DeepMind

Perspectives on the current status of and future prospects for ML in lattice QFT

Sébastien Racanière
sracaniere@google.com

Lattice 21

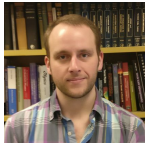
28/07/2021

The team


Michael S. Albergo


Denis Boyda


Kyle Cranmer


Daniel C. Hackett


Gurtej Kanwar


Sébastien Racanière


Danilo J. Rezende


Phiala E. Shanahan


Julian M. Urban

Massachusetts Institute of Technology

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

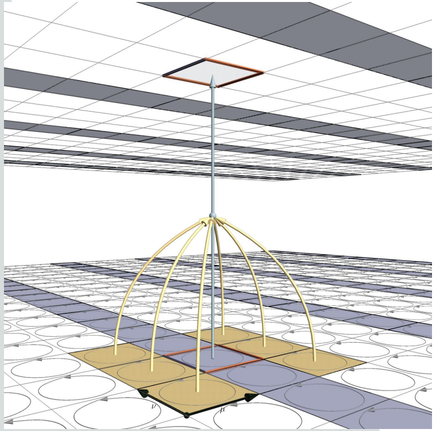
NEW YORK UNIVERSITY

Agenda

1. ML and Physics
2. Some DL successes
3. Case study
4. Conclusion

DeepMind

ML and Physics



Credit: Danilo J. Rezende

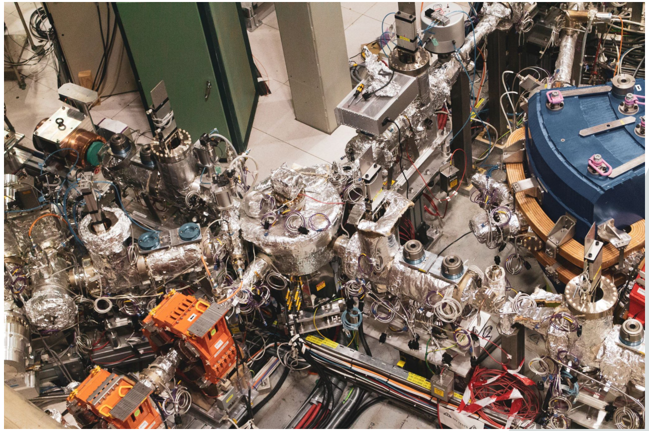
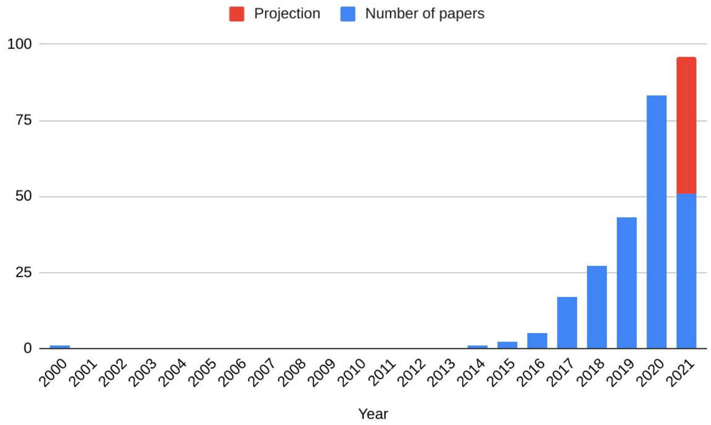


Photo by Devon Rogers on Unsplash

ML and Physics

Long history of ML and anomaly detection in particle accelerators

e.g.: TMVA - Toolkit for Multivariate Data Analysis, arXiv:0703039



HEP Theory papers with ML

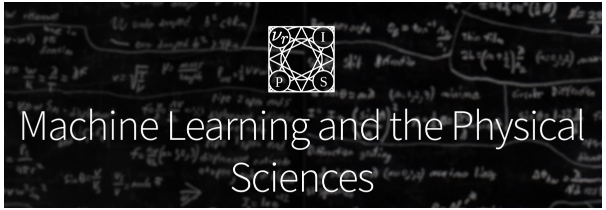
arxiv papers per year with cat: (hep-lat | hep-th | hep-ph) & (cs.LG | stat.ML)

IAIFI iaifi.org/

Physics ∩ ML

a virtual hub at the interface of theoretical physics and deep learning.

www.physicsmeetsml.org/



NeurIPS Workshop https://ml4physicalsciences.github.io/2020/

ML and Lattices (non-exhaustive)

Efficient computations of correlation functions / observables

Yoon, Bhattacharya, Gupta, Phys. Rev. D 100, 014504 (**2019**)

Zhang et al., 1909.10990 (**2019**)

Karpie et al., 1901.05408 (**2019**)

Matsumoto et al., 1909.06238 (**2019**)

Zhang et al, Phys. Rev. D 101, 034516 (**2020**)

Nicoli et al., 2007.07115 (**2020**)

Sign-problem avoidance via contour deformation of path integrals

Mori et al., 1705.05605 (**2017**)

Alexandru et al., 1709.01971 (**2017**)

Alexandru et al., Phys Rev. Lett. 121 (**2020**)

Detmold et al. 2003.05914 (**2020**)

Alexandru et al., 2007.05436 (**2020**)

Detmold, Kanwar et al., 2101.12668 (**2021**)

Lawrence et al., 2101.05755 (**2021**)

Analysis, order parameters, insights

Tanaka and Tomiya, Journal of the Physical Society of Japan, 86 (**2017**)

Wetzel and Scherzer, Phys. Rev. B 96 (**2017**)

Li et al., 1703.02369 (**2017**)

Shanahan et al., 1801.05784 (**2018**)

S. Bluecher et al., Phys. Rev. D 101 (**2020**)

Boyda et al., 2009.10971 (**2020**)

Chernodub et al., 2006.09113 (**2020**)

Bachtis et al., 2004.14341 (**2020**)

Field configuration generation

Tanaka and Tomiya, 1712.03893 (**2017**)

Zhou et al., Phys. Rev. D 100 (**2019**)

Li et al., PRX 10 (**2020**)

Pawlowski and Urban, 1811.03533 (**2020**)

Nagai, Tanaka, Tomiya 2010.11900 (**2020**)

Luo, Clarkes, Stokes, 2012.05232 (**2020**)

Favoni et al., 2012.12901 (**2020**)

Medvidovic et al., 2012.01442(**2020**)

Luo et al., 2101.07243 (**2021**)

Del Debbio et al., 2105.12481 (**2021**)

Thanks to Phiala Shanahan and Gurtej Kanwar for providing the above references.

ML and Physics at Lattice 21 - Part I

Akio Tomiya (Tue 13:00): Smearing is a neural network

Neill Warrington (Tue 13:15): Contour Deformations for Lattice Field Theory

Gurtej Kanwar (Tue 13:30): Observifolds: Taming the observable signal-to-noise problem via path integral contour deformations

Shuzhe Shi (Tue 13:45): From lattice QCD to heavy-flavor in-medium potential via deep learning

Sunkyu Lee (Tue 22:45): Deep learning study on the Dirac eigenvalue spectrum of staggered quarks

Yukari Yamauchi (Wed 21:00): Normalizing flows for the real-time sign problem

Chen ShiYang (Wed 22:00): Machine learning Hadron Spectral Functions in Lattice QCD

Fu-Jiun Jiang (Wed 22:15): A universal neural network for learning phases and criticalities

ML and Physics at Lattice 21 - Part II

Dimitrios Bachtis (Thur 05:00): Machine learning with quantum field theories

David Muller (Thur 05:15): Lattice Gauge Symmetry in Neural Networks

Matteo Favoni (Thur 05:30): Generalization capabilities of neural networks in lattice applications

Gert Aarts (Thur 05:45): Interpreting machine learning functions as physical observables

Kim Nicoli (Thur 06:00): Machine Learning for Thermodynamic Observables

Marina Marinkovic (Thur 06:15): Machine learning phase transitions in a scalable manner

Michael Albergo (Thur 13:30): Flow-based sampling for fermionic field theories

Xiao-Yong Jin (Thur 13:45): Neural Network Field Transformation and Its Application in HMC

Denis Boyda (Thur 14:00): Sampling lattice gauge theory in four dimensions with normalizing flows

Sam Foreman (Thur 14:45): LeapFrogLayers: A Trainable Framework for Effective Topological Sampling

Boram Yoon (Thur 22:15): Prediction and compression of lattice QCD data using machine learning algorithms on quantum annealer

ML and Physics

Designing self-assembling kinetics with differentiable statistical physics models

Carl P. Goodrich[a,b,1,2], Ella M. King[c,1], Samuel S. Schoenholz[d], Ekin D. Cubuk[d], and Michael P. Brenner[a,c,d]

[a]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; [b]Institute of Science and Technology Austria, A-3400 Klosterneuburg, Austria; [c]Physics Department, Harvard University, Cambridge, MA 02138; and [d]Brain Team, Google Research, Mountain View, CA 94043

Edited by Steve Granick, Institute for Basic Science, Ulju-gun, Ulsan, South Korea, and approved January 25, 2021 (received for review November 20, 2020)

The inverse problem of designing component interactions to target emergent structure is fundamental to numerous applications in biotechnology, materials science, and statistical physics. Equally important is the inverse problem of designing emergent kinetics, but this has received considerably less attention. Using recent advances in automatic differentiation, we show how kinetic pathways can be precisely designed by directly differentiating through statistical physics models, namely free energy calculations and molecular dynamics simulations. We consider two systems that are crucial to our understanding of structural self-assembly: bulk crystallization and small nanoclusters. In each case, we are able to assemble precise dynamical features. Using gradient information, we manipulate interactions among constituent particles to tune the rate at which these systems yield specific structures of interest. Moreover, we use this approach to learn nontrivial features about the high-dimensional design space, allowing us to accurately predict when multiple kinetic features can be simultaneously and independently controlled. These results provide a concrete and generalizable foundation for studying nonstructural self-assembly, including kinetic properties as well as other complex emergent properties, in a vast array of systems.

for solving nonlinear partial differential equations (19), the discovery of molecules for drug development (20), and greatly improved predictions of protein structure (21).

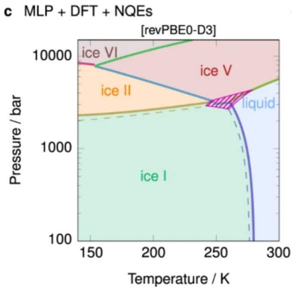
Here, we begin to explore a class of complex inverse design problems that traditionally has been hard to access. Using AD to train well-established statistical physics-based models, we design materials for dynamic, rather than structural, features. We also use AD to gain theoretical insights into this design space, allowing us to predict the extent of designability of different properties. AD is an essential component of this approach (22) because we rely on gradients to connect physical parameters to complex emergent behavior. While there are other approaches for obtaining gradient information (e.g., finite difference approximations), AD calculates exact derivatives and more importantly, can efficiently handle large numbers of parameters. Furthermore, the theoretical insights we develop rely on accurate calculations of the Hessian matrix of second derivatives, for which finite difference approaches are insufficient.

We start in *Tuning Assembly Rates of Honeycomb Crystals* by considering the bulk crystallization of identical particles into both honeycomb and triangular lattices. By differentiating over entire molecular dynamics (MD) trajectories with respect to

Quantum-mechanical exploration of the phase diagram of water

Reinhardt, A., Cheng, B. Quantum-mechanical exploration of the phase diagram of water. Nat Commun 12, 588 (2021). https://doi.org/10.1038/s41467-020-20821-w

Abstract: The set of known stable phases of water may not be complete, and some of the phase boundaries between them are fuzzy. Starting from liquid water and a comprehensive set of 50 ice structures, we compute the phase diagram at three hybrid density-functional-theory levels of approximation, accounting for thermal and nuclear fluctuations as well as proton disorder. **Such calculations are only made tractable because we combine machine-learning methods and advanced free-energy techniques.** The computed phase diagram is in qualitative agreement with experiment, particularly at pressures ≲ 8000 bar, and the discrepancy in chemical potential is comparable with the subtle uncertainties introduced by proton disorder and the spread between the three hybrid functionals. None of the hypothetical ice phases considered is thermodynamically stable in our calculations, suggesting the completeness of the experimental water phase diagram in the region considered. Our work demonstrates the feasibility of predicting the phase diagram of a polymorphic system from first principles and provides a thermodynamic way of testing the limits of quantum-mechanical calculations.



ML and Physics

Topological Obstructions to Autoencoding

Joshua Batson, C. Grace Haaf, Yonatan Kahn,
Daniel A. Roberts, arXiv:2102.08380

ABSTRACT: Autoencoders have been proposed as a powerful tool for model-independent anomaly detection in high-energy physics. The operating principle is that events which do not belong to the space of training data will be reconstructed poorly, thus flagging them as anomalies. We point out that in a variety of examples of interest, the connection between large reconstruction error and anomalies is not so clear. In particular, for data sets with nontrivial topology, there will always be points that erroneously seem anomalous due to global issues. Conversely, neural networks typically have an inductive bias or prior to locally interpolate such that undersampled or rare events may be reconstructed with small error, despite actually being the desired anomalies. Taken together, these facts are in tension with the simple picture of the autoencoder as an anomaly detector. Using a series of illustrative low-dimensional examples, we show explicitly how the intrinsic and extrinsic topology of the dataset affects the behavior of an autoencoder and how this topology is manifested in the latent space representation during training. We ground this analysis in the discussion of a mock "bump hunt" in which the autoencoder fails to identify an anomalous "signal" for reasons tied to the intrinsic topology of n-particle phase space.

E(n) Equivariant Normalizing Flows

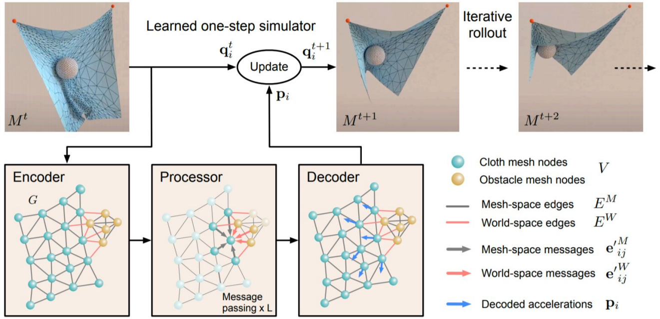
Victor Garcia Satorras[1*], Emiel Hoogeboom[1*], Fabian B. Fuchs[2], Ingmar Posner[2], Max Welling[1]

Abstract

This paper introduces a generative model equivariant to Euclidean symmetries: E(n) Equivariant Normalizing Flows (E-NFs). To construct E-NFs, we take the discriminative E(n) graph neural networks and integrate them as a differential equation to obtain an invertible equivariant function: a continuous-time normalizing flow. We demonstrate that E-NFs considerably outperform baselines and existing methods from the literature on particle systems such as DW4 and LJ13, and on molecules from QM9 in terms of log-likelihood. To the best of our knowledge, this is the first flow that jointly generates molecule features and positions in 3D.

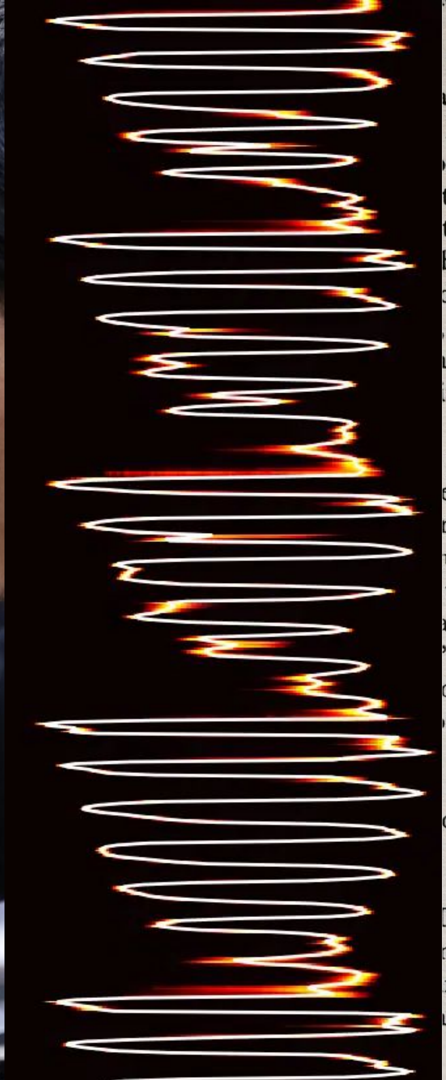
Learning Mesh-Based Simulation with Graph Networks

Tobias Pfaff*, Meire Fortunato*, Alvaro Sanchez-Gonzalez*, Peter W. Battaglia



DeepMind

Some DL successes



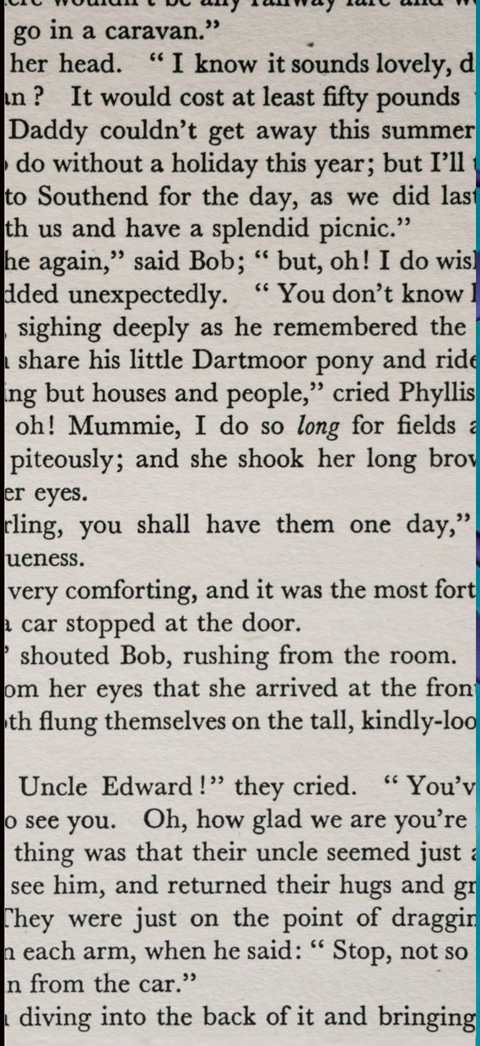


Photo by Annie Spratt on Unsplash

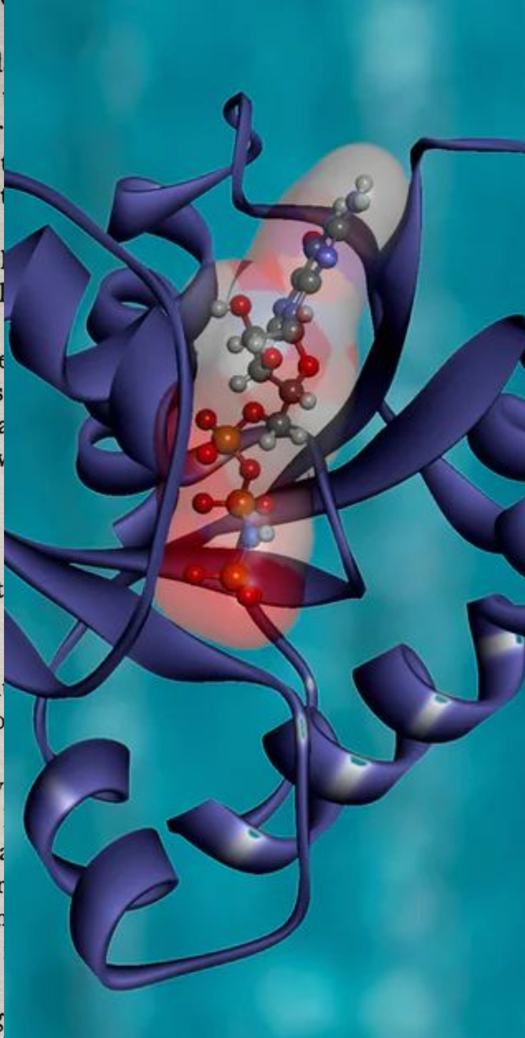
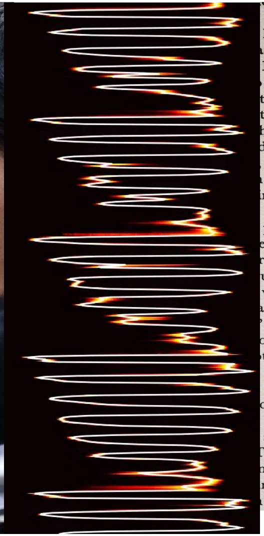
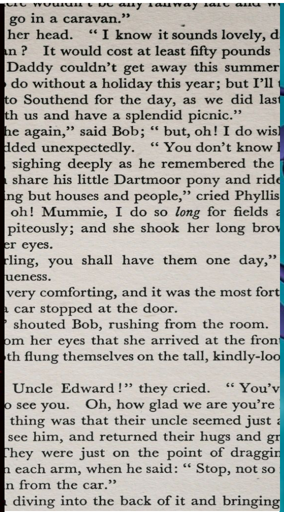


Photo by National Cancer Institute on Unsplash

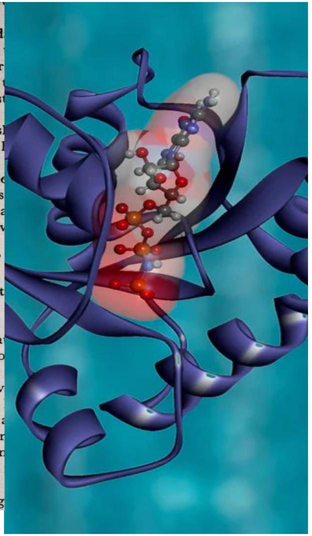
Characteristics of DL successes

- Plenty of data
- Well defined objective
- Huge compute power
- End-to-end optimization
- Inductive bias (i.e. assumptions built into the model)

go in a caravan."
her head. "I know it sounds lovely, d
n? It would cost at least fifty pounds
Daddy couldn't get away this summer
do without a holiday this year; but I'll
to Southend for the day, as we did last
th us and have a splendid picnic."
he again," said Bob; "but, oh! I do wis
dded unexpectedly. "You don't know
sighing deeply as he remembered the
share his little Dartmoor pony and ride
ng but houses and people," cried Phyllis
oh! Mummie, I do so *long* for fields
piteously; and she shook her long bro
er eyes.
rling, you shall have them one day,"
ueness.
very comforting, and it was the most fort
car stopped at the door.
shouted Bob, rushing from the room.
om her eyes that she arrived at the fron
th flung themselves on the tall, kindly-loo

Uncle Edward!" they cried. "You'v
o see you. Oh, how glad we are you're
thing was that their uncle seemed just
see him, and returned their hugs and gr
They were just on the point of draggin
each arm, when he said: "Stop, not so
n from the car."
diving into the back of it and bringing

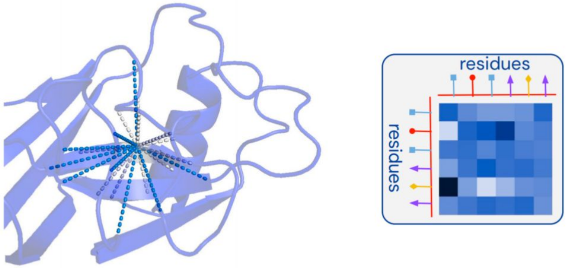


Characteristics of DL successes

- Plenty of data
- Well defined objective
- Huge compute power
- End-to-end optimization
- Inductive bias (i.e. assumptions built into the model)

Putting our protein knowledge into the model

- Physical insights are built into the network structure, not just a process around it
- End-to-end system directly producing a structure instead of inter-residue distances
- Inductive biases reflect our knowledge of protein physics and geometry
  - The positions of residues in the sequence are de-emphasized
  - Instead residues that are close in the folded protein need to communicate
  - The network iteratively learns a graph of which residues are close, while reasoning over this implicit graph as it is being built



DeepMind, CASP14 presentation

Mainstream DL tools

- Data type: 2D images, 3D videos, text, sound
- Symmetries: translations, permutations, SE(3)

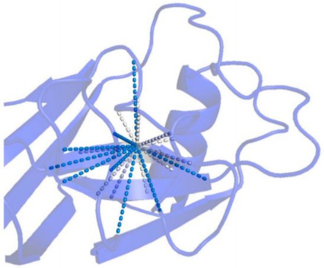
Cifar100

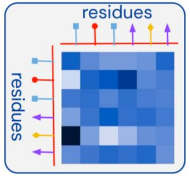
Characteristics of DL successes

- Plenty of data
- Well defined objective
- Huge compute power
- End-to-end optimization
- Inductive bias (i.e. assumptions built into the model)

Putting our ~~protein~~ lattice knowledge into the model

- Physical insights are built into the network structure, not just a process around it
- End-to-end system directly producing a structure instead of inter-residue distances
- Inductive biases reflect our knowledge of protein physics and geometry
  - The positions of residues in the sequence are de-emphasized
  - Instead residues that are close in the folded protein need to communicate
  - The network iteratively learns a graph of which residues are close, while reasoning over this implicit graph as it is being built





DeepMind, CASP14 presentation

DeepMind

Case study (a.k.a. stuff I've worked on)

Paper trail

- Machine learning action parameters in lattice quantum chromodynamics, arXiv:1801.05784, Phiala E. Shanahan, Amalie Trewartha, William Detmold
- Flow-based generative models for Markov chain Monte Carlo in lattice field theory, arXiv:1904.12072, M. S. Albergo, G. Kanwar, P. E. Shanahan

---

- Equivariant flow-based sampling for lattice gauge theory, arXiv:2003.06413, Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, Phiala E. Shanahan
- Sampling using SU(N) gauge equivariant flows, arXiv:2008.05456, Denis Boyda, Gurtej Kanwar, Sébastien Racanière, Danilo Jimenez Rezende, Michael S. Albergo, Kyle Cranmer, Daniel C. Hackett, Phiala E. Shanahan
- Introduction to Normalizing Flows for Lattice Field Theory, arXiv:2101.08176, Michael S. Albergo, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Kyle Cranmer, Sébastien Racanière, Danilo Jimenez Rezende, Phiala E. Shanahan
- Flow-based sampling for fermionic lattice field theories, arXiv:2106.05934, Michael S. Albergo, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Julian M. Urban, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Phiala E. Shanahan

---

- Flow-based sampling for multimodal distributions in lattice field theory, arXiv:2107.00734, Daniel C. Hackett, Chung-Chun Hsieh, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, Phiala E. Shanahan

Paper trail

- Machine learning action parameters in lattice quantum chromodynamics, arXiv:1801.05784, Phiala E. Shanahan, Amalie Trewartha, William Detmold
- Flow-based generative models for Markov chain Monte Carlo in lattice field theory, arXiv:1904.12072, M. S. Albergo, G. Kanwar, P. E. Shanahan

---

- Equivariant flow-based sampling for lattice gauge theory, arXiv:2003.06413, Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, Phiala E. Shanahan
- Sampling using SU(N) gauge equivariant flows, arXiv:2008.05456, Denis Boyda, Gurtej Kanwar, Sébastien Racanière, Danilo Jimenez Rezende, Michael S. Albergo, Kyle Cranmer, Daniel C. Hackett, Phiala E. Shanahan
- Introduction to Normalizing Flows for Lattice Field Theory, arXiv:2101.08176, Michael S. Albergo, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Kyle Cranmer, Sébastien Racanière, Danilo Jimenez Rezende, Phiala E. Shanahan
- Flow-based sampling for fermionic lattice field theories, arXiv:2106.05934, Michael S. Albergo, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Julian M. Urban, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Phiala E. Shanahan

---

- Flow-based sampling for multimodal distributions in lattice field theory, arXiv:2107.00734, Daniel C. Hackett, Chung-Chun Hsieh, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, Phiala E. Shanahan

Crash course on (normalizing) flows

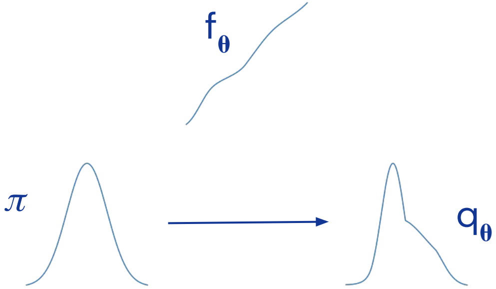
The problem[1]:

Learn a target distribution from a simple base distribution and a learnt transformation.

- Manifold M with a simple distribution π
- Target density p on M (for ex: $p = e^{-S}$)
- Smooth family of invertible maps $f_\theta: M \to M$
- Define q_θ density of random variable $x = f_\theta(z)$, $z \sim \pi$

  $$q_\theta(x) = \pi(z)|\det \partial f_\theta/\partial z|^{-1}$$

- Learn $q_\theta = \mathrm{argmin}_\theta\, KL(q_\theta || p) = \mathrm{argmin}_\theta\, E_{x\sim q\theta} \log q_\theta(x) / \log p(x)$



[1] Think "learnt trivialising map". For trivialising maps, see for example "Trivializing maps, the Wilson flow and the HMC algorithm", Martin Lüscher, arxiv:0907.5491

Flows and symmetry invariance

Assume a group G acts on M

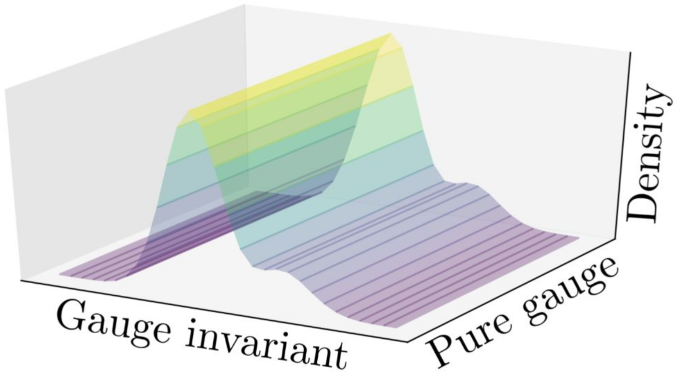
g in G, x in M $\rightarrow g \cdot x$

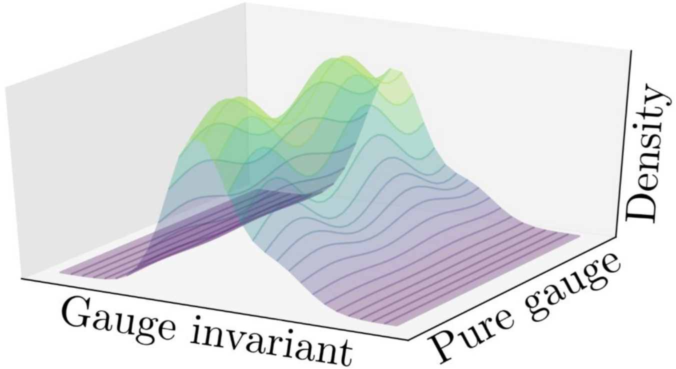
Assume the base density is invariant with respect to G.

Assume f_θ is equivariant: $f_\theta(g \cdot z) = g \cdot f_\theta(z)$

Then q_θ is invariant with respect to G

$q_\theta(g \cdot x) = q_\theta(x)$





Flows for lattices

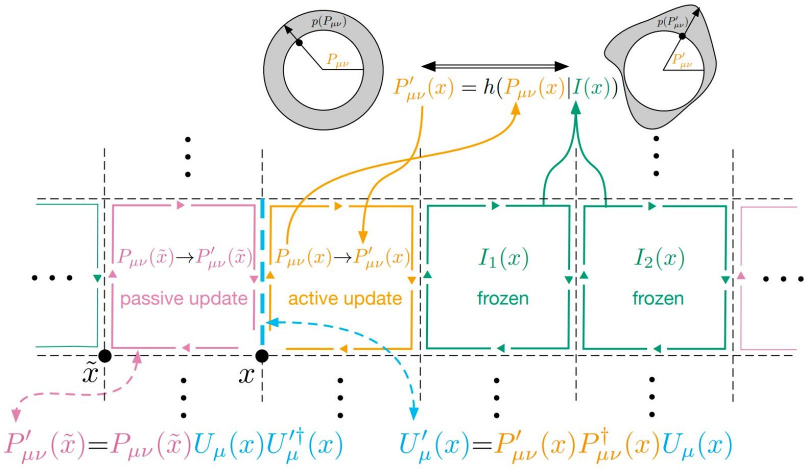
Generate configurations for Lattice QFT

Provide density of sampled configurations

Provably-exact algorithms

Equivariant flow-based sampling for lattice gauge theory

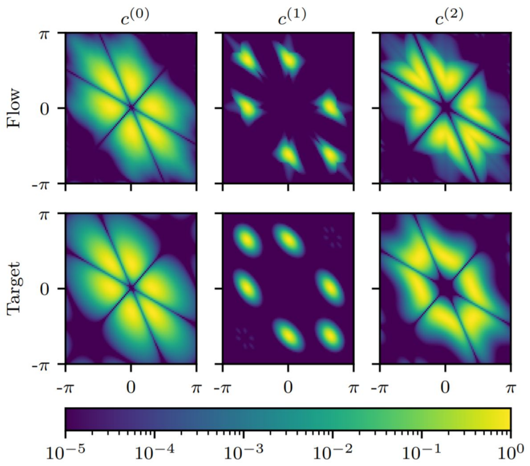
- (1+1)d pure-gauge U(1)
- Gauge-equivariant flows
  - Build flow on untraced plaquettes
  - Gauge equivariance ← Conjugation equivariance

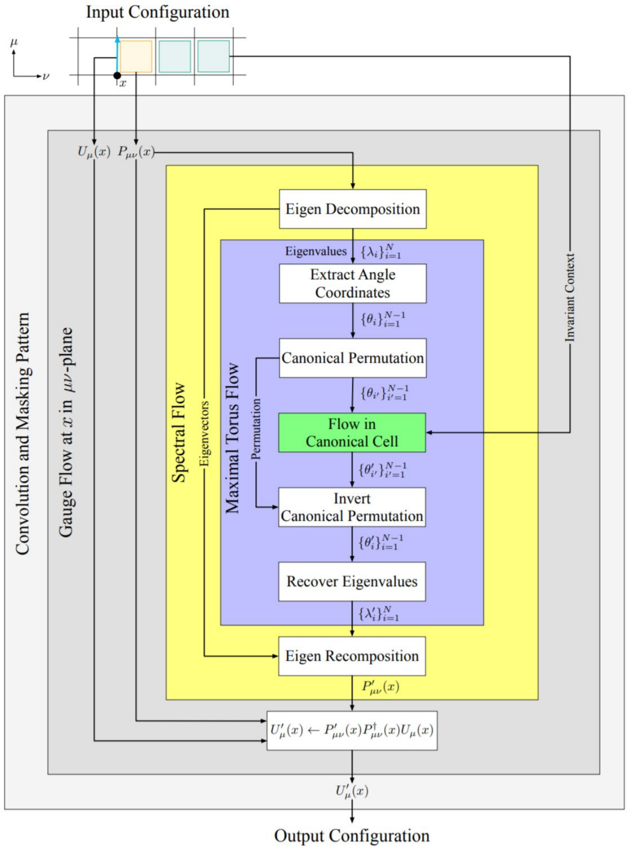


See also *Sampling lattice gauge theory in four dimensions with normalizing flows*, Denis Boyda 29 July, 14.00

Sampling using SU(N) gauge equivariant flows

- (1+1)d pure-gauge SU(3)
- Conjugation equivariant flows on SU(N)
  - Build flow by transforming eigenvalues





Flow-based sampling for fermionic lattice field theories

- (1+1)d Yukawa
- Flows that respect symmetries of pseudo-fermions: periodic & anti-periodic boundary conditions

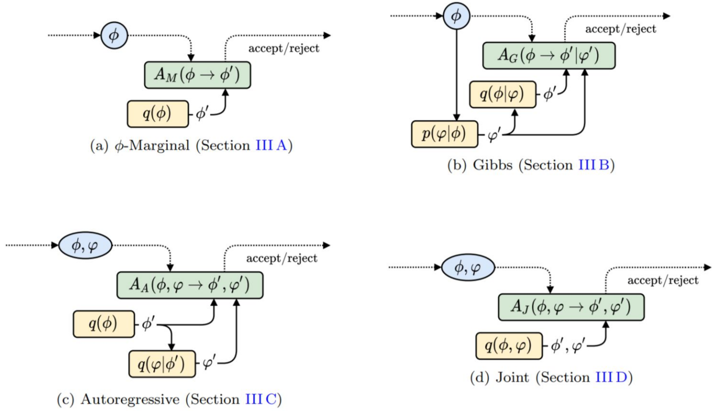


FIG. 1. Diagrams illustrating the four types of sampling schemes described in Section III. Blue circles/ellipses depict the current state of the Markov chain. Yellow boxes depict exactly sampleable densities either produced from generative models or by Equation (16). Green boxes correspond to Metropolis accept/reject steps using the acceptance probabilities defined in the text. Dotted lines indicate the Markov chain, whereas solid lines correspond to the internal operations of each Markov chain step.

Challenges

- (3+1)d
- SU(3) with fermions
- Scalability
    - Training with large lattices
    - What models: masked flows, linear flows, convex potential flows...

Related work

Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks

Jan M. Pawlowski and Julian M. Urban
arXiv:1811.03533

Short autocorrelation times are essential for a reliable error assessment in Monte Carlo simulations of lattice systems. In many interesting scenarios, the decay of autocorrelations in the Markov chain is prohibitively slow. Generative samplers can provide statistically independent field configurations, thereby potentially ameliorating these issues. In this work, the applicability of neural samplers to this problem is investigated. Specifically, we work with a generative adversarial network (GAN).

Deep Learning Hamiltonian Monte Carlo

Sam Foreman, Xiao-Yong Jin & James C. Osborn
arXiv:2105.03418

We generalize the Hamiltonian Monte Carlo algorithm with a stack of neural network layers, and evaluate its ability to sample from different topologies in a two dimensional lattice gauge theory. We demonstrate that our model is able to successfully mix between modes of different topologies, significantly reducing the computational cost required to generate independent gauge field configurations. Our implementation is available at https://github.com/saforem2/l2hmc-qcd.

Efficient Modelling of Trivializing Maps for Lattice ϕ^4 Theory Using Normalizing Flows: A First Look at Scalability

Luigi Del Debbio, Joe Marsh Rossney, Michael Wilson
arXiv:2105.12481

General-purpose MCMC sampling algorithms suffer from a dramatic reduction in efficiency as the system being studied is driven towards a critical point. Recently, a series of seminal studies suggested that normalizing flows -- a class of deep generative models -- can form the basis of a sampling strategy that does not suffer from this 'critical slowing down'. [...] We pick up this thread, with the aim of quantifying how well we can expect this approach to scale as we increase the number of degrees of freedom in the system.

Conclusion

Conclusion

- ML + Physics: Fast growing field
- Lots of powerful tools in DL
- Need to be adapted to problem: not black-box
- Symmetries are our friends
- Fruitful cross-disciplines interactions