas by 1.

$$Z_n(l) \rightarrow Z_n(l+1)$$

- Direct estimation using **normalizing flows (NFs)** based sampling!

Adapted from Bulgarelli and Panero, JHEP (2024)



Related Works: **NKA** et al., Phys. Rev. Lett. (2021) , Caselle et al., JHEP (2022)

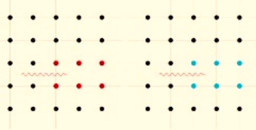
Flows for Entanglement Entropy

💡 **<u>Idea</u>**: learn localized defects of the lattice using normalizing flows

Our Protocol: learn transport between target distributions $p_l \rightarrow p_{l+1}$ with increasing length of the cut l

Thermalized Samples
$\phi_l \sim p_l$
Heatbath

NFs/SNFs →
Use of special couplings



Biased samples
$\phi_{l+1} \sim q_\theta \approx p_{l+1}$

Reweigthing →

Unbiased samples
$\hat{\phi}_{l+1} \sim p_{l+1}$

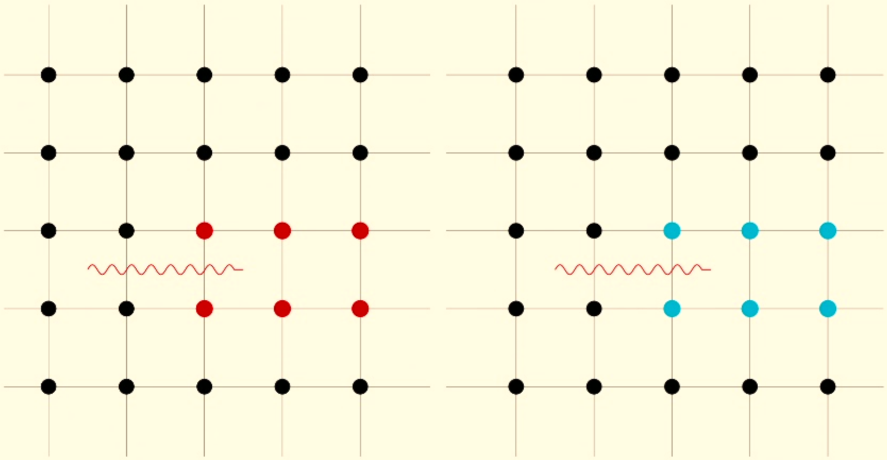
Training: Iterate the protocol for minimizing the objective $\mathrm{KL}(q_\theta || p_{l+1})$ with prior p_l.

Sampling: Start with thermalized samples from p_l and transform them into samples from p_{l+1}.

Flows for Entanglement Entropy

- The coupling layers act **locally** only on a portion of the defect (example with 2 replicas)

Defect Aware Coupling Layers



Frozen Replica

Active Replica

- ● Environment (always frozen)
- 🔴 Frozen (input to CL)
- 🔵 Active (transformed by CL)

Related work: Abbott et al., PoS LATTICE2023 (2024)

The Active sites (blue) are transformed, taking as input the Frozen sites (red).

Study of entropic c-functions for ϕ^4

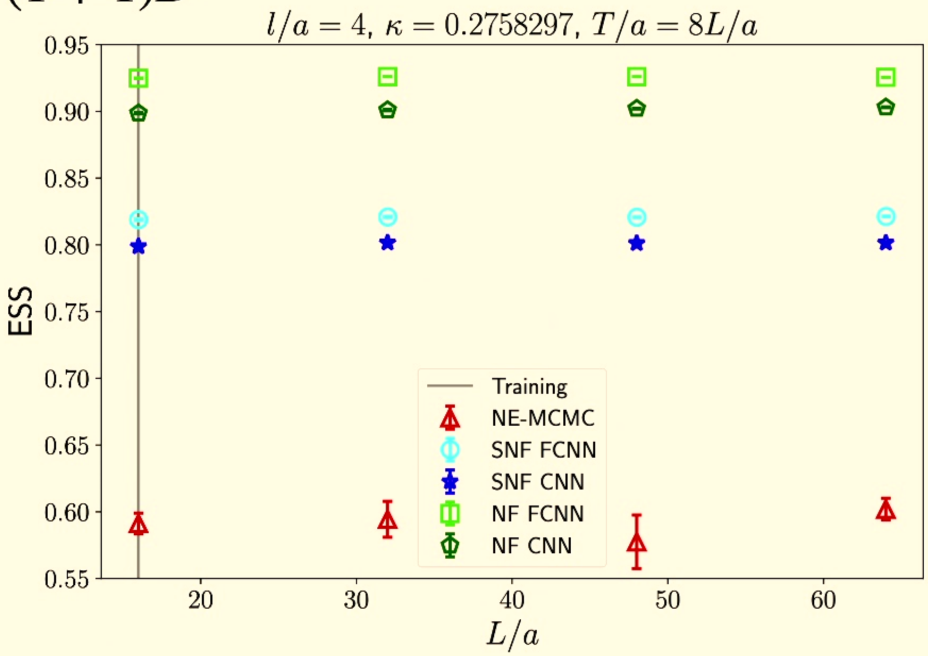
- We study the entropic c-functions C_2 in the ϕ^4 **scalar field theory** with action:

$$S(\phi) = \sum_{x\in\Lambda} \left[-2\kappa \sum_{\mu=1}^{D} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\phi(x)^2 + \lambda\phi(x)^4 \right]$$

  - We always train at $\kappa = \kappa_c$ **(critical point of the theory)** Bosetti et al., Phys. Rev. D (2015)
  - Generalization to **arbitrary n replicas** (in this work $n = 2$ for simplicity)
  - **Zero temperature** (temporal extent ≫ spatial extent) and $d = \{1,2\}$.
  - Flow trained for fixed Lattice L/a and cut length l/a **(cheap and fast training!)**
  - Model evaluation for different κ, L/a and l/a **without no retraining (Transfer learning)!**

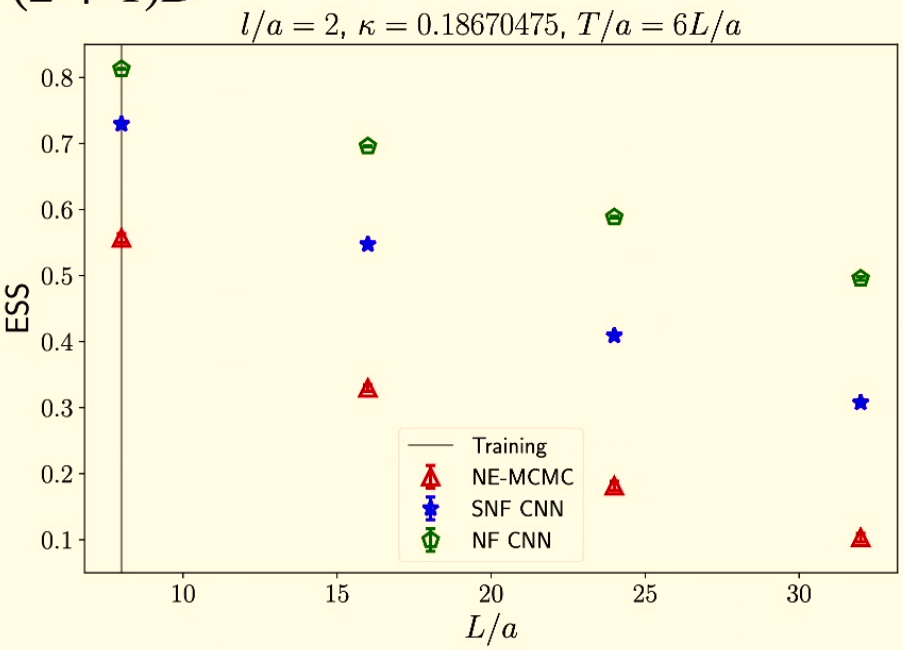
Transfer in the Volume (Target theory: ϕ^4)

$(1+1)D$



- **Training:** $\Lambda = 16 \times 128\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$

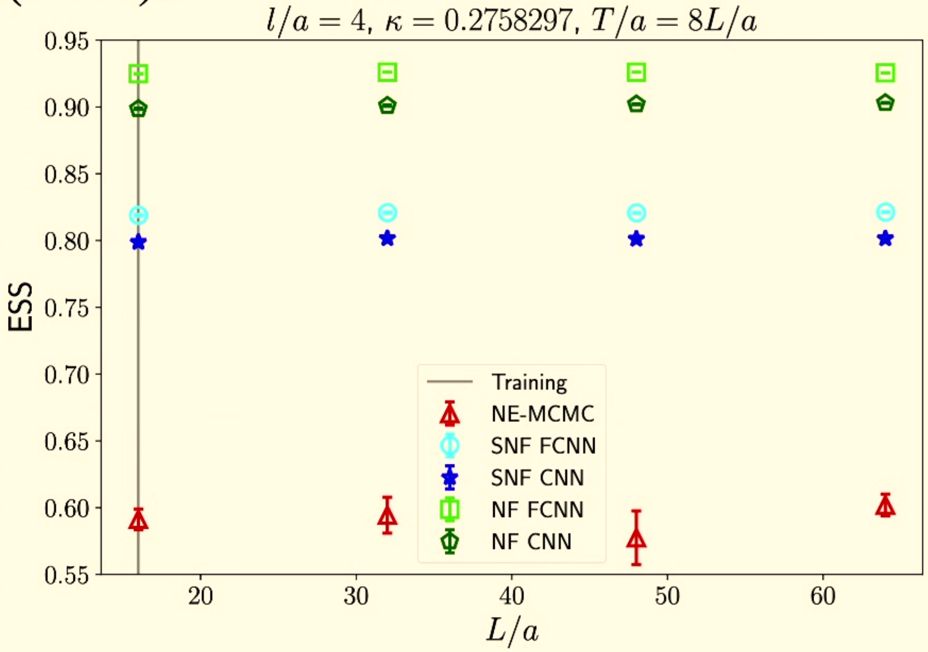
$(2+1)D$



- **Training:** $\Lambda = 8^2 \times 32\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$

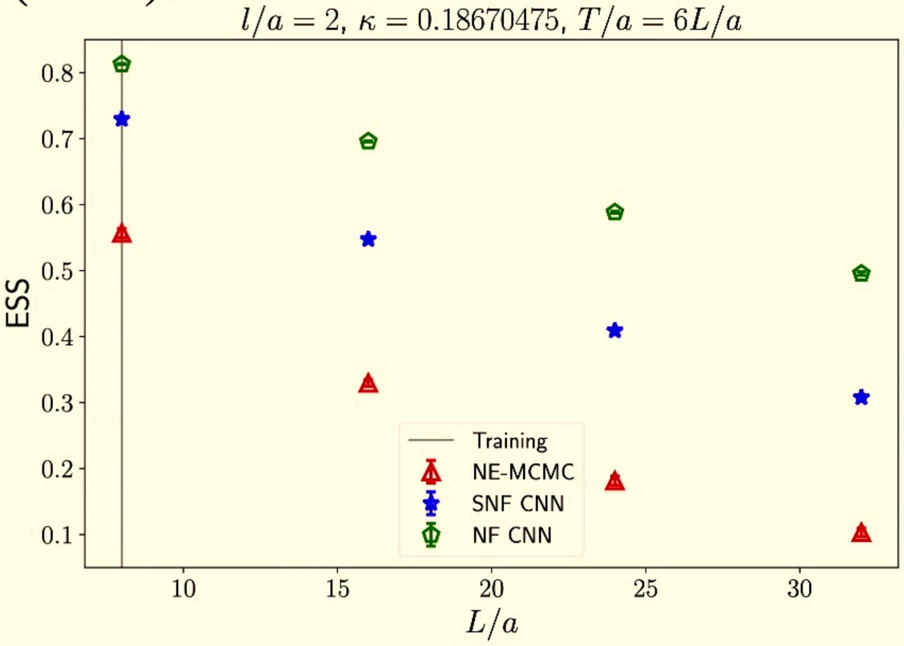
Transfer in the Volume (Target theory: ϕ^4)

$(1+1)D$



- **Training:** $\Lambda = 16 \times 128\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$

$(2+1)D$

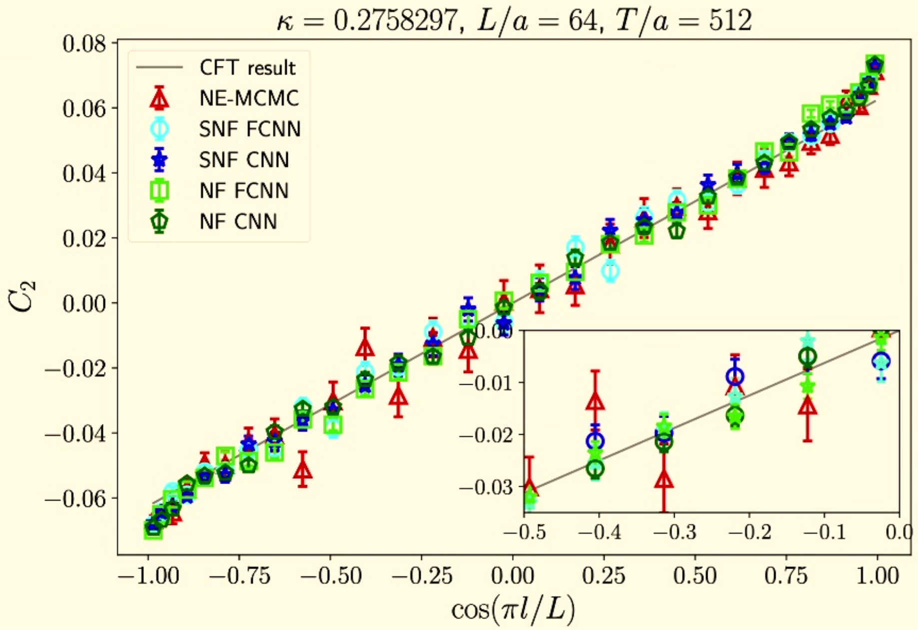


- **Training:** $\Lambda = 8^2 \times 32\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$

We can train at **very small lattices** and sample (almost) arbitrarily **large lattices without retraining**.

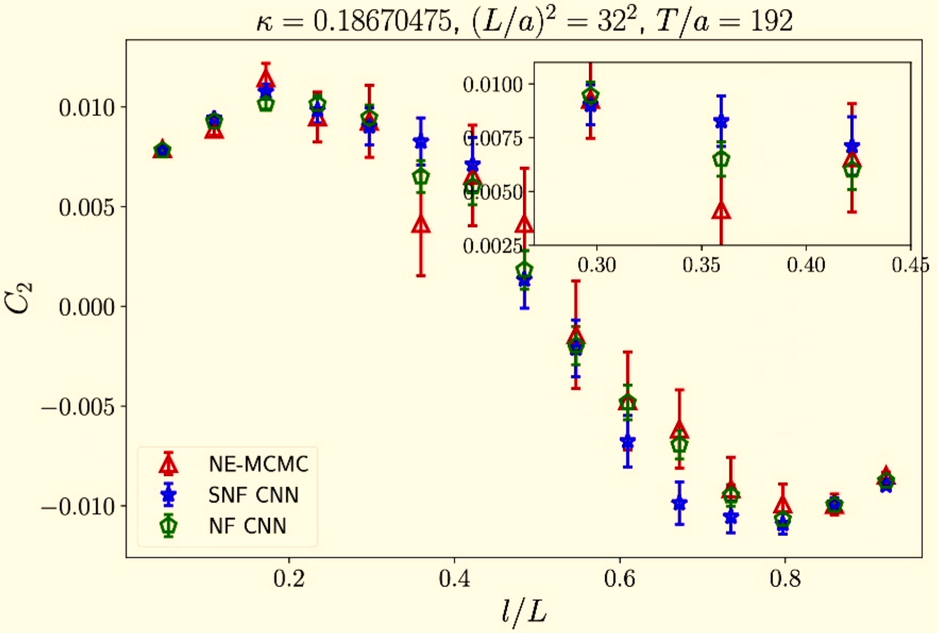
Entropic C_2 function (Target theory: ϕ^4)

$(1+1)D$



- **Training:** $\Lambda = 16 \times 128\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$
- Analytic solution from CFT in $(1+1)D$.

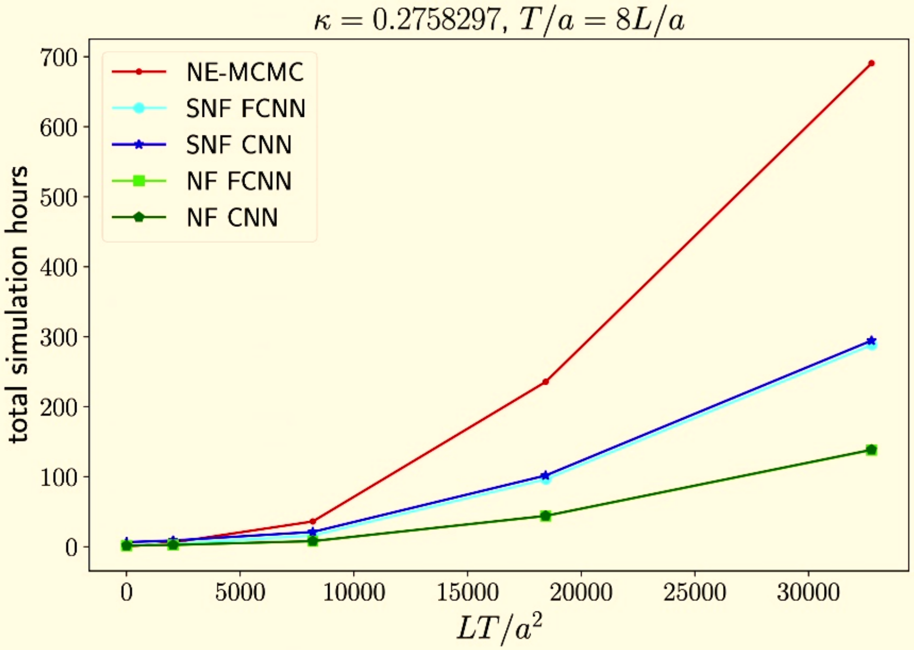
$(2+1)D$



- **Training:** $\Lambda = 8^2 \times 32\ ,\ l = 1,\ \kappa = \kappa_c,\ a = 1$
- No analytic solution is available $(2+1)D$.

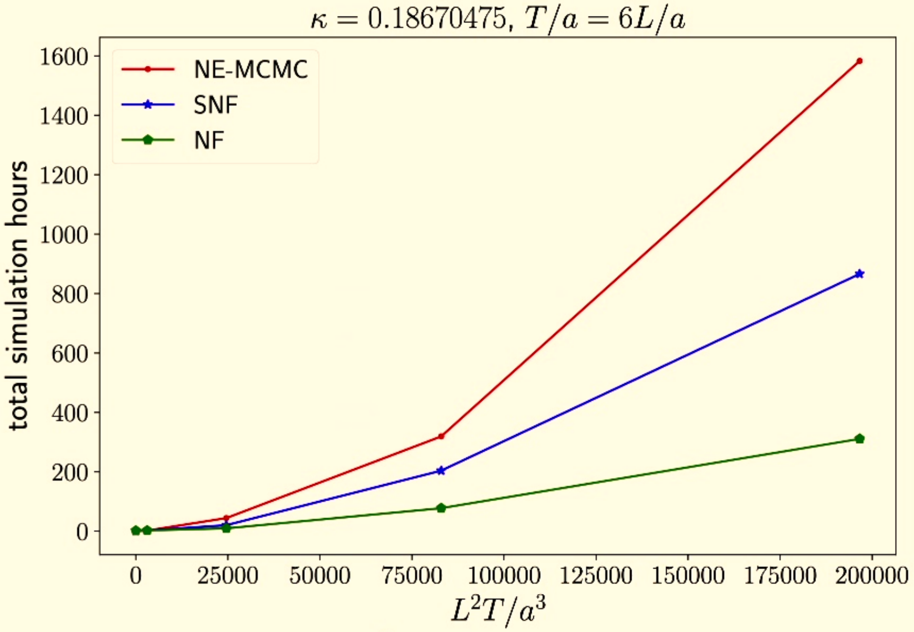
Wall-clock Time Comparison (Target theory: ϕ^4)

$(1+1)D$


$\kappa = 0.2758297,\ T/a = 8L/a$

- **Training:** $\Lambda = 16 \times 128\ ,\ l = 1,\ \kappa = \kappa_c$

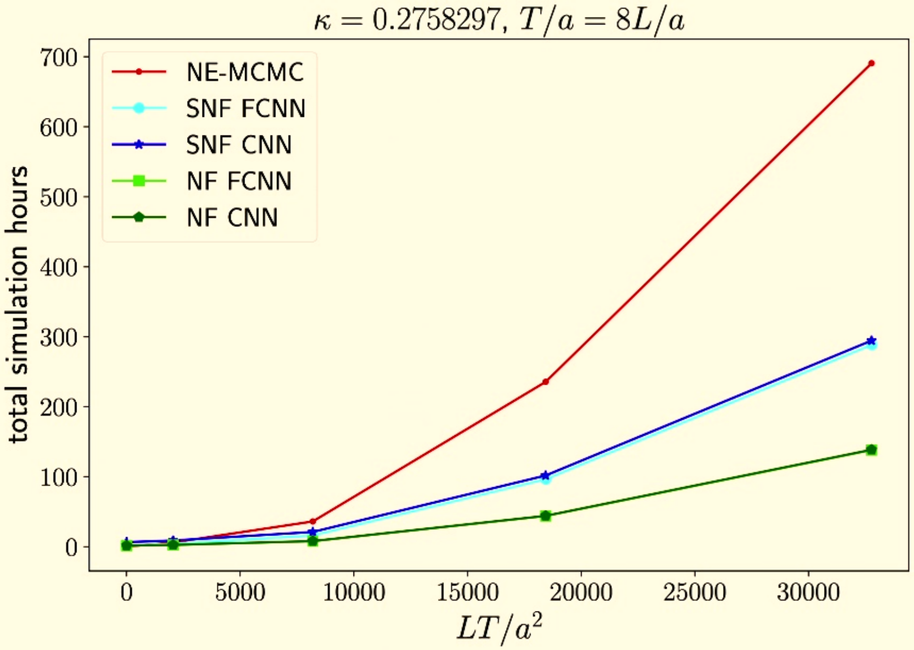
$(2+1)D$


$\kappa = 0.18670475,\ T/a = 6L/a$

- **Training:** $\Lambda = 8^2 \times 32\ ,\ l = 1,\ \kappa = \kappa_c$

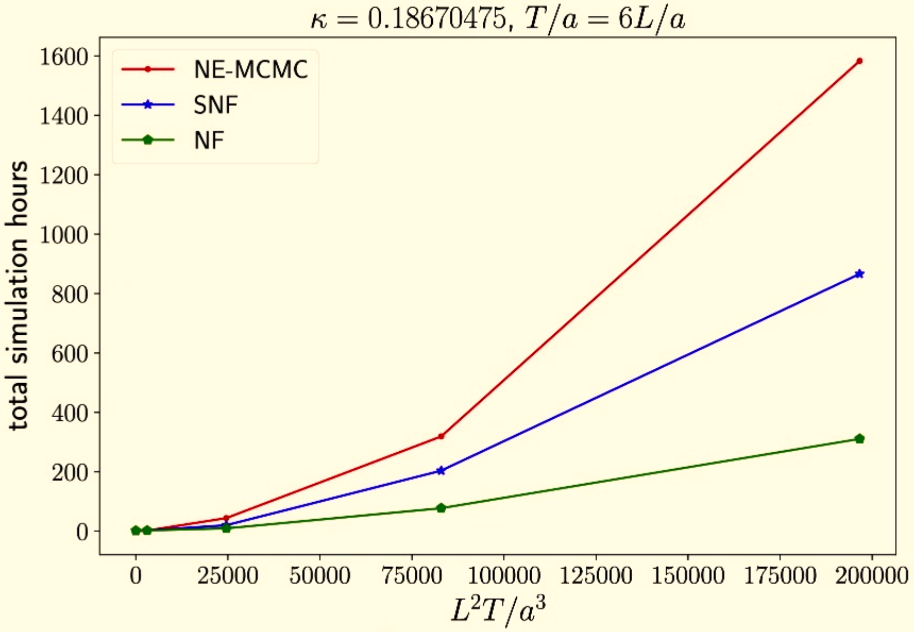
Wall-clock Time Comparison (Target theory: ϕ^4)

$(1+1)D$



- **Training:** $\Lambda = 16 \times 128\ ,\ l = 1,\ \kappa = \kappa_c$

$(2+1)D$



- **Training:** $\Lambda = 8^2 \times 32\ ,\ l = 1,\ \kappa = \kappa_c$

Substantial computational advantage for sampling at larger volumes.

Simulating the Hubbard Model with Normalizing Flows

In collaboration with: **D. Schuh**, **J. Kreit**, E. Berkowitz, L. Funcke, T. Luu, M. Rodekamp

The Hubbard Model

- ➡ High practical relevance.
- ➡ Used to describe carbon nanomaterial, e.g. graphene.

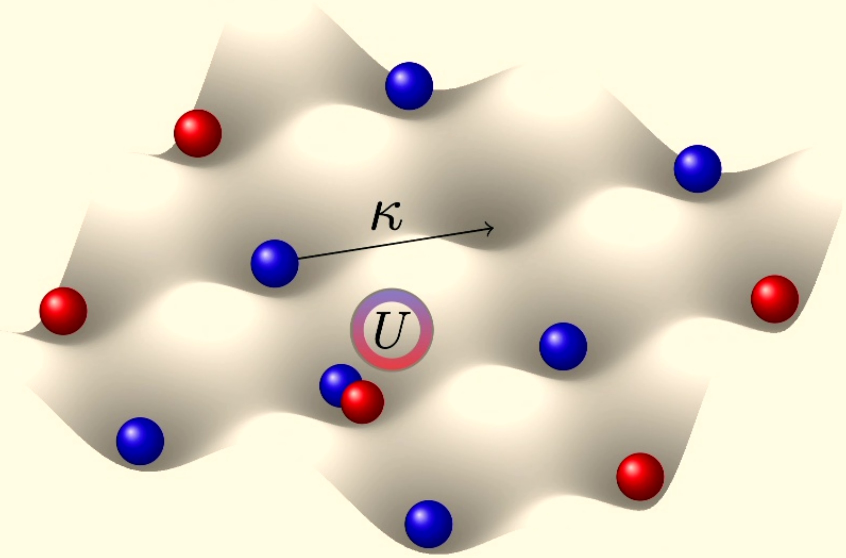
$$H = \underbrace{-\kappa \sum_{\langle x,y \rangle} (a^\dagger_{x,\uparrow} a_{y,\uparrow} + a^\dagger_{x,\downarrow} a_{y,\downarrow})}_{\text{tigh}-\text{binding}} \underbrace{- \frac{U}{2} \sum_{x} (n_{x,\uparrow} - n_{x,\downarrow})^2}_{\text{on}-\text{site interaction}}$$

κ → Hopping parameter
U → Interaction strength

Hubbard-Stratonovich transform allows to describe the system with bosonic auxiliary fields ϕ

$\tilde{U} = U \cdot \beta \cdot N_t^{-1}$ → Interaction strength
$\tilde{\kappa} = \kappa \cdot \beta \cdot N_t^{-1}$ → Hopping parameter
$M[\phi]$ → Fermion matrix

$$S = \frac{1}{2\tilde{U}} \sum_{x,t} \phi^2_{x,t} - \log \det M[\phi] - \log \det M[-\phi]$$

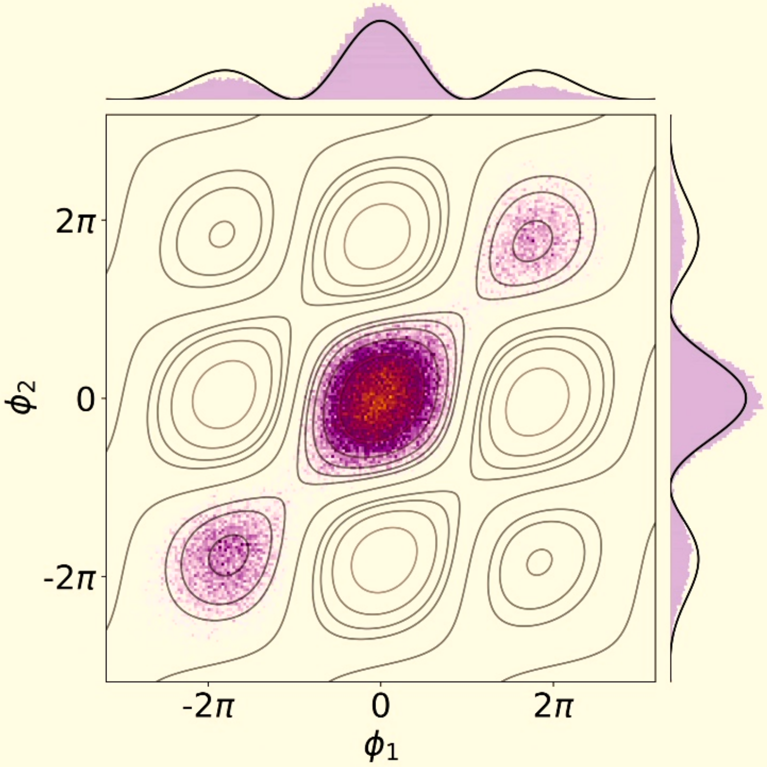


See also Wynen et al., Phys. Rev. B (2019)

Computational Bottlenecks

Sampling is **challenging** (similar MCMC drawbacks as before).

Let's consider the Hubbard Model in $(1+1)D$ with: $N_x = 2,\ N_t = 1$

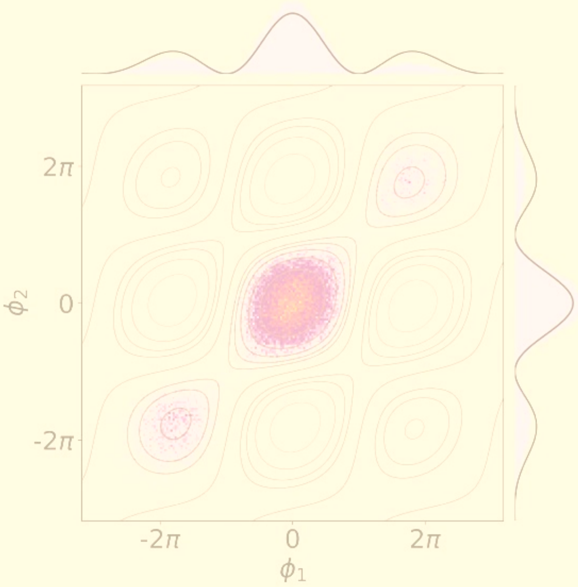


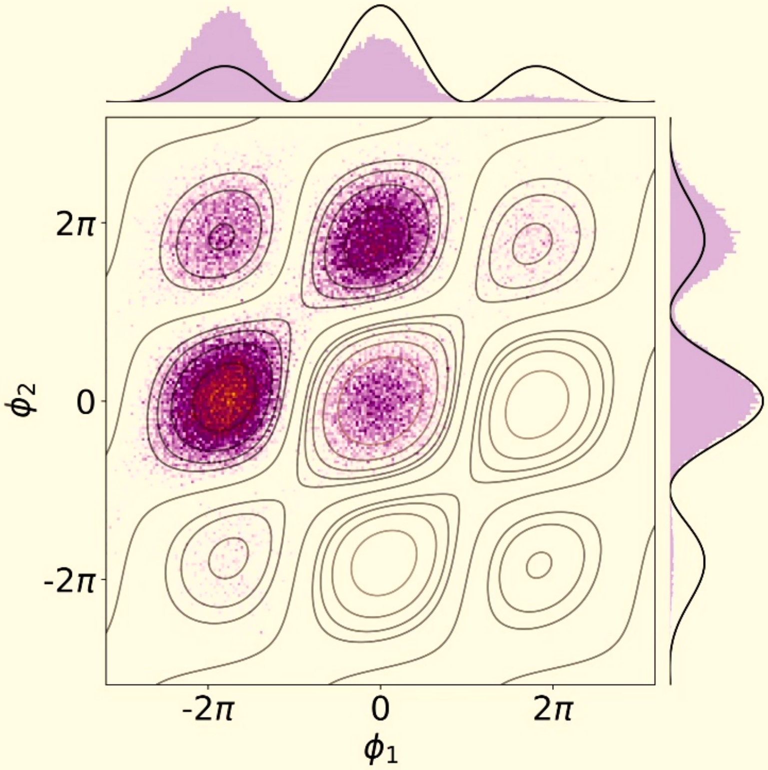
MD step $\Longrightarrow \ \epsilon \approx 0.1$

Computational Bottlenecks

Sampling is **challenging** (similar MCMC drawbacks as before).

Let's consider the Hubbard Model in $(1+1)D$ with: $N_x = 2,\ N_t = 1$



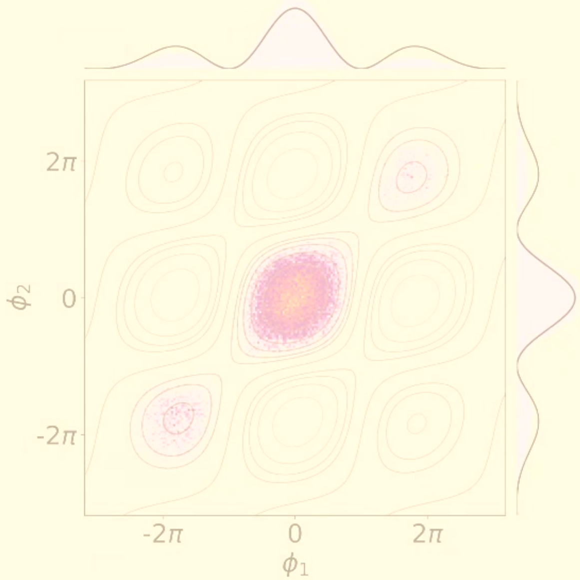


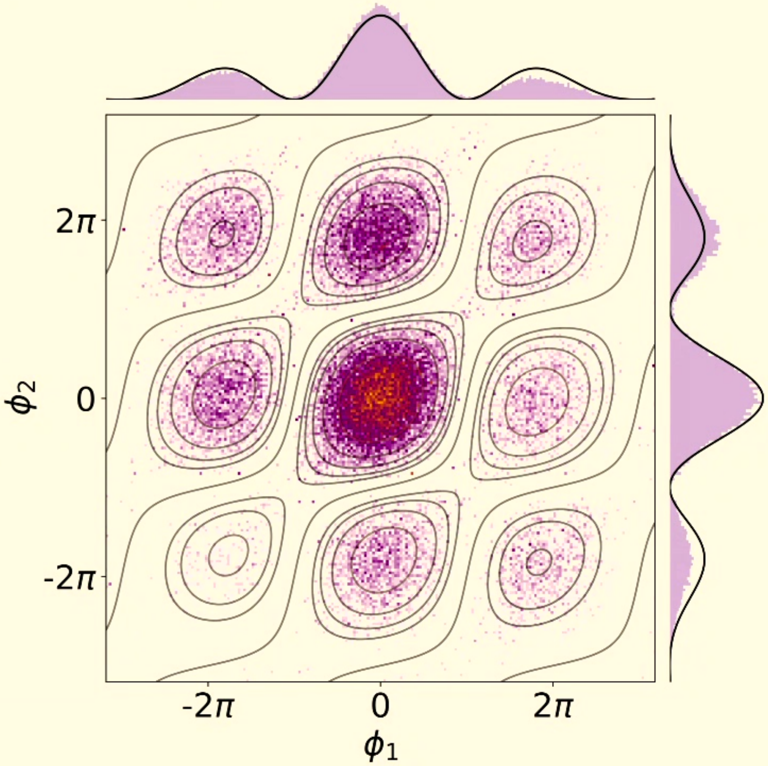
MD step $\Longrightarrow$ $\epsilon \approx 0.5$

Computational Bottlenecks

Sampling is **challenging** (similar MCMC drawbacks as before).

Let's consider the Hubbard Model in $(1+1)D$ with: $N_x = 2,\ N_t = 1$



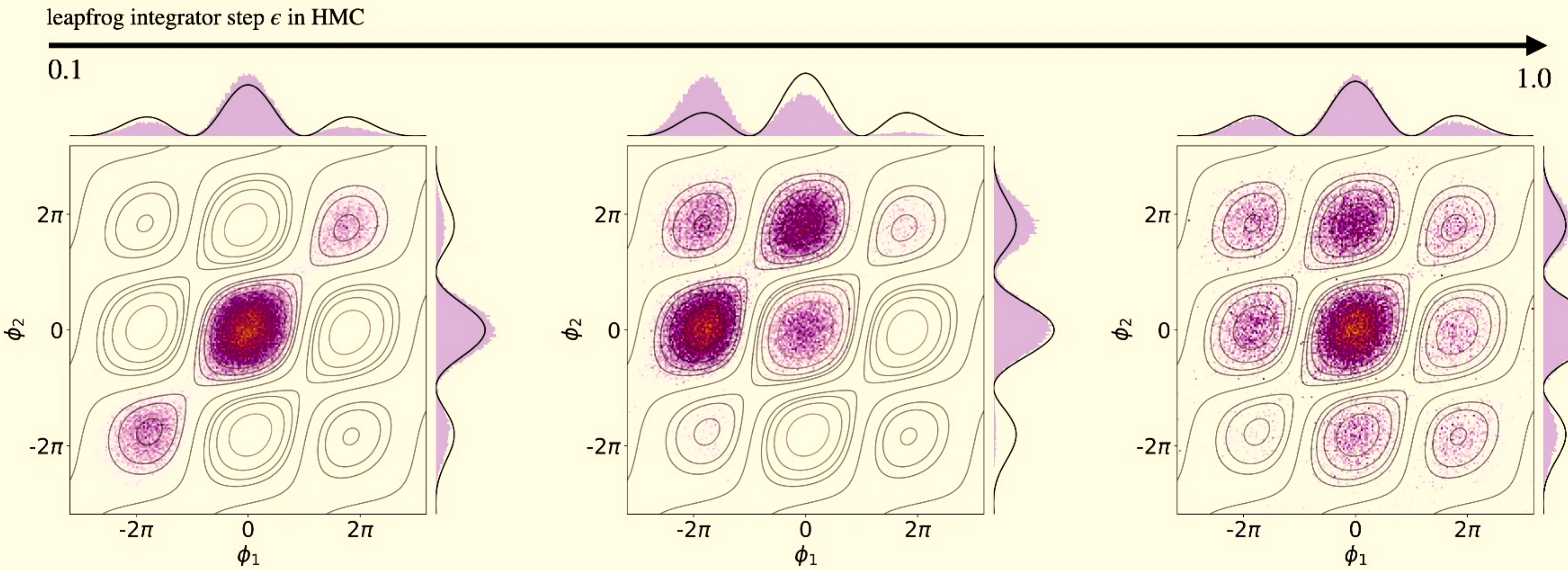


MD step $\Longrightarrow\ \epsilon \approx 1.0$

Computational Bottlenecks

Sampling is **challenging** (similar MCMC drawbacks as before).

Let's consider the Hubbard Model in $(1+1)D$ with: $N_x = 2,\ N_t = 1$



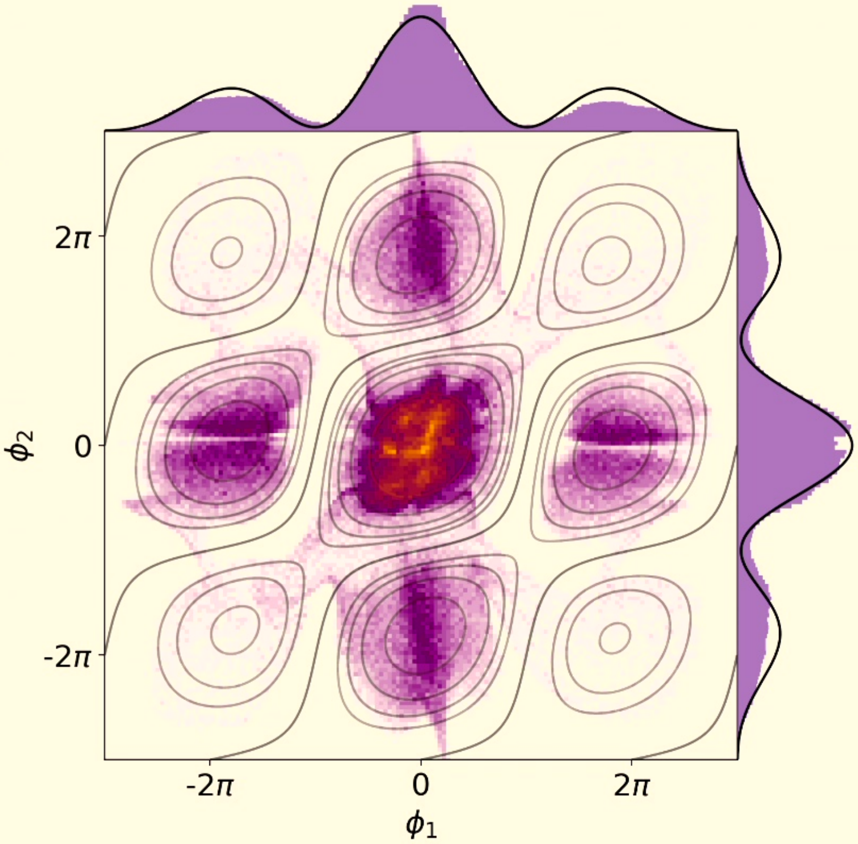
- Acc. Rate: **High** 🥳
- Ergodicity: **Low** 😭

- Acc. Rate: **Low** 😭
- Ergodicity: **High** 🥳

Learning the Hubbard Model with Normalizing Flows

Lattice shape: $N_x = 2,\ N_t = 1$

- Normalizing flow trained with **Reverse KL**
- **Direct** Sampling from the flow is **biased**



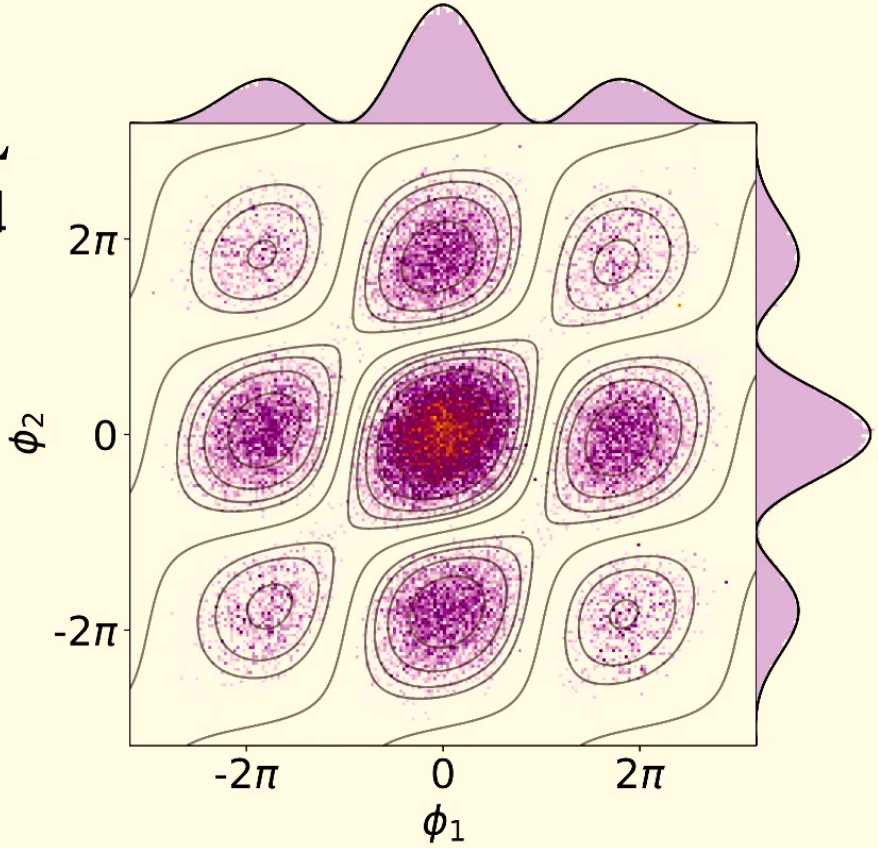
Learning the Hubbard Model with Normalizing Flows

Lattice shape: $N_x = 2,\ N_t = 1$

- Normalizing flow trained with **Reverse KL**
- Bias removed with **NIS** or **NeuralHMC** [1]

- Effective Sampling Size (Flow): 70.1%
- Acceptance Rate (Flow): 74.2%
- Int. Autocor. Time (Flow): $\tau = 1.522 \pm 0.04$ 🤩

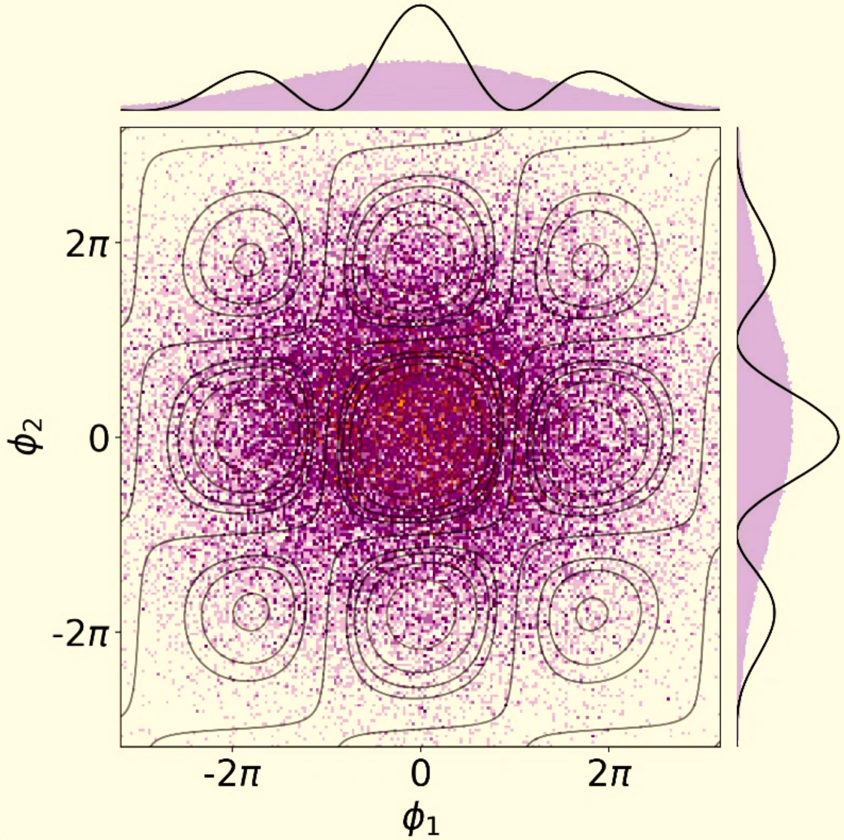
- Int. Autocor. Time (HMC): $\tau_{\text{HMC}} = 443 \pm 136$ 😭



[1] **NKA**, Nakajima, Strodthoff, Samek, Müller, and Kessel, Phys. Rev. E (2020)

Scaling to Larger Lattices

Lattice shape: $N_x = 2,\ N_t = 2$



Our approach starts to fall apart….

Equivariant Normalizing Flows

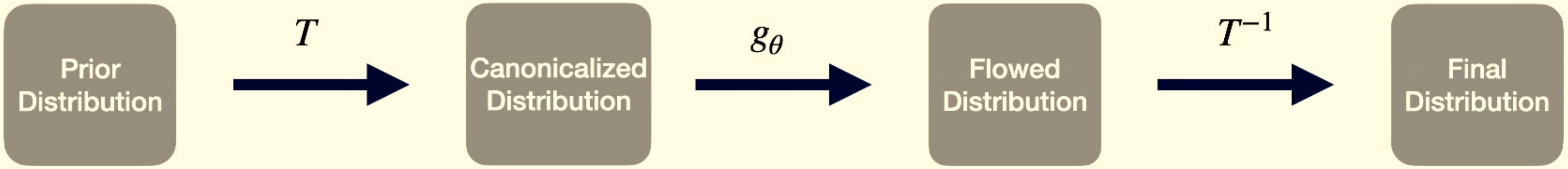
- The Hubbard Model is rich in **symmetries,** which can be leveraged to enhance the training process.
- We need the flow to be **equivariant** with respect to such symmetries.
- The most general way to achieve this is via the so-called **canonicalization**.

Prior Distribution

See also: Köhler et al., ICML (2020), Boyda et al., Phys. Rev D (2021)

Equivariant Normalizing Flows

- The Hubbard Model is rich in **symmetries,** which can be leveraged to enhance the training process.
- We need the flow to be **equivariant** with respect to such symmetries.
- The most general way to achieve this is via the so-called **canonicalization**.



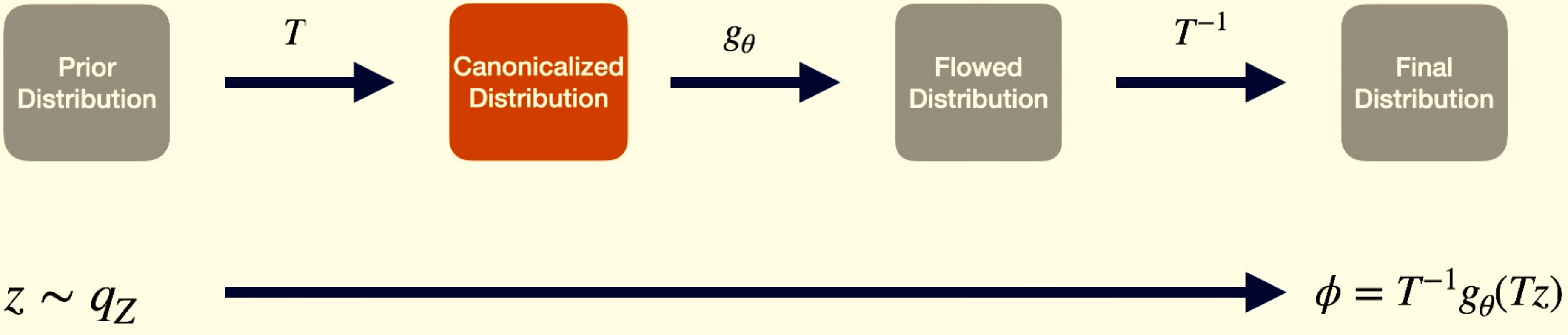
Prior Distribution $\xrightarrow{T}$ Canonicalized Distribution $\xrightarrow{g_\theta}$ Flowed Distribution $\xrightarrow{T^{-1}}$ Final Distribution

$$z \sim q_Z \longrightarrow \phi = T^{-1} g_\theta(Tz)$$

See also: Köhler et al., ICML (2020), Boyda et al., Phys. Rev D (2021)

Equivariant Normalizing Flows

- ➡ The Hubbard Model is rich in **symmetries,** which can be leveraged to enhance the training process.
- ➡ We need the flow to be **equivariant** with respect to such symmetries.
- ➡ The most general way to achieve this is via the so-called **canonicalization**.

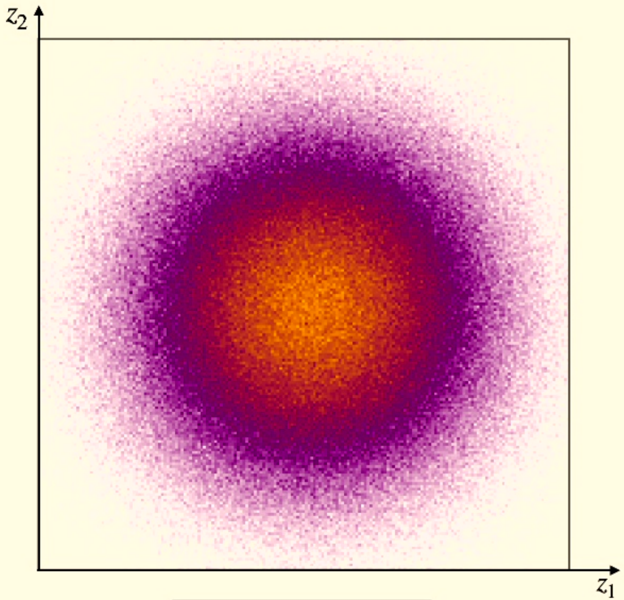


Prior Distribution $\xrightarrow{T}$ **Canonicalized Distribution** $\xrightarrow{g_\theta}$ **Flowed Distribution** $\xrightarrow{T^{-1}}$ **Final Distribution**

$$z \sim q_Z \longrightarrow \phi = T^{-1} g_\theta(Tz)$$

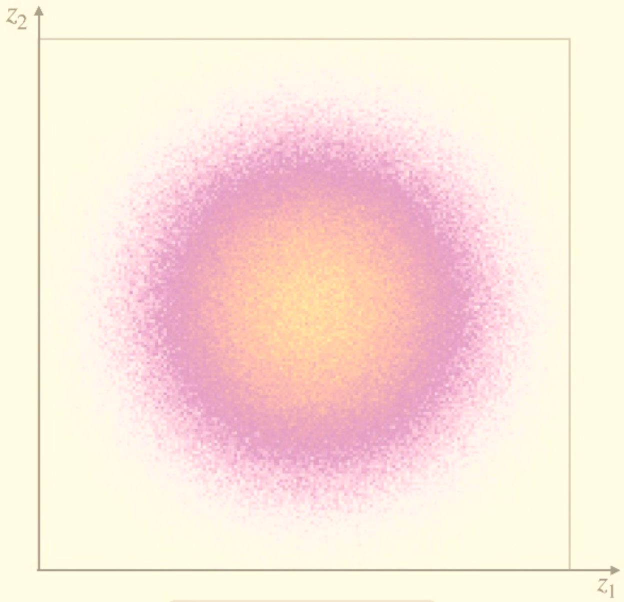
See also: Köhler et al., ICML (2020), Boyda et al., Phys. Rev D (2021)

Canonicalization

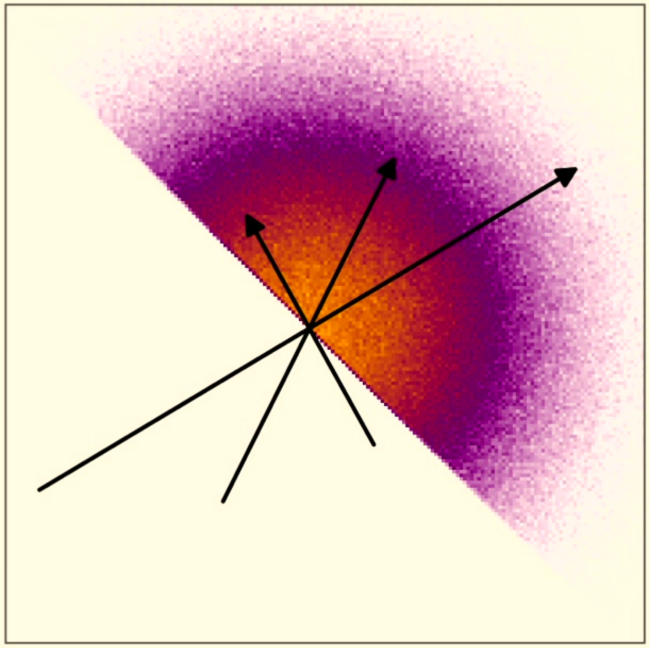


Prior distribution

Canonicalization



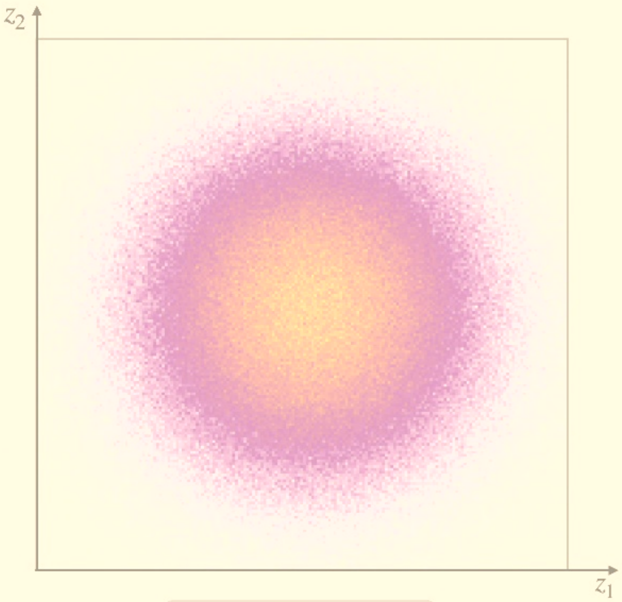
Prior distribution



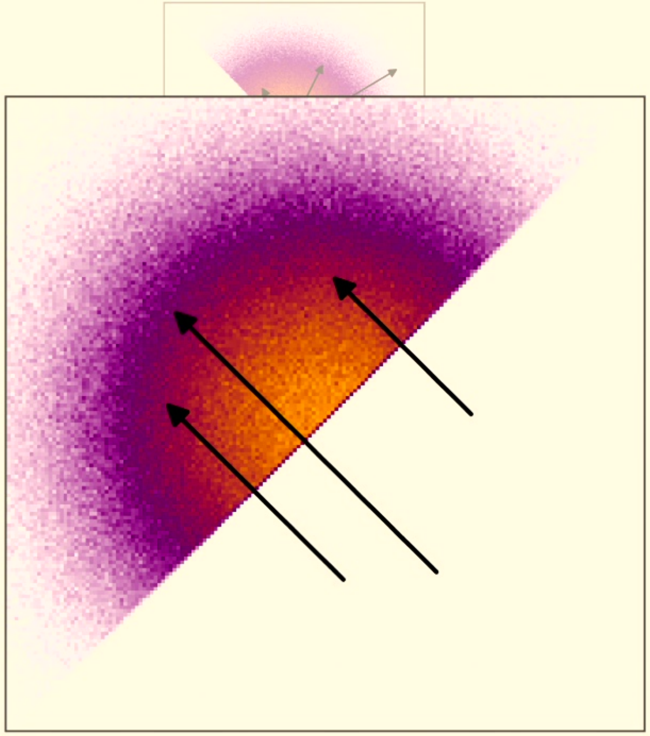
$\mathbb{Z}_2$ Symmetry

$$z \to \begin{cases} z & \text{if } z_1 + z_2 \geq 0 \\ -z & \text{else} \end{cases}$$

Canonicalization



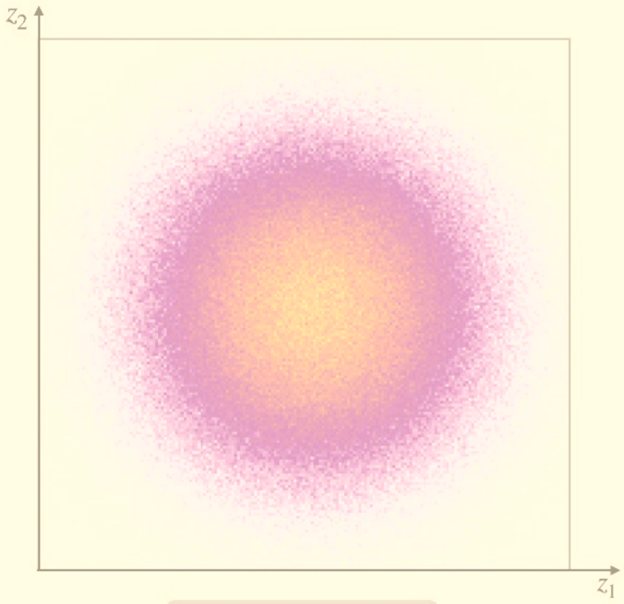
Prior distribution



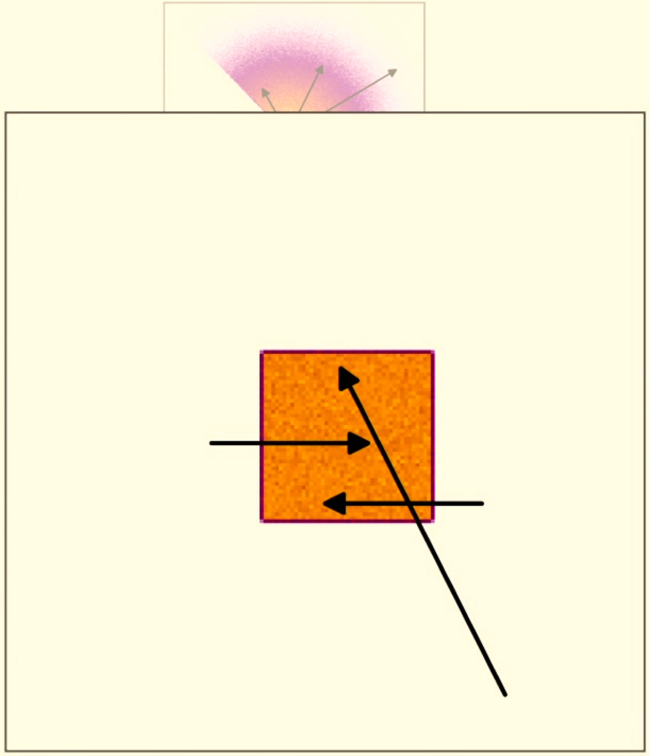
Spacetime Symmetry

$$(z_1, z_2) \to \begin{cases} (z_1, z_2) & \text{if } z_1 - z_2 \leq 0 \\ (z_2, z_1) & \text{else} \end{cases}$$

Canonicalization



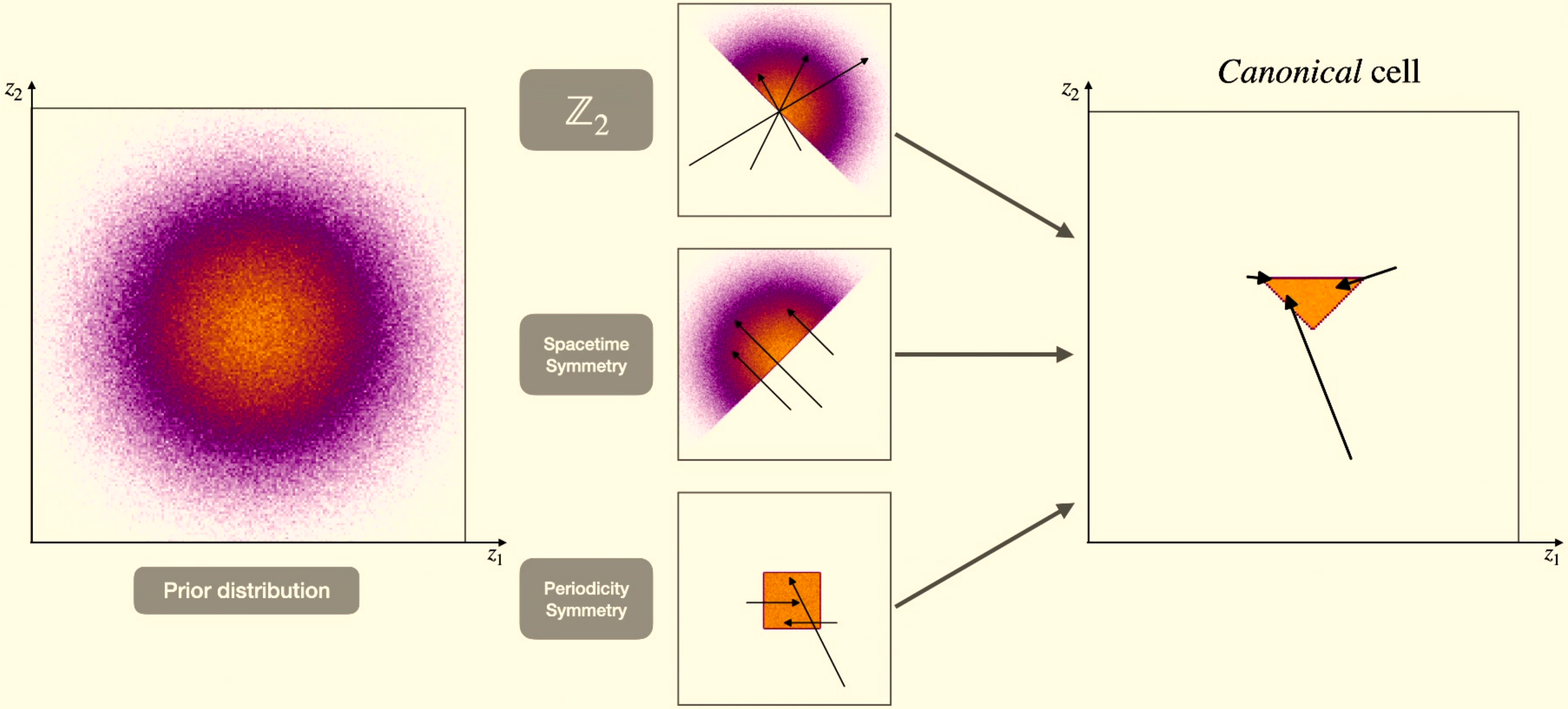
Prior distribution



Periodicity Symmetry

$$z \to z - 2\pi \cdot \mathrm{round}\left(\frac{z}{2\pi}\right)$$

Canonicalization



Prior distribution

$\mathbb{Z}_2$

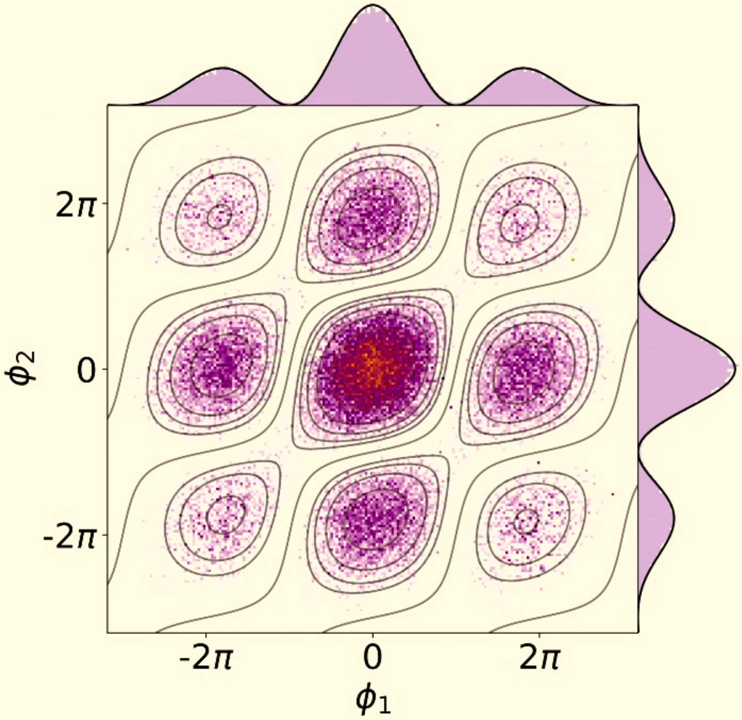
Spacetime Symmetry

Periodicity Symmetry

Canonical cell

Hubbard Equivariant Flow: $N_x = 2,\ N_t = 1$

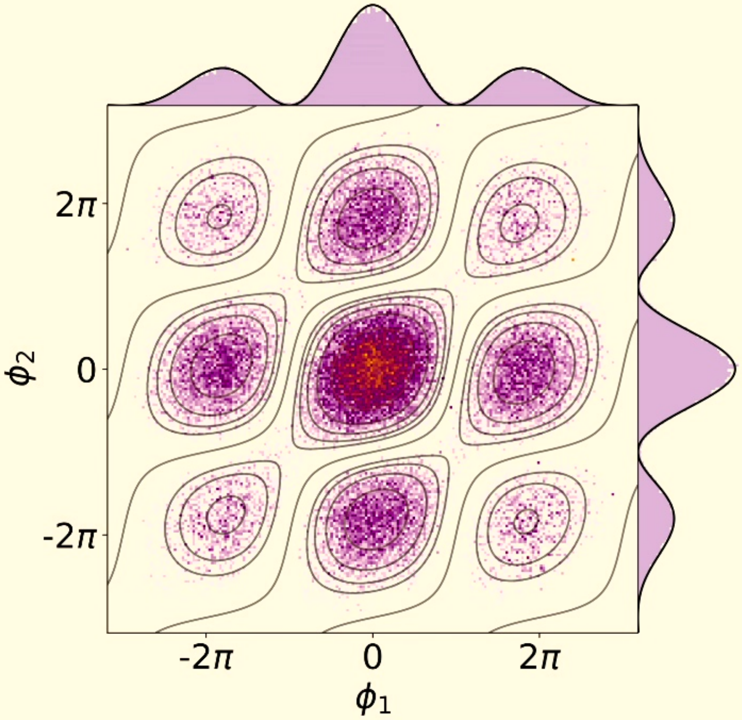
- Acceptance Rate: 74.2%
- $\tau = 1.522 \pm 0.04$



Non-canonicalized Flow

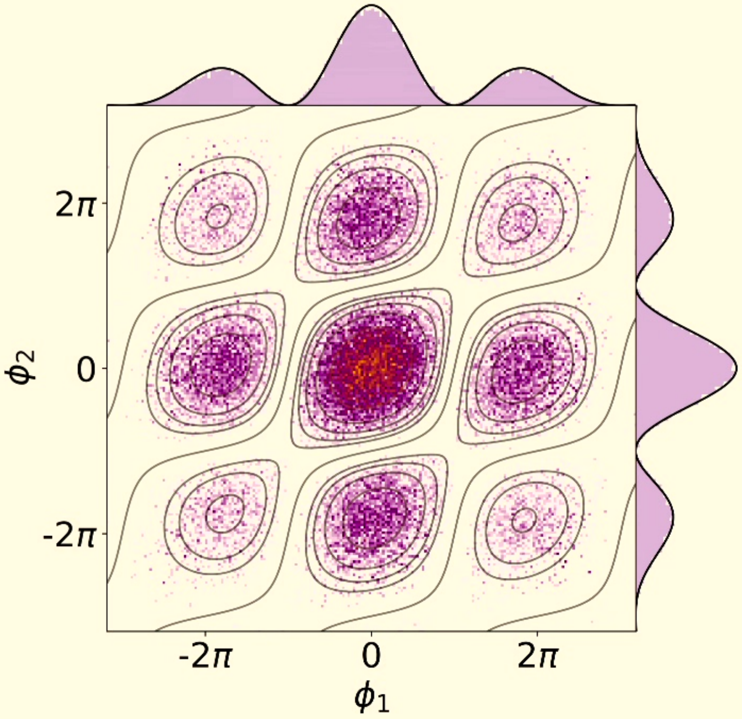
Hubbard Equivariant Flow: $N_x = 2,\ N_t = 1$

- Acceptance Rate: 74.2%
- $\tau = 1.522 \pm 0.04$



Non-canonicalized Flow

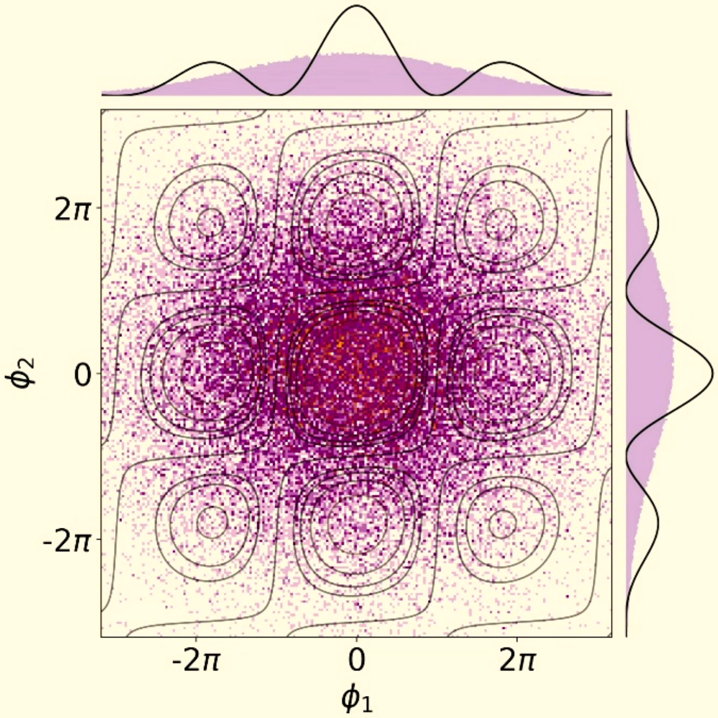
- Acceptance Rate: 84,8% ↑
- $\tau = 0.71 \pm 0.02$ ↓



Canonicalized Flow (Unbiased)

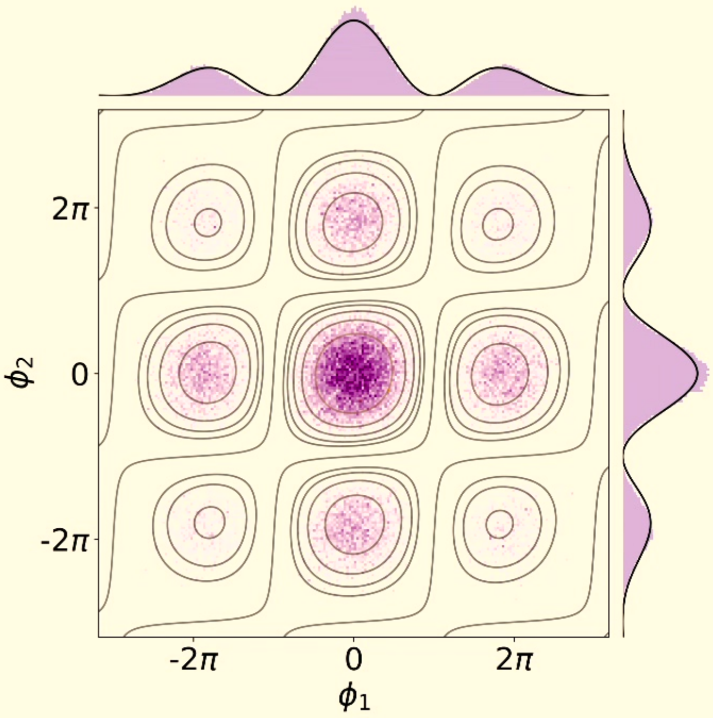
Hubbard Equivariant Flow: $N_x = 2,\ N_t = 2$

- Acceptance Rate: N.A.
- $\tau =$ N.A.



Non-canonicalized Flow

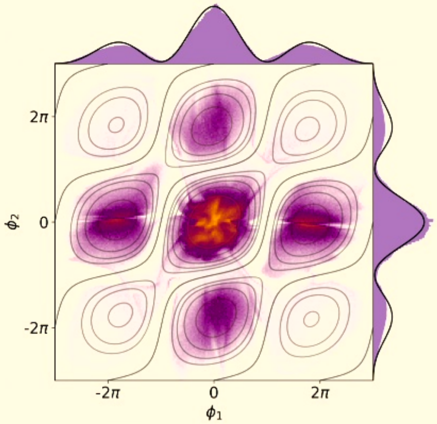
- Acceptance Rate: 69.4%
- $\tau = 1.17 \pm 0.03$



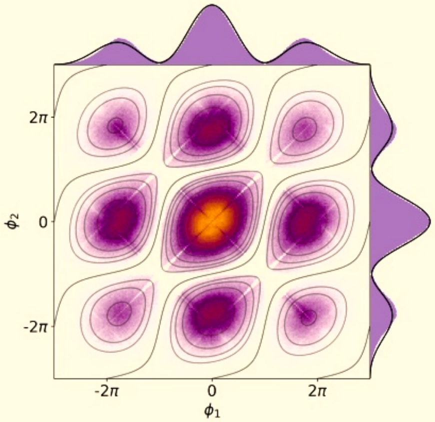
Canonicalized Flow (Unbiased)

Advantages of (Physics Informed) Equivariant Flows

➡ **More efficient training**

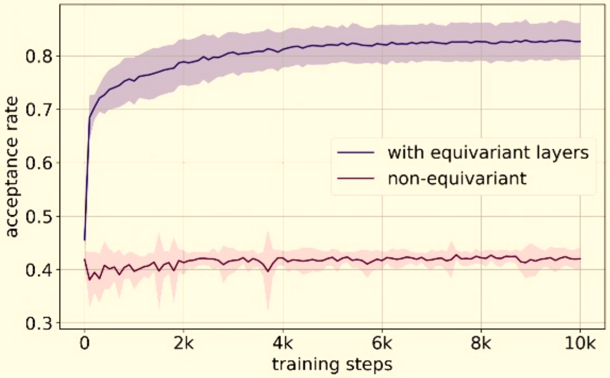


- Non-Equivariant
- Acceptance Rate: 75%
- Training Time: **25h** 🚜



- Equivariant
- Acceptance Rate: 85%
- Training Time: **16min** 🚀

➡ **Higher acceptance rate**



- Comparison:
  - 20 equivariant & 20 non-equivariant models
  - **Mean** and **standard deviation** of **acceptance rate**

- **Non-equivariant**: Comparable AR after 500k training steps
- **Equivariant layers**: Computational overhead less than 10%

Software Development: NeuLat

- **Idea**: Software for flow-based simulation of LFT.
- **Vision**: software is meant to be accessible, **modular**, and easy to **extend** and **maintain**.
- **Goal**: remove the **overhead** of re-implementing existing code between different formats.
- The first **release** of the software is planned for the **upcoming months**.
- NeuLat is aimed to be a **community-wide effort**. Get in touch if you would like to contribute.

Summary and Conclusions

- **Exponential growth** of works in the recent years (HEP, condensed matter etc.).
- **New tools** have been developed yet not fully exploited (e.g., Diffusion models, SNFs).
- Successful applications in **condensed matter physics and beyond** (e.g., Hubbard Model, Entanglement).
- Applications where DGMs **overcome the performance** of standard methods simulations **at scale**.
- Community is growing, and a **reliable**, **established** software will be important for future endeavors.

Summary and Conclusions (Part 2)

TL;DR

Many challenges are yet to be addressed, but many new applications of generative models have proven to be successful.

It's <u>not</u> the end… and there's an exciting future ahead!

Thank You for listening!

"It is nice to know that the computer understands the problem. But I would like to understand it too."
- E. Wigner