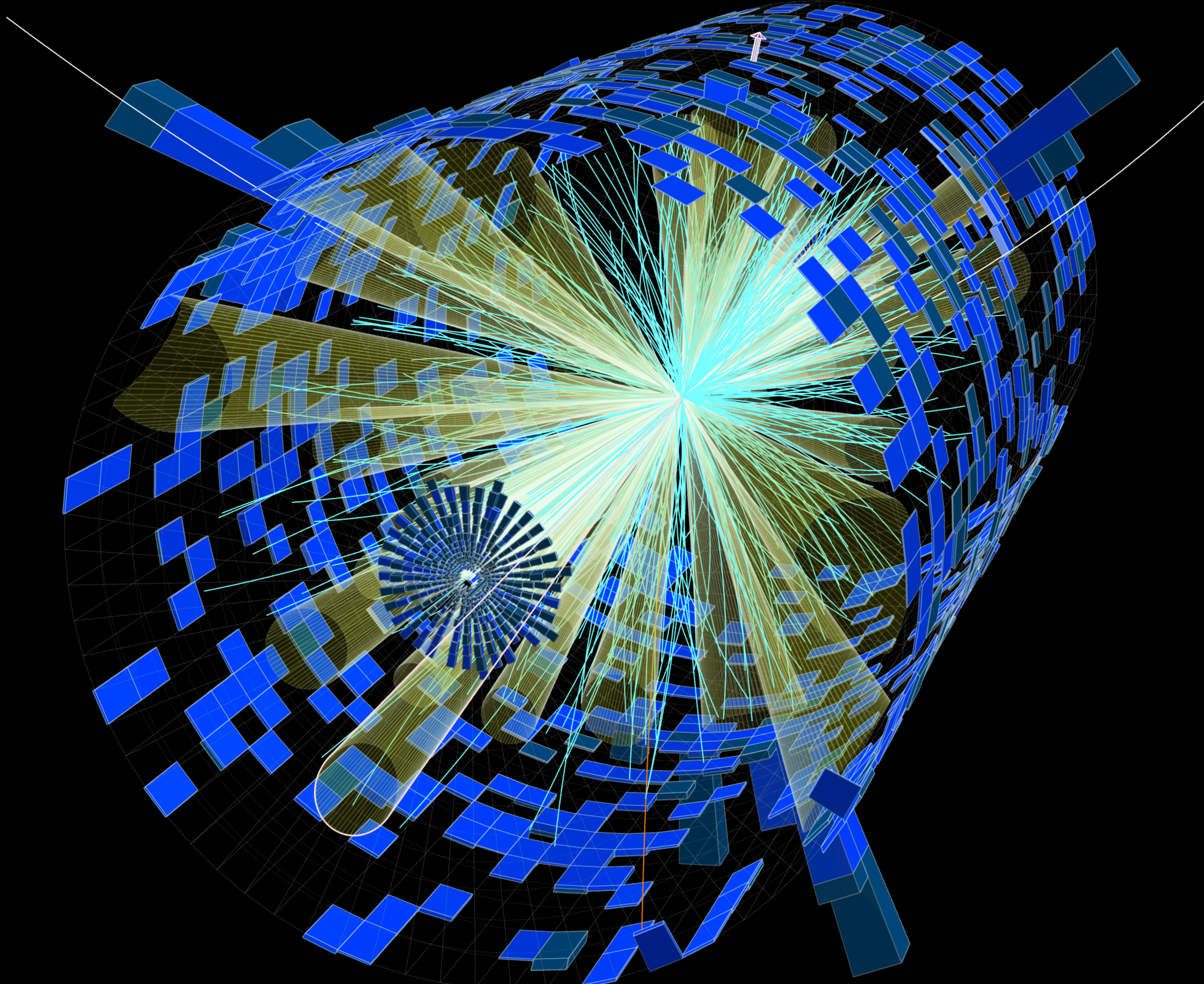
NYU Center for Data Science

Center for Cosmology and Particle Physics

EXPLORATIONS AT THE PHYSICS ∩ ML INTERFACE



@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

Collaborators + Many More


Gilles Louppe
U. Liège


Kyunghyun Cho


Joan Bruna


Sebastian Macaluso


Meghan Frate


Juan Pavez


Miles Cranmer


@HServiansky


Nimrod Segol


Johann Brehmer


Craig Greenberg


Nicholas Monath


Peter Battaglia


Alvaro Sanchez Gonzalez

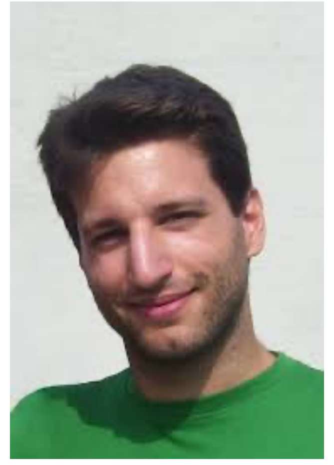

Shirley Ho

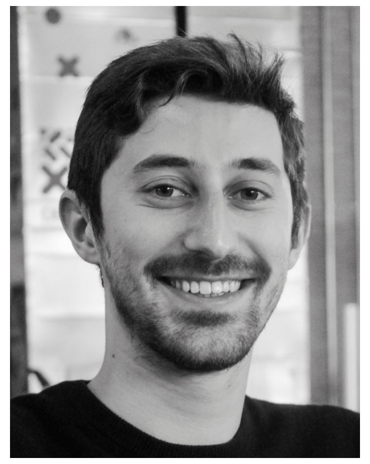

Irina Espejo


@HaggaiMaron


Yaron Lipman


Peter Sadowski


George Papamakarios


Michael Albergo


Danilo Rezende


Atılım Güneş Baydin
University of Oxford


Michela Paganini
FAIR


Wahid Bhimji
NERSC, Berkeley Lab


Frank Wood
U. Victoria


Jonathan Shlomi


Eilam Gross


Phiala Shanahan


William Detmold


Gurtej Kanwar


Dan Hackett


Denis Boyda


Sébastien Racanière


Savannah Thais
Princeton


Ruth Angus
Columbia University

NSF

MSDSE
Moore - Sloan
Data Science Environments

My recent Physics ∩ ML work

Diverse set of projects

- QCD-aware jet classifiers for the LHC with Tree and Graph based NNs
- ML for precision measurements at the LHC data analysis, e.g. effective field theory
- Probing dark matter substructure with strong gravitational lenses
- Probabilistic programing for LHC data analysis
- ML to model the Quantum Density Matrix
- ML for Lattice Field Theory
- Hamiltonian + ODE + GraphNNs for dynamical systems
- GraphNNs + Symbolic regression for dynamical systems
- Dynamical programming and RL for probabilistic treatment of hierarchical clustering

What do these projects have in common?

My recent Physics ⋂ ML work

What do these projects have in common?

- Several are aimed at import physics questions, but…
- Many are explorations to refine (my) understanding of the pros and cons and different approaches to modeling / inference / science / AI
  - Traditional physics-based approaches (mechanistic)
  - Black box machine learning (function approximation, prediction)

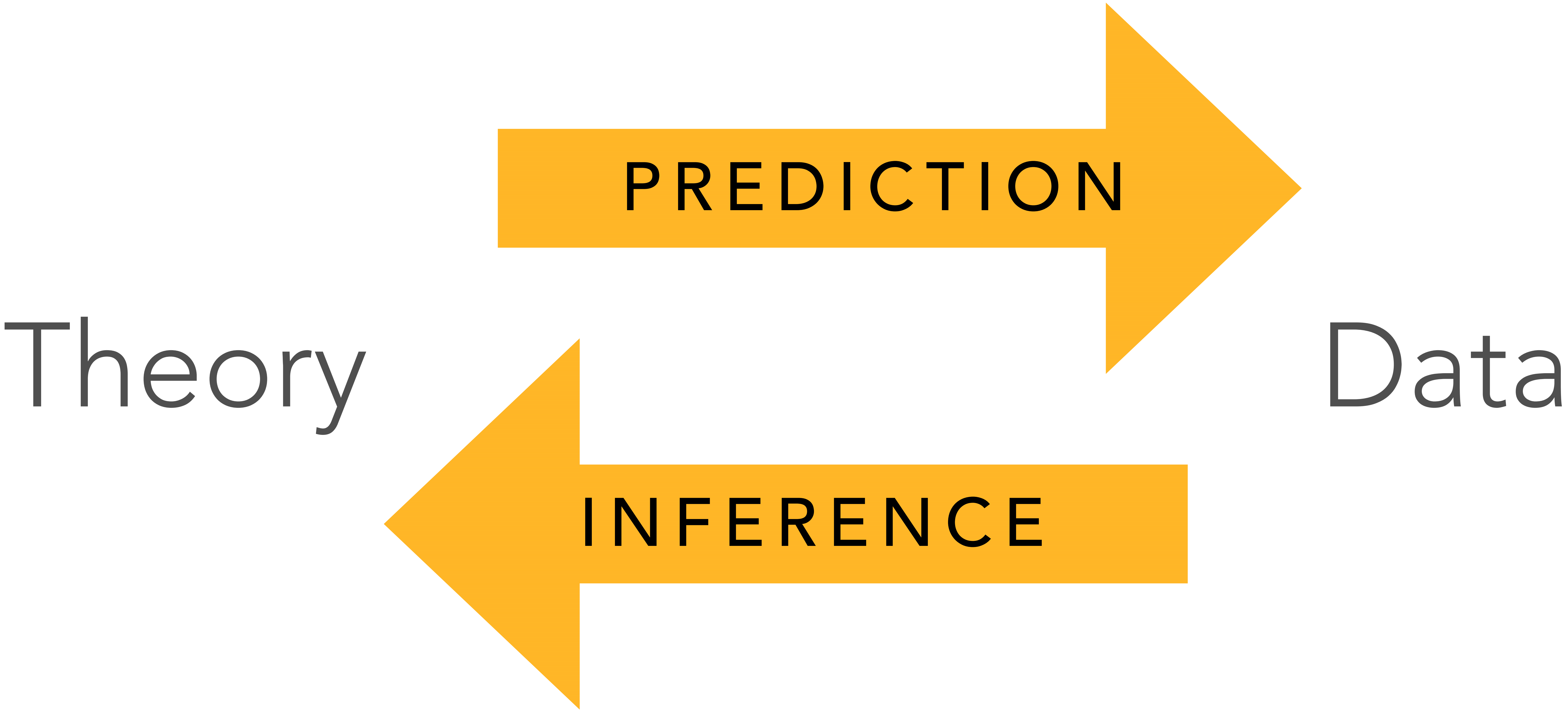
This exploration has

- Led to some interesting insights into AI / ML, and
- Refined my thinking about physics (and the philosophy of science)

Abstract

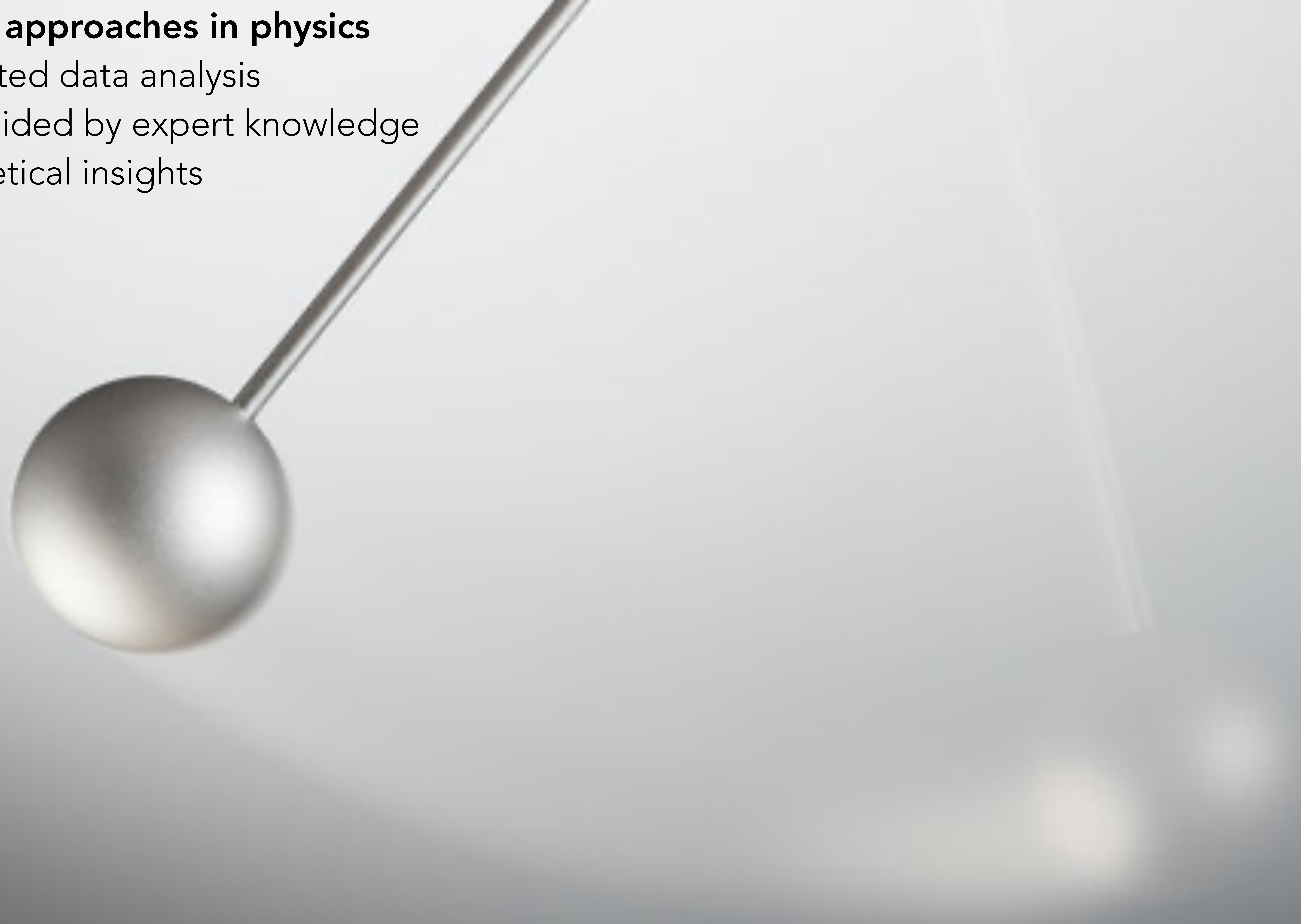
Instead of focusing on a specific application, I will discuss a few projects that explore the Physics ∩ ML Interface.

- How do we incorporate our physical insight into the underlying causal mechanism into the inductive bias of machine learning architectures?
  - Is that helpful or necessary?
- Why do we care if a model is interpretable?
- Where do we stand on the spectrum between ML-supercharged data analysis and an AI / robot scientist?
- How does this line of thinking influence research in AI and ML?


PREDICTION
Theory
Data
INFERENCE

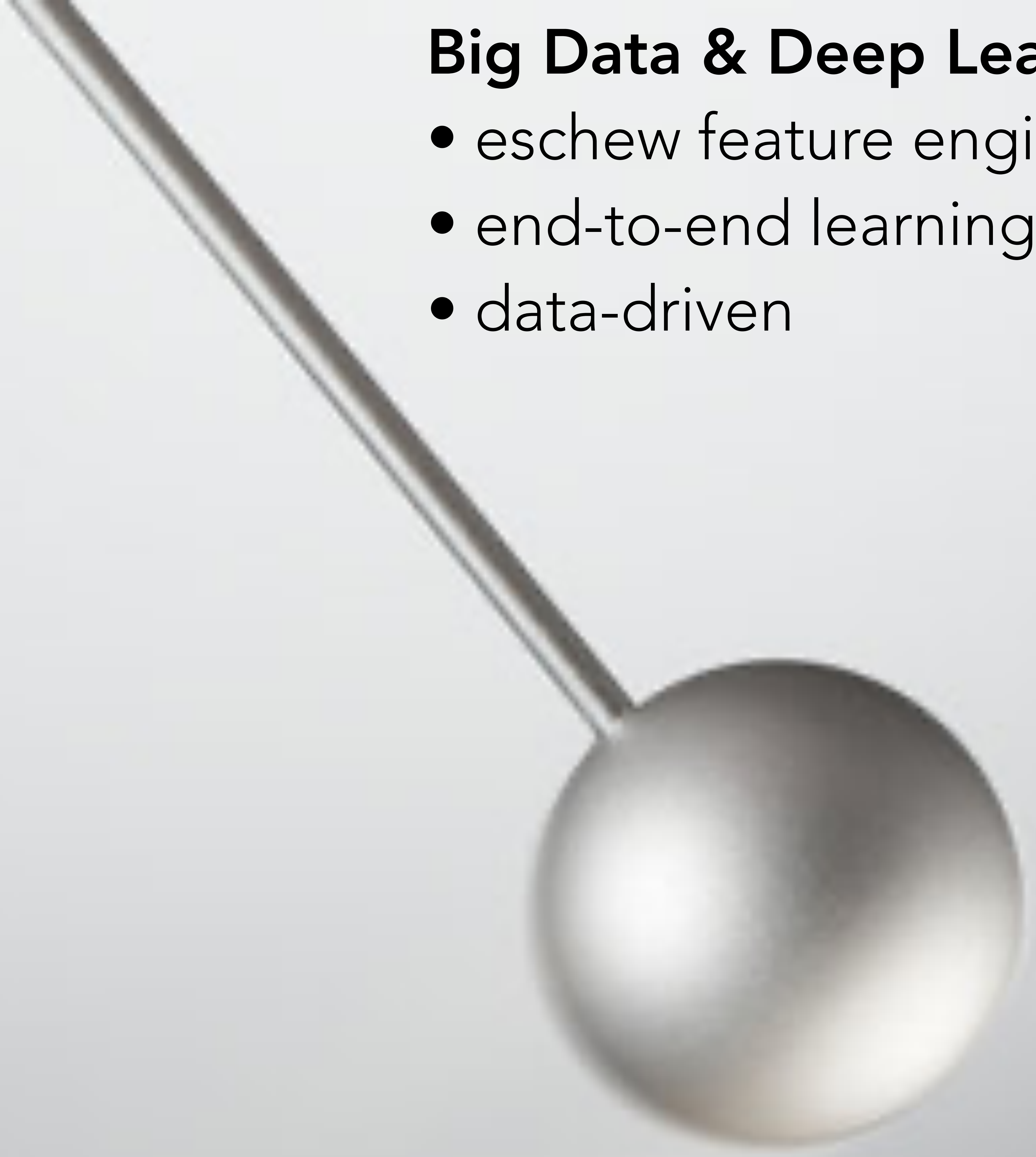
Traditional approaches in physics

- hand-crafted data analysis
- largely guided by expert knowledge and theoretical insights



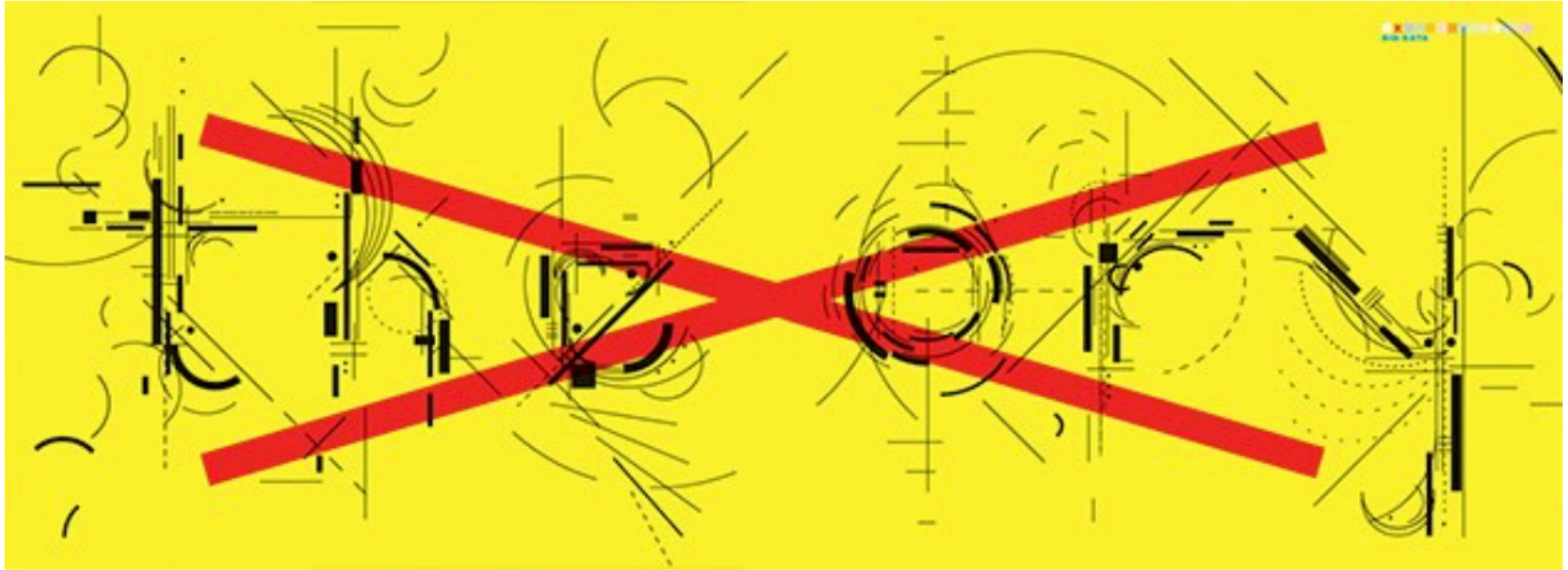
Big Data & Deep Learning

- eschew feature engineering
- end-to-end learning
- data-driven



CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE




All models are wrong,
and increasingly you can succeed without them.
— Peter Norvig

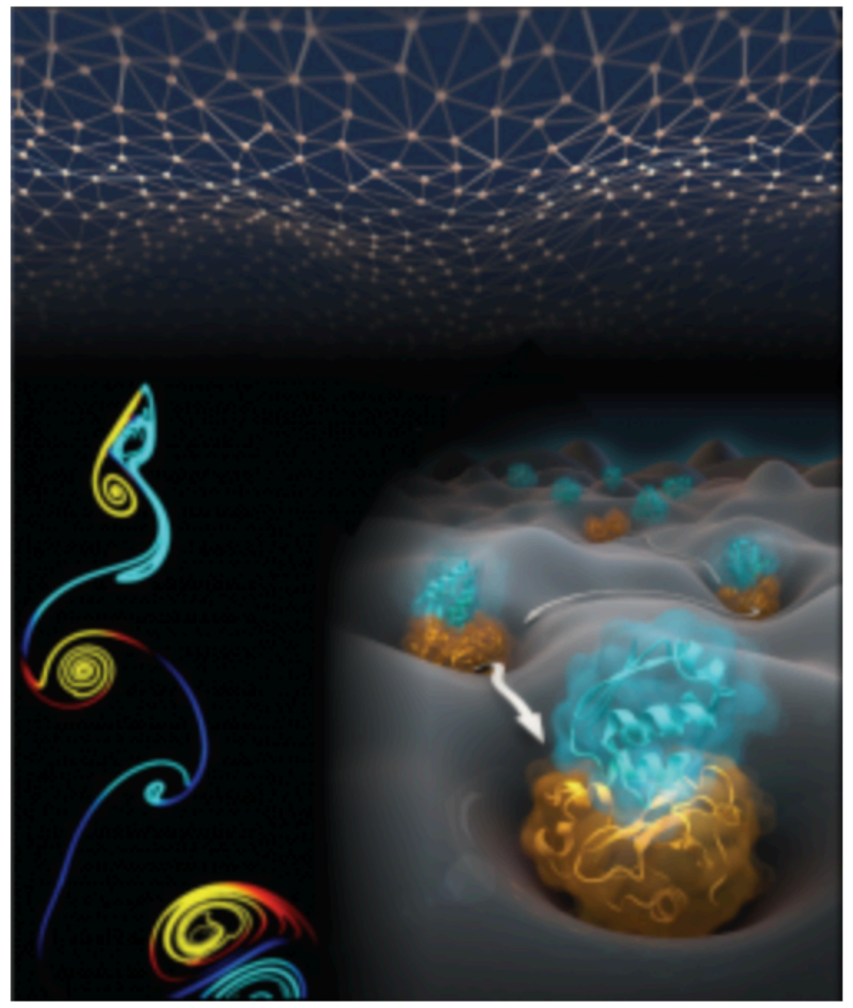
Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview

Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.



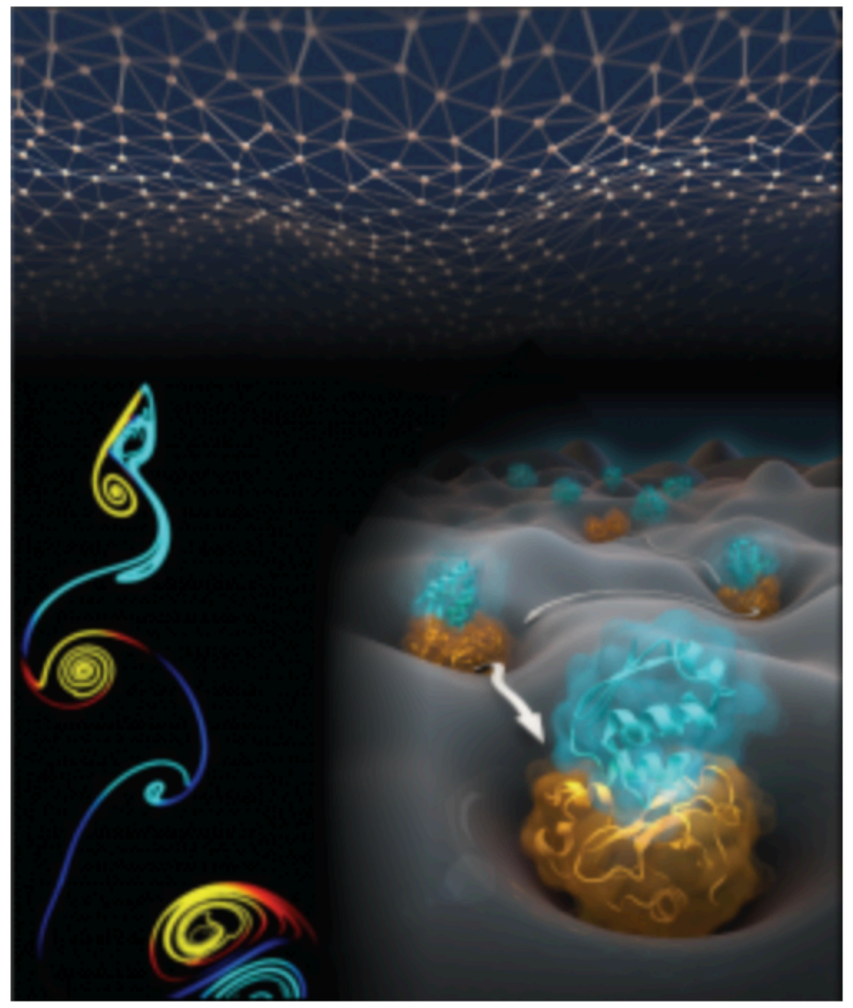
As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

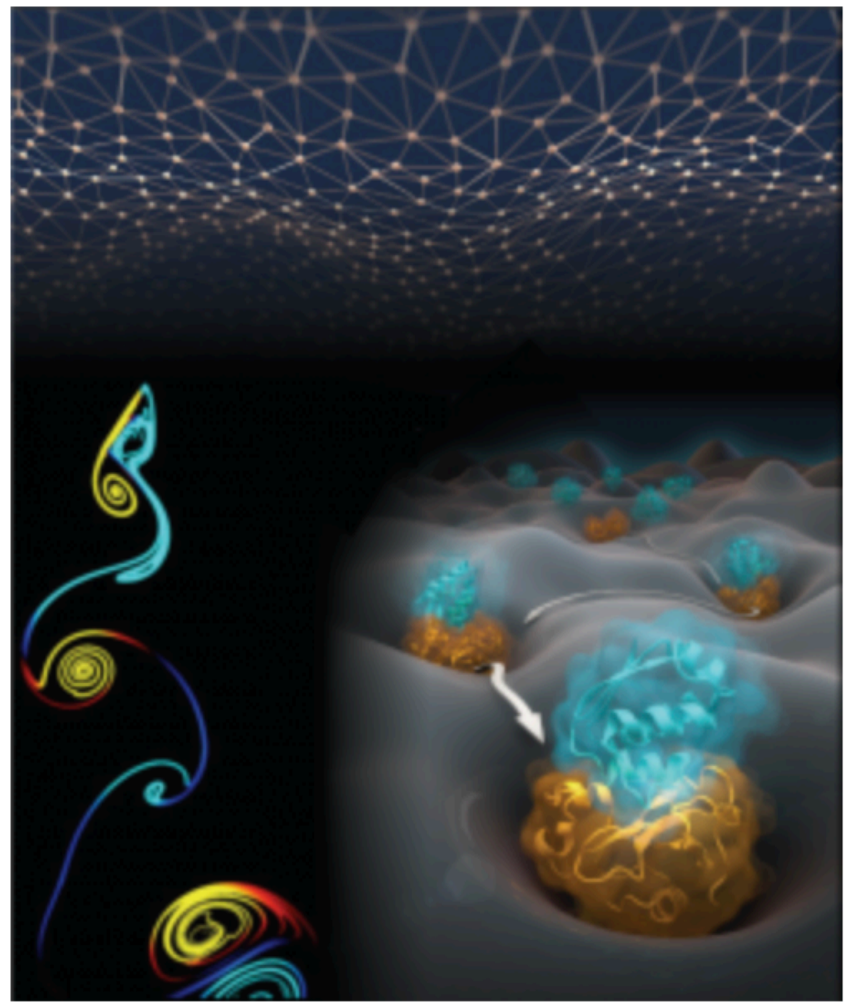
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

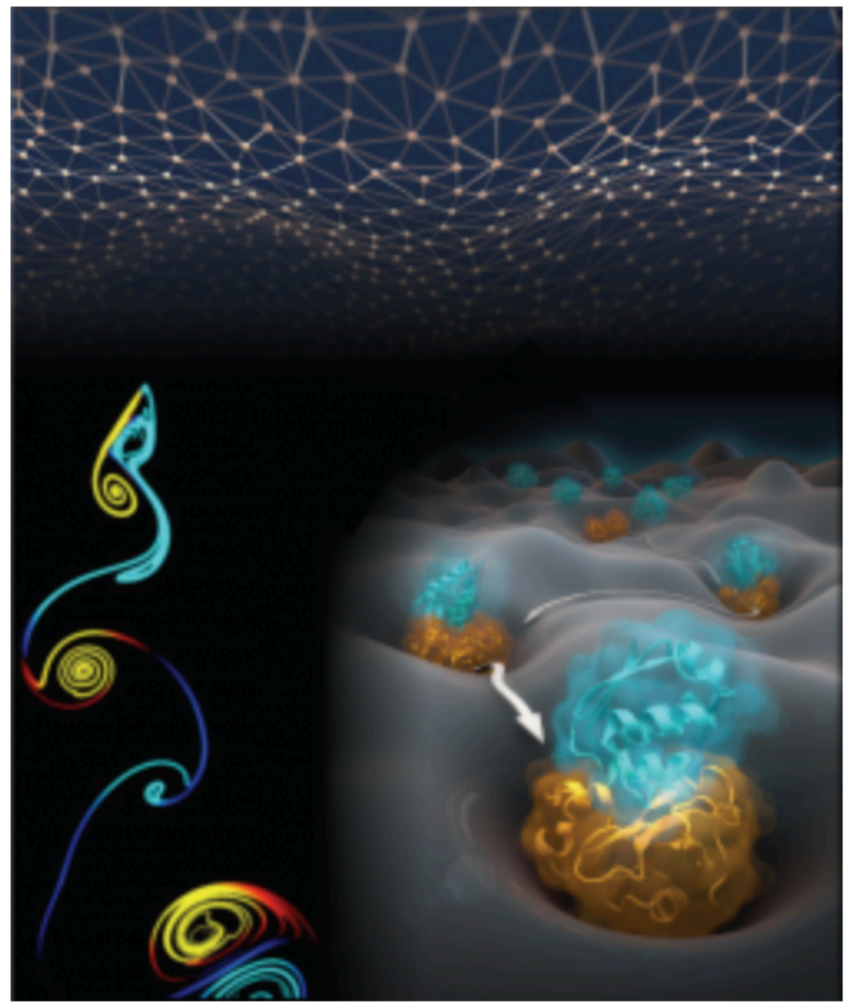
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

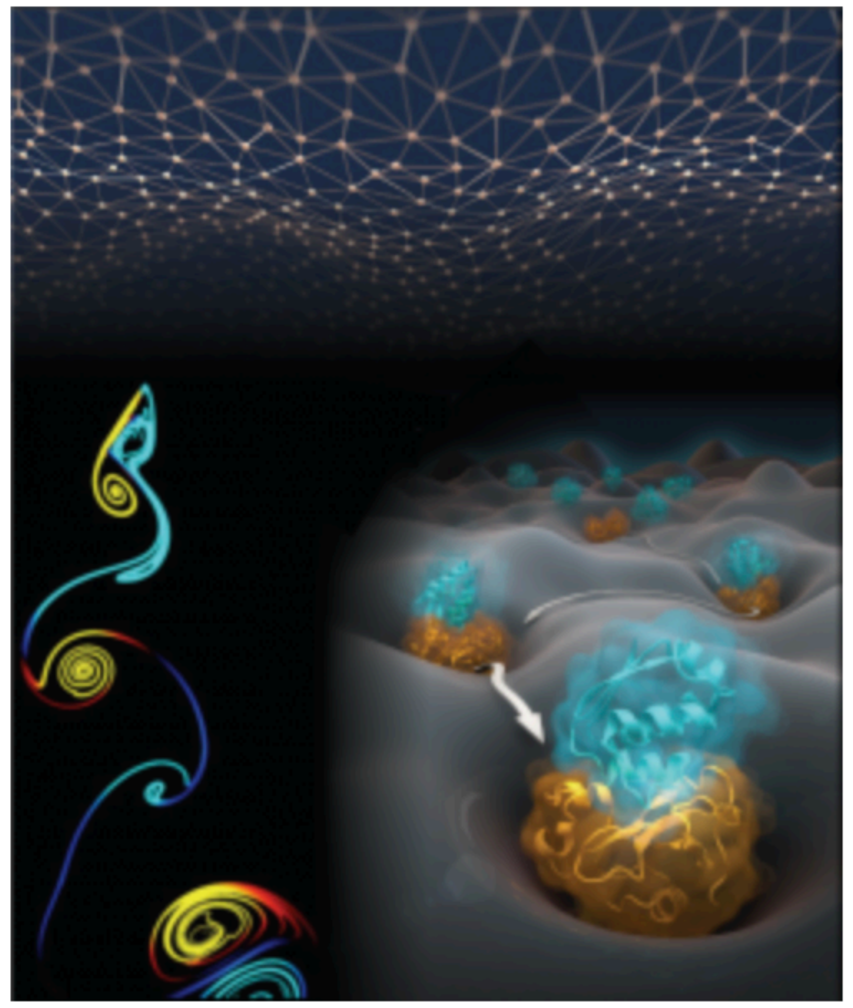
As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

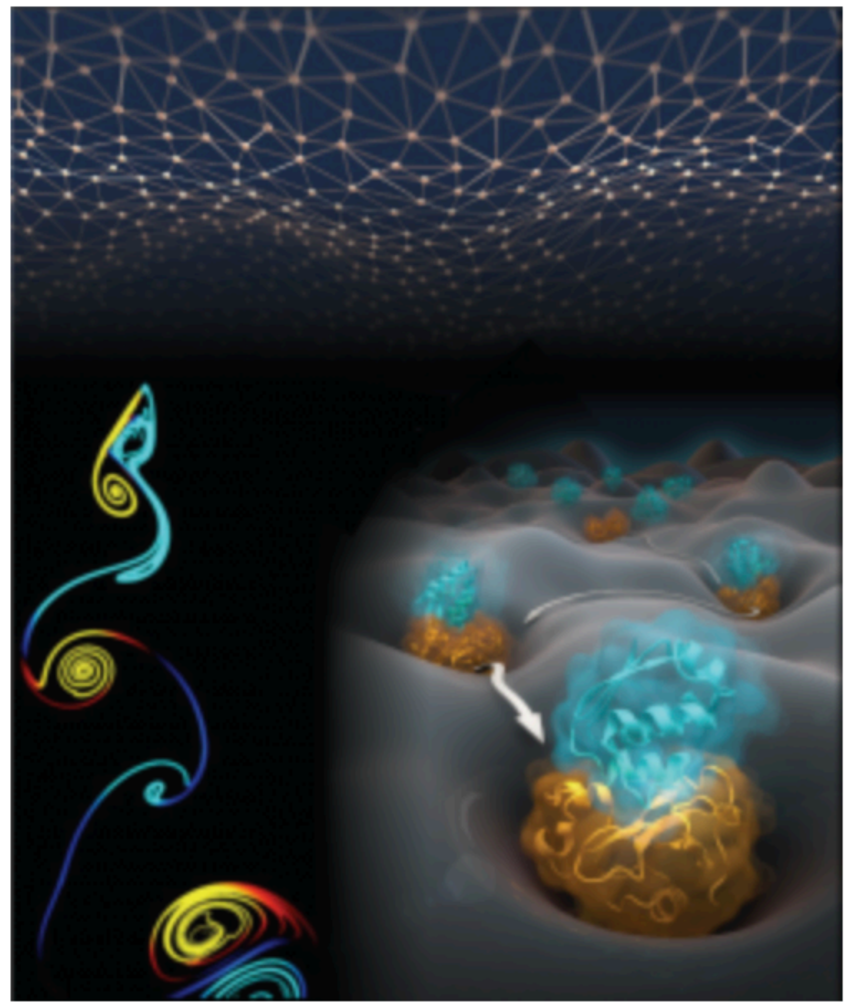
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

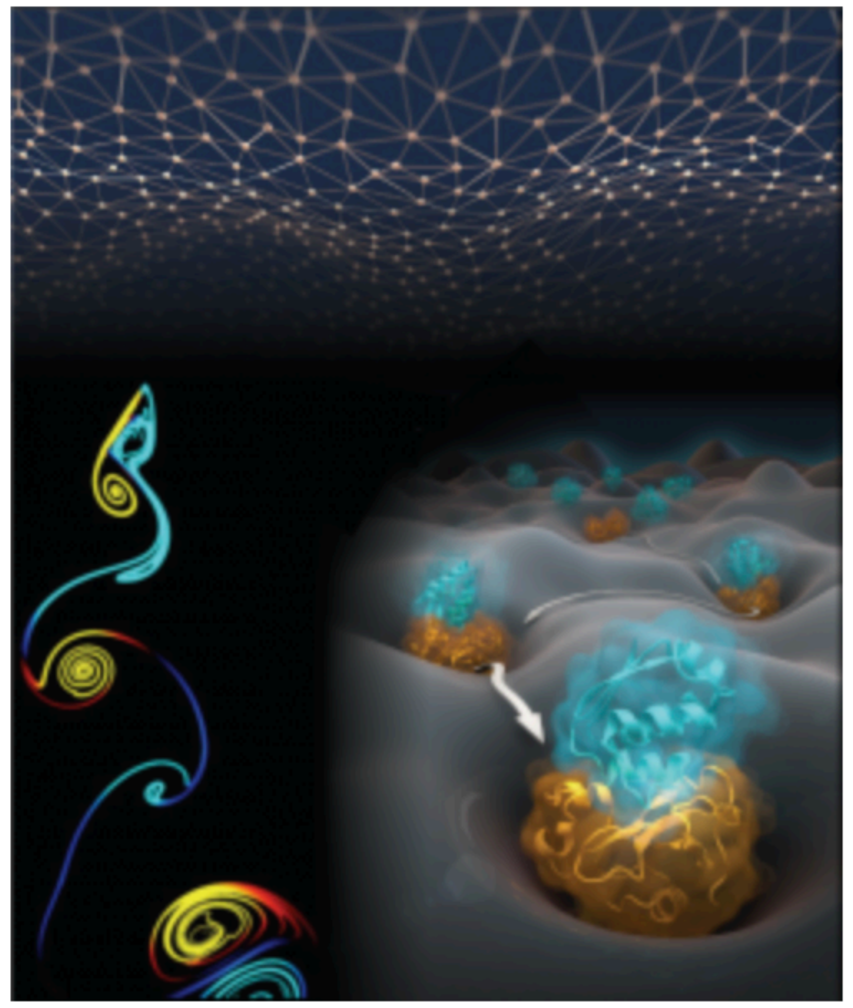
As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

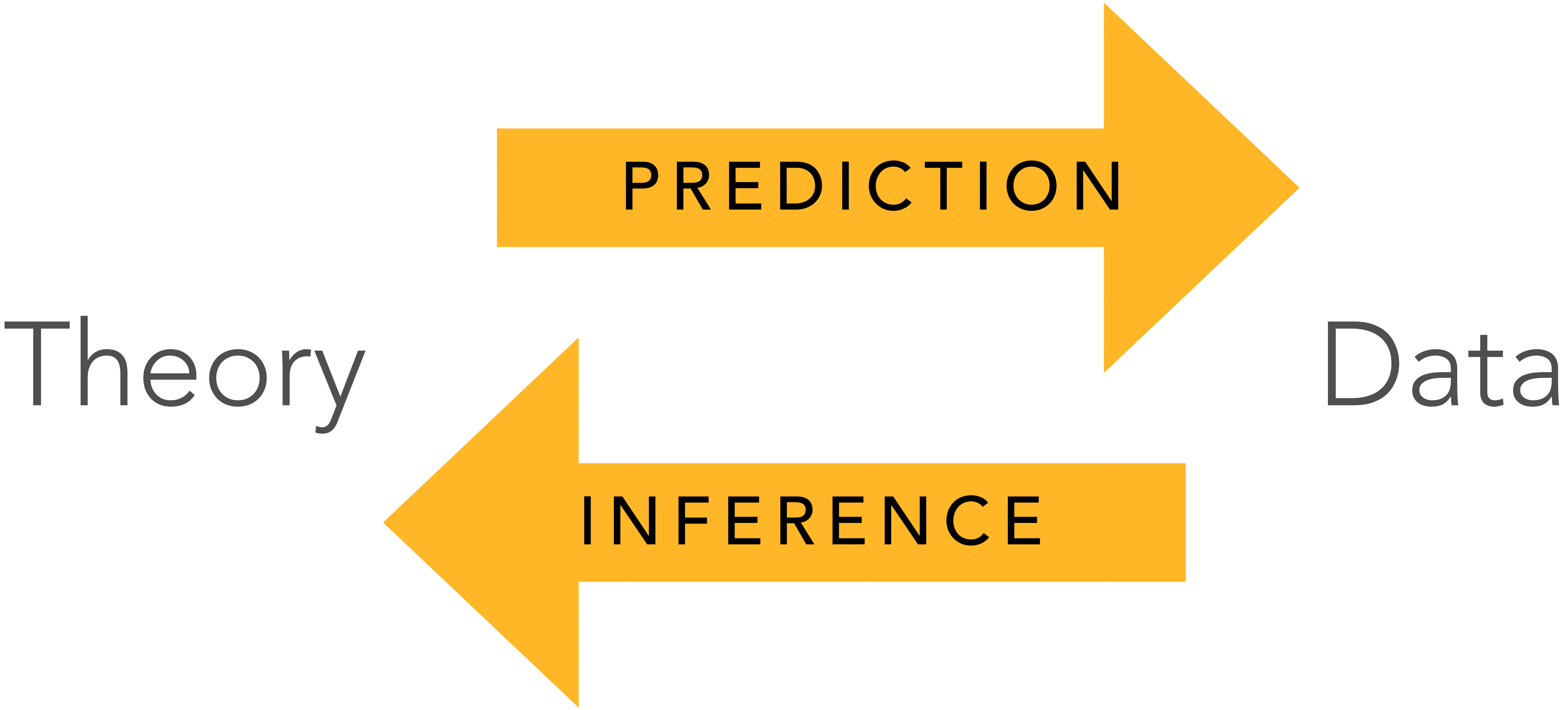
Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

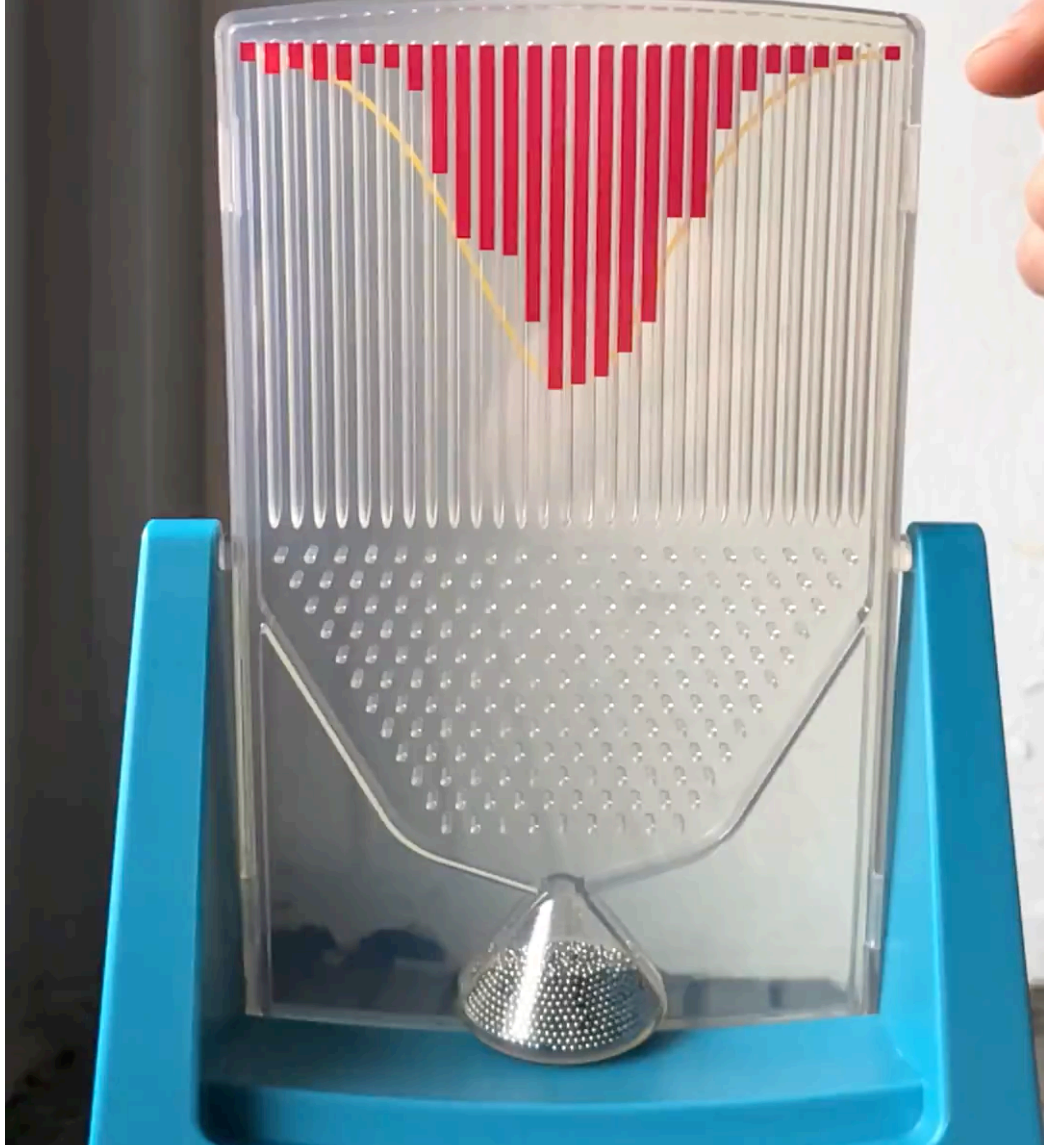
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.


Theory
PREDICTION
INFERENCE
Data

A toy example

Imagine the entire board is slightly tilted, which biases the probability to bounce left/right.

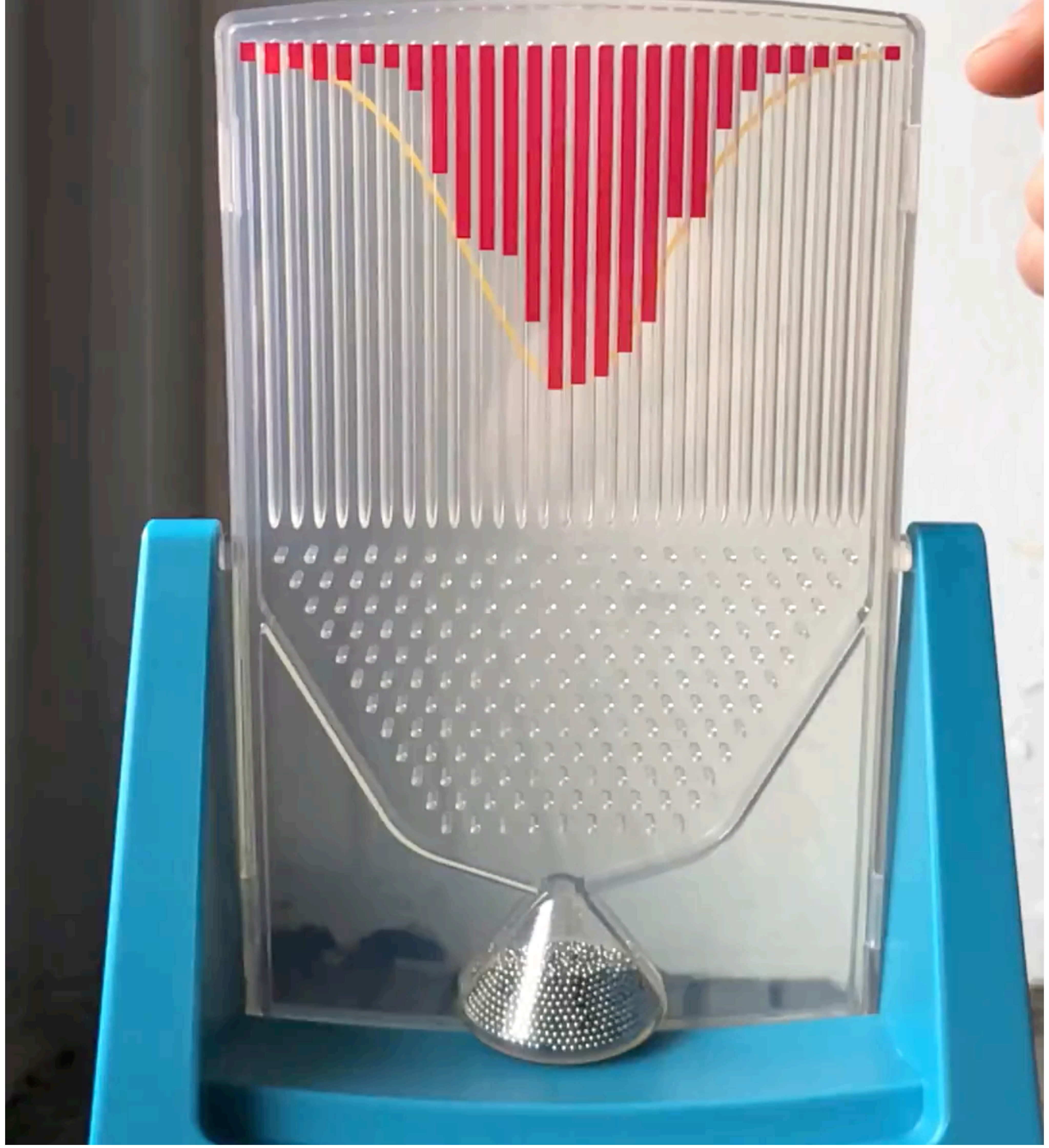
Say we want to infer θ, the probability to bounce right based on distribution of x



A toy example

Imagine the entire board is slightly tilted, which biases the probability to bounce left/right.

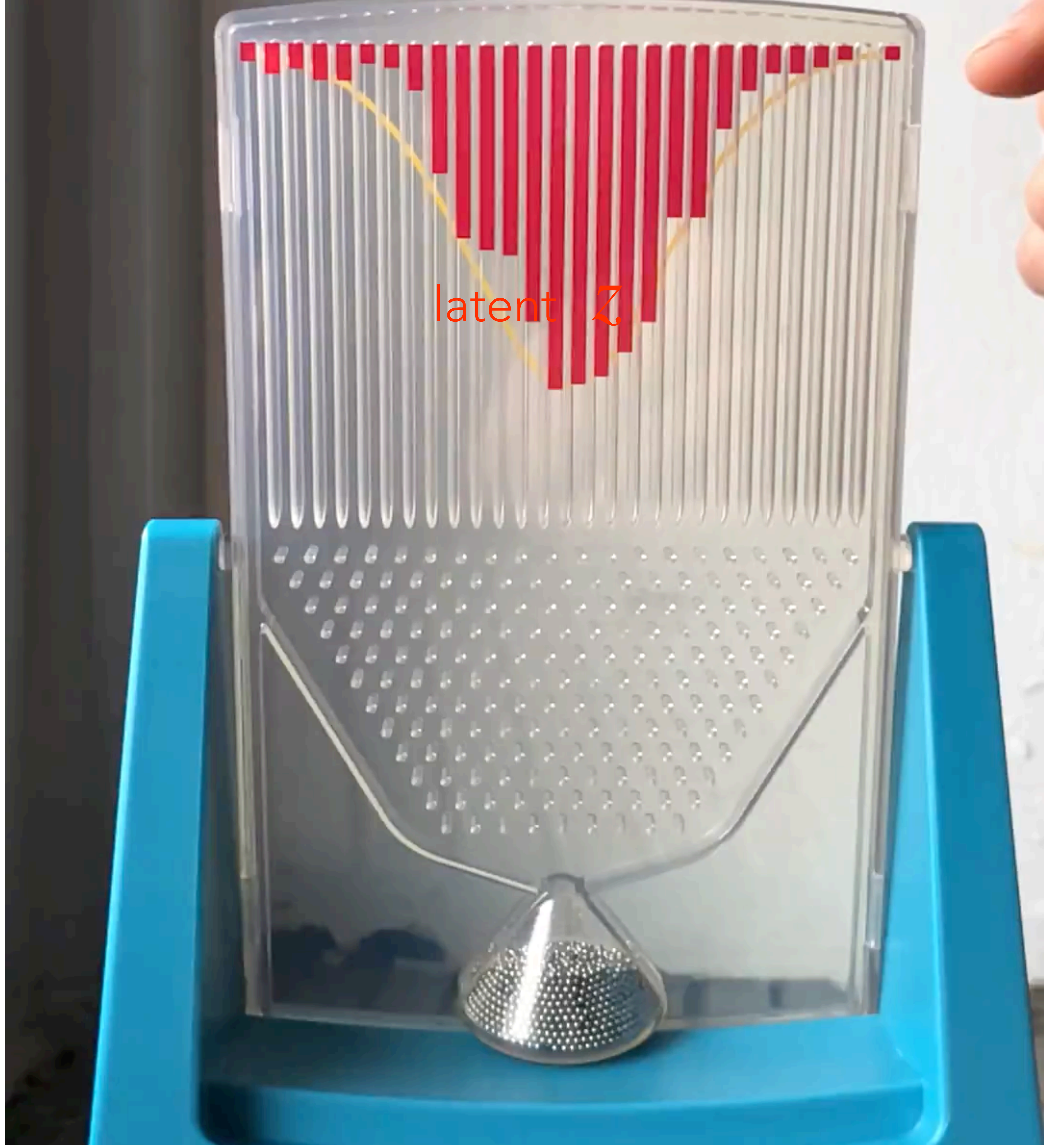
Say we want to infer θ, the probability to bounce right based on distribution of x



A toy example

Imagine the entire board is slightly tilted, which biases the probability to bounce left/right.

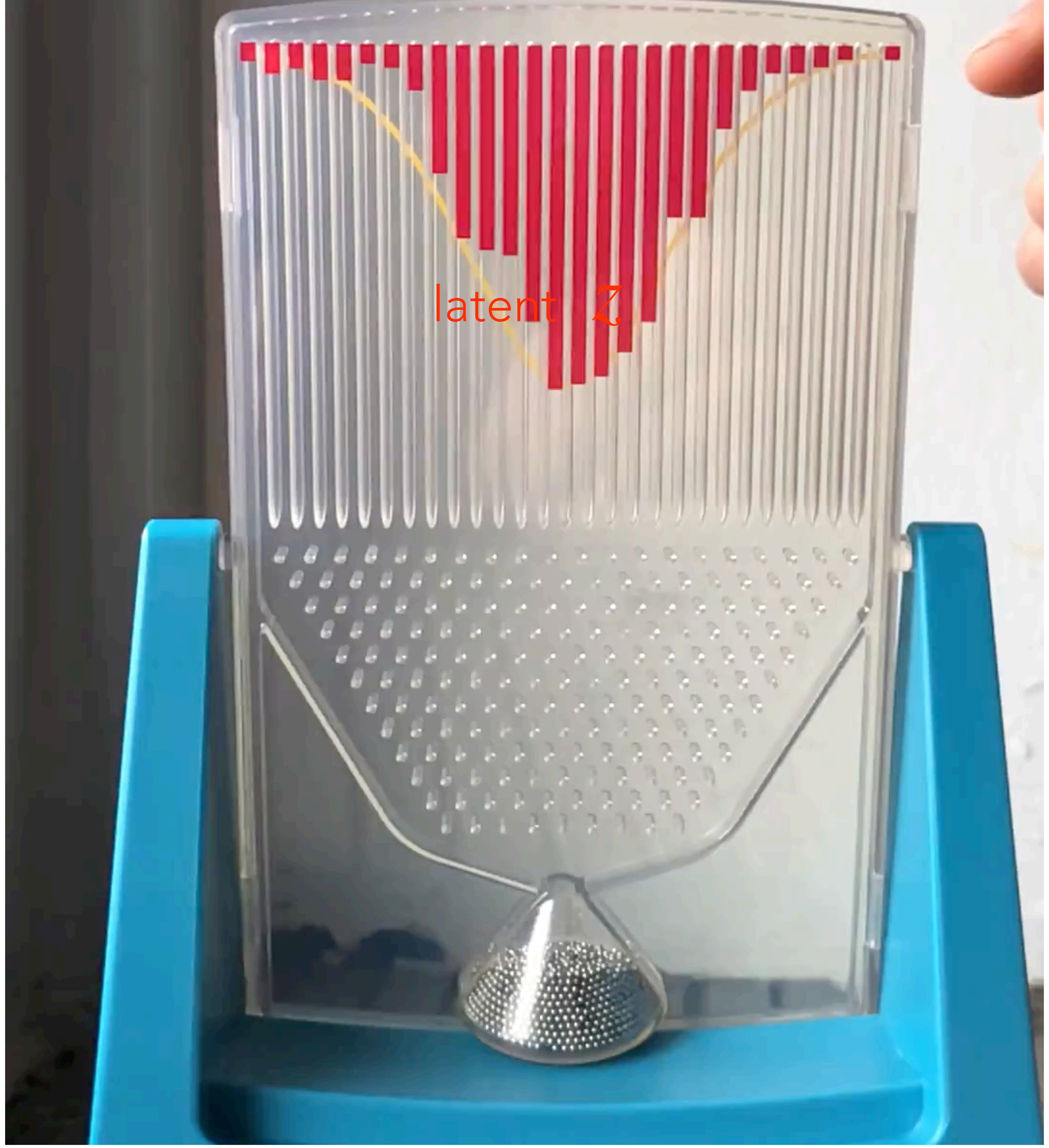
Say we want to infer θ, the probability to bounce right based on distribution of x



A toy example

Imagine the entire board is slightly tilted, which biases the probability to bounce left/right.

Say we want to infer θ, the probability to bounce right based on distribution of x

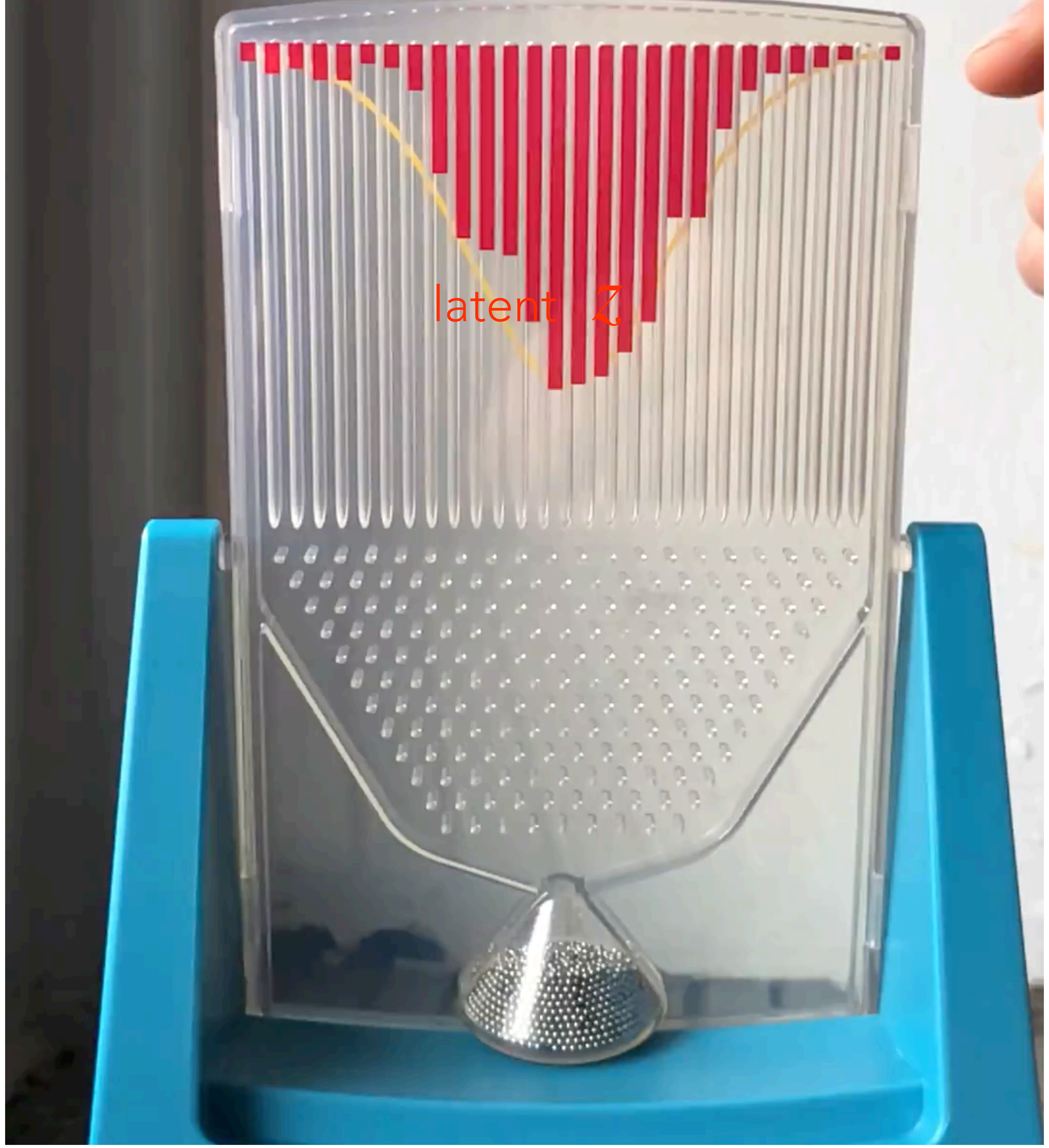


observe x

A toy example

Imagine the entire board is slightly tilted, which biases the probability to bounce left/right.

Say we want to infer θ, the probability to bounce right based on distribution of x



The probability of ending in bin x corresponds to the total probability of all the paths z from start to x.

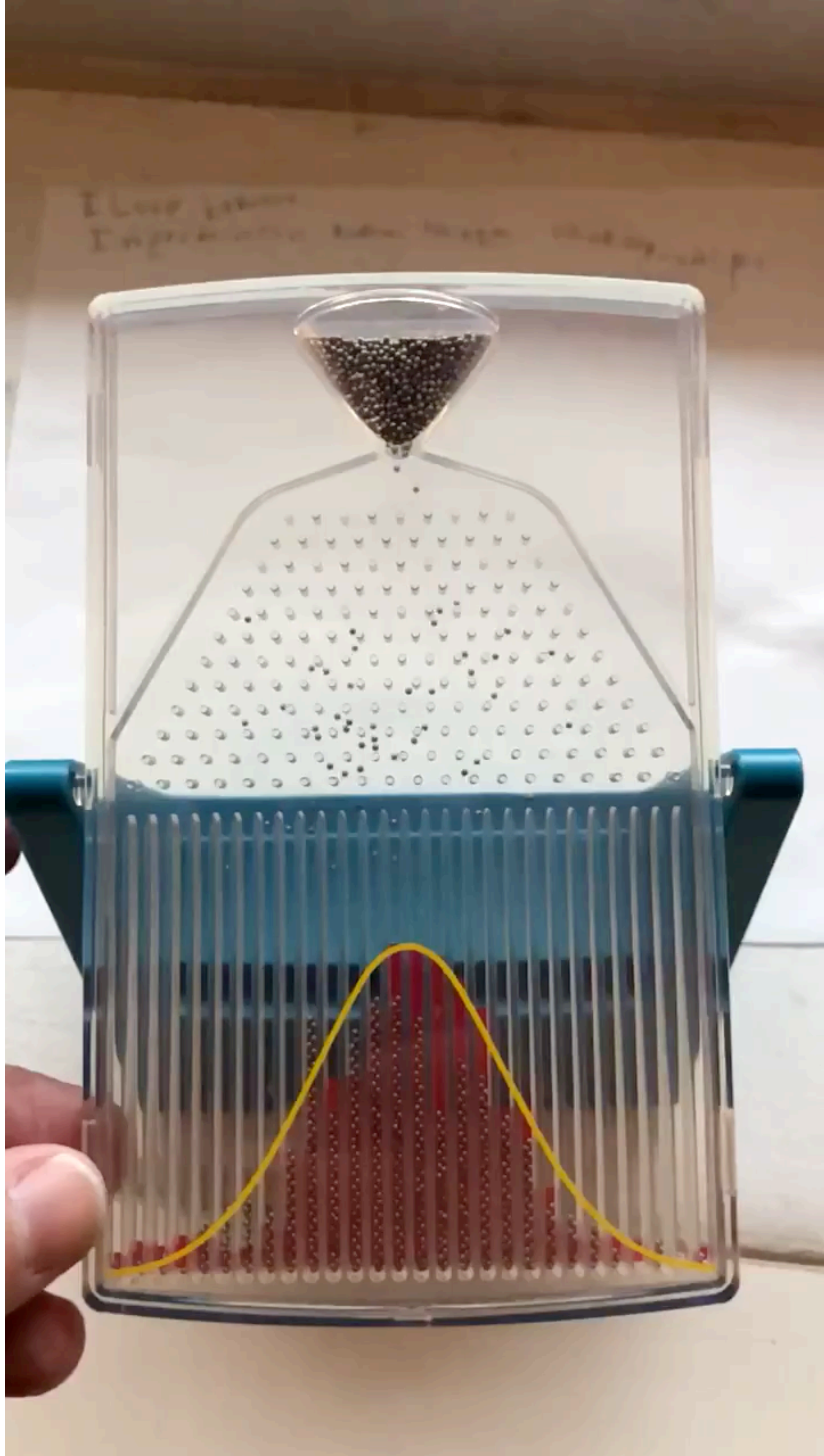
$$p(x|\theta) = \int p(x, z|\theta) dz = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Uh oh!

The actual situation is much more complicated.

It's not a Binomial distribution!

What is it?

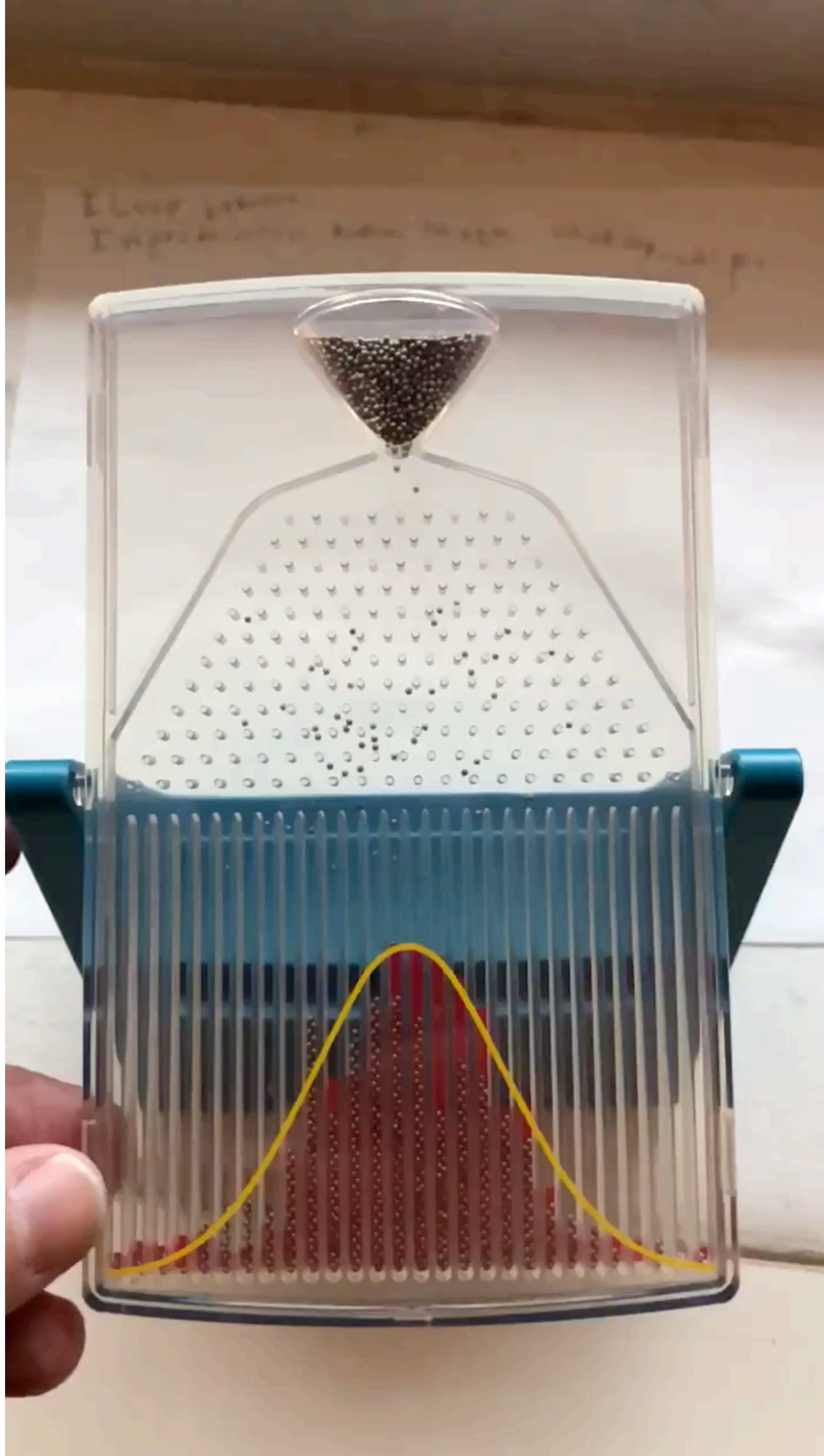


Uh oh!

The actual situation is much more complicated.

It's not a Binomial distribution!

What is it?



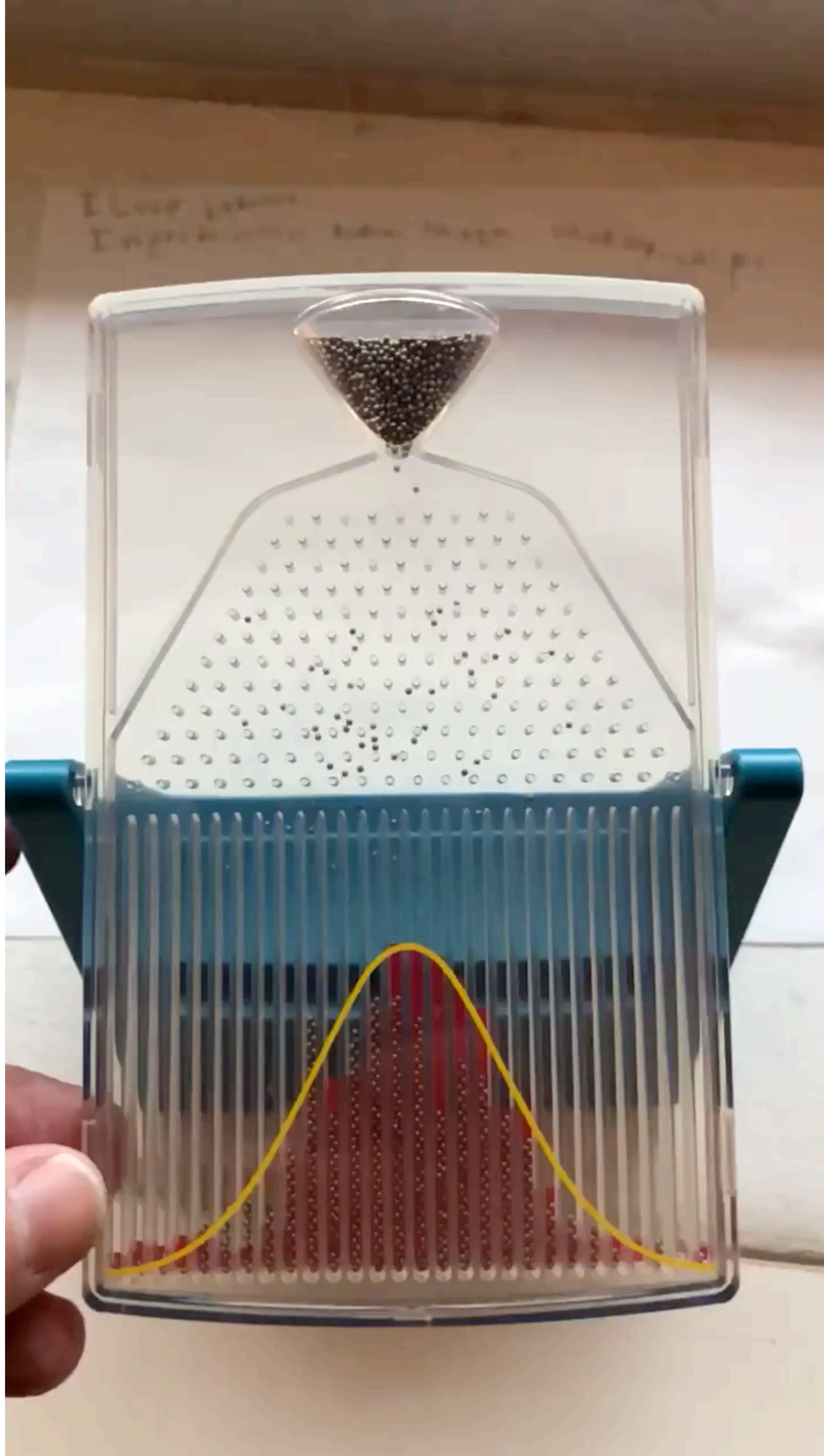
Uh oh!

The actual situation is much more complicated.

It's not a Binomial distribution!

What is it?

I have no idea, but I can **simulate it!**



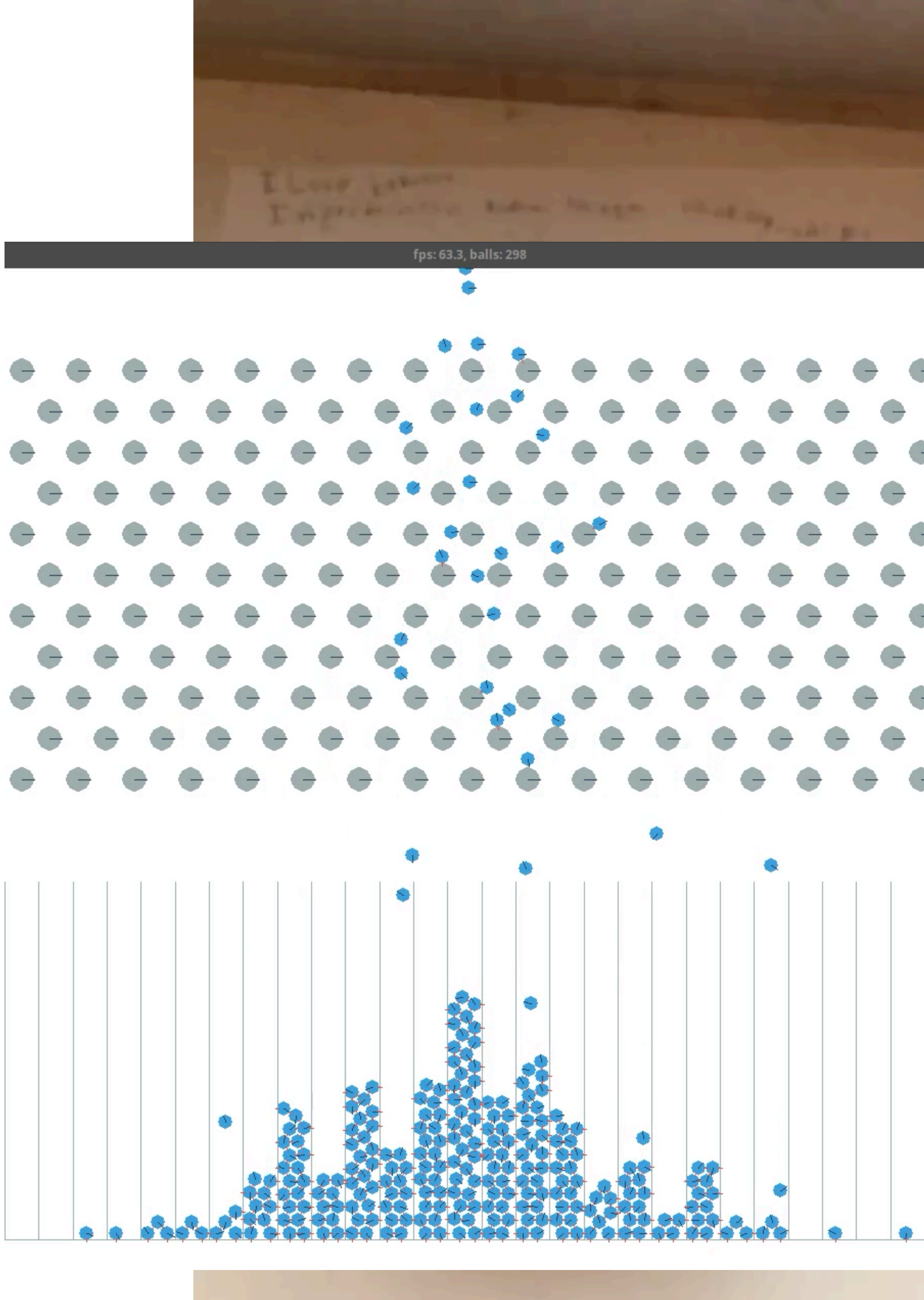
Uh oh!

The actual situation is much more complicated.

It's not a Binomial distribution!

What is it?

I have no idea, but I can **simulate it!**



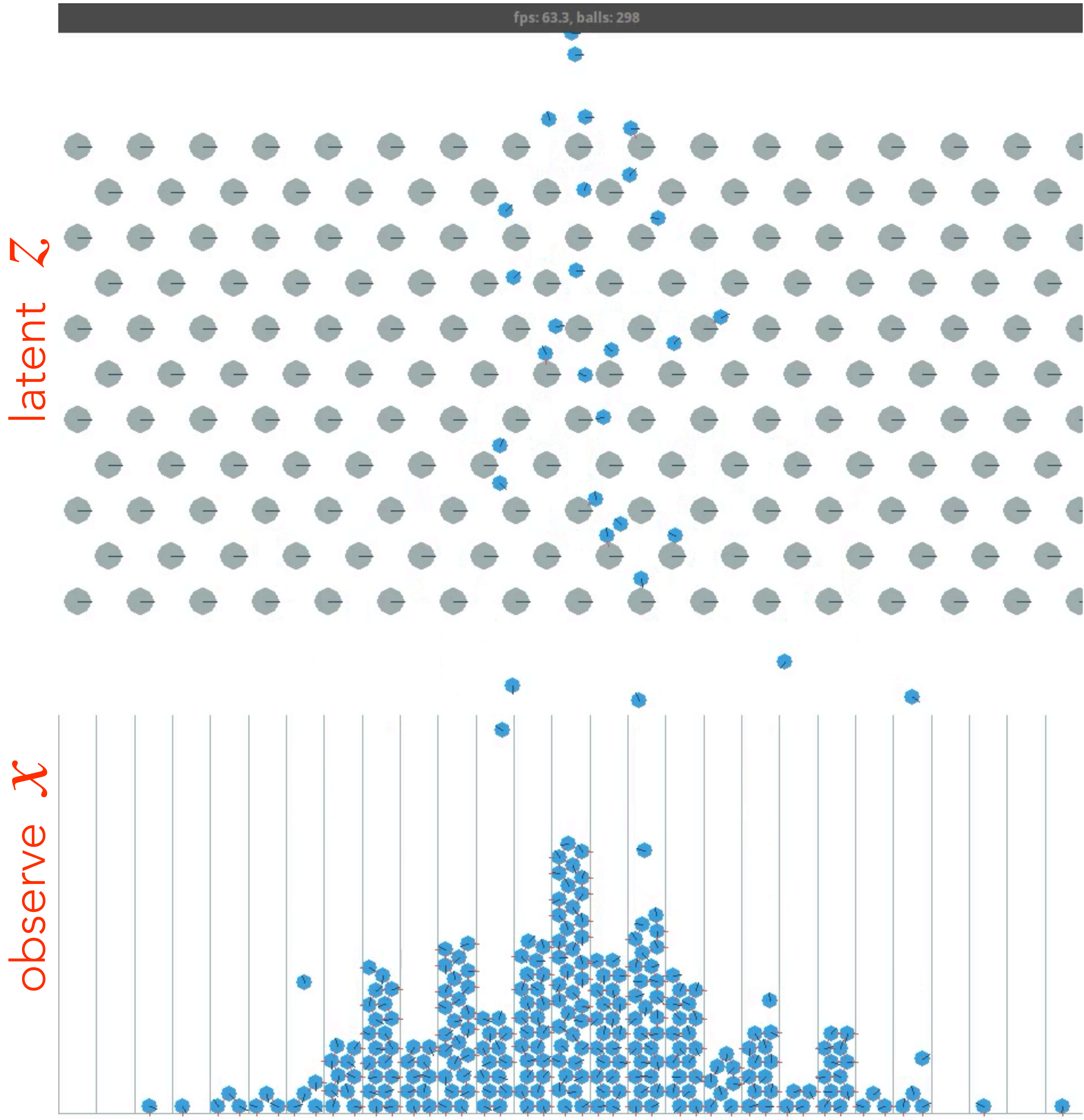
An example

ANIMATION BY ATILIM GÜNEŞ BAYDIN

The probability of landing in a bin x corresponds to cumulative probability of all the latent paths z that end in x

$$p(x|\theta) = \int p(x, z \mid \theta)\mathrm{d}z$$

- But the integral (sum) can no longer be simplified analytically
- As the latent space grows, the number of paths grows rapidly.
- The integral becomes **intractable**
- But generating synthetic observations remains easy



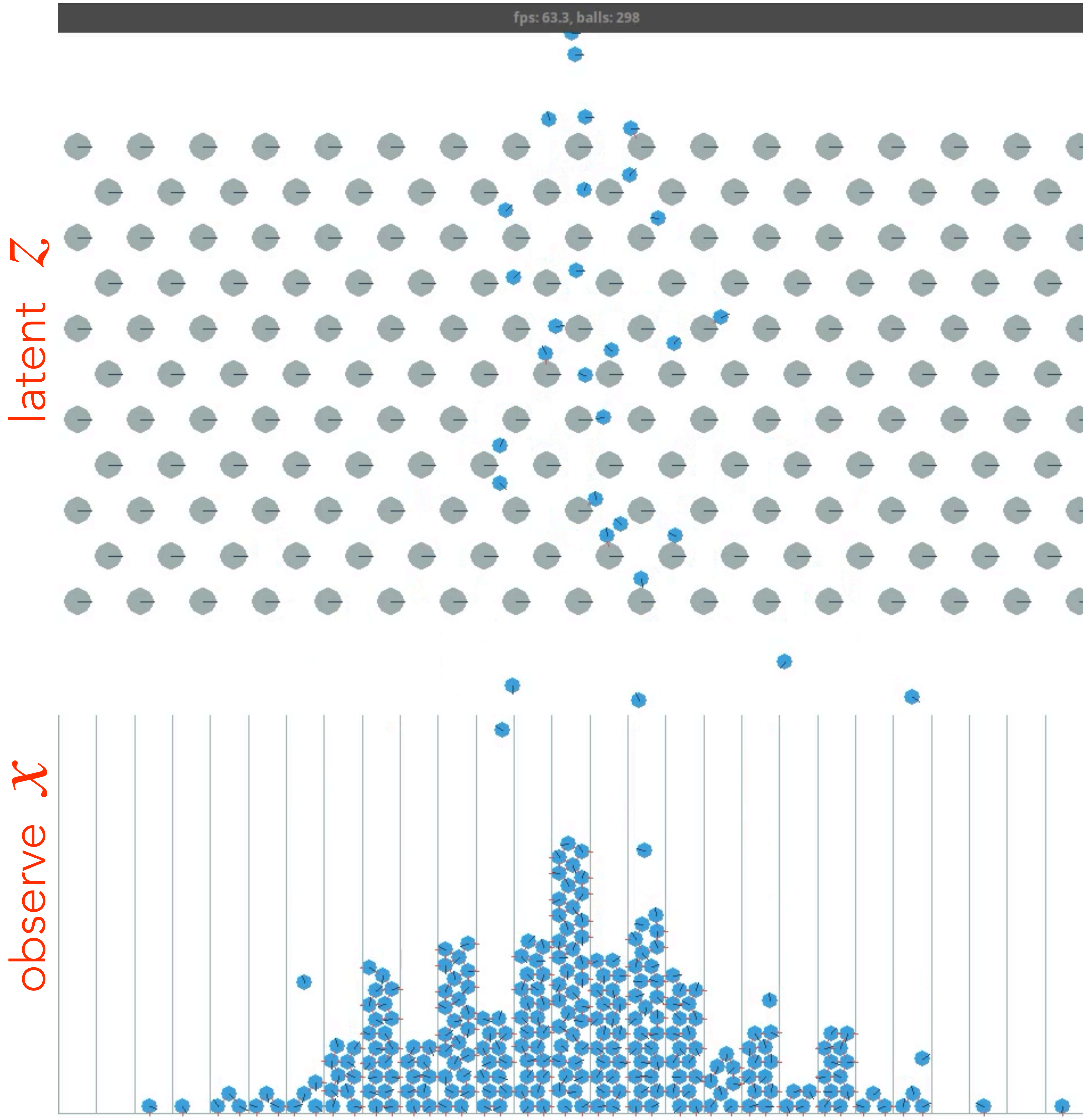
An example

The probability of landing in a bin x corresponds to cumulative probability of all the latent paths z that end in x

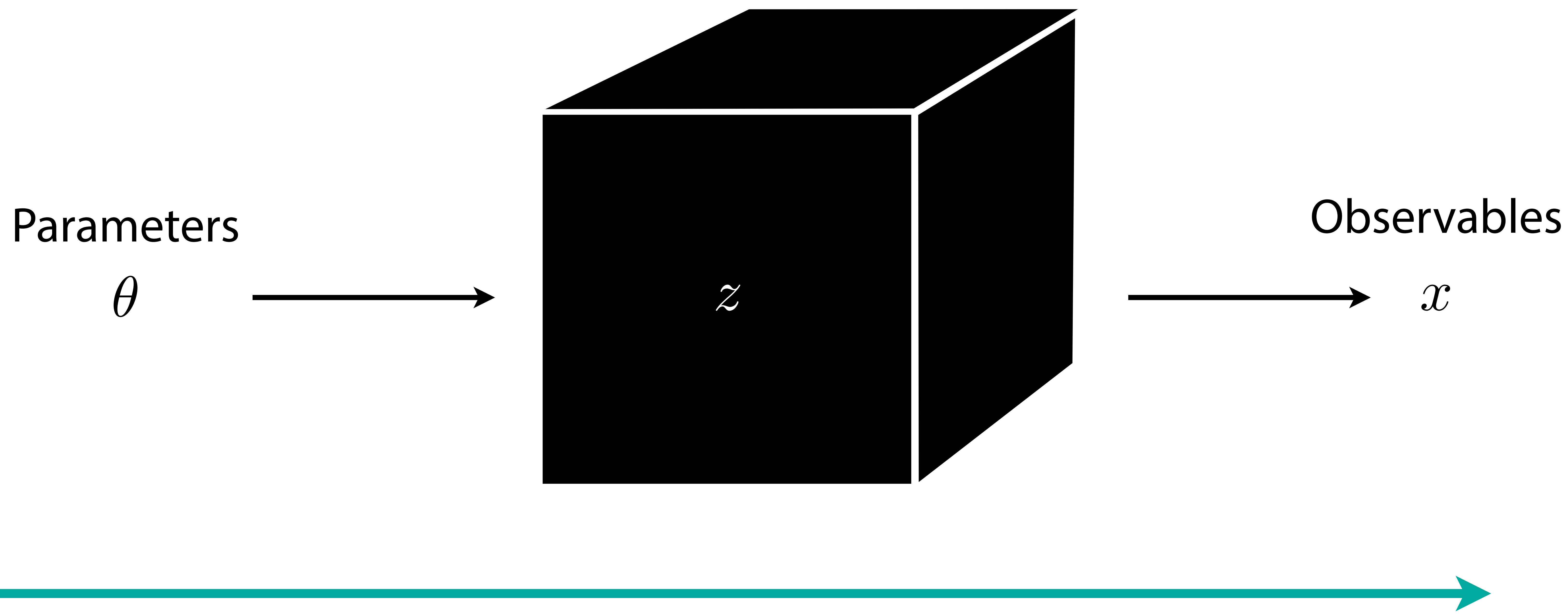
$$p(x|\theta) = \int p(x, z \mid \theta) \mathrm{d}z$$

- But the integral (sum) can no longer be simplified analytically
- As the latent space grows, the number of paths grows rapidly.
- The integral becomes **intractable**
- But generating synthetic observations remains easy

ANIMATION BY ATILIM GÜNEŞ BAYDIN



Simulation—based inference



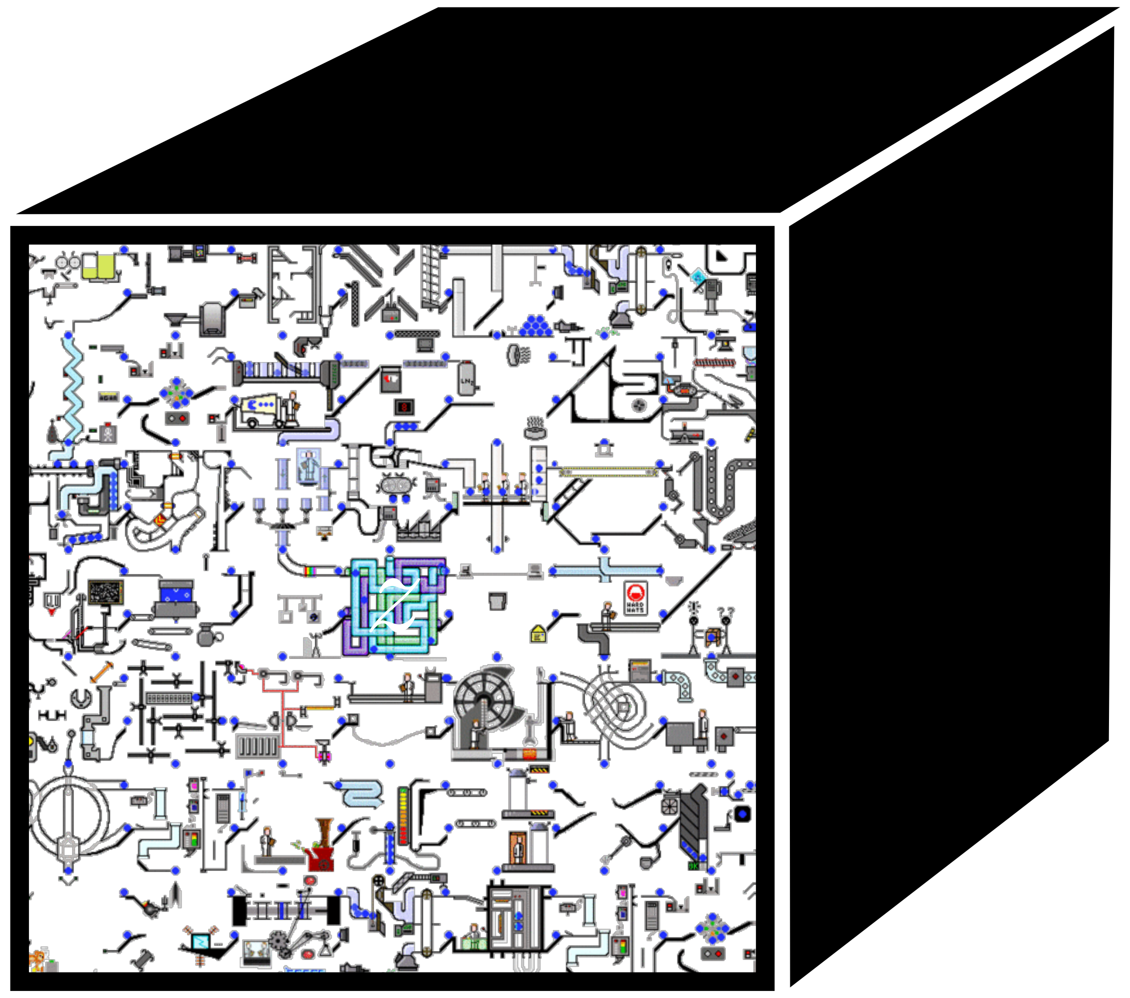
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Simulation—based inference

Parameters

θ



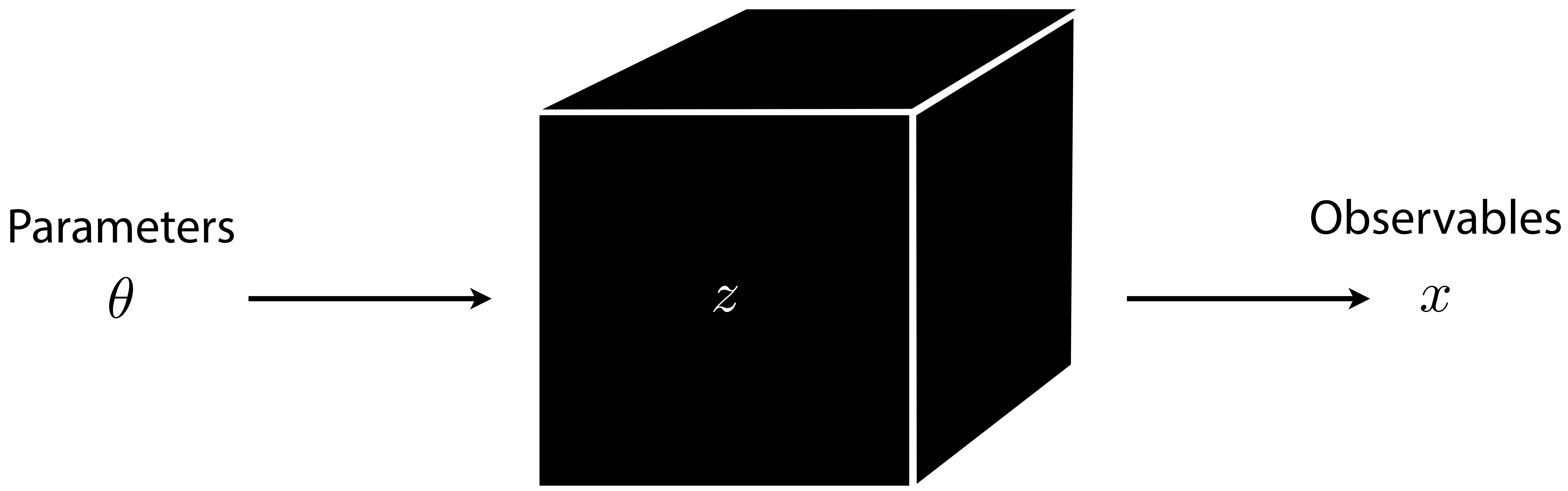
Observables

x

Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Simulation—based inference



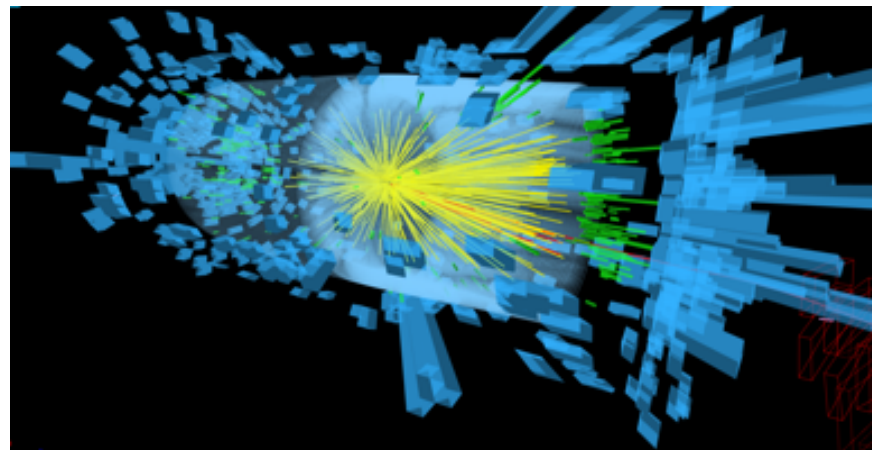
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

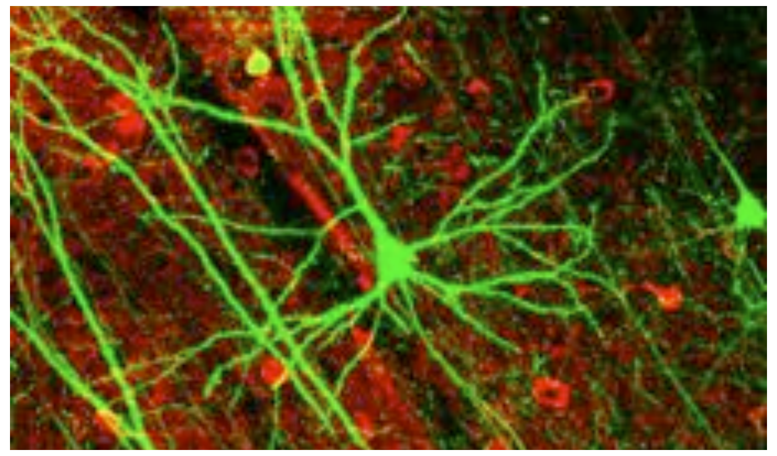
Inference:

- Likelihood function $p(x|\theta)$ is intractable
- Goal: estimator $\hat{p}(x|\theta)$

Science is replete with high-fidelity simulators



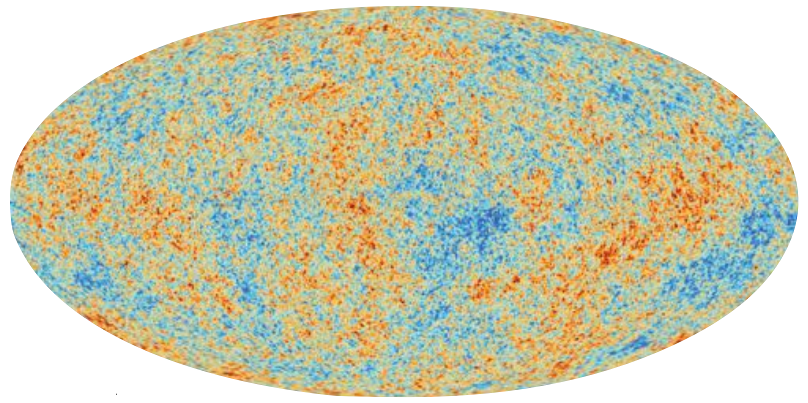
Particle colliders



Neuron activity

Epidemics

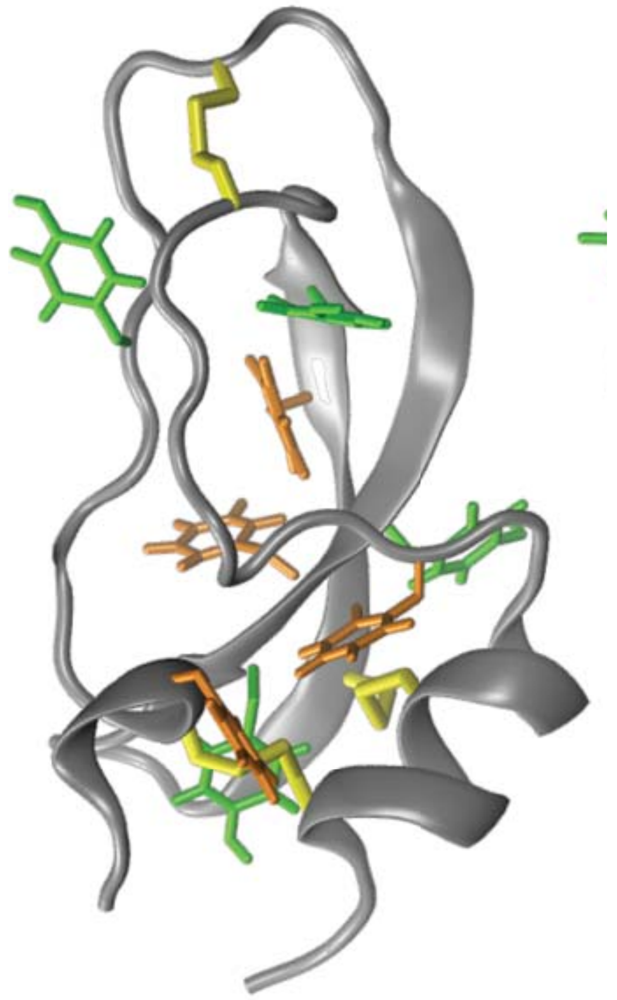
Gravitational lensing

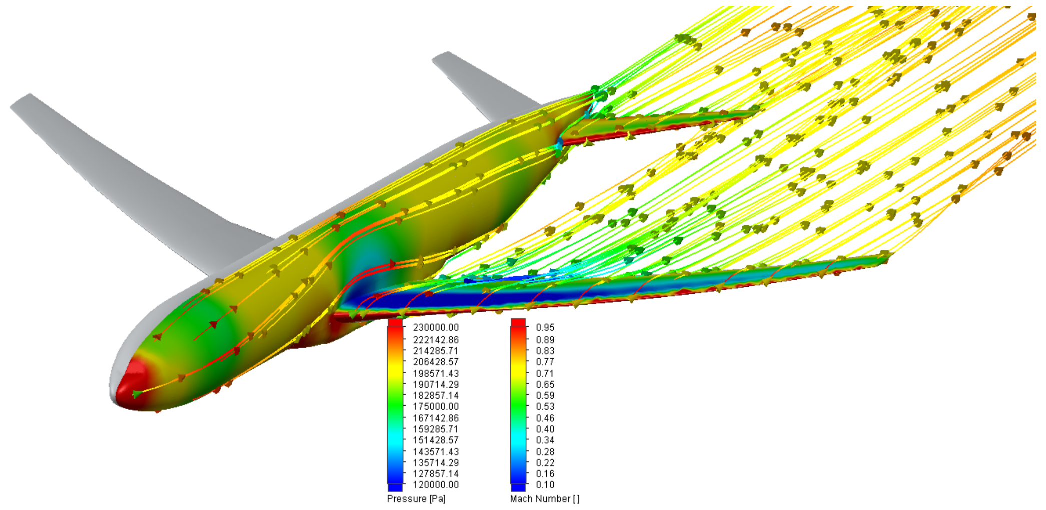


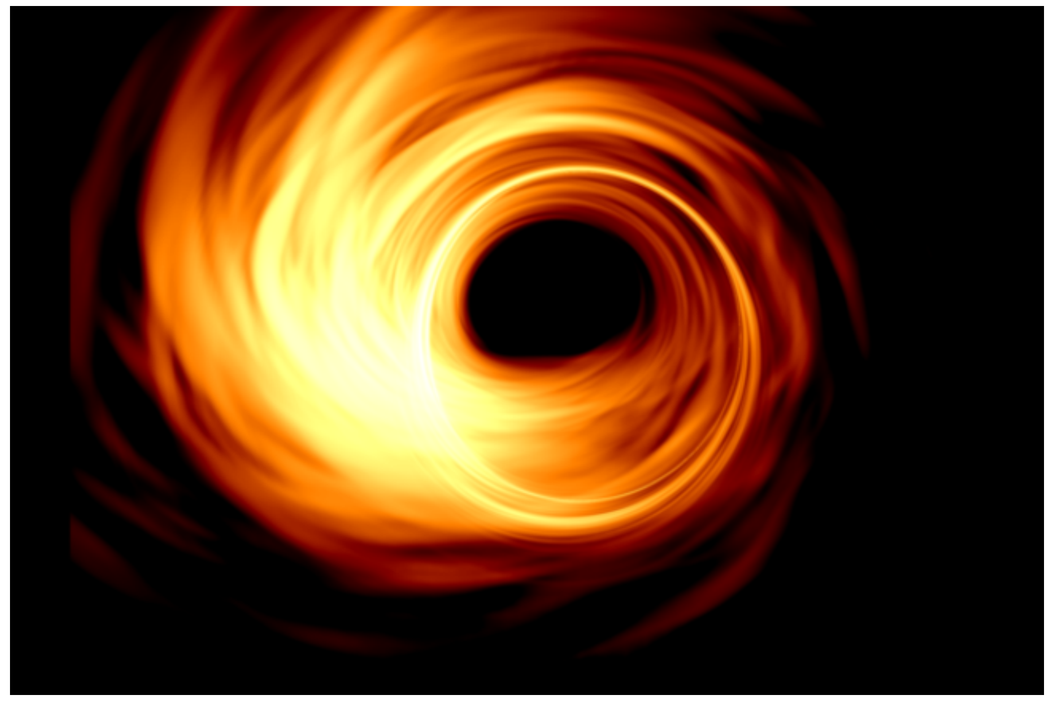
Evolution of the Universe

10^{-18} 10^{-15} 10^{-12} 10^{-9} 10^{-6} 10^{-3} 10^{0} 10^{3} 10^{6} 10^{9} 10^{12} 10^{15} 10^{18} 10^{21} 10^{24} 10^{27}

Length scale [m]







Simulators are implicit, causal, generative models that can produce synthetic data

[Cranmer, Brehmer, Louppe PNAS (2020), arXiv:1911.01429]

ICML 2017 Workshop on Implicit Models

Workshop Aims

Probabilistic models are an important tool in machine learning. They form the basis for models that generate realistic data, uncover hidden structure, and make predictions. Traditionally, probabilistic models in machine learning have focused on prescribed models. Prescribed models specify a joint density over observed and hidden variables that can be easily evaluated. The requirement of a tractable density simplifies their learning but limits their flexibility --- several real world phenomena are better described by simulators that do not admit a tractable density. Probabilistic models defined only via the simulations they produce are called implicit models.

Arguably starting with generative adversarial networks, research on implicit models in machine learning has exploded in recent years. This workshop's aim is to foster a discussion around the recent developments and future directions of implicit models.

Implicit models have many applications. They are used in ecology where models simulate animal populations over time; they are used in phylogeny, where simulations produce hypothetical ancestry trees; they are used in physics to generate particle simulations for high energy processes. Recently, implicit models have been used to improve the state-of-the-art in image and content generation. Part of the workshop's focus is to discuss the commonalities among applications of implicit models.

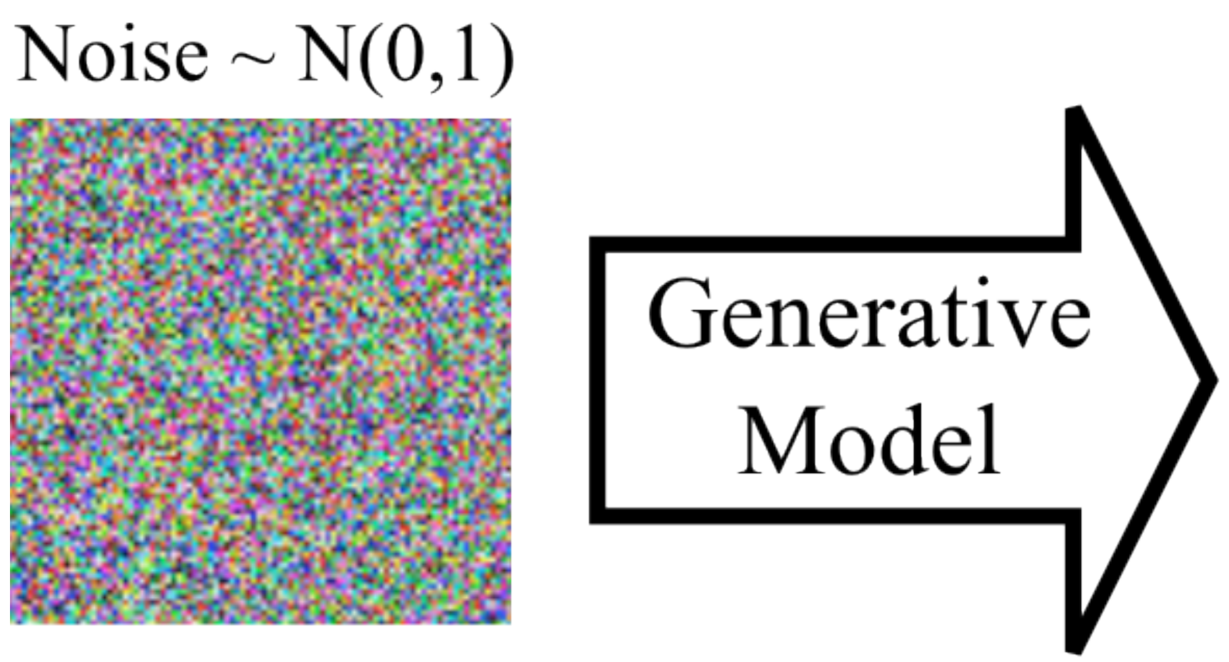
Of particular interest at this workshop is to unite fields that work on implicit models. For example:

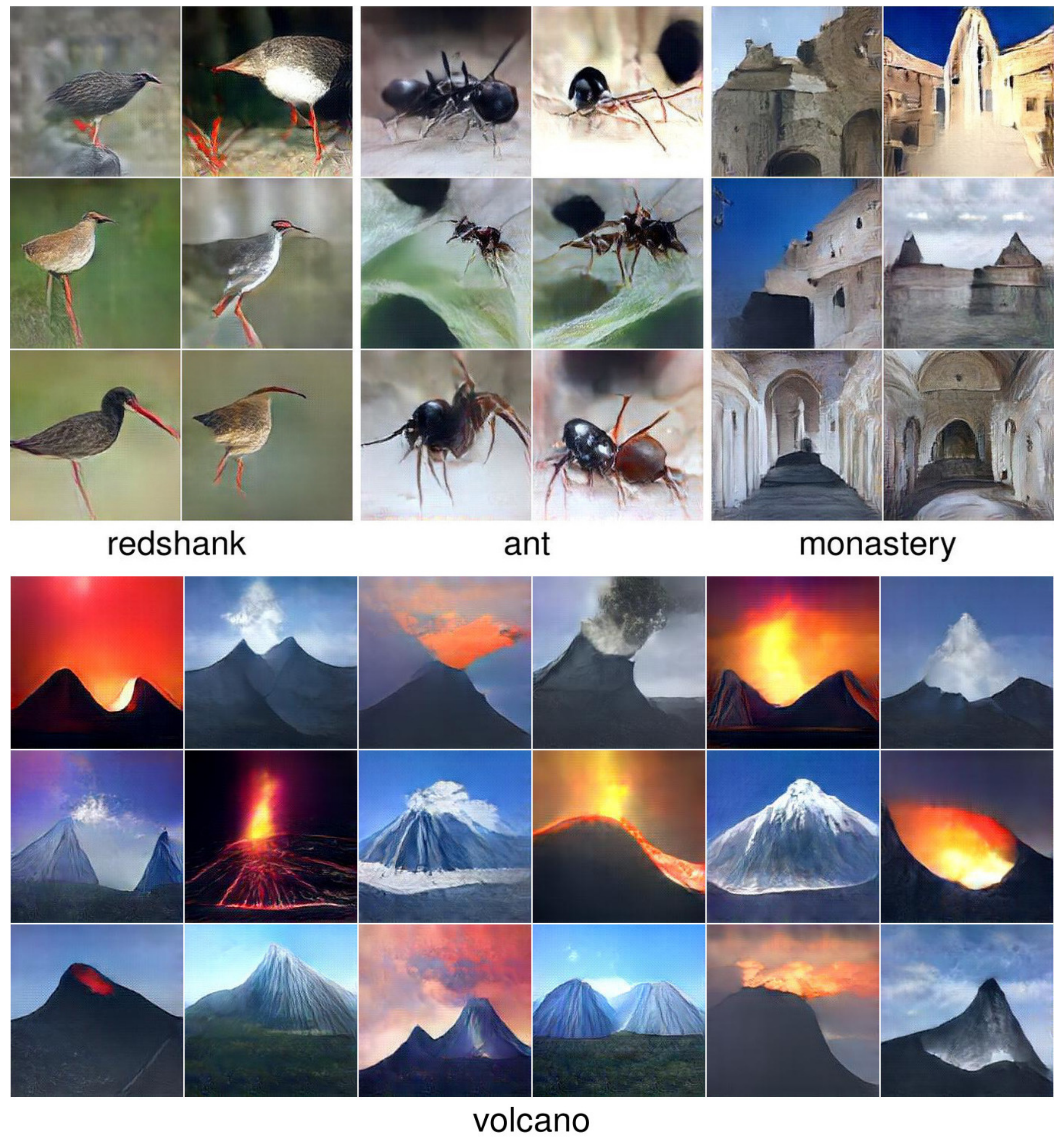
- **Generative adversarial networks** (a NIPS 2016 workshop) are implicit models with an adversarial training scheme.
- Recent advances in **variational inference** (a NIPS 2015 and 2016 workshop) have leveraged implicit models for more accurate approximations.
- **Approximate Bayesian computation** (a NIPS 2015 workshop) focuses on posterior inference for models with implicit likelihoods.
- Learning implicit models is deeply connected to **two sample testing, density ratio and density difference** estimation.

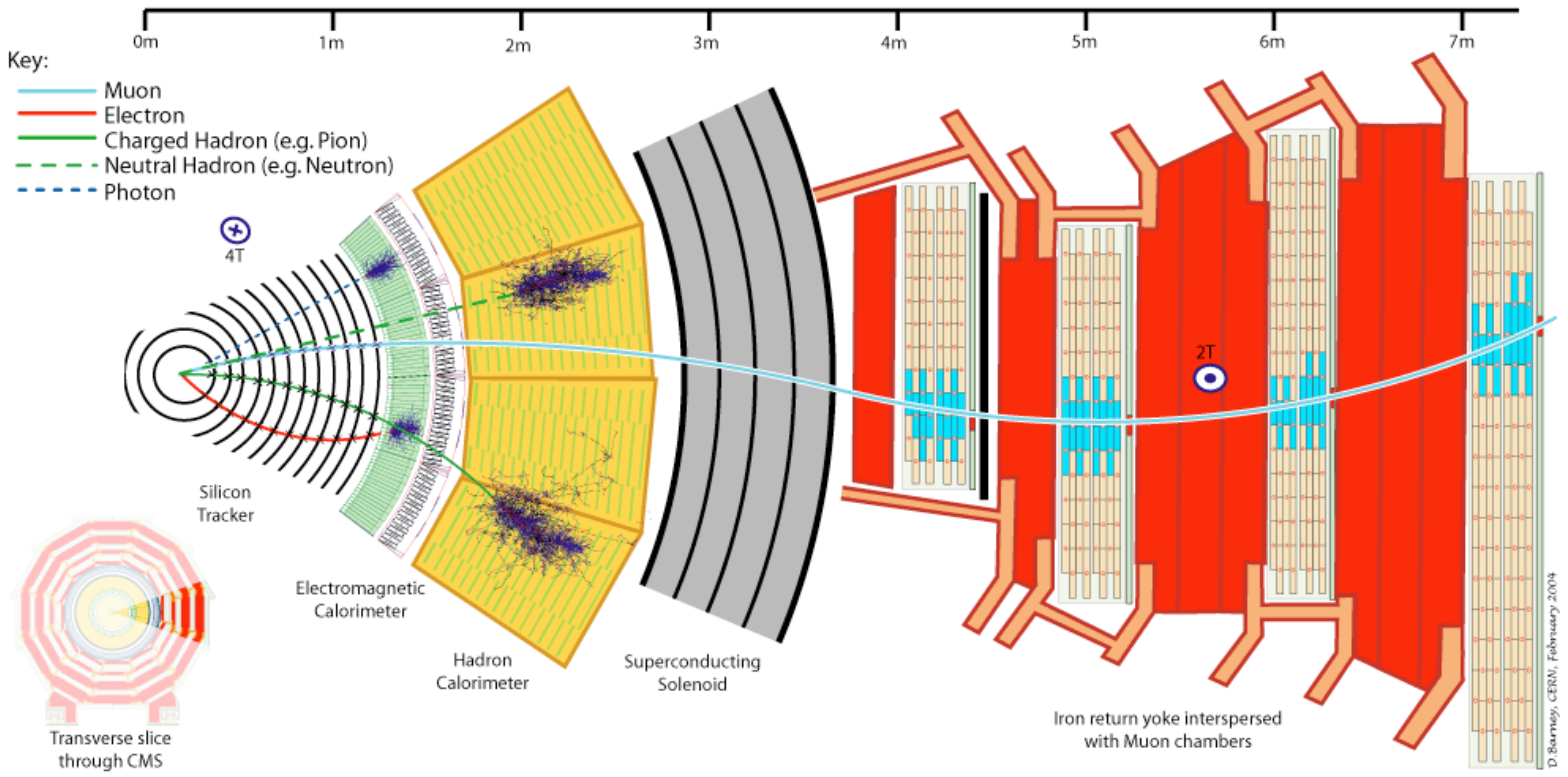
We hope to bring together these different views on implicit models, identifying their core challenges and combining their innovations.

Can we fit the simulator like a GAN?

z x







Adversarial Variational Optimization

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe, University of Liège
Joeri Hermans, University of Liège
Kyle Cranmer, New York University

Leo is G Tom is D

$$U_d(\phi) = \mathbb{E}_{\theta \sim q(\theta|\psi)}[\mathcal{L}_d(\phi)]$$
$$U_g(\psi) = \mathbb{E}_{\theta \sim q(\theta|\psi)}[\mathcal{L}_g(\theta)]$$

Similar to GAN setup, but instead of using a neural network as the generator, use the actual simulation

Continue to use a neural network discriminator / critic.

$$\mathcal{L}_d(\phi) = \mathbb{E}_{\mathbf{x}\sim p_r(\mathbf{x})}[-\log(d(\mathbf{x};\phi))] + \mathbb{E}_{\tilde{\mathbf{x}}\sim p(\mathbf{x}|\theta)}[-\log(1-d(\tilde{\mathbf{x}};\phi))]$$

$$\mathcal{L}_g(\theta) = \mathbb{E}_{\tilde{\mathbf{x}}\sim p(\mathbf{x}|\theta)}[\log(1-d(\tilde{\mathbf{x}};\phi))],$$

Difficulty: the simulator isn't differentiable, but there's a **trick**!
Variational optimization + REINFORCE gradients

$$\min_{\theta} f(\theta) \leq \mathbb{E}_{\theta\sim q(\theta|\psi)}[f(\theta)] = U(\psi)$$
$$\nabla_\psi U(\psi) = \mathbb{E}_{\theta\sim q(\theta|\psi)}[f(\theta)\nabla_\psi \log q(\theta|\psi)]$$

Proposal $q(\theta|\psi)$ concentrates around θ^*. Allows us to efficiently **fit the simulation** with stochastic gradient techniques!

But point estimates only, not posterior $p(\theta|x)$ 😞

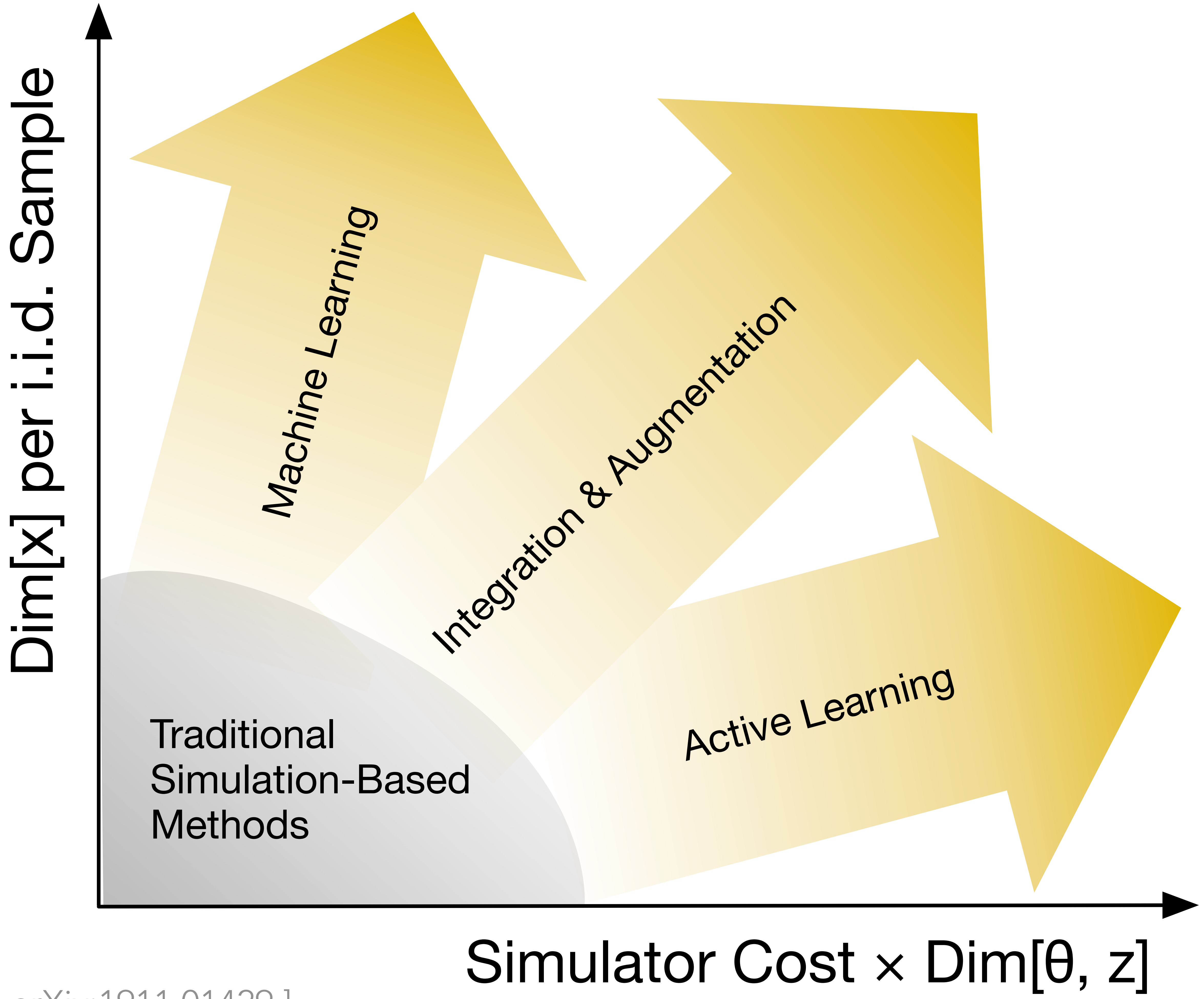
Published in Proceedings of the National Academy of Sciences

The frontier of simulation-based inference

Kyle Cranmer[a,b,1], **Johann Brehmer**[a,b], **and Gilles Louppe**[c]

[a]Center for Cosmology and Particle Physics, New York University, USA; [b]Center for Data Science, New York University, USA; [c]Montefiore Institute, University of Liège, Belgium

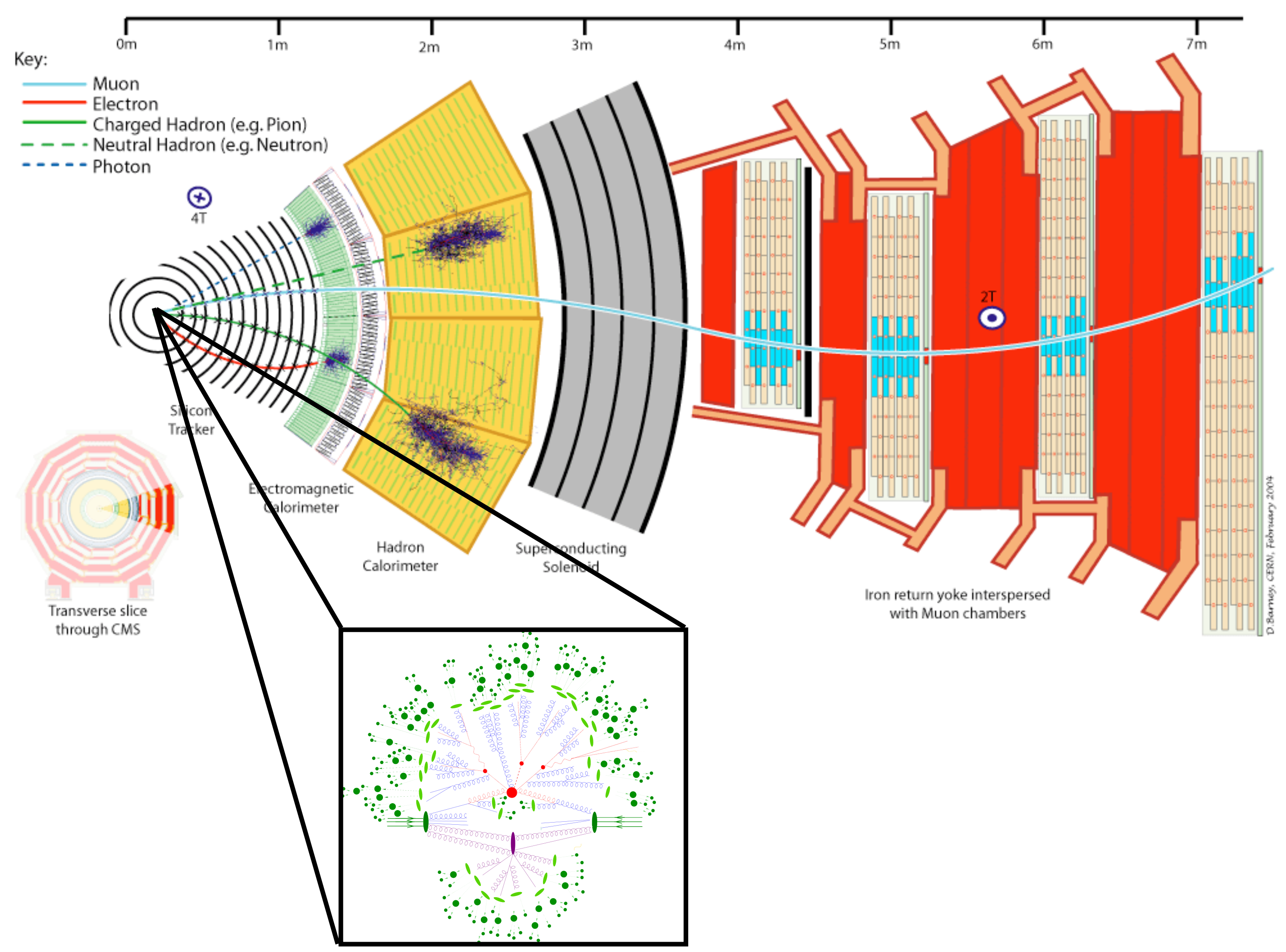
April 3, 2020



[Cranmer, Brehmer, Louppe PNAS (2020), arXiv:1911.01429]

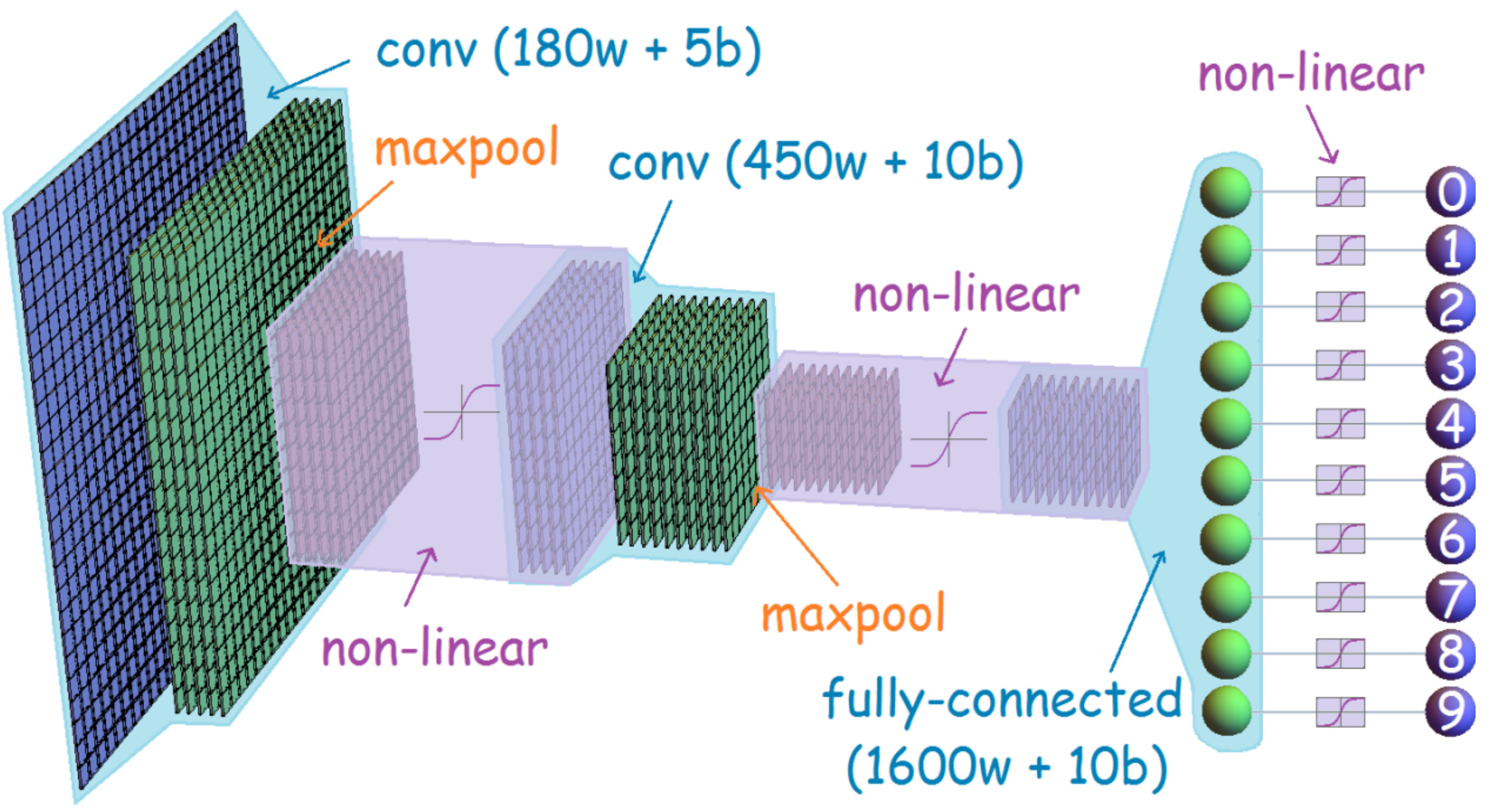
Two approaches simulation-based inference

Use simulator
(much more efficiently)



- Approximate Bayesian Computation (ABC)
- Probabilistic Programming
- Adversarial Variational Optimization

Learn simulator
(with deep learning)

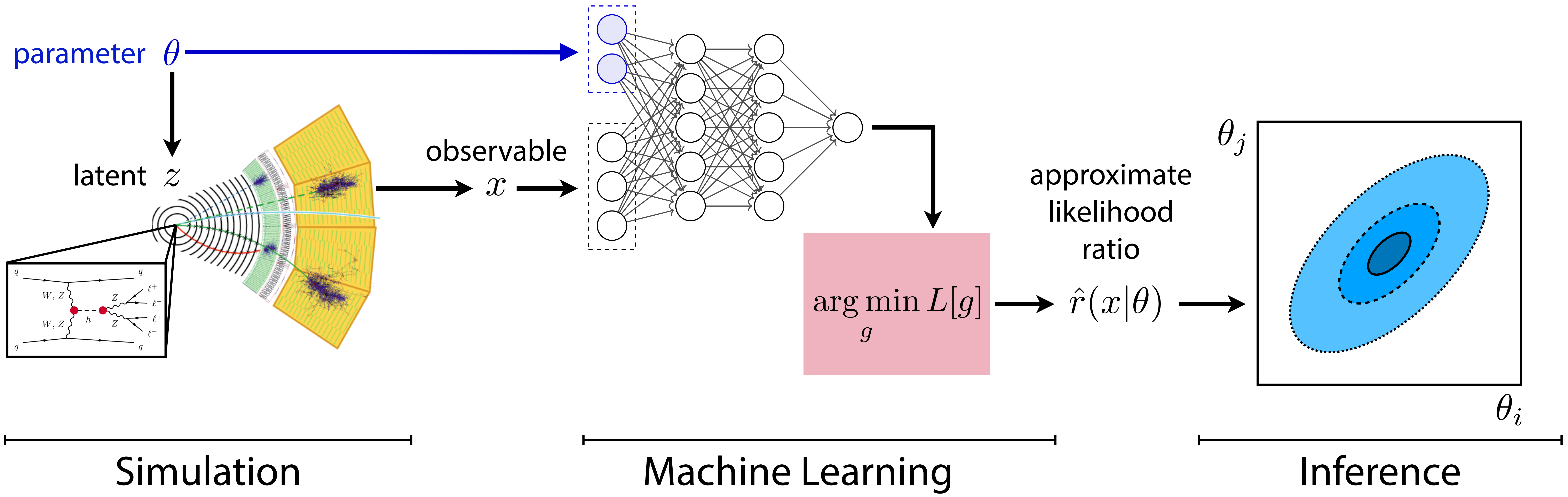


- Likelihood ratio trick (with classifiers)
- Conditional density estimate (with normalizing flows)
- Learned summary statistics

[image credit: A.P. Goucher]

Going beyond engineered summary statistics

arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
arXiv:1808.00973
physics.aps.org/articles/v11/90



The **surrogate for the likelihood (ratio)** used for inference

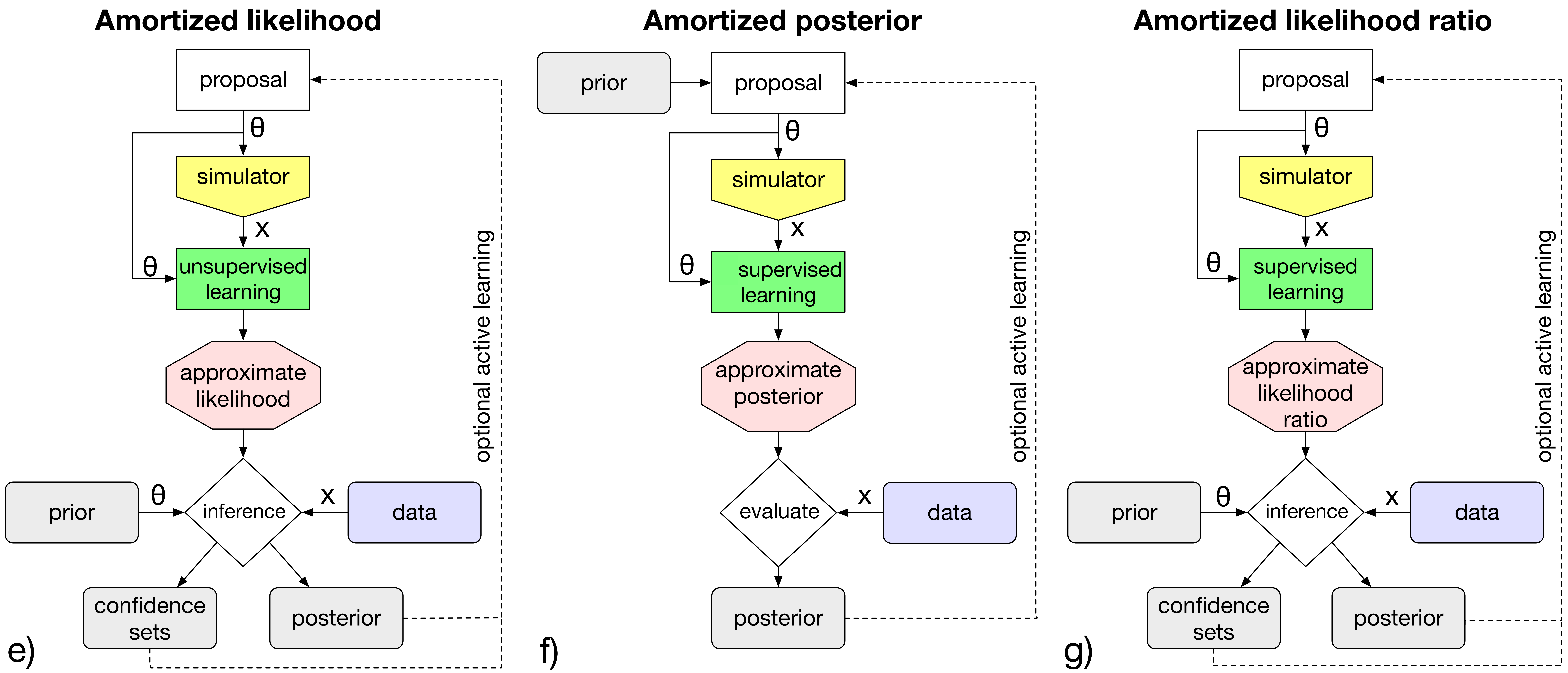
A 2-stage process:

1. learning surrogate (**amortized**)
2. Inference on parameters of simulator

Likelihoods, likelihood ratios, and posteriors

We can target different quantities to be used as a surrogate for inference

- **Unsupervised learning** typically used to learn likelihood with density estimation
- **Supervised learning** can be used to learn likelihood ratio (or posterior)



Examples

Simulating particle physics processes

$$
\begin{aligned}
\mathcal{L}_{SM} = \quad & \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{kinetic energies and self-interactions of the gauge bosons}} \\
+ \; & \underbrace{\bar{L}\gamma^\mu(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)L + \bar{R}\gamma^\mu(i\partial_\mu - \frac{1}{2}g'YB_\mu)R}_{\text{kinetic energies and electroweak interactions of fermions}} \\
+ \; & \underbrace{\frac{1}{2}\left|(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)\,\phi\right|^2 - V(\phi)}_{W^\pm, Z, \gamma, \text{and Higgs masses and couplings}} \\
+ \; & \underbrace{g''(\bar{q}\gamma^\mu T_a q)\,G^a_\mu}_{\text{interactions between quarks and gluons}} + \underbrace{(G_1\bar{L}\phi R + G_2\bar{L}\phi_c R + h.c.)}_{\text{fermion masses and couplings to Higgs}}
\end{aligned}
$$

Simulating particle physics processes

Simulating particle physics processes

Theory parameters

θ

Evolution

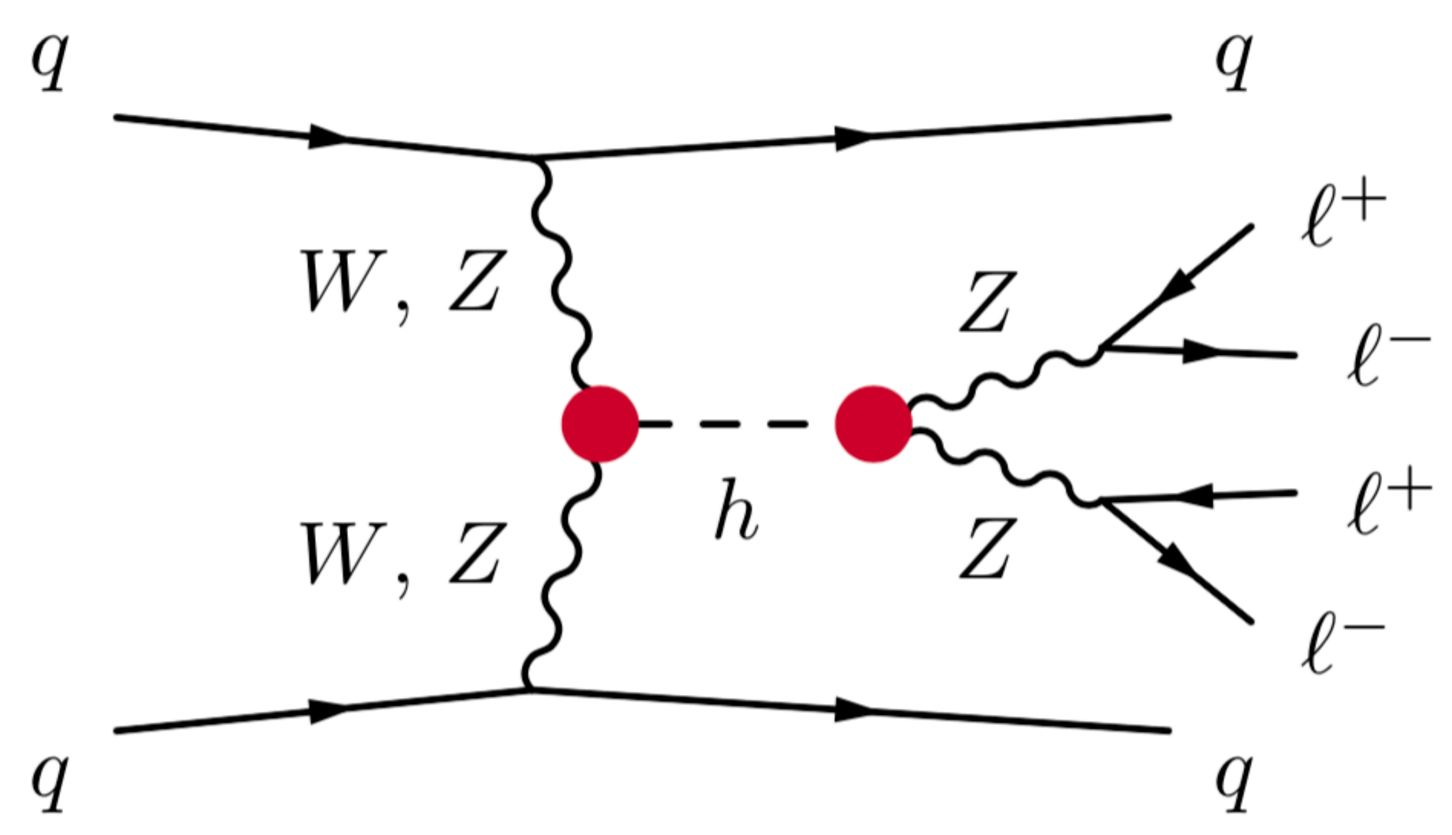
Simulating particle physics processes

Latent variables

Parton-level momenta

Theory parameters

$z_p \longleftarrow \theta$

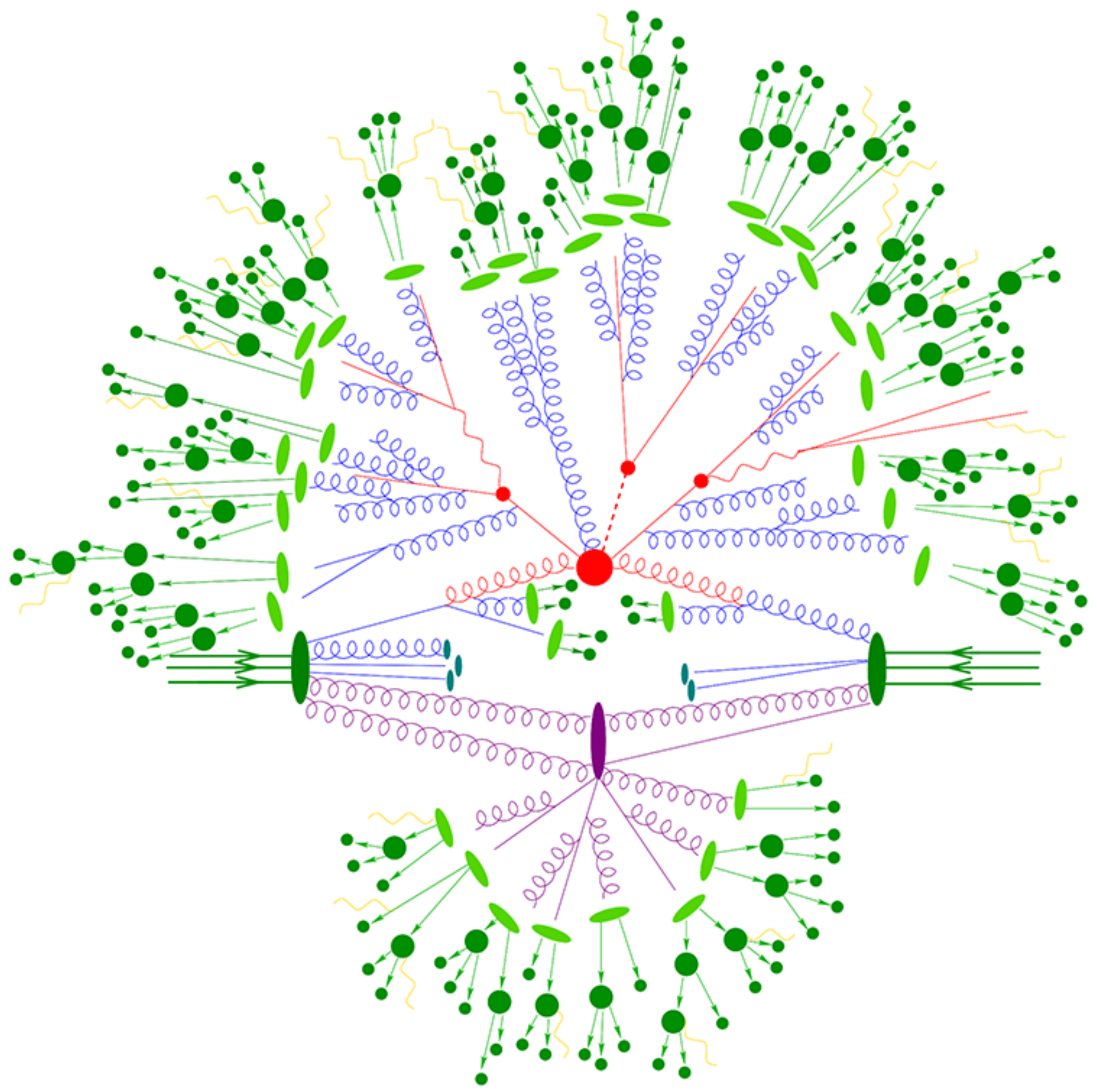


Evolution

Simulating particle physics processes

Latent variables

Shower splittings		Parton-level momenta		Theory parameters
z_s	$\longleftarrow$	z_p	$\longleftarrow$	θ



[F. Krauss]

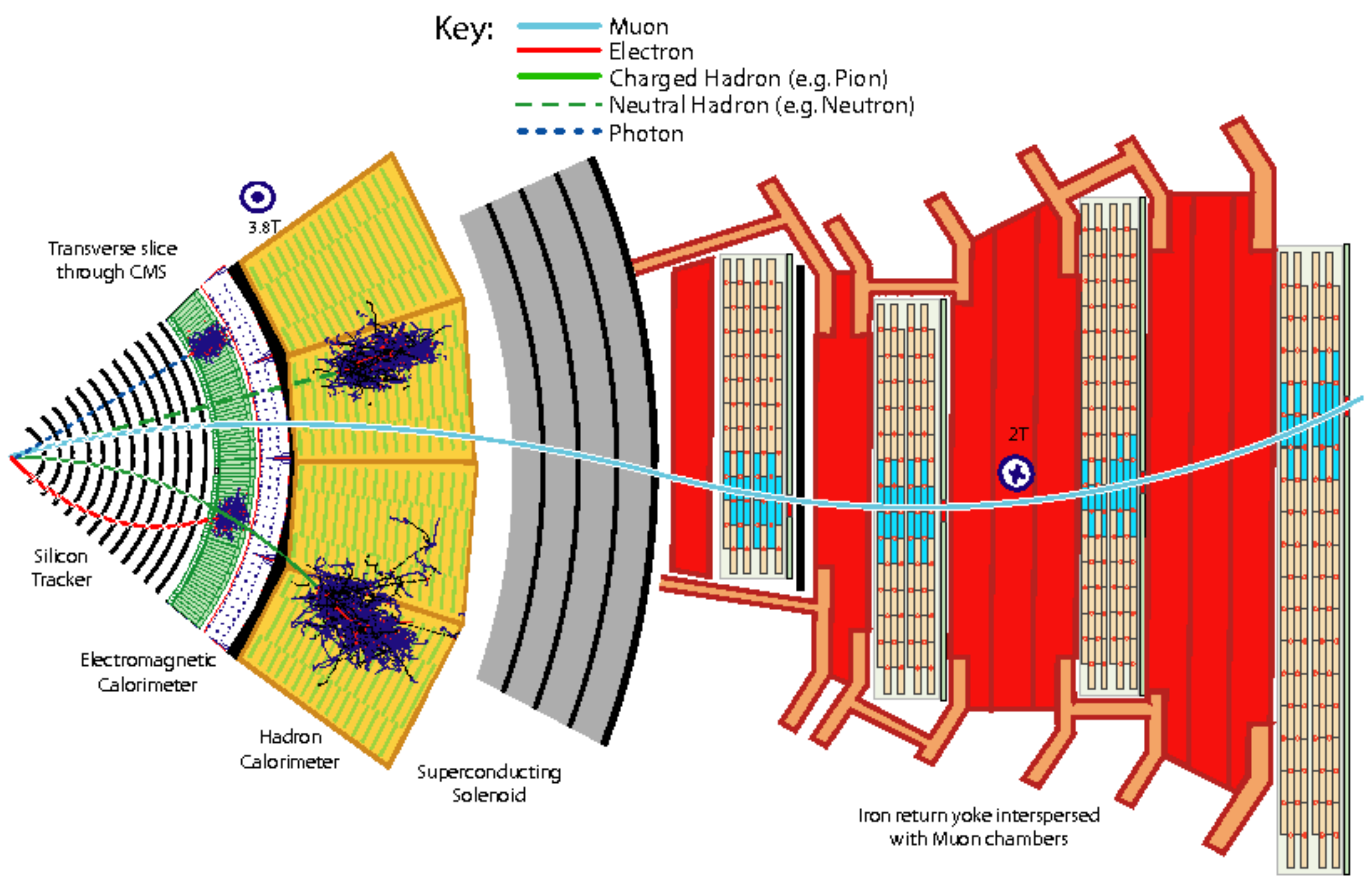
$\longleftarrow$

Evolution

Simulating particle physics processes

Latent variables

Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
z_d	$\leftarrow$	z_s	$\leftarrow$	z_p	$\leftarrow$	θ



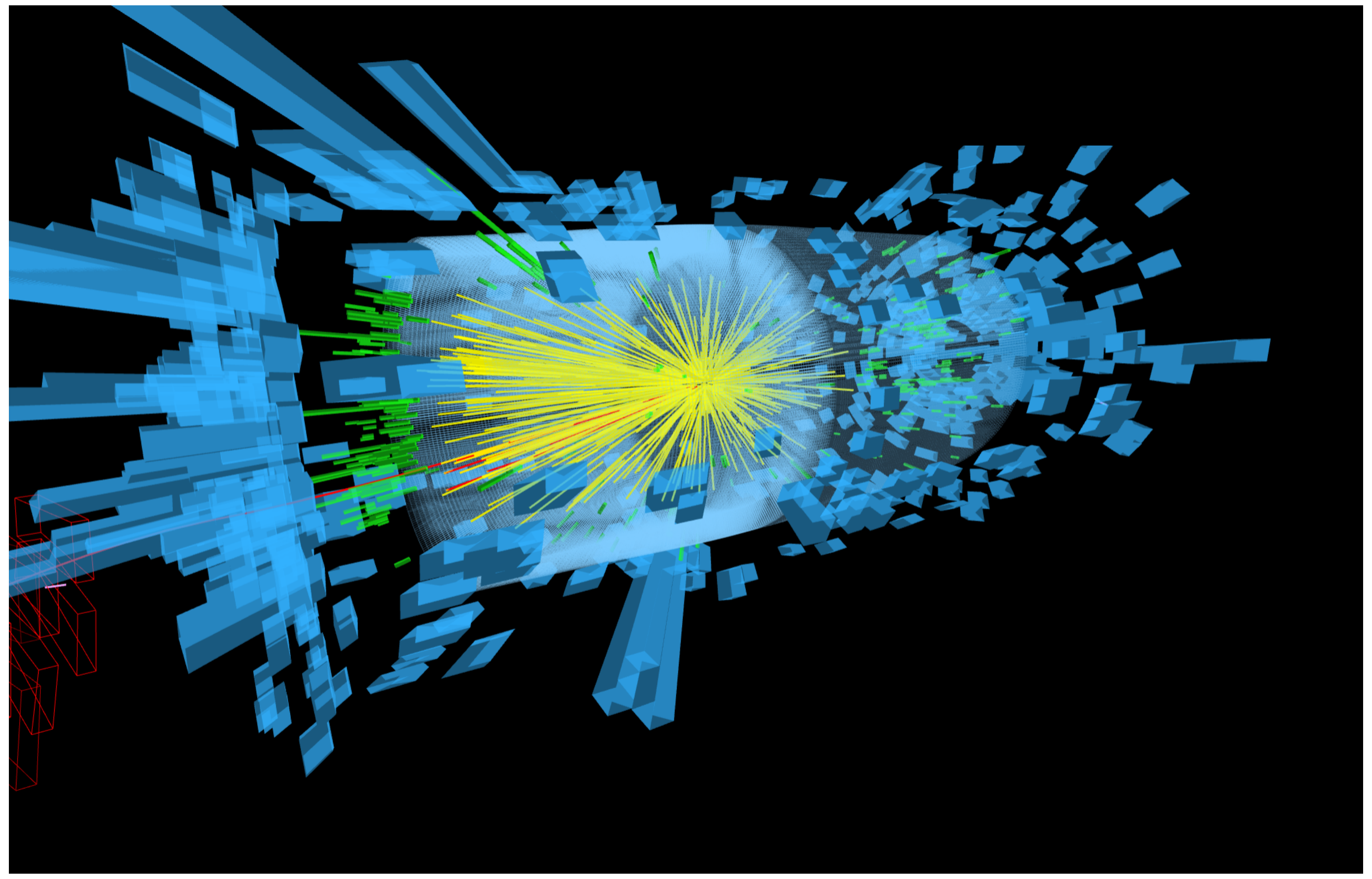
[CMS]

$\longleftarrow$ Evolution

Simulating particle physics processes

Latent variables

Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
x ←	z_d ←	z_s ←	z_p ←	θ



← Evolution

Simulating particle physics processes

Latent variables

Observables		Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
x	$\longleftarrow$	z_d $\longleftarrow$	z_s $\longleftarrow$	z_p $\longleftarrow$	θ
Sample from	$p(x\|z_d)$	$p(z_d\|z_s)$	$p(z_s\|z_p)$	$p(z_p\|\theta)$	

GEANT4
A SIMULATION TOOLKIT

MADGRAPH5_aMC@NLO

Prediction (simulation)

Simulating particle physics processes

Latent variables

Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters

$$x \longleftarrow z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \quad p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

Inference →

Simulating particle physics processes

Latent variables

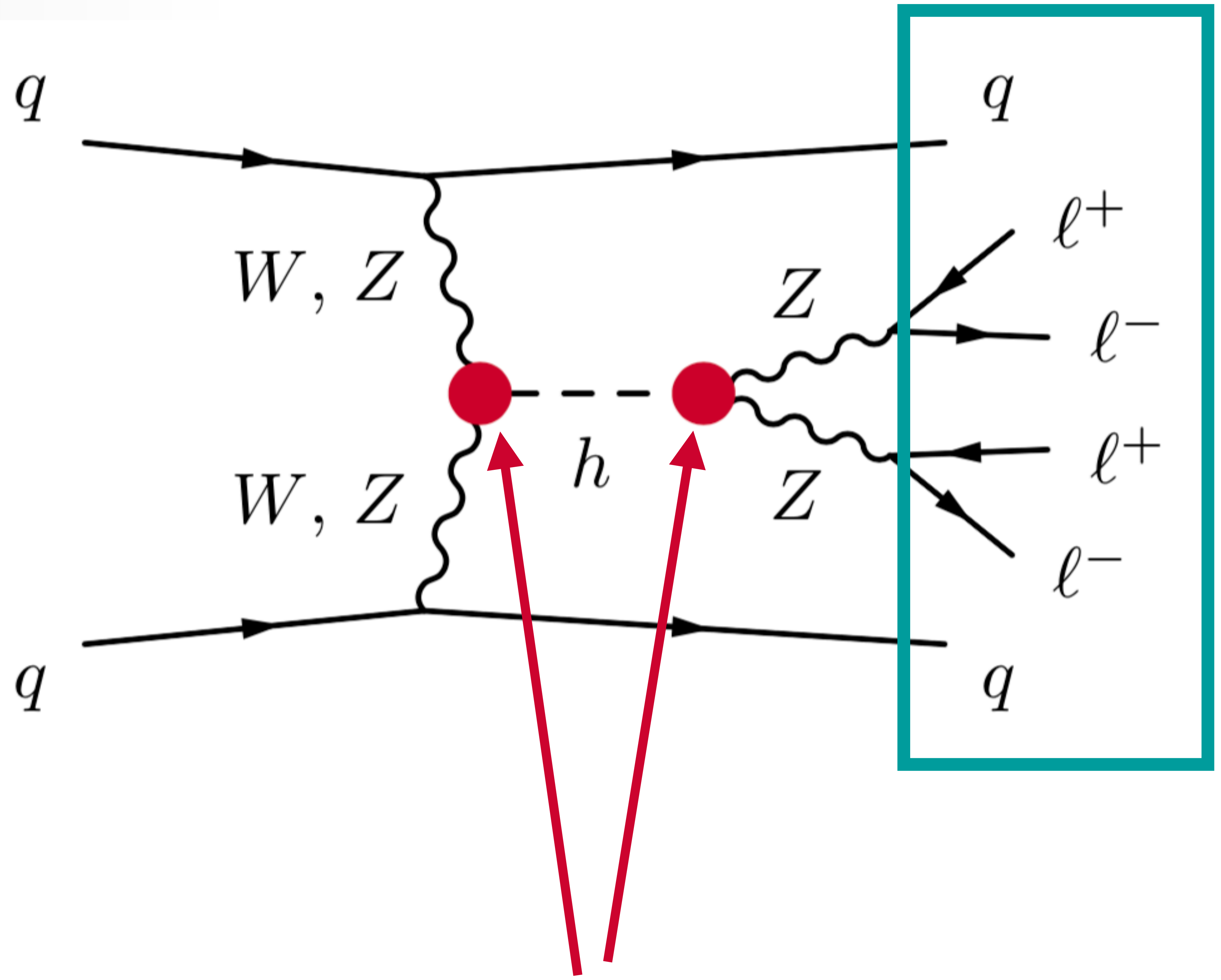
Observables		Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
x	←	z_d	←	z_s	←	z_p	←	θ

$$p(x|\theta) = \boxed{\int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p} \; p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

It's infeasible to calculate the integral over this enormous space!

→ Inference

Impact on Studies of The Higgs Boson



42-Dim observable **x**

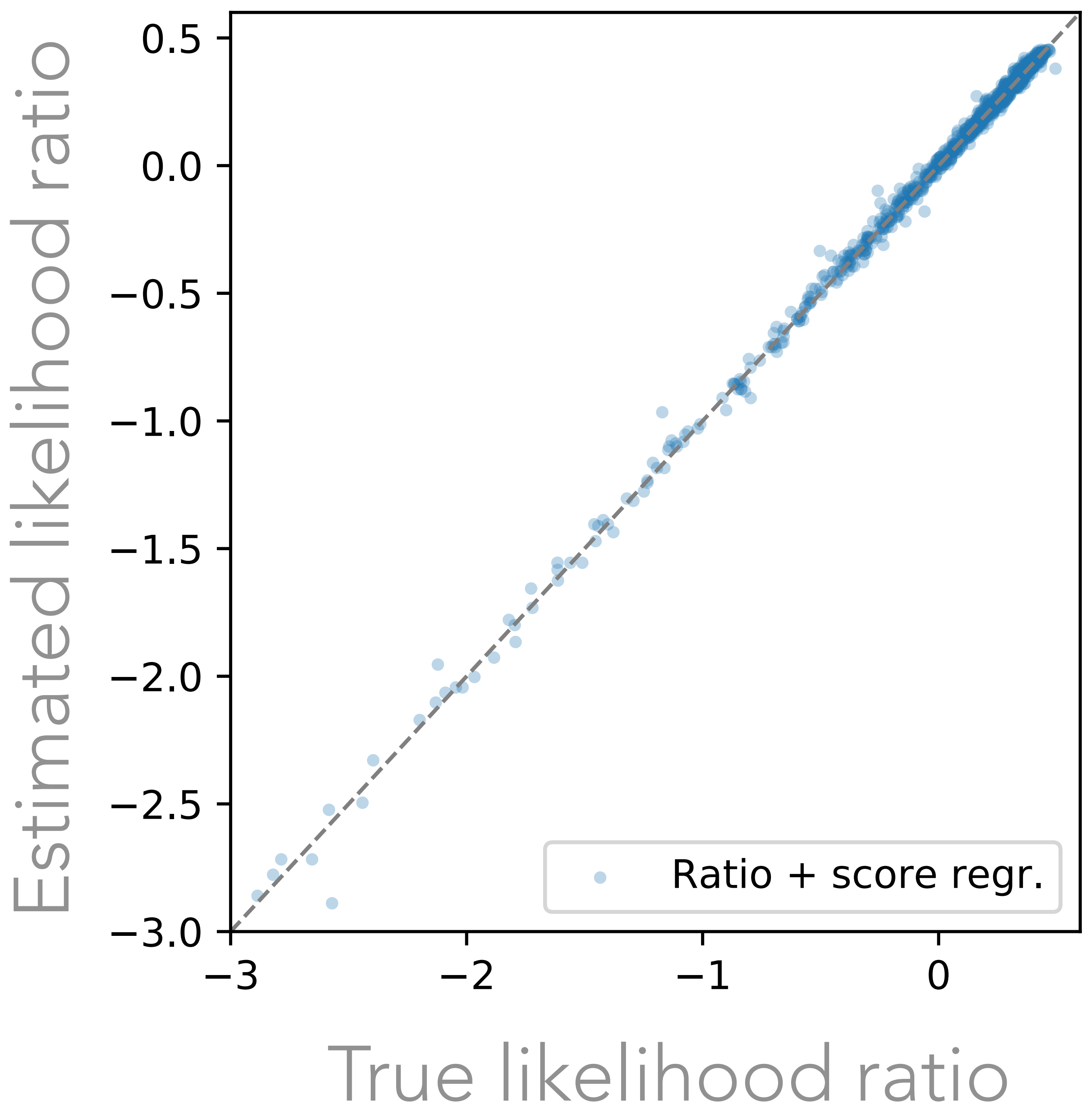
Exciting new physics might hide here!
We parameterize it with two coefficients:

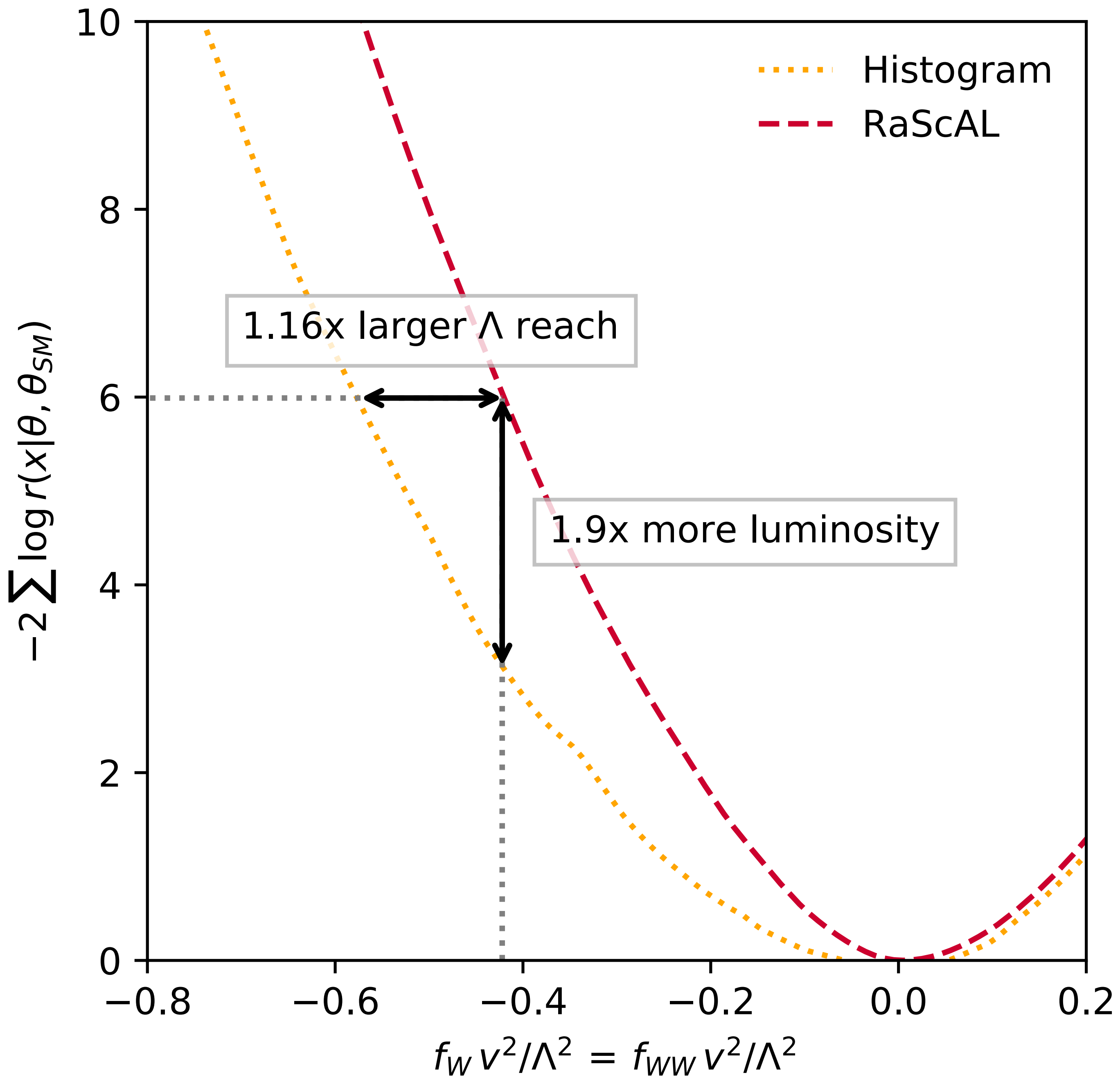
$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \boxed{\frac{f_W}{\Lambda^2}} \underbrace{\frac{\mathrm{i}g}{2} \, (D^\mu \phi)^\dagger \, \sigma^a \, D^\nu \phi \, W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}} \underbrace{\frac{g^2}{4} \, (\phi^\dagger \phi) \, W^a_{\mu\nu} \, W^{\mu\nu\, a}}_{\mathcal{O}_{WW}}$$

J Brehmer, J Pavez, G Louppe, K.C. PRL & PRD 2018 [arXiv:1805.00013 & arXiv:1805.00020], CARL [arxiv:1506.02169]

Impact on Studies of The Higgs Boson

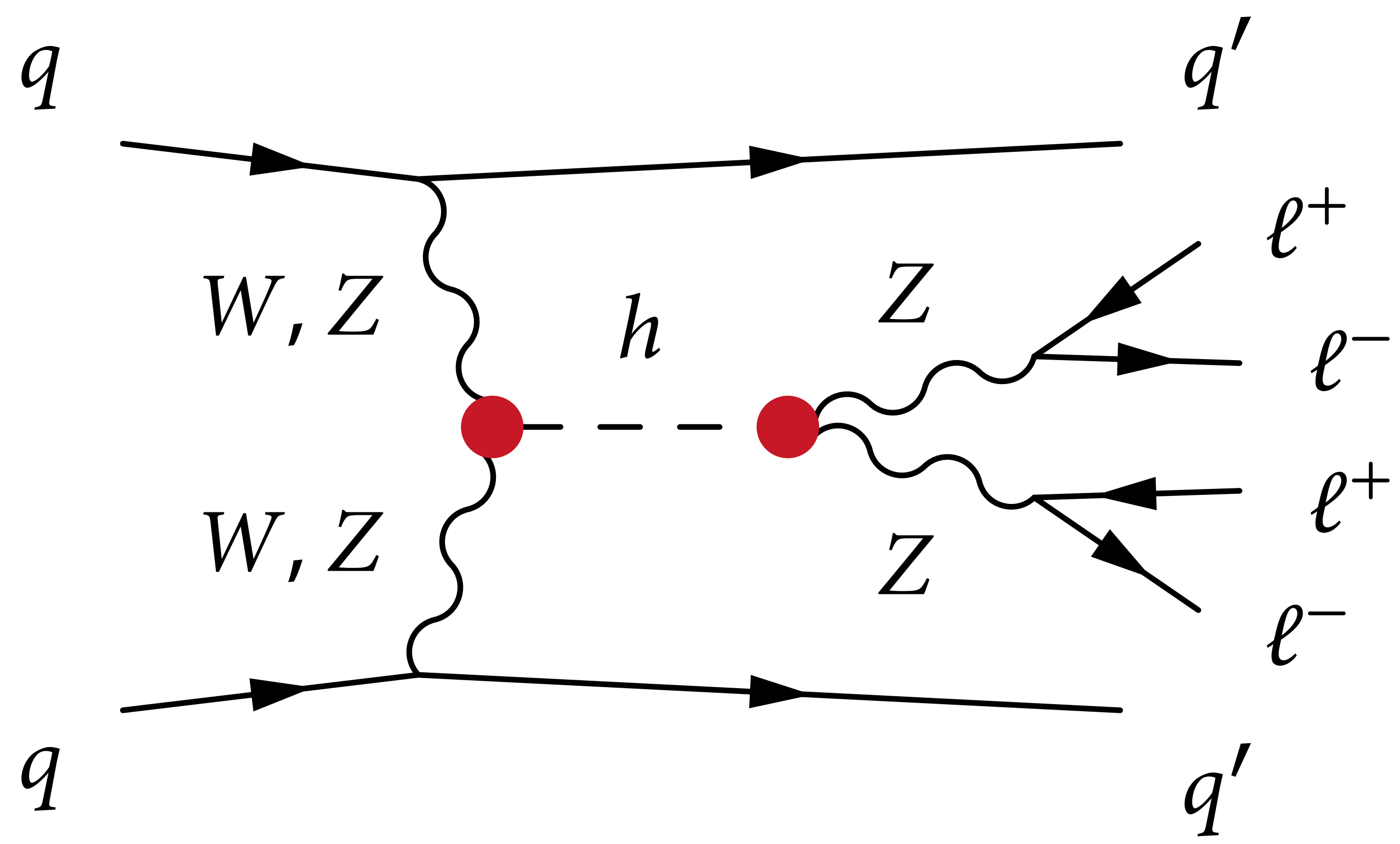
(based on a 42-Dim observation $\boldsymbol{x}$)





Accurate likelihood ratio estimates without the need for summary statistics improves sensitivity significantly

- Equivalent to 90% more LHC data!

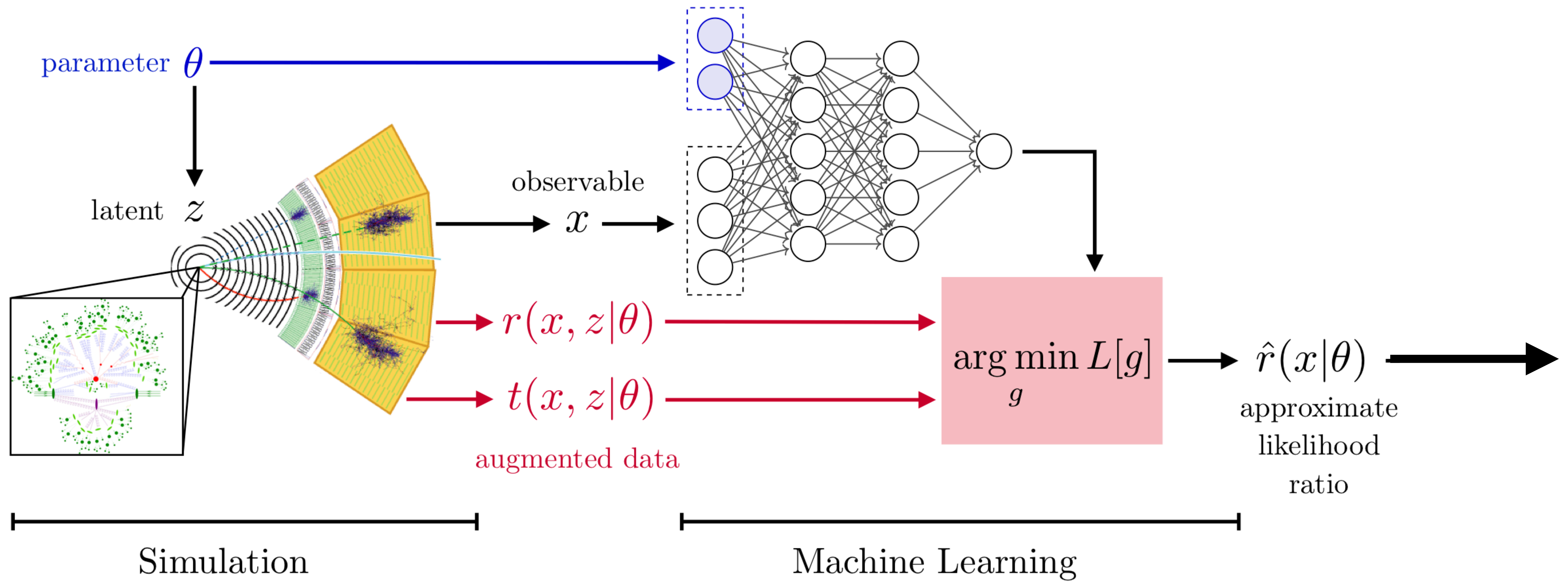


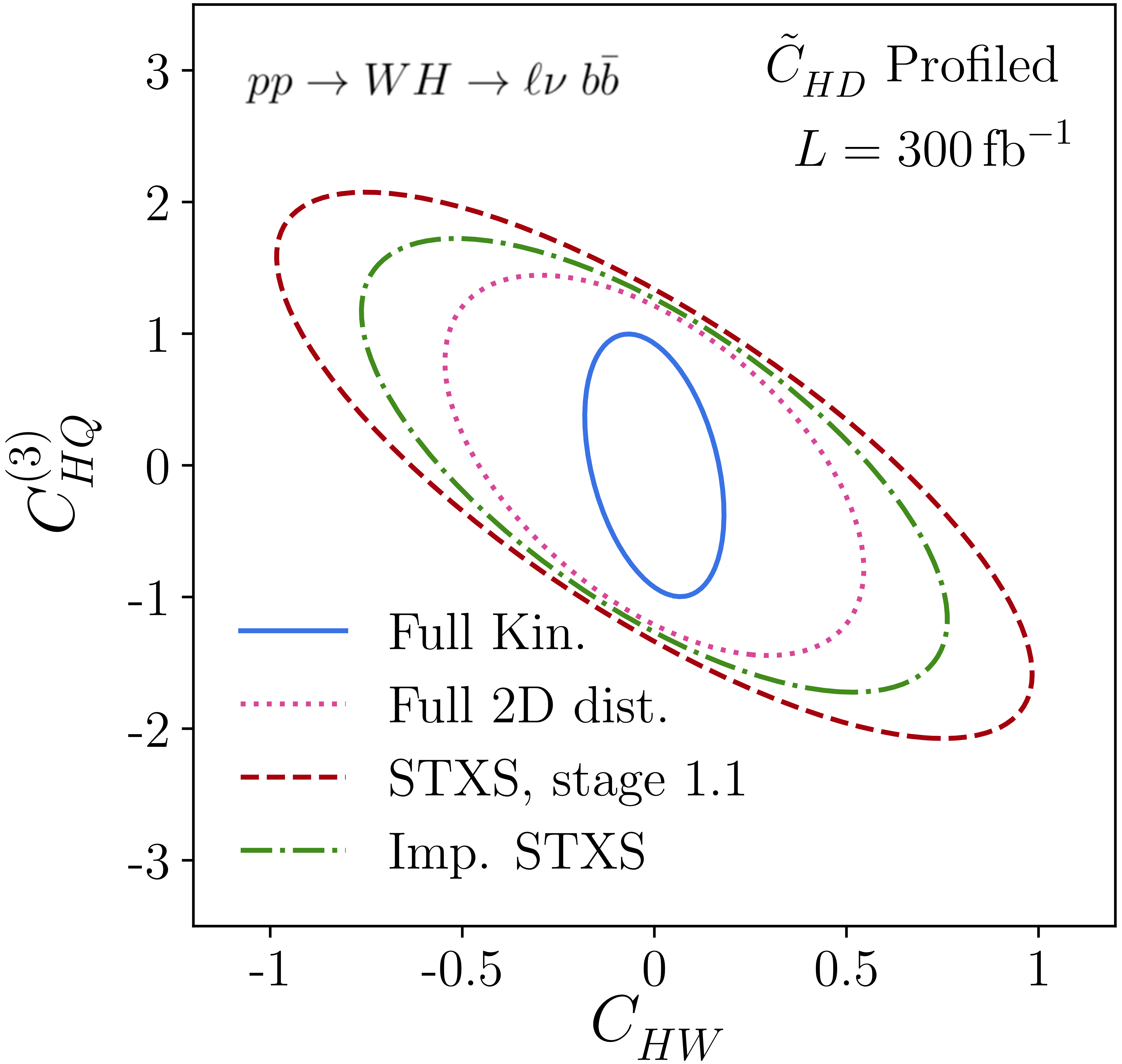
J Brehmer, J Pavez, G Louppe, K.C. PRL & PRD 2018 [arXiv:1805.00013 & arXiv:1805.00020], CARL [arxiv:1506.02169]

Impact on Studies of The Higgs Boson

Massive gains in precision of a flagship measurement at the LHC !

Equivalent increasing data collected by LHC by several factors



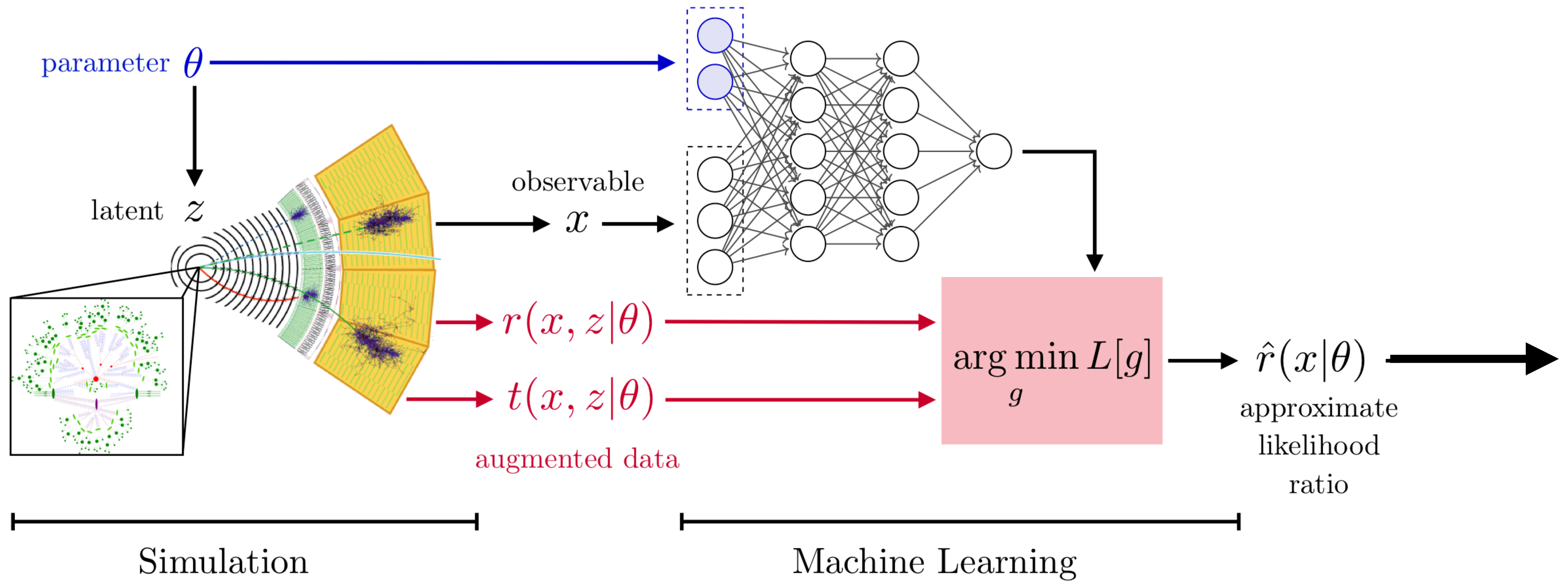


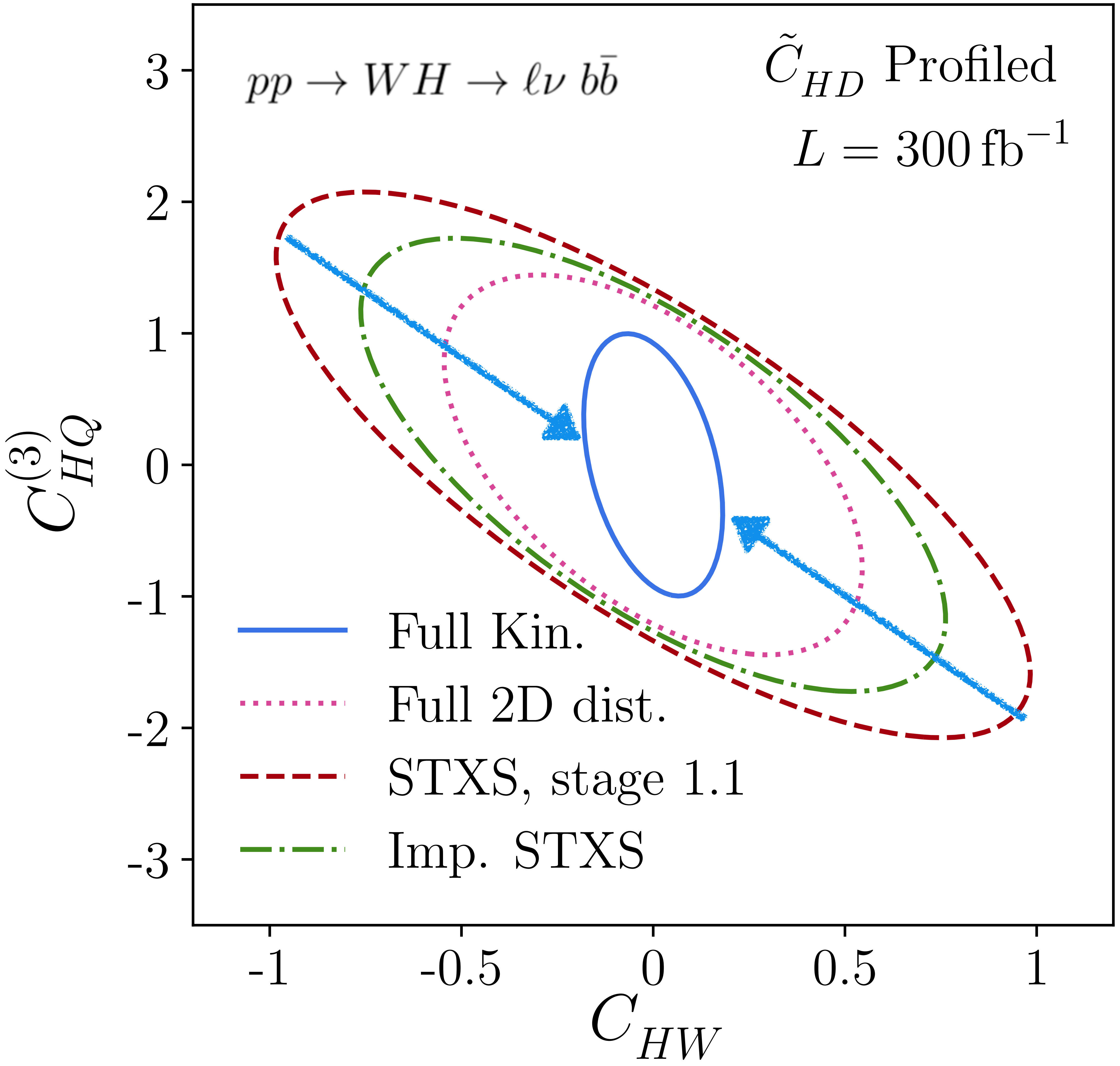
[J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn 1908.06980]

[J. Brehmer, F. Kling, I. Espejo, K. Cranmer 1907.10621]

Massive gains in precision of a flagship measurement at the LHC !

Equivalent increasing data collected by LHC by several factors



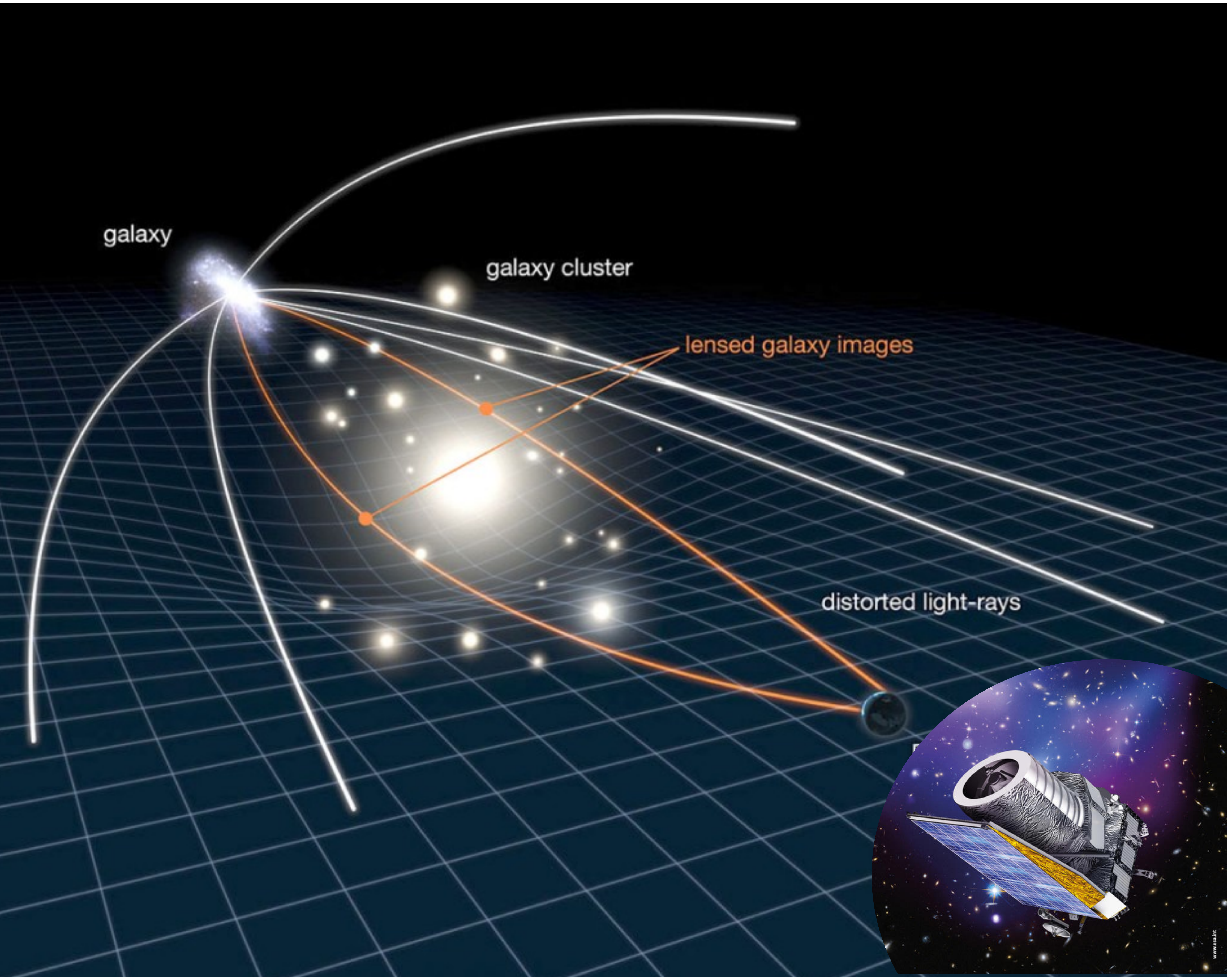


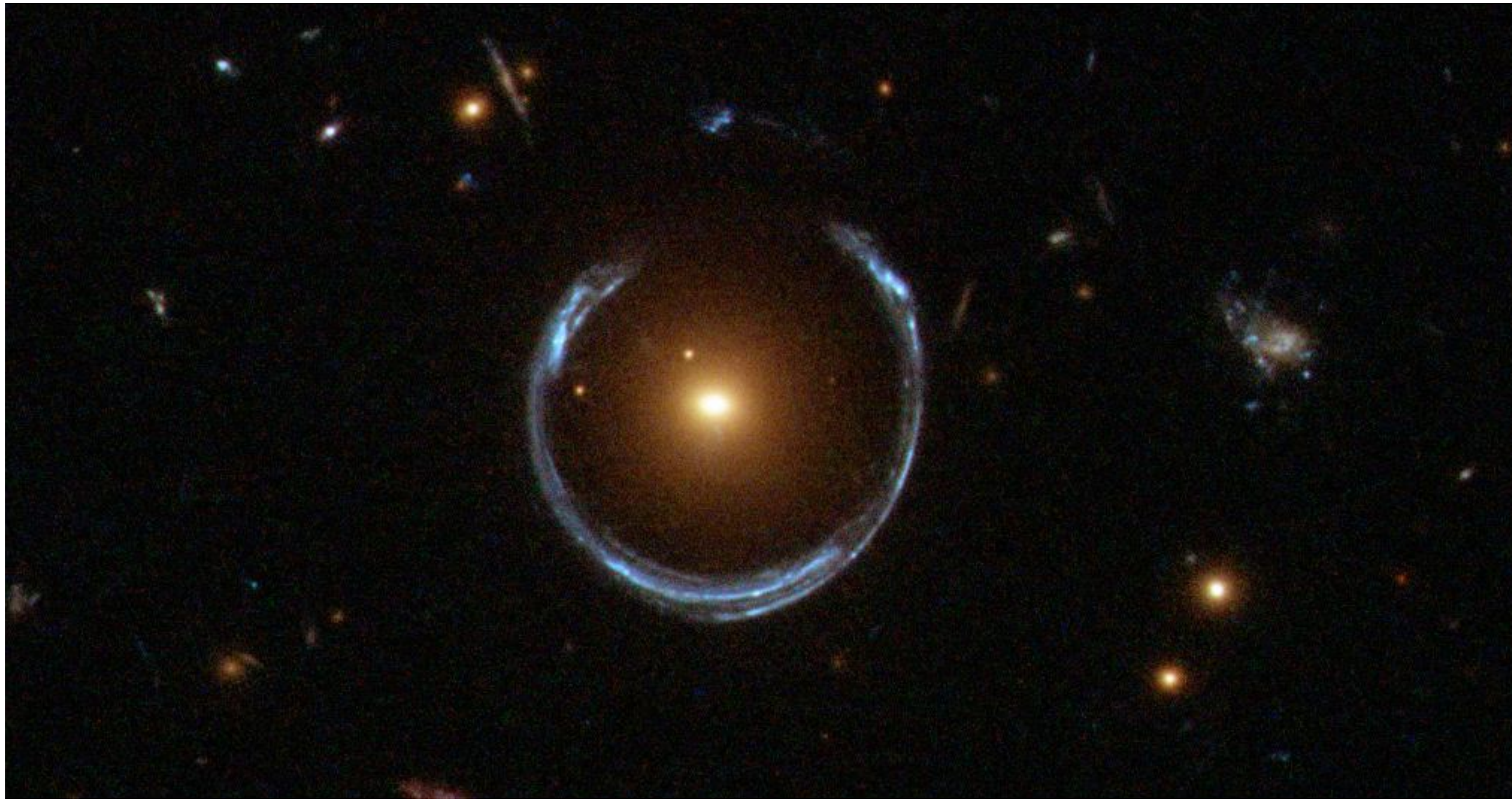
[J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn 1908.06980]

[J. Brehmer, F. Kling, I. Espejo, K. Cranmer 1907.10621]

Impact on science: Dark Matter

We can learn about the particle nature of dark matter by looking for subtle statistical signatures in images of gravitationally lensed galaxies

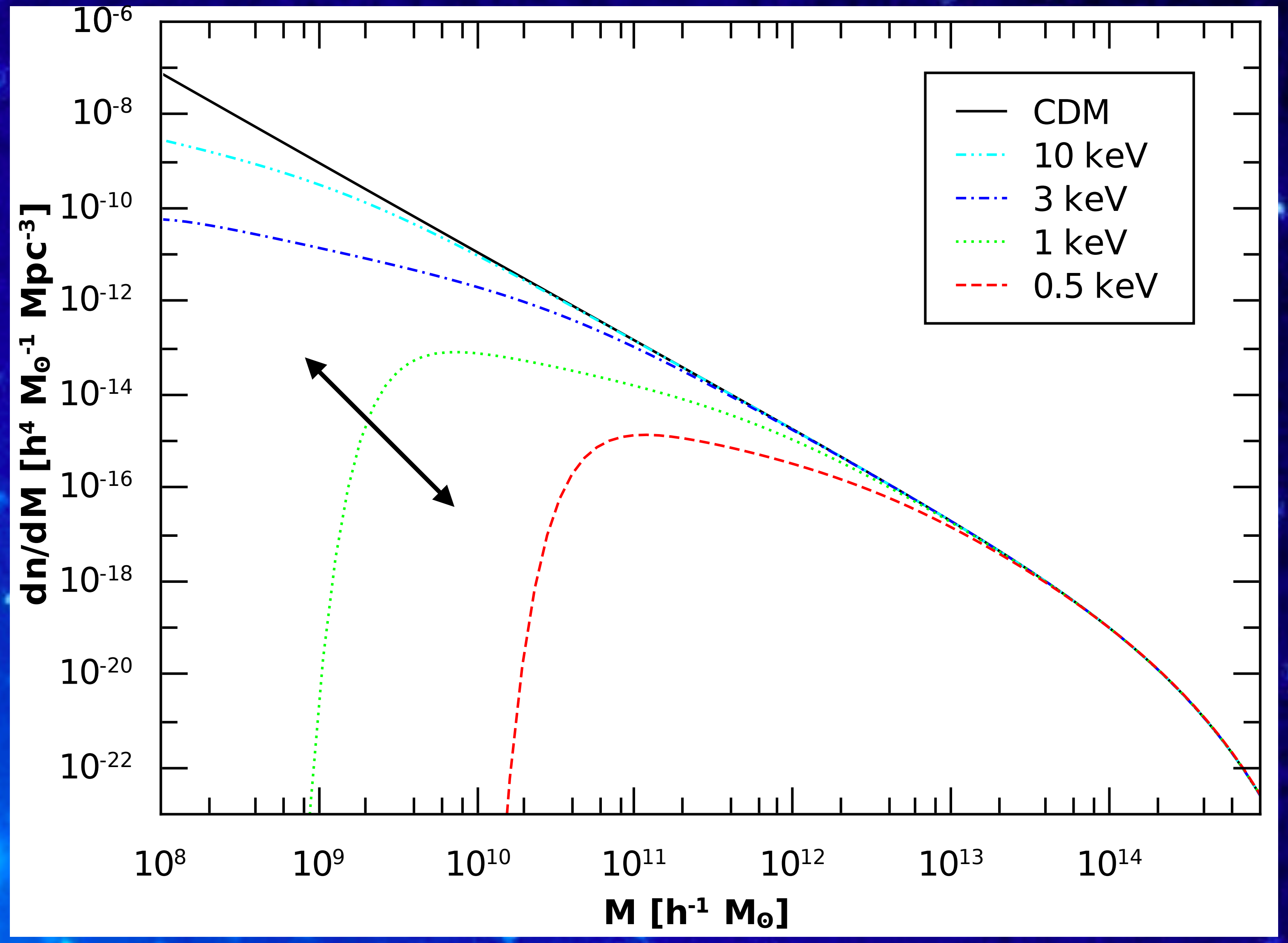




Sid Mishra-Sharma Johann Brehmer Gilles Louppe Joeri Hermans , K.C. published in ApJ, 886 (2019) arXiv:1909.02005]

Dark Matter Substructure

Abundance of DM subhalos vs mass:

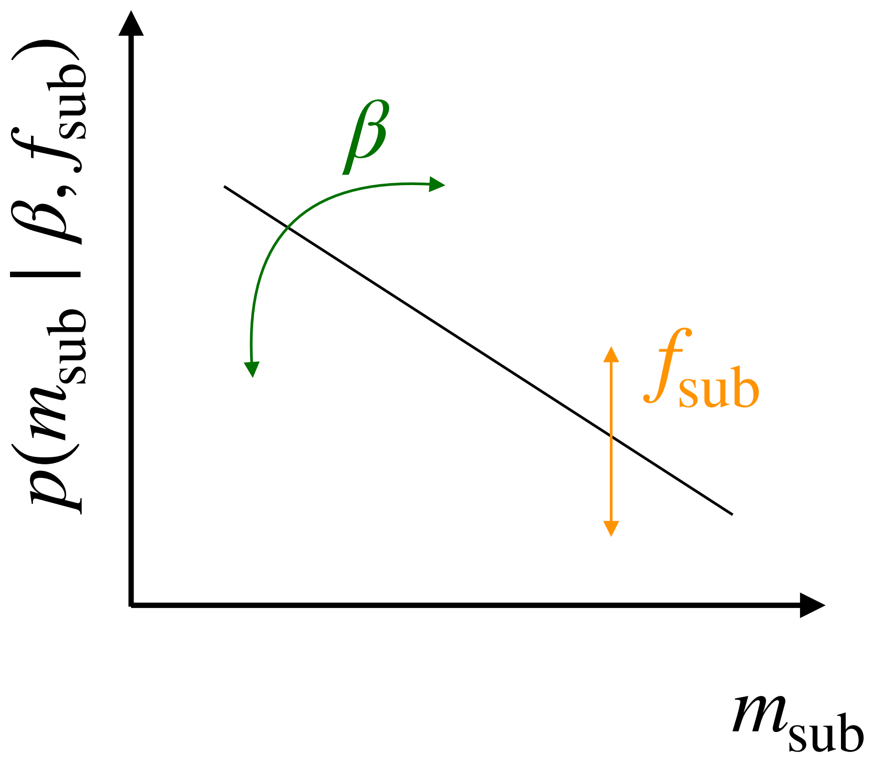


[R. *Dunstan et al* 1109.6291]

[T. Brown, J.Tumlinson]

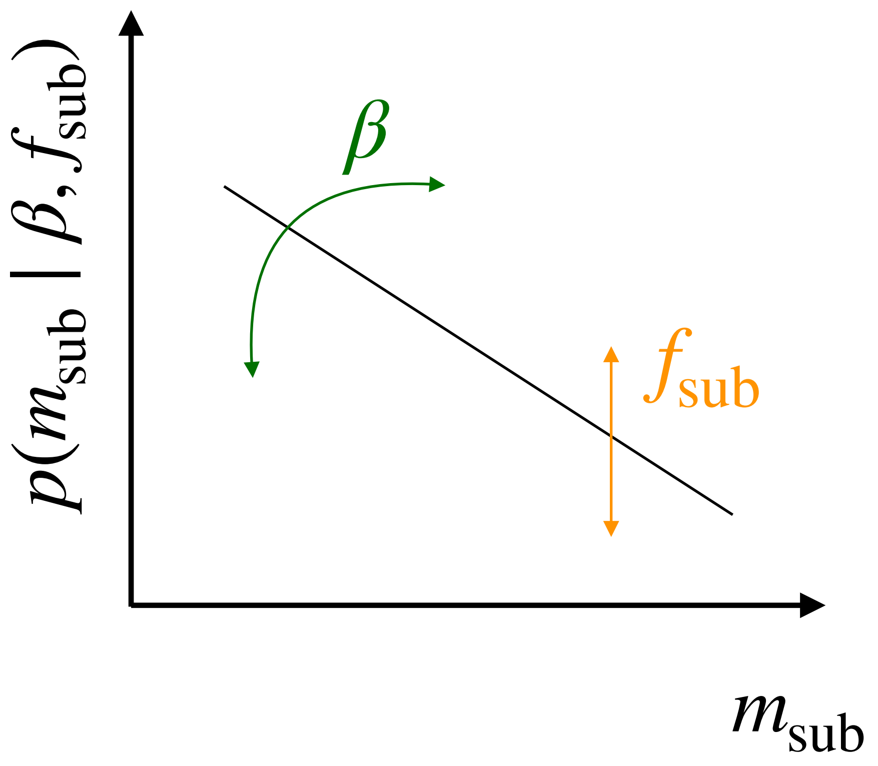
Simulation-based inference for strong lensing

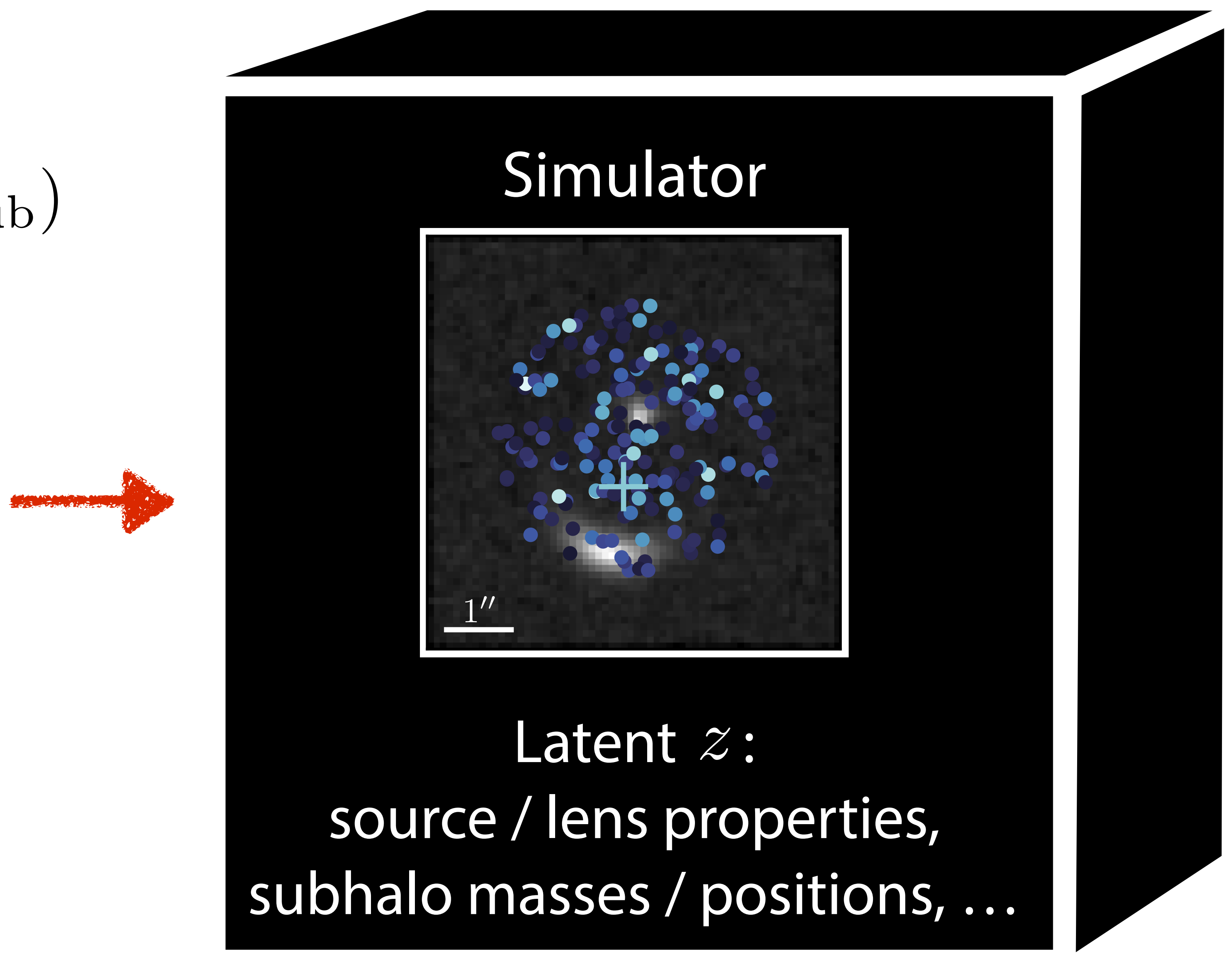
2 parameters $\theta = (\beta, f_{\mathrm{sub}})$



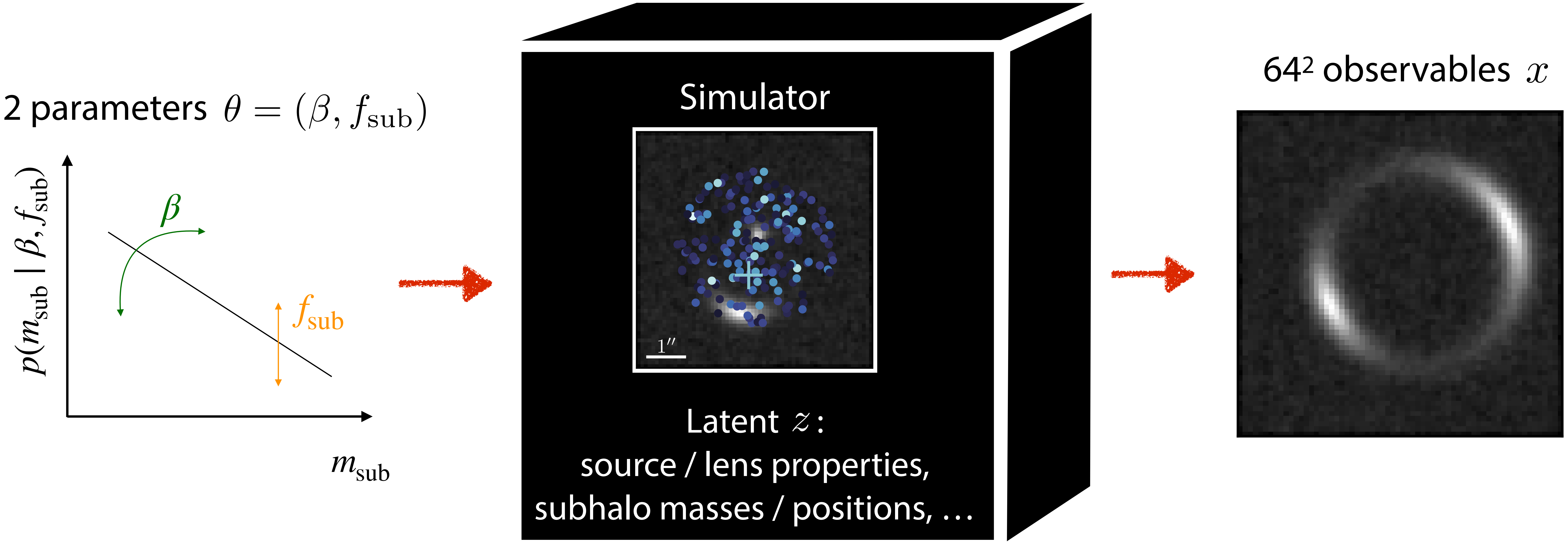
Simulation-based inference for strong lensing

2 parameters $\theta = (\beta, f_{\rm sub})$

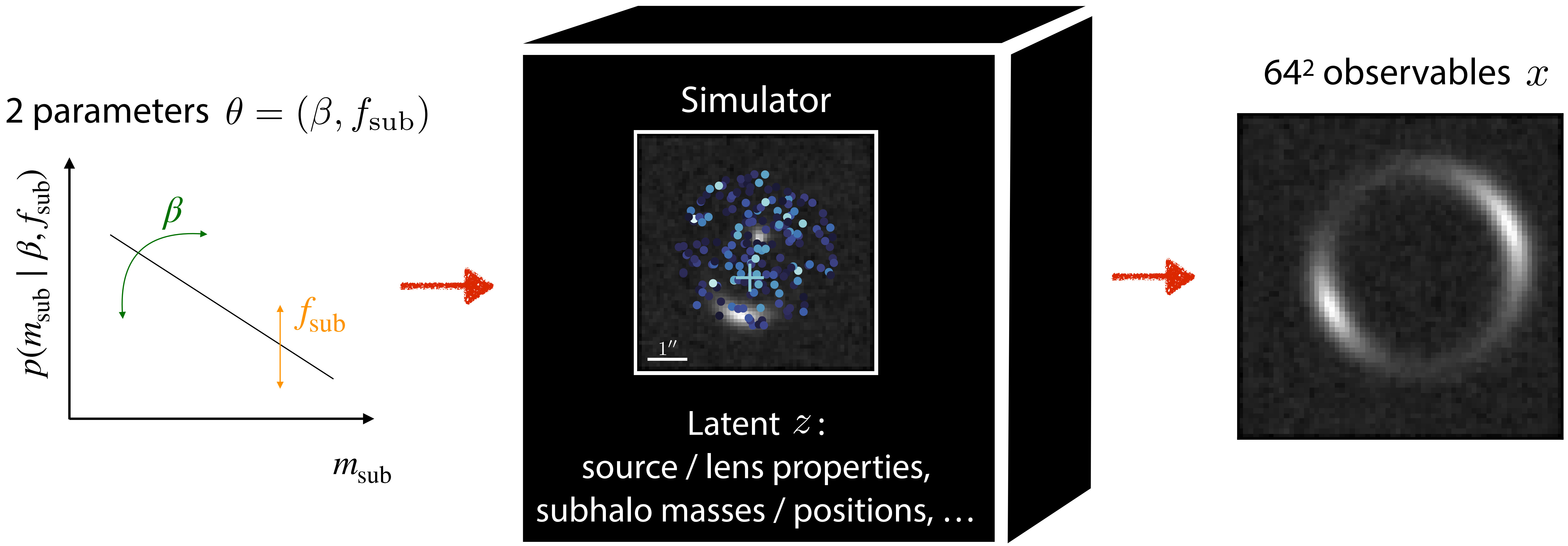




Simulation-based inference for strong lensing

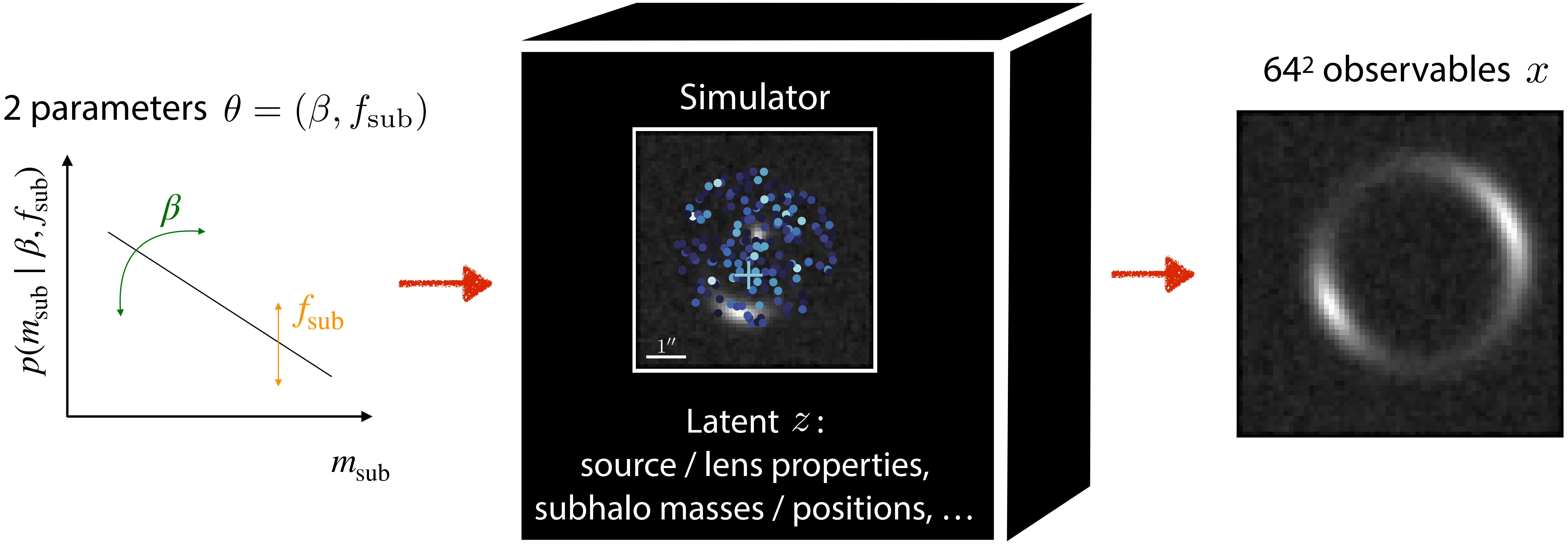


Simulation-based inference for strong lensing



Prediction: We construct a simulator that can sample $x \sim p(x|\theta)$

Simulation-based inference for strong lensing

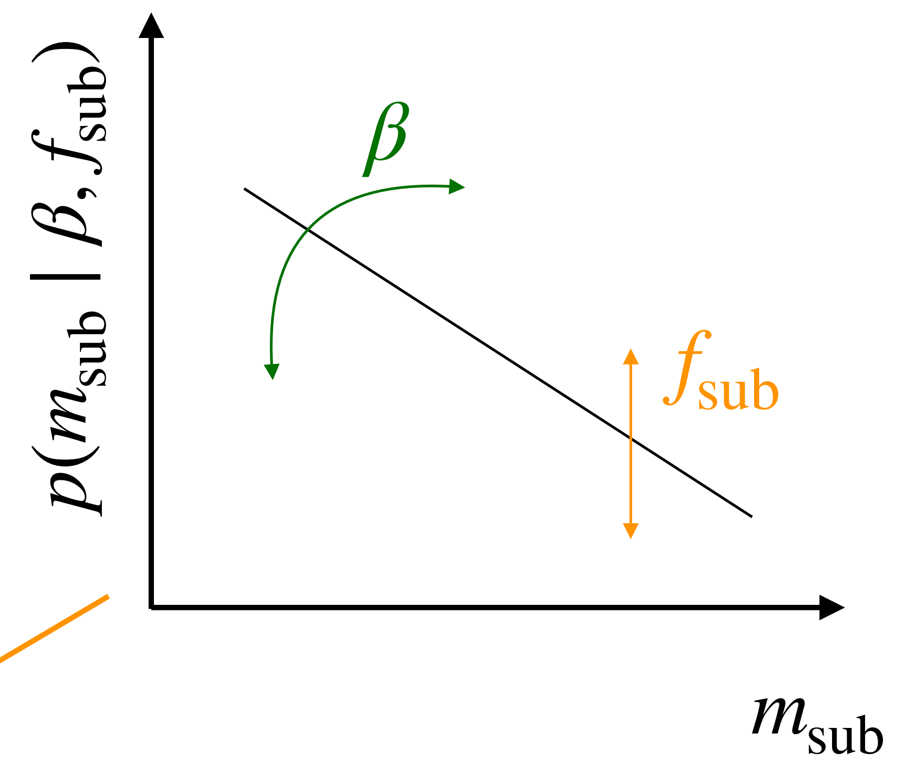


Prediction: We construct a simulator that can sample $x \sim p(x|\theta)$

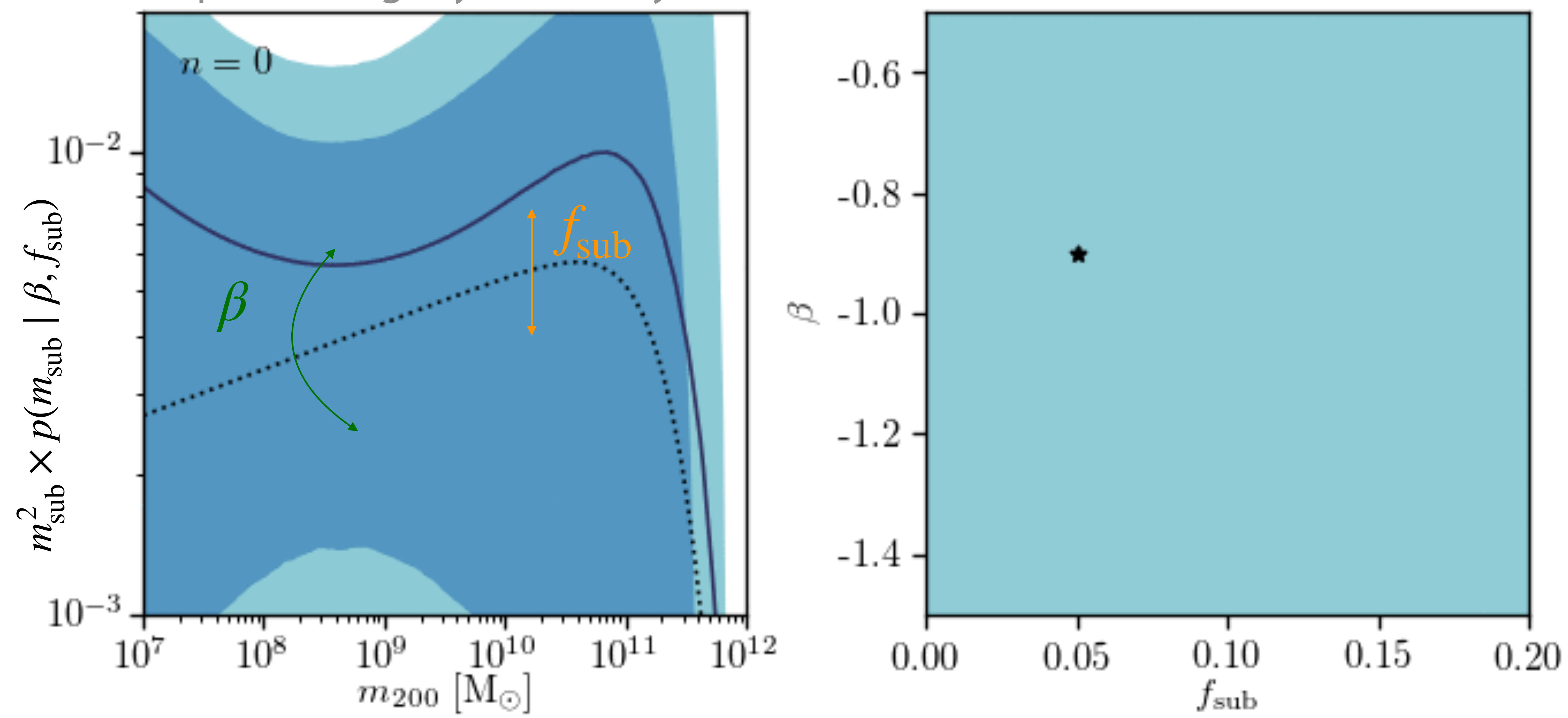
Inference: We train neural likelihood ratio estimators $\hat{r}(x|\theta)$

Posterior from amortized likelihood ratio

Watch how the posterior for two population parameters concentrate around true value used to generate mock data.



(plotted slightly differently)

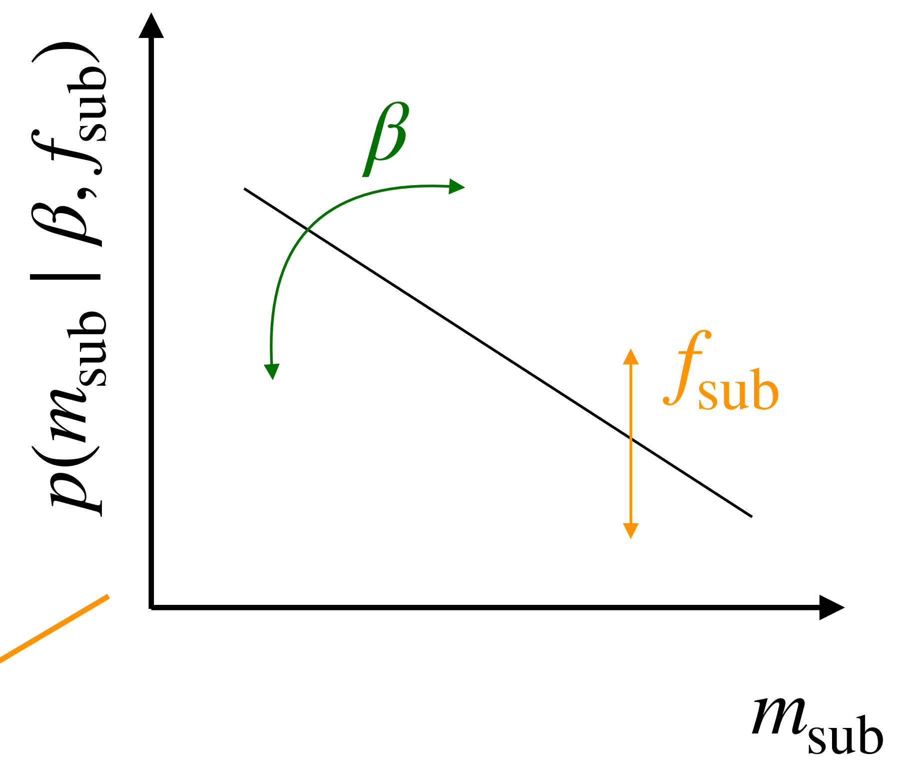




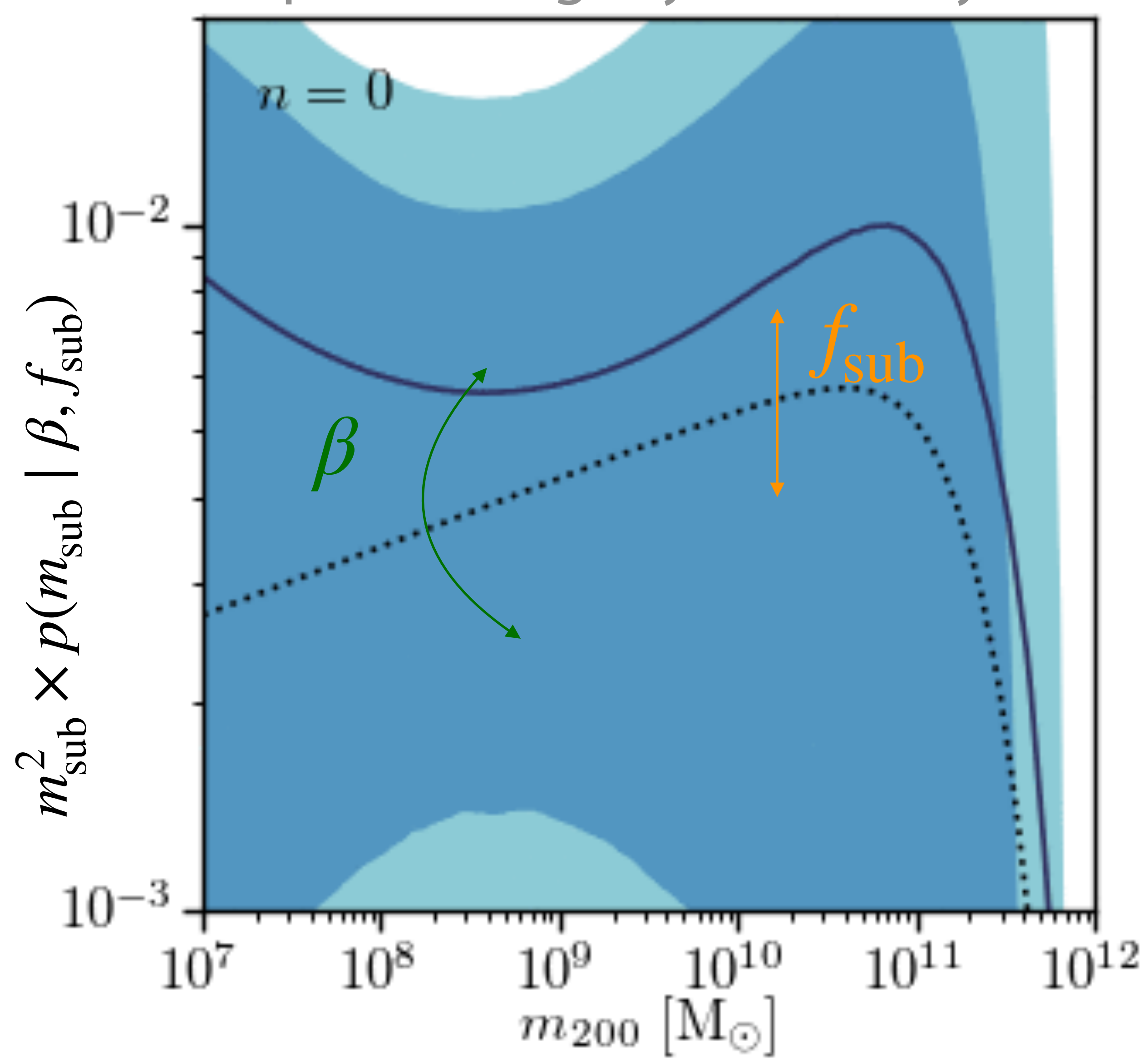

Sid Mishra-Sharma Johann Brehmer Gilles Louppe Joeri Hermans, K. Cranmer. published in ApJ, 886 (2019) arXiv:1909.02005

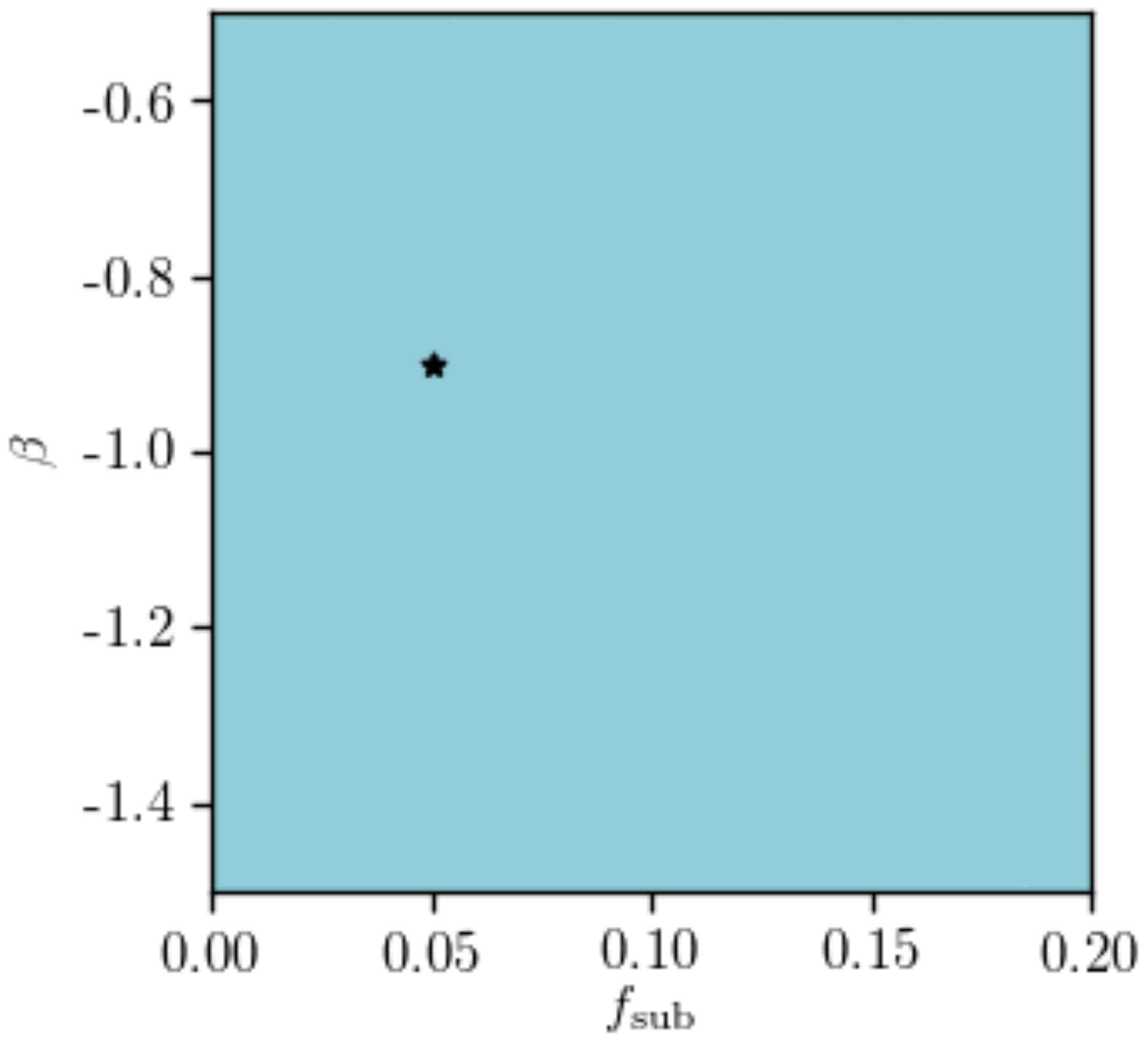
Posterior from amortized likelihood ratio

Watch how the posterior for two population parameters concentrate around true value used to generate mock data.



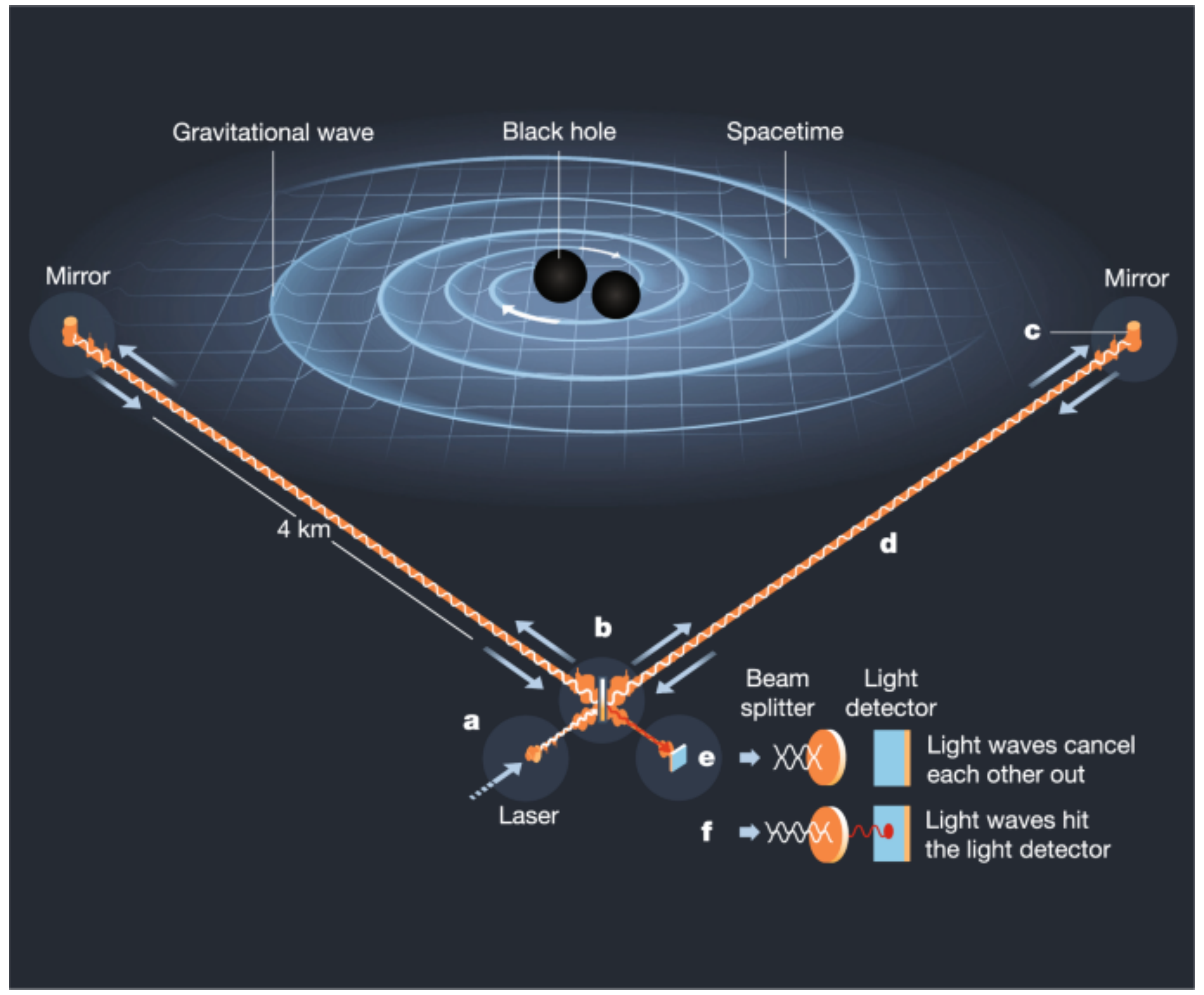
(plotted slightly differently)

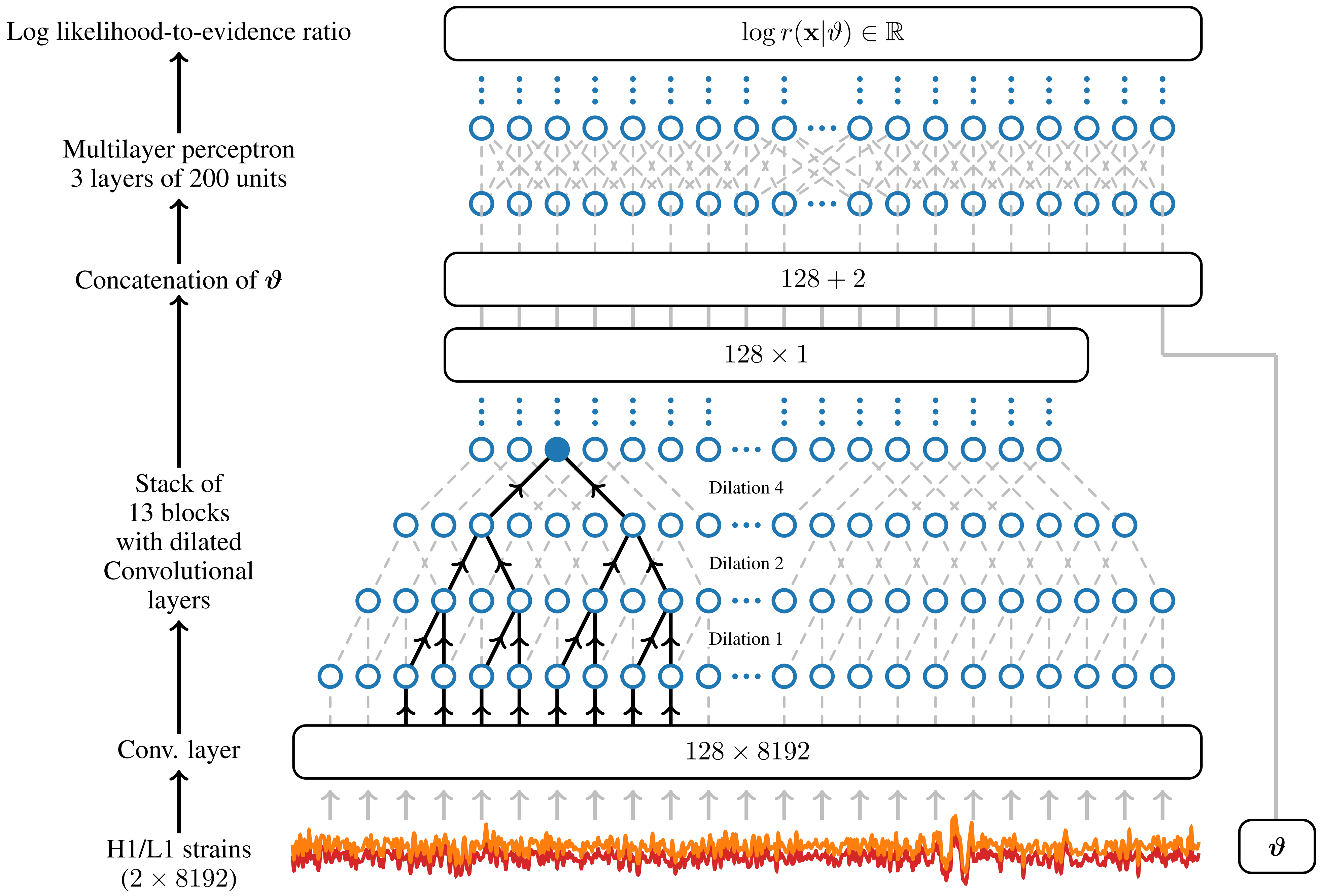




Sid Mishra-Sharma Johann Brehmer Gilles Louppe Joeri Hermans, K. Cranmer. published in ApJ, 886 (2019) arXiv:1909.02005

Gravitational Wave Astronomy

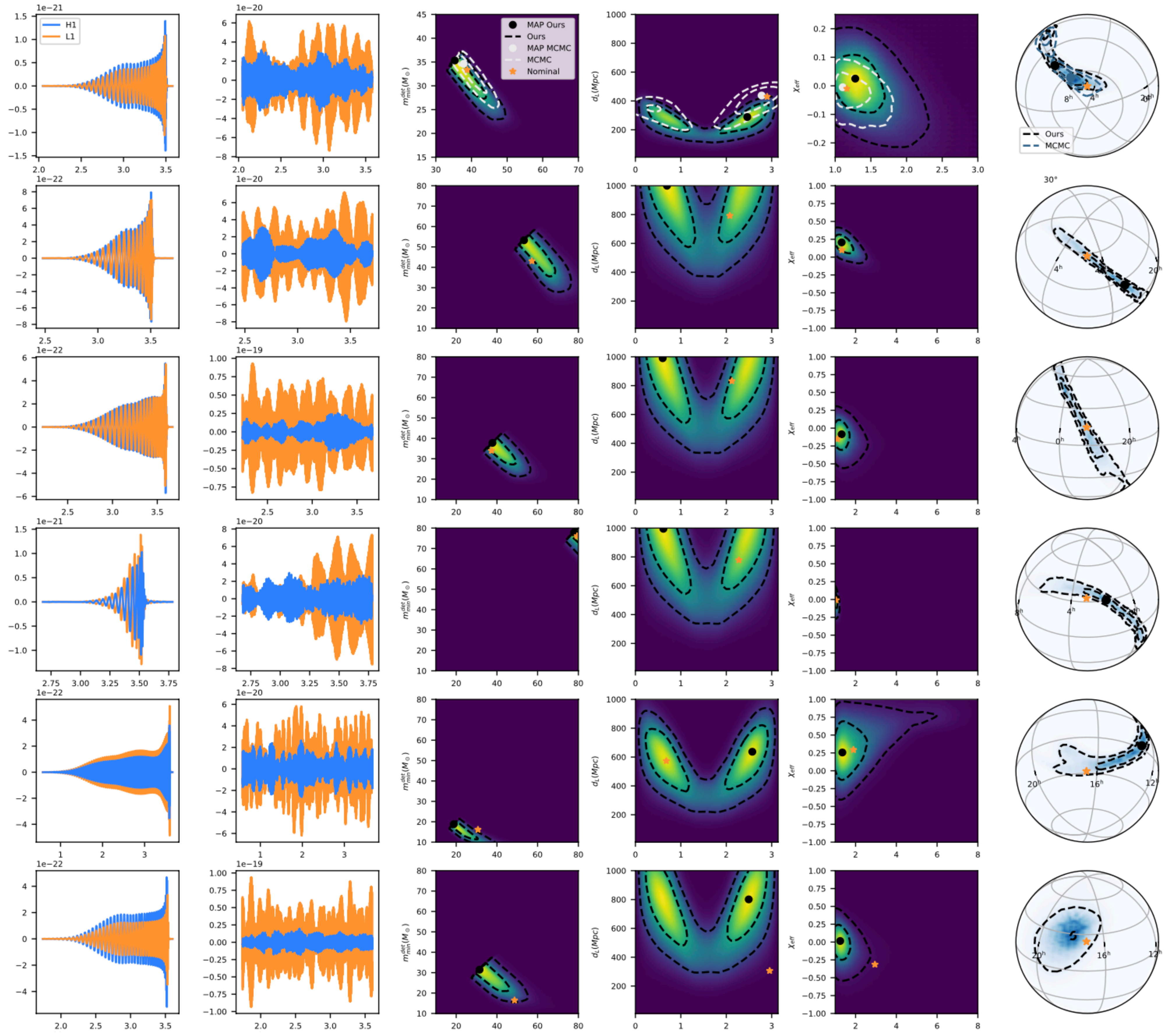




Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization

Delaunoy, Wehenkel, Hinderer, Nissanke, Weniger, Williamson, Louppe

[arXiv:2010.12931]



In cosmology nd astrophysics

Physics ∩ ML

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

10
Feb 2021

Physics meets ML to solve cosmological inference

Ben Wandelt, Institut d’Astrophysique de Paris / Institut Lagrange, Sorbonne University and Center for Computational Astrophysics, Flatiron Institute, New York, 12:00 EDT

Abstract: The goal of cosmological inference is to learn about the origin, composition, evolution, and fate of the cosmos from all accessible sources of astronomical data, such as the cosmic microwave background, galaxy surveys, or electromagnetic and gravitational wave transients. Traditionally, the field has progressed by designing and modeling intuitive summaries of the data, such as n-point correlations. This traditional approach has a number of risks and limitations: how do we know if we computed the most informative statistics? Did we forget any summaries that would have provided additional information or break parameter degeneracies? Did we take into account all the ways the model is affecting the data? To be feasible, the traditional approach imposes approximations on the statistical modeling (e.g. the likelihood form) and on the physical modeling. I will discuss a new mode of cosmological inference: simulation-based, full-physics modeling, made feasible through multiple advances in 1) machine-learning, 2) in the way we design and run simulations of cosmological observables, and 3) in how we compare models to data. The goal is to use current and next generation data to reconstruct the cosmological initial conditions and constrain cosmological physics much more completely than has been feasible in the past. I will discuss current status, and ways to meet the new challenges inherent in this approach, including robustness to model misspecification.

Physics ∩ ML

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

04
Nov 2020

Flow-based likelihoods for non-Gaussian inference.

Ana Diaz Rivero, Harvard University, 12:00 EDT

Abstract: We investigate the use of data-driven likelihoods to bypass a key assumption made in many scientific analyses, which is that the true likelihood of the data is Gaussian. In particular, we suggest using the optimization targets of flow-based generative models, a class of models that can capture complex distributions by transforming a simple base distribution through layers of nonlinearities. We call these flow-based likelihoods (FBL). We analyze the accuracy and precision of the reconstructed likelihoods on mock Gaussian data, and show that simply gauging the quality of samples drawn from the trained model is not a sufficient indicator that the true likelihood has been learned. We nevertheless demonstrate that the likelihood can be reconstructed to a precision equal to that of sampling error due to a finite sample size. We then apply FBLs to mock weak lensing convergence power spectra, a cosmological observable that is significantly non-Gaussian (NG). We find that the FBL captures the NG signatures in the data extremely well, while other commonly-used data-driven likelihoods, such as Gaussian mixture models and independent component analysis, fail to do so. This suggests that works that have found small posterior shifts in NG data with data-driven likelihoods such as these could be underestimating the impact of non-Gaussianity in parameter constraints. By introducing a suite of tests that can capture different levels of NG in the data, we show that the success or failure of traditional data-driven likelihoods can be tied back to the structure of the NG in the data. Unlike other methods, the flexibility of the FBL makes it successful at tackling different types of NG simultaneously. Because of this, and consequently their likely applicability across datasets and domains, we encourage their use for inference when sufficient mock data are available for training.

A profound shift

Scientific simulators are based on well-motivated mechanistic models

- However, the aggregate effect of many interactions between their low-level components leads to intractable inverse problems

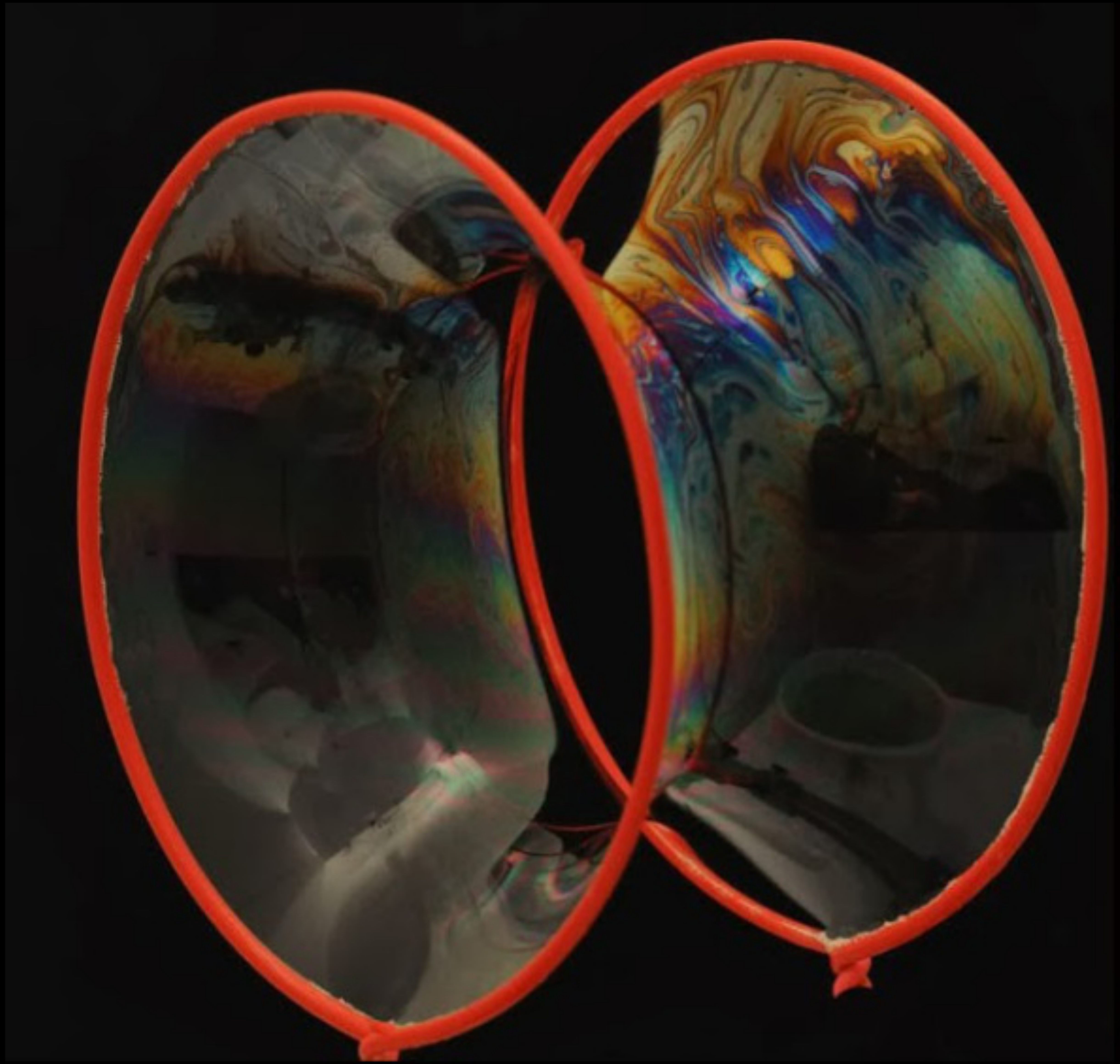
The developments in machine learning have the potential to effectively bridge the microscopic - macroscopic divide & aid in these inverse problems

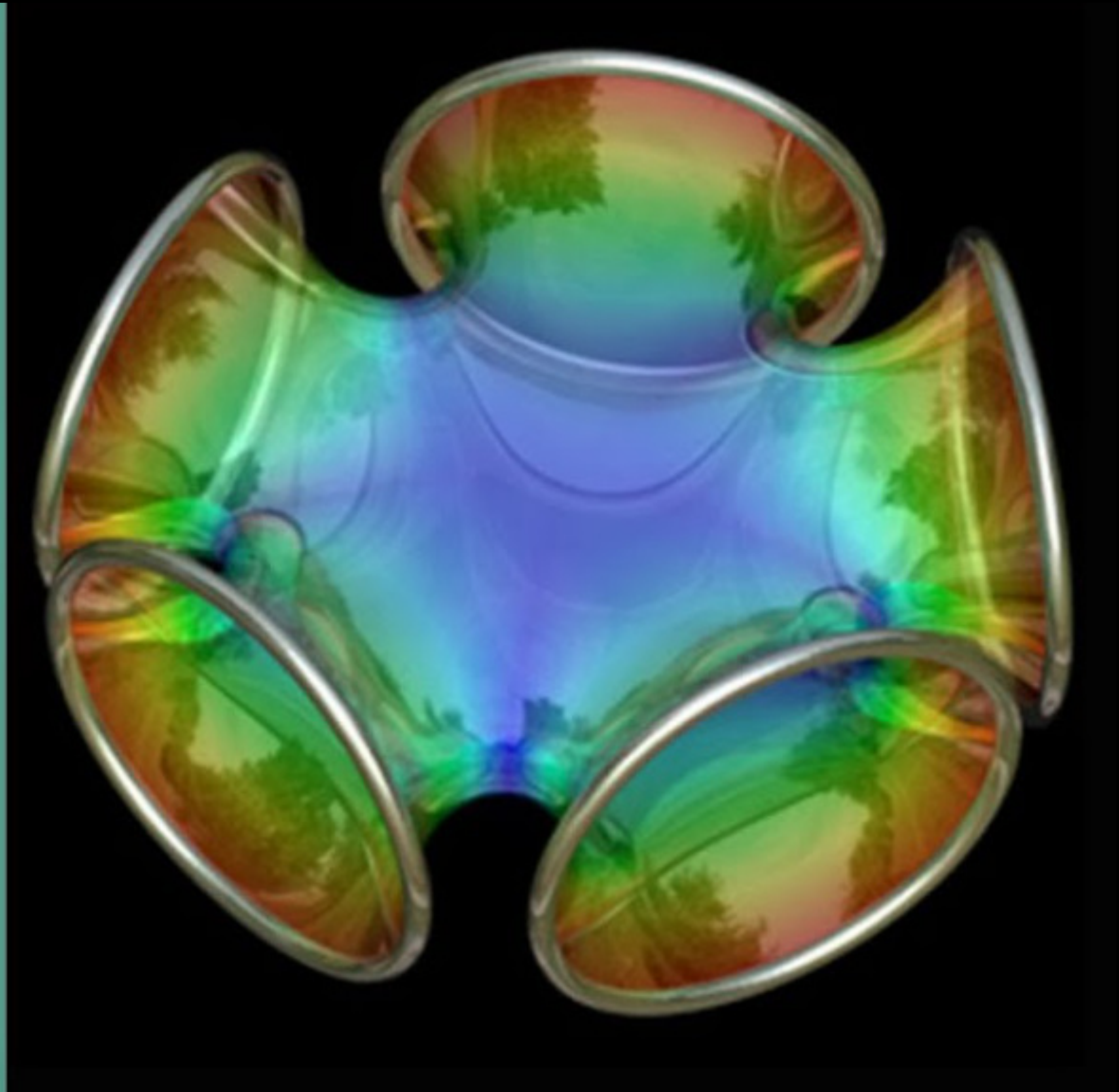
- They can provide effective models for macroscopic (emergent) phenomena that are tied back to the low-level microscopic (reductionist) model

Notice in the next few slides I will be agnostic to the architectures of the ML models!

Do we care?

Machine Learning = Applied Calculus of Variations



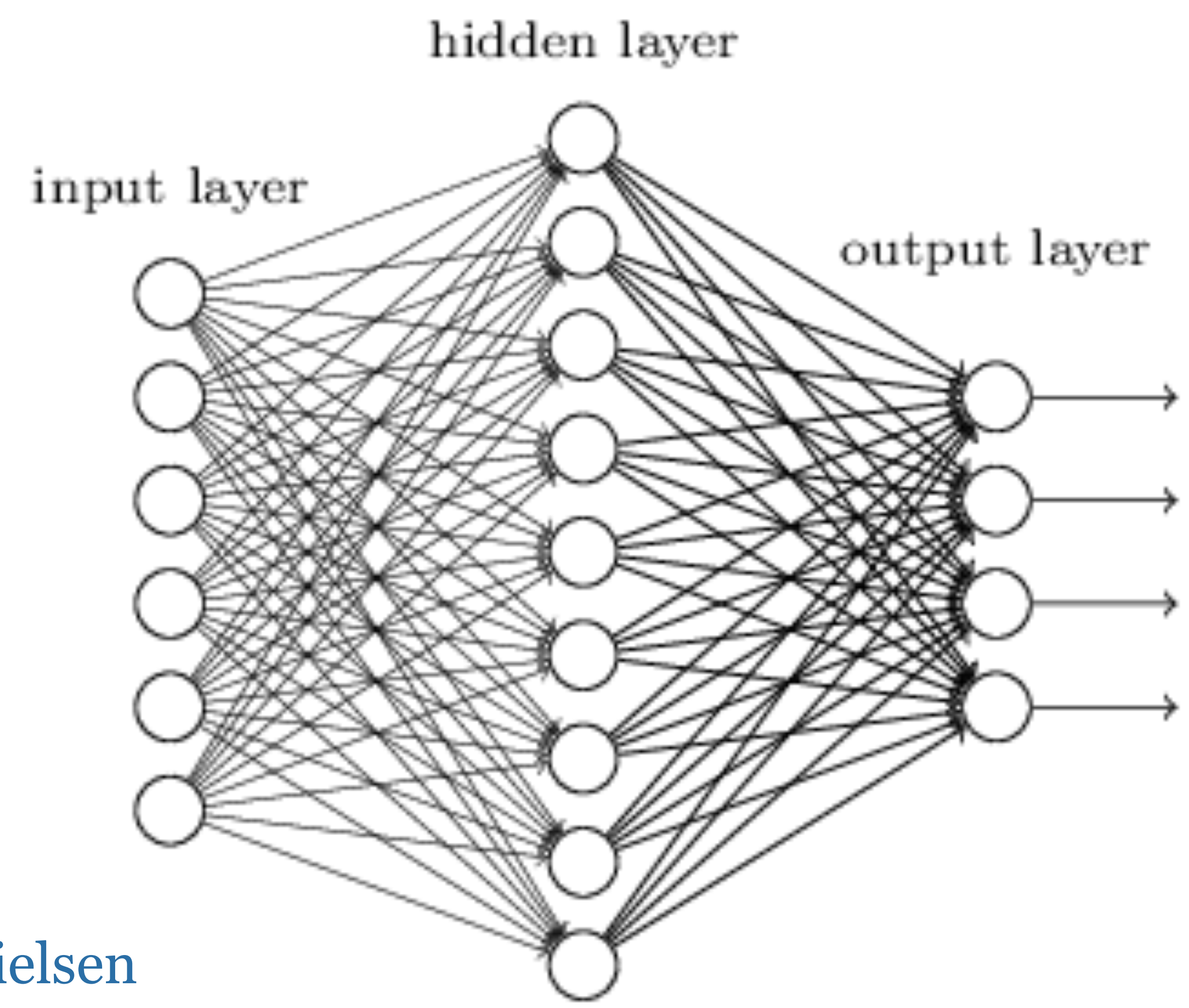


NN = A highly Flexible Family of Functions

In calculus of variations, the optimization is over all functions: $\hat{s} = \underset{s}{\operatorname{argmin}}\, L[s]$

- In applied calculus of variations, we consider a highly flexible family of functions s_ϕ and optimize: i.e. $\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\, L[s_\phi]$ and $\hat{s} \approx s_{\hat{\phi}}$
- Think of neural networks as a highly flexible family of functions
- Machine learning also includes non-convex optimization algorithms that are effective even with millions of parameters!

" **Shallow neural network**



Deep neural network

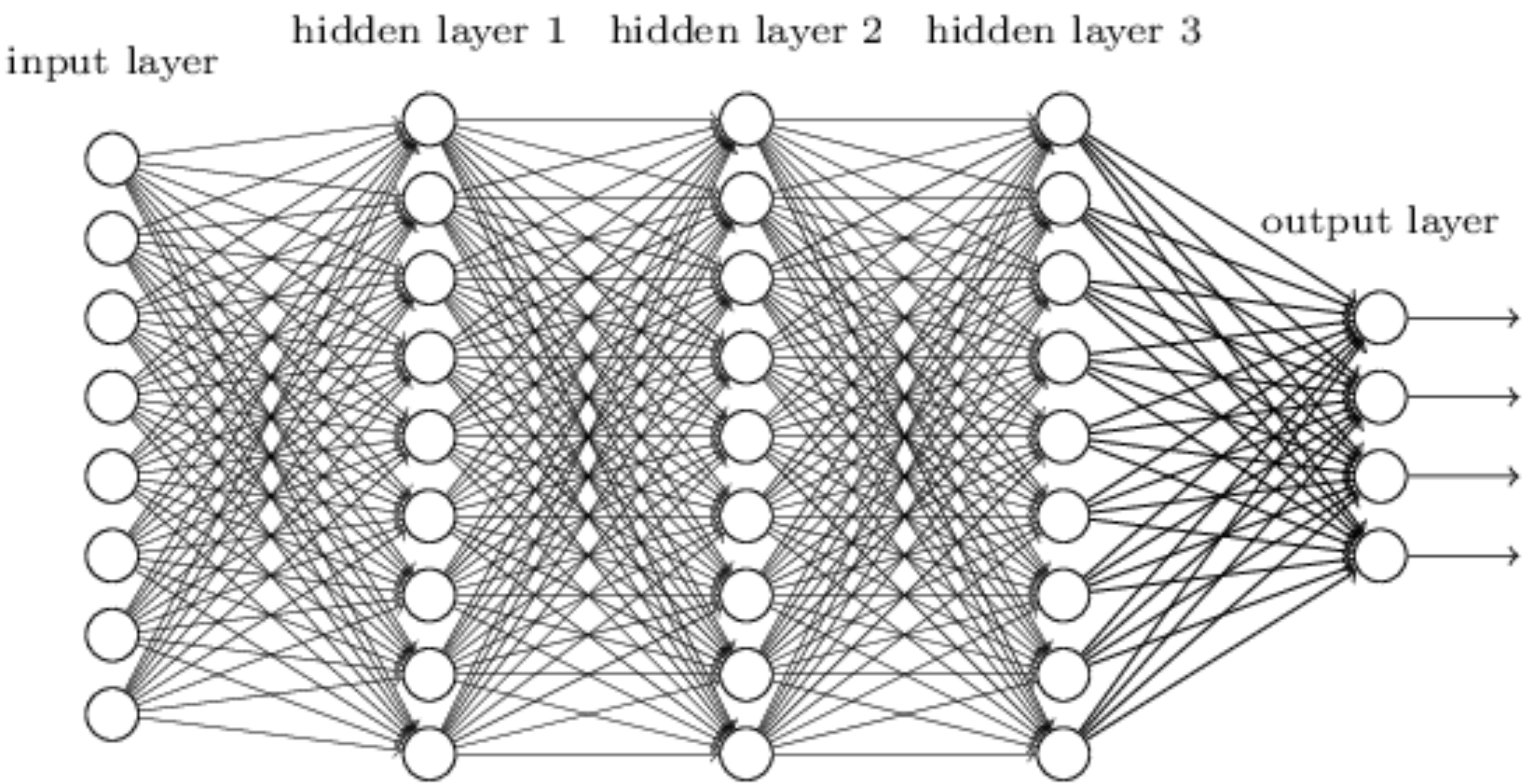


image credit: Michael Nielsen

41

Machine learning = Applied Calculus of Variations

Kyle Cranmer added **3 new photos** — with **Sarah Demers Konezny** and **Paul Tipton**.
April 20, 2016 · New Haven, CT ·

Seminar at Yale today. Felt good to talk about new ideas... Equally confusing for theorists and experimentalists 😜

Machine Learning = Applied Calculus of Variations

2 Deriving BP using the Hamiltonian/Lagrangian formalism

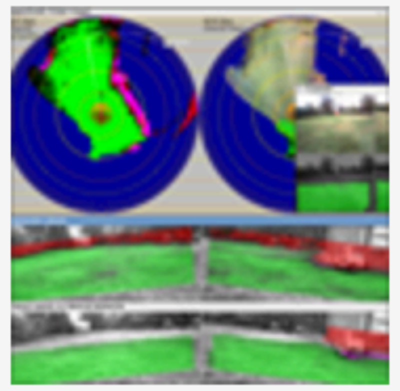
2.1 Notations

For the sake of clarity, we will introduce the formalism in a simple case. A more general formulation will be presented afterwards. It will be assumed that the network is composed of a number of layers connected in a feed-forward manner. Furthermore, we make the assumption that connections cannot skip layers. These assumptions can be easily relaxed [le Cun, 1987].

Yann LeCun Deep learning = calculus of variations

Backprop is like the Langrangian formulation of classical mechanics.

Y. LeCun: A theoretical framework for Back-Propagation, in Touretzky, D. and Hinton, G. and Sejnowski, T. (Eds), Proceedings of the 1988 Connectionist Models Summer School, 21-28, Morgan Kaufmann, CMU, Pittsburgh, Pa, 1988.
http://yann.lecun.com/exdb/publis/index.html#lecun-88



[bib2web] Yann LeCun's Publications

YANN.LECUN.COM

Like · Reply · Remove Preview · 2 · April 20, 2016 at 2:30am

Kyle Cranmer I guess this counts as an endorsement for this point of view 😆

Many physicists (particularly theoretical ones) are skeptical of machine learning because it usually is explained to them in some ad hoc way (neurons, etc). But minimizing a loss function(al) is much more palatable.

Like · Reply · 2 · April 20, 2016 at 2:39am · Edited

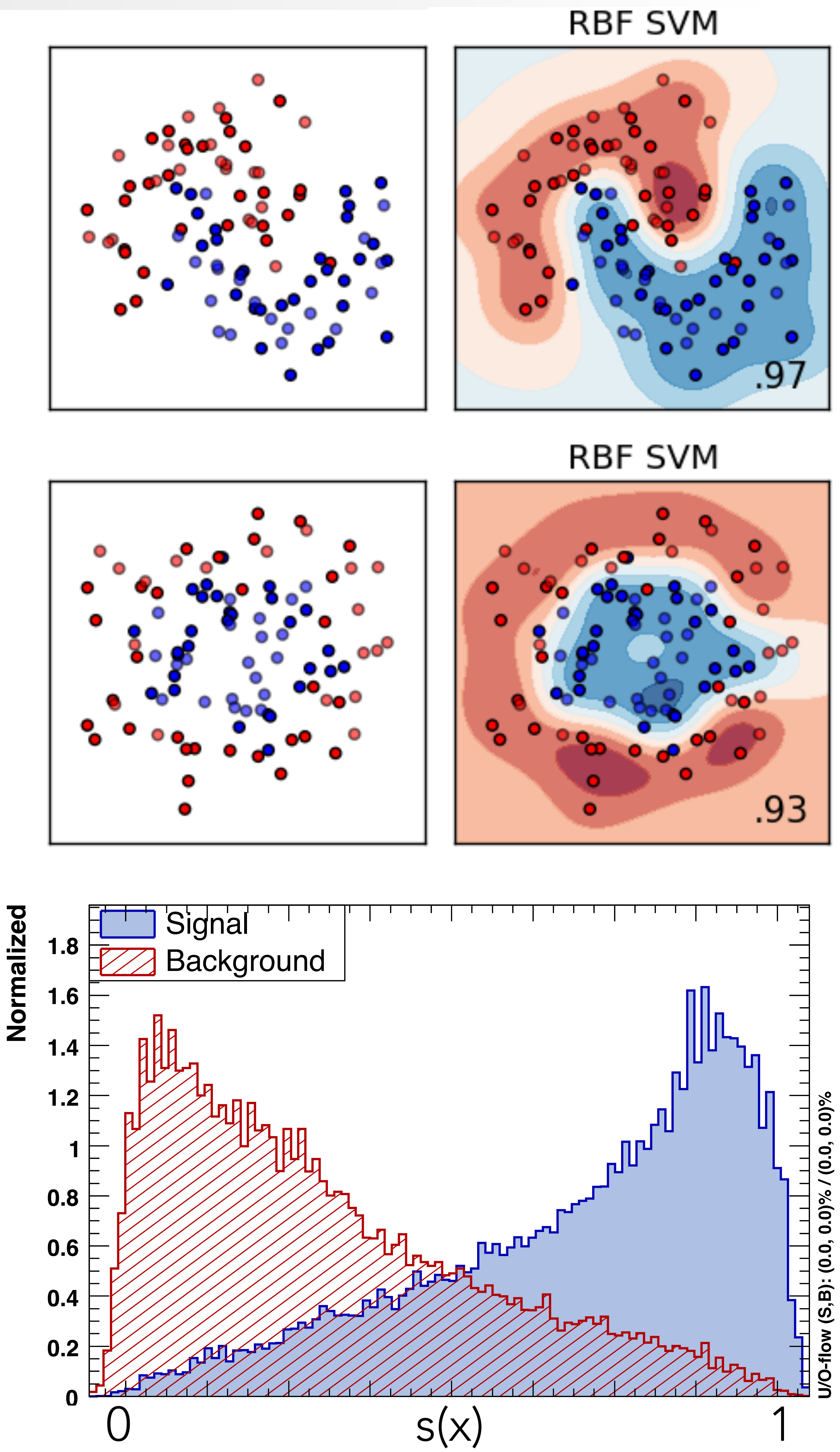
From the review

[Cranmer, J. Brehmer, G. Louppe, PNAS (2020), arXiv:1911.01429]

Fig. 3. Overview of different approaches to simulation-based inference.

Likelihood Ratio Trick

[Hastie et al., 2001; Sugiyama et al., 2012; Cranmer et. al., 2015]



- **binary classifier**: find function $s(x)$ that minimizes ***loss***:

$$L[s] = \mathbb{E}_{p(x|H_1)}[-\log s(x)] + \mathbb{E}_{p(x|H_0)}[-\log(1 - s(x))]$$

- i.e. approximate the optimal classifier

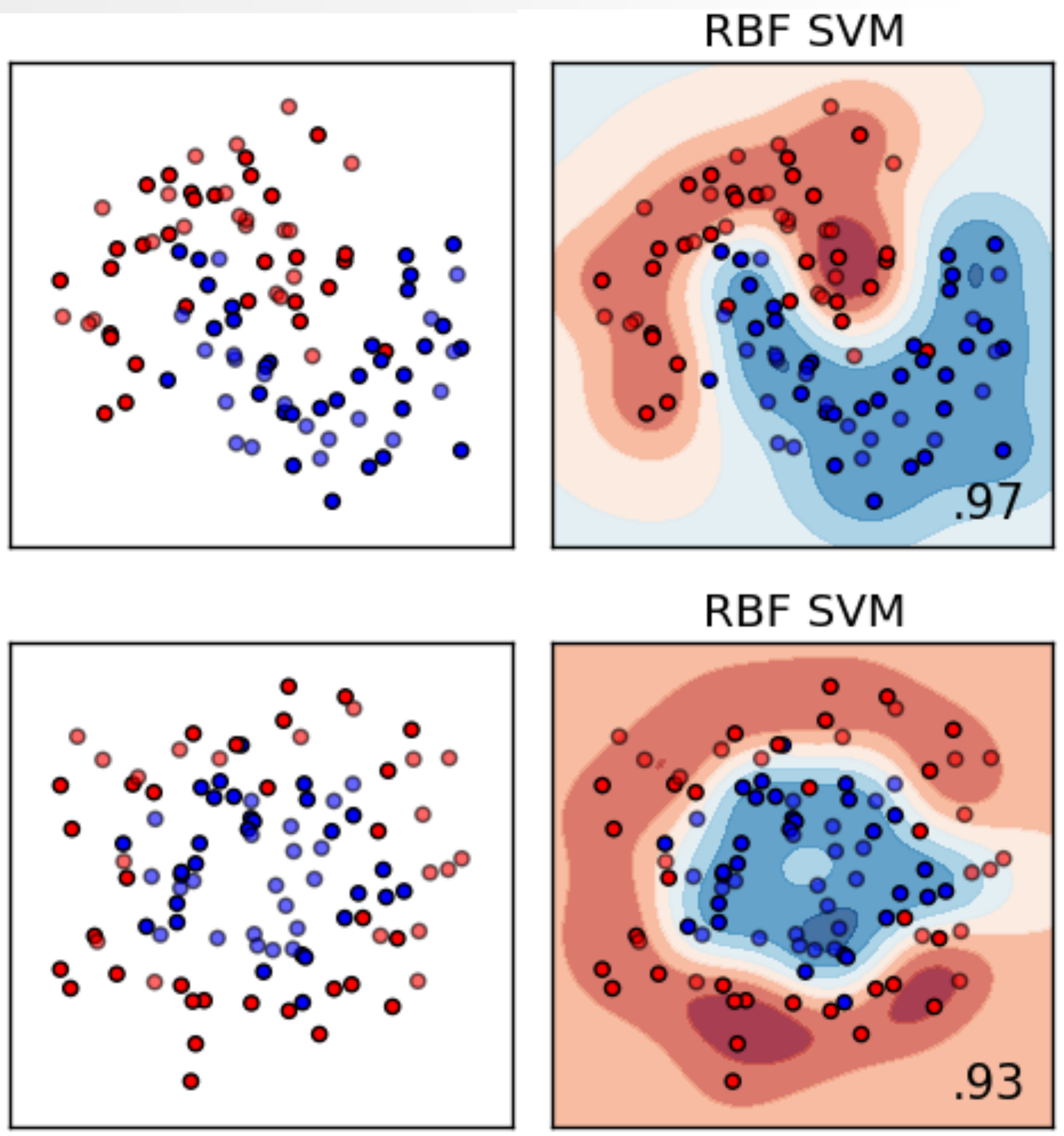
$$s(x) = \frac{p(x|H_1)}{p(x|H_0) + p(x|H_1)}$$

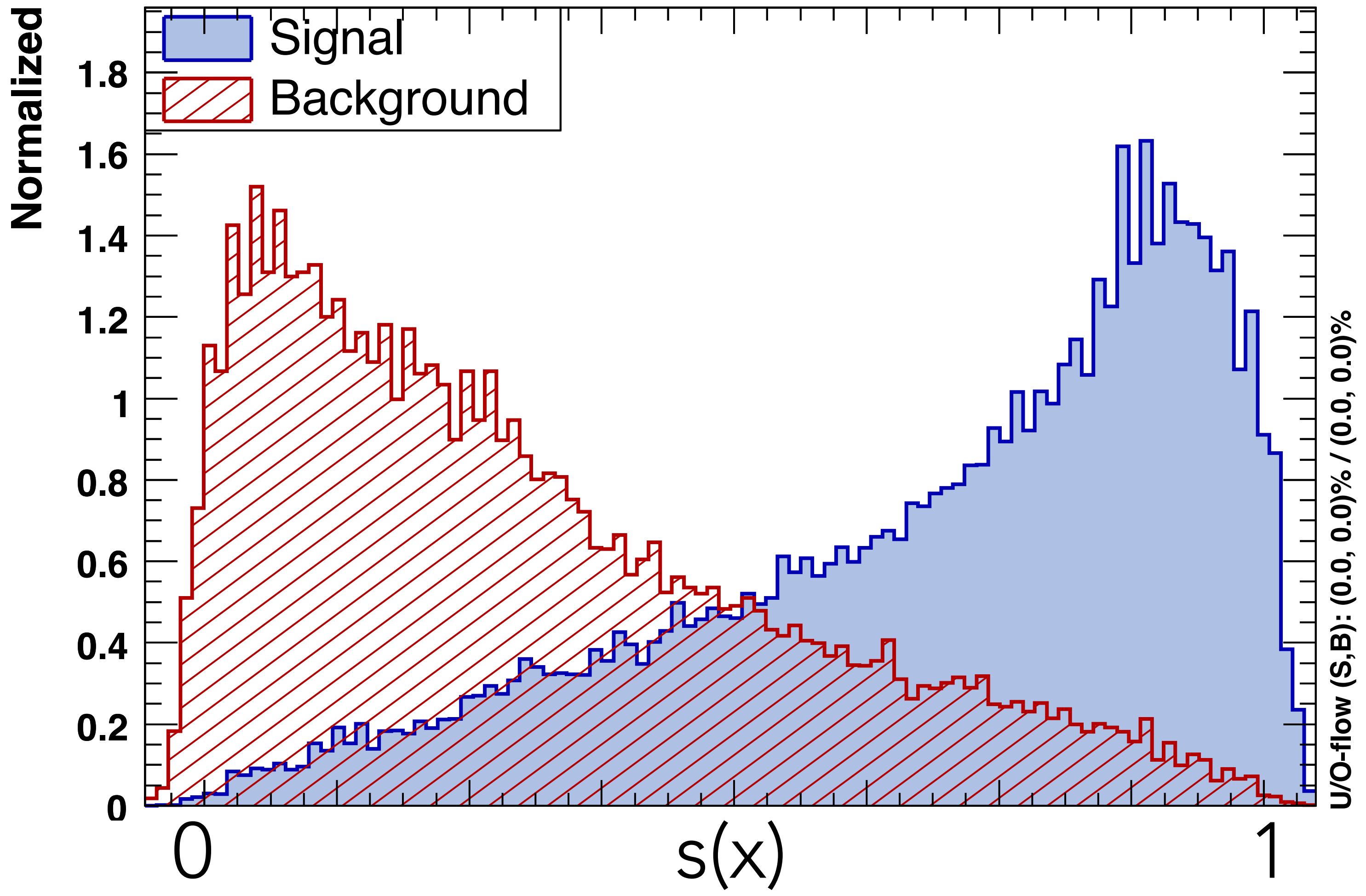
- which is 1-to-1 with the likelihood ratio

$$r(x) = \frac{p(x|H_1)}{p(x|H_0)} = 1 - \frac{1}{s(x)}$$

Likelihood Ratio Trick

[Hastie et al., 2001; Sugiyama et al., 2012; Cranmer et. al., 2015]





- **binary classifier**: find function $s(x)$ that minimizes ***loss***:

$$L[s] = \mathbb{E}_{p(x|H_1)}[-\log s(x)] + \mathbb{E}_{p(x|H_0)}[-\log(1-s(x))]$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} -y_i \log s(x_i) - (1-y_i)\log(1-s(x_i))$$

- i.e. approximate the optimal classifier

$$s(x) = \frac{p(x|H_1)}{p(x|H_0)+p(x|H_1)}$$

- which is 1-to-1 with the likelihood ratio

$$r(x) = \frac{p(x|H_1)}{p(x|H_0)} = 1 - \frac{1}{s(x)}$$

Parametrizing the Likelihood Ratio Trick

Can do the same thing for any two points θ_0 & θ_1 in parameter space $\boldsymbol{\Theta}$.

$$r(x;\theta_0,\theta_1) = \frac{p(x \mid \theta_0)}{p(x \mid \theta_1)} = 1 - \frac{1}{s(x;\theta_0,\theta_1)}$$

Or train to classify data from $p(x|\theta)$ versus some fixed reference $p_{\mathrm{ref}}(x)$

$$r(x;\theta) = \frac{p(x|\theta)}{p_{\mathrm{ref}}(x)} = 1 - \frac{1}{s(x;\theta)}$$

I call this a **parametrized classifier.**

K.C., G. Louppe, J. Pavez: Approximating Likelihood Ratios with Calibrated Discriminative Classifiers [arXiv:1506.02169]

From the review

[Cranmer, J. Brehmer, G. Louppe, PNAS (2020), arXiv:1911.01429]

Fig. 3. Overview of different approaches to simulation-based inference.

Neural likelihood

Cranmer, Louppe (2016). DOI:10.5281/zenodo.198541
Papamakarios, Sterratt, Murray AISTATS 2019 [arXiv:1805.07226]

Based on (θ_n, x_n) pairs with $x_n \sim p(x \mid \theta_n)$ estimate likelihood with a conditional density estimator $q_\phi(x \mid \theta)$

- Can sample $\theta_n \sim \tilde{p}(\theta)$ from any proposal distribution with appropriate support
- Leveraging advances in **normalizing flows** and neural density estimation

Unifying generative models and exact likelihood-free inference with conditional bijections

By Kyle Cranmer, Gilles Louppe 2016

Sequential Neural Likelihood:
Fast Likelihood-free Inference with Autoregressive Flows

George Papamakarios, University of Edinburgh
David C. Sterratt, University of Edinburgh
Iain Murray, University of Edinburgh

AISTATS 2019

Minimize the loss

$$L[q] = -\int p(x) \log q(x) dx$$

Subject to $\int q(x) dx = 1$

Yields $\hat{q}(x) = p(x)$

Neural likelihood

Cranmer, Louppe (2016). DOI:10.5281/zenodo.198541
Papamakarios, Sterratt, Murray AISTATS 2019 [arXiv:1805.07226]

Based on (θ_n, x_n) pairs with $x_n \sim p(x \mid \theta_n)$ estimate likelihood with a conditional density estimator $q_\phi(x \mid \theta)$

- Can sample $\theta_n \sim \tilde{p}(\theta)$ from any proposal distribution with appropriate support
- Leveraging advances in **normalizing flows** and neural density estimation

Unifying generative models and exact likelihood-free inference with conditional bijections

By Kyle Cranmer, Gilles Louppe 2016

Sequential Neural Likelihood:
Fast Likelihood-free Inference with Autoregressive Flows

George Papamakarios, University of Edinburgh
David C. Sterratt, University of Edinburgh
Iain Murray, University of Edinburgh

AISTATS 2019

Minimize the loss

$$L[q] = -\int p(x) \log q(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} \log q_\phi(x_i)$$

Subject to $\int q(x) dx = 1$

Yields $\hat{q}(x) = p(x)$

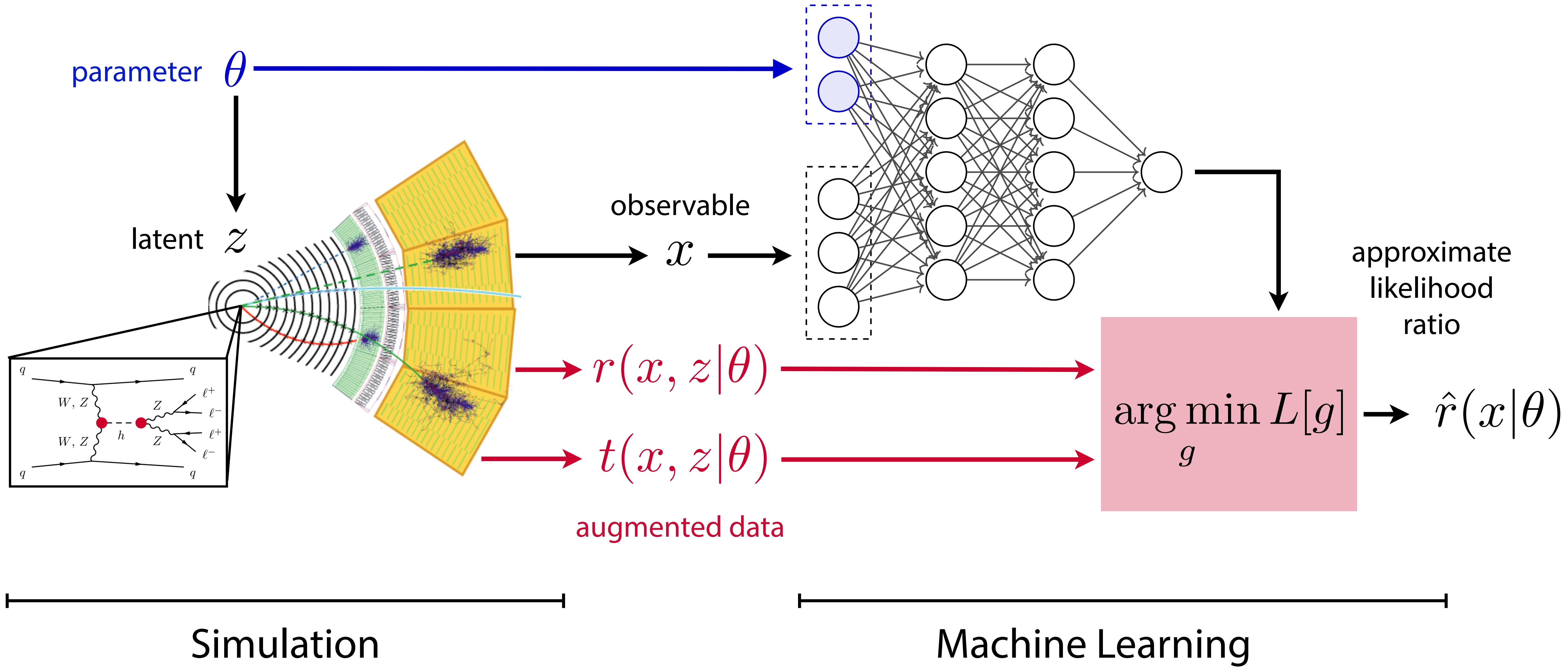
Gold mining: augmenting the training data

Sample efficiency is a major concern for these methods as many simulators are computationally expensive

Gold mining: augmenting the training data

Sample efficiency is a major concern for these methods as many simulators are computationally expensive

Recently, we realized we can **extract more from the simulator**.
We can use **augmented data** to improve training



Johann Brehmer

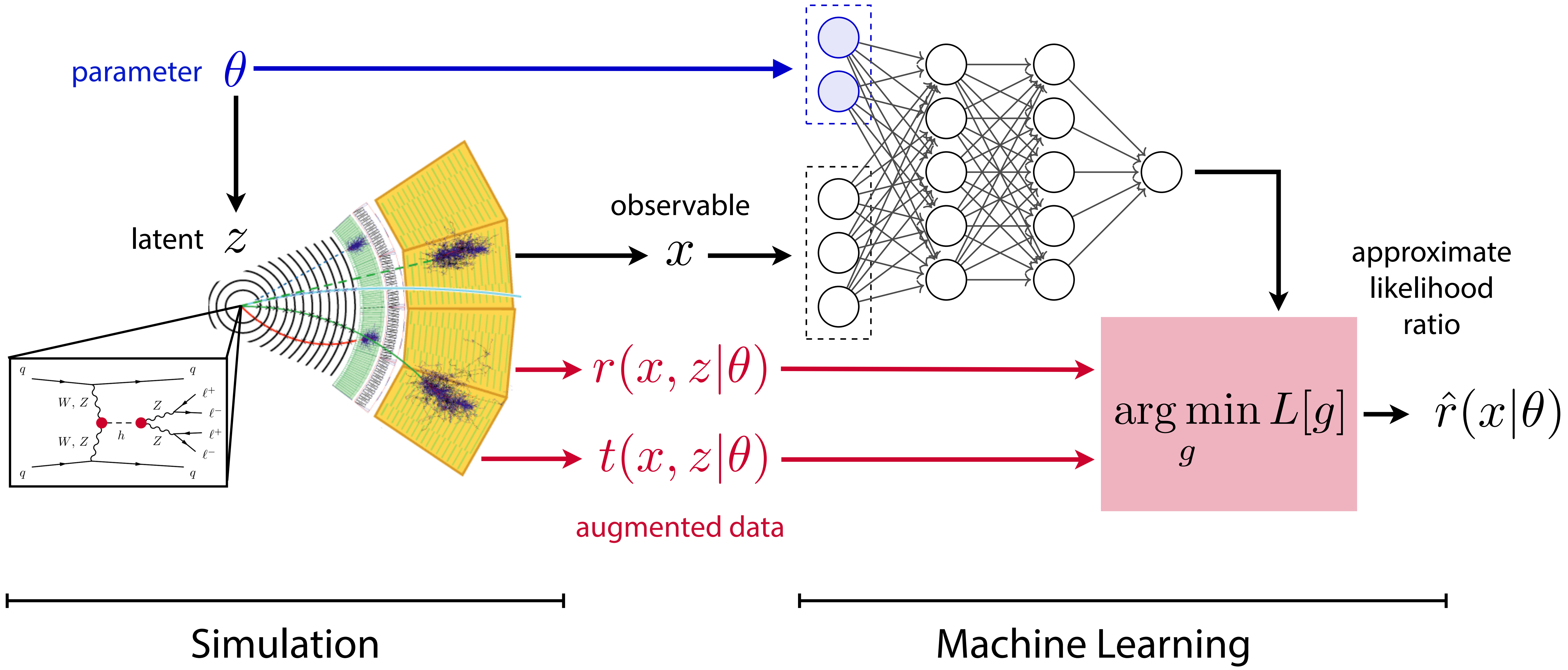
Gilles Louppe

Brehmer, Louppe, Pavez, KC, PNAS (2019), arXiv:1805.12244
See also Wenliang, Moskovitz, Kanagawa, Sahani, ICML2020

Gold mining: augmenting the training data

Sample efficiency is a major concern for these methods as many simulators are computationally expensive

Recently, we realized we can **extract more from the simulator**.
We can use **augmented data** to improve training



While implicit density is intractable

$$p(x|\theta) = \int dz p(x, z|\theta)$$

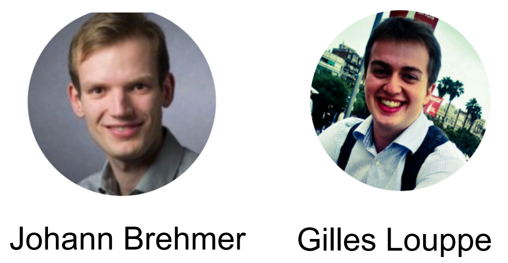
We can **augment the simulator** to calculate some quantities conditioned on latent z, which are tractable:

Joint likelihood ratio:

$$r(x, z|\theta_0, \theta_1) = \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)}$$

and joint score:

$$t(x, z|\theta_0) = \frac{\nabla_\theta p(x, z|\theta)|_{\theta_0}}{p(x, z|\theta_0)} = \nabla_\theta \log p(x, z|\theta)|_{\theta_0}$$



Brehmer, Louppe, Pavez, KC, PNAS (2019), arXiv:1805.12244
See also Wenliang, Moskovitz, Kanagawa, Sahani, ICML2020

The value of gold

We can calculate the joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

("How much more likely is this simulated event, including all intermediate states, for θ_0 compared to θ_1?")

↓

We want the likelihood ratio function

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

("How much more likely is the observation x for θ_0 compared to θ_1?")

The value of gold

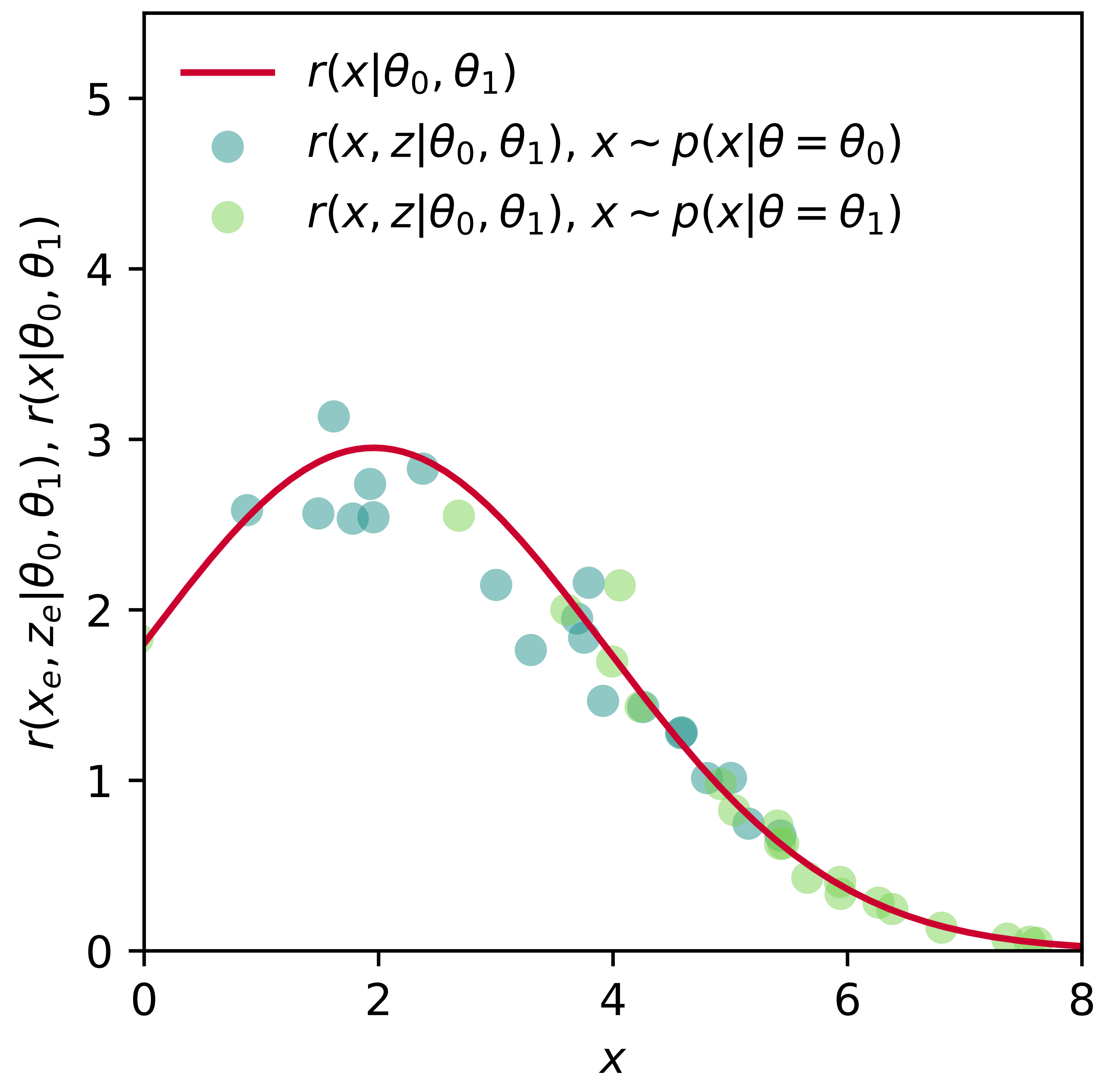
We can calculate the joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

$r(x, z|\theta_0, \theta_1)$ are scattered around $r(x|\theta_0, \theta_1)$

We want the likelihood ratio function

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$



The value of gold

We can calculate the joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

↓

We want the likelihood ratio function

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

With $r(x, z|\theta_0, \theta_1)$, we define a functional like

$$L_r[\hat{r}(x|\theta_0, \theta_1)] = \int \mathrm{d}x \int \mathrm{d}z \; p(x, z|\theta_1) \left[\left(\hat{r}(x|\theta_0, \theta_1) - r(x, z|\theta_0, \theta_1) \right)^2 \right].$$

It is minimized by

$$r(x|\theta_0, \theta_1) = \underset{\hat{r}(x|\theta_0, \theta_1)}{\arg\min} \, L_r[\hat{r}(x|\theta_0, \theta_1)]!$$

(And we can sample from $p(x, z|\theta)$ by running the simulator.)

The value of gold

We can calculate the joint likelihood ratio

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)}$$

$\downarrow$

We want the likelihood ratio function

$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

With $r(x, z|\theta_0, \theta_1)$, we define a functional like

$$L_r[\hat{r}(x|\theta_0, \theta_1)] = \int \mathrm{d}x \int \mathrm{d}z \; p(x, z|\theta_1) \left[\left(\hat{r}(x|\theta_0, \theta_1) - r(x, z|\theta_0, \theta_1) \right)^2 \right].$$

It is minimized by

$$r(x|\theta_0, \theta_1) = \arg\min_{\hat{r}(x|\theta_0, \theta_1)} L_r[\hat{r}(x|\theta_0, \theta_1)]!$$

(And we can sample from $p(x, z|\theta)$ by running the simulator.)

.... and then magic ...

$$\mathbb{E}_{z \sim p(z|x, \theta_1)} \left[r(x, z|\theta_0, \theta_1) \right] = \int \mathrm{d}z \; p(z|x, \theta_1) \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)}$$

$$= \int \mathrm{d}z \; \frac{p(x, z|\theta_1)}{p(x|\theta_1)} \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)}$$

$$= r(x|\theta_0, \theta_1) \;!$$

Learning the score

Similar to the joint likelihood ratio, from the simulator we can extract the joint score

$$t(x, z|\theta_0) \equiv \nabla_\theta \log p(x, z_d, z_s, z_p|\theta)\Big|_{\theta_0}$$

↓

We want the score

$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$$

Learning the score

Similar to the joint likelihood ratio, from the simulator we can extract the joint score

$$t(x, z|\theta_0) \equiv \nabla_\theta \log p(x, z_d, z_s, z_p|\theta)\Big|_{\theta_0}$$

We want the score

$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$$

Given $t(x, z|\theta_0)$,
we define the functional

$$L_t[\hat{t}(x|\theta_0)] = \int \mathrm{d}x \int \mathrm{d}z \;\; p(x, z|\theta_0) \, \left[\left(\hat{t}(x|\theta_0) - t(x, z|\theta_0)\right)^2\right].$$

One can show it is minimized by

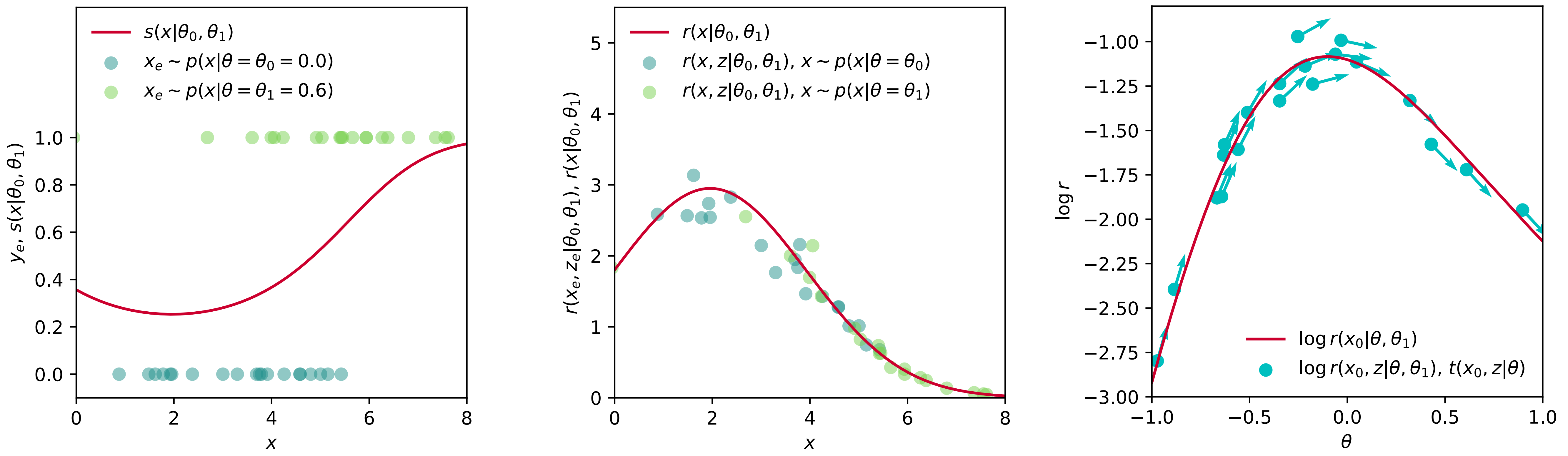
$$t(x|\theta_0) = \underset{\hat{t}(x|\theta_0)}{\arg\min} \, L_t[\hat{t}(x|\theta_0)] .$$

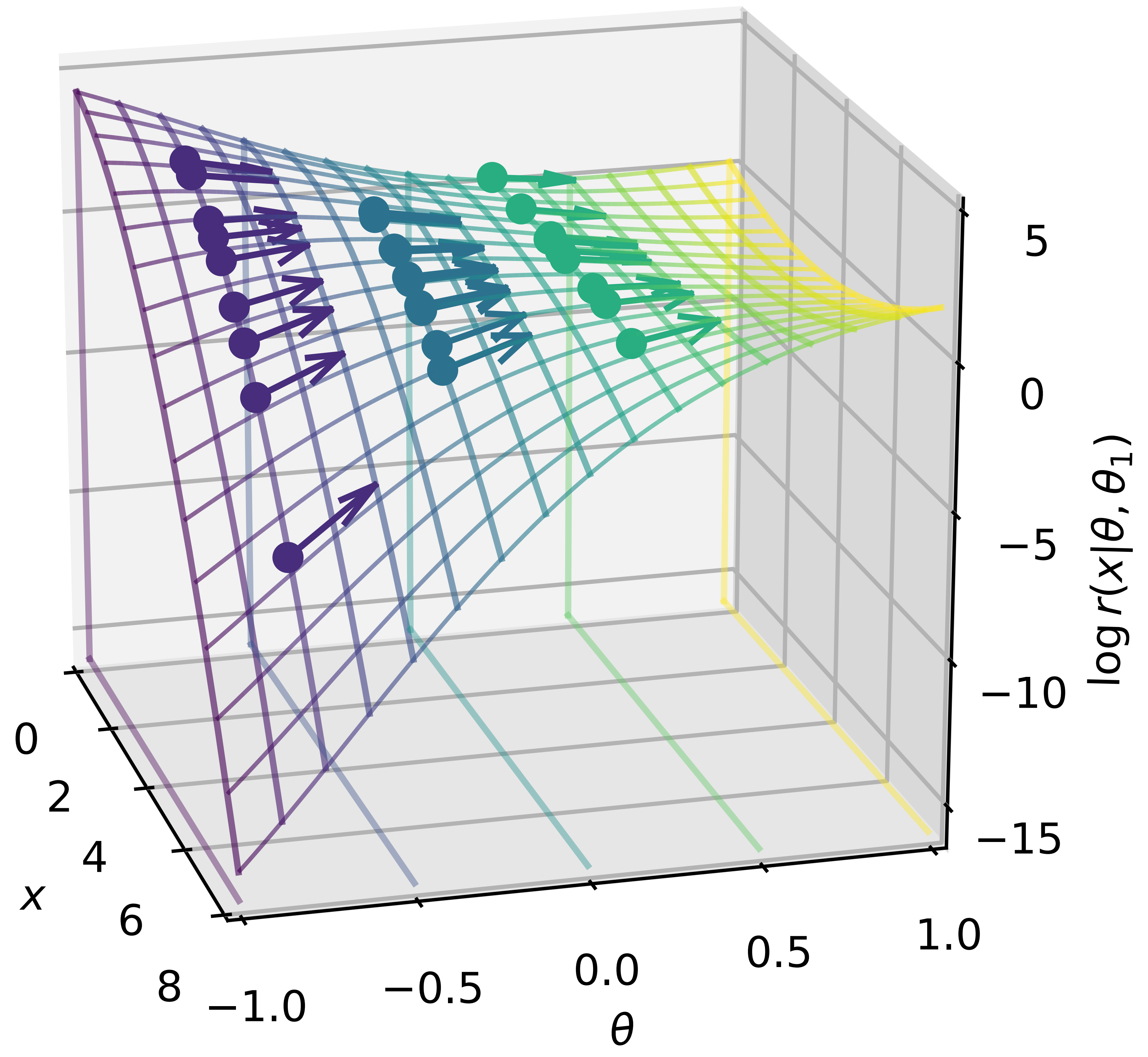
Again, we implement this minimization through machine learning.

Gold mining: augmenting the training data

The augmented training data converts supervised classification into supervised regression with lower variance

- improvement in training efficiency



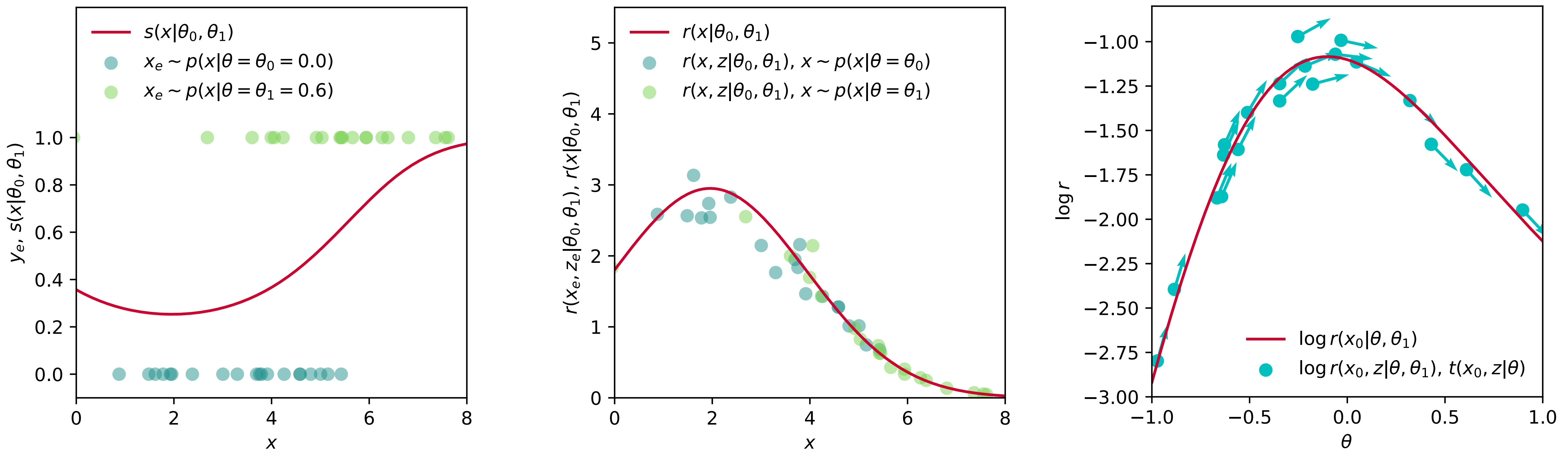


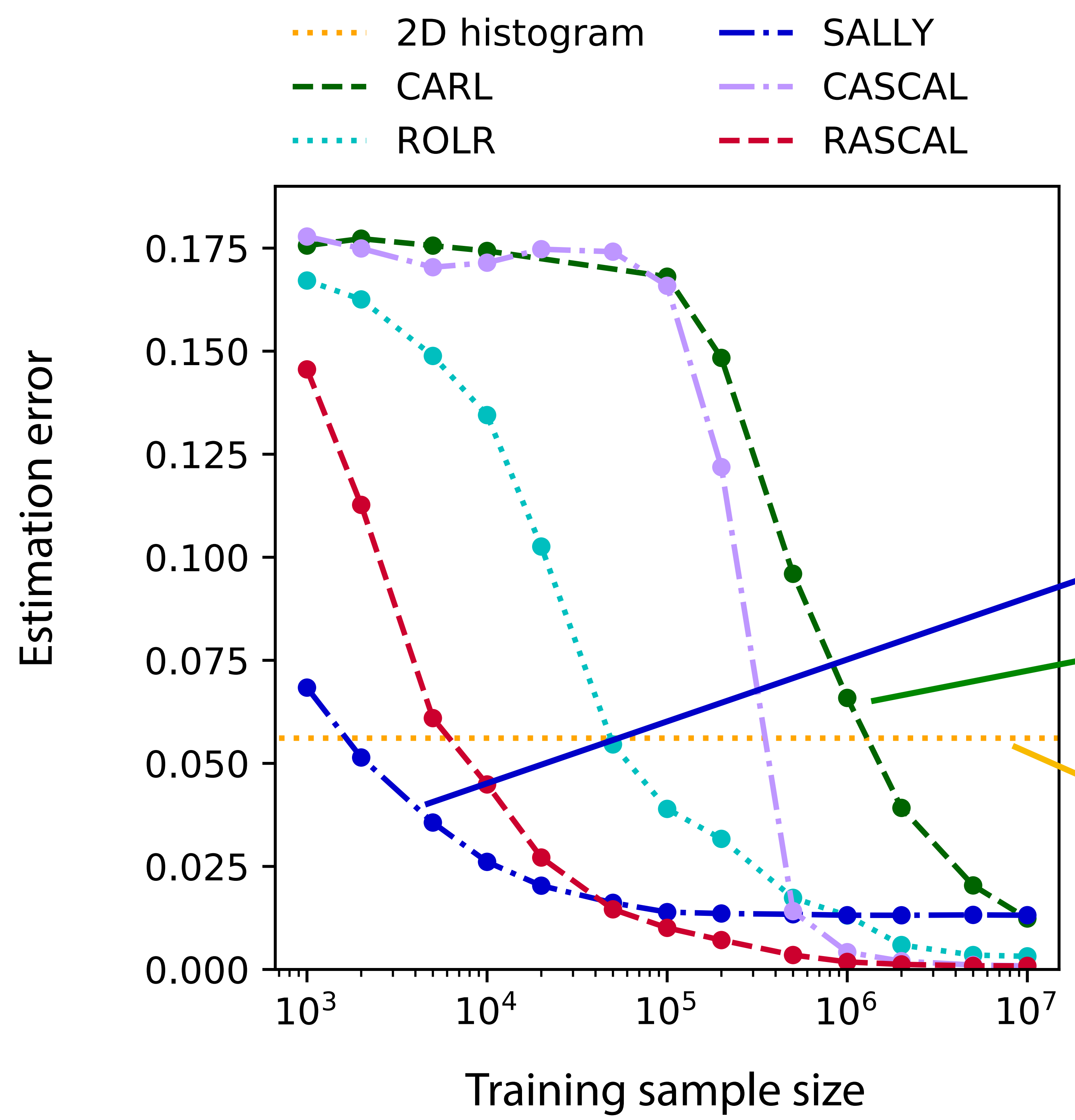
Brehmer, Louppe, Pavez, KC, PNAS (2019), arXiv:1805.12244

Gold mining: augmenting the training data

The augmented training data converts supervised classification into supervised regression with lower variance

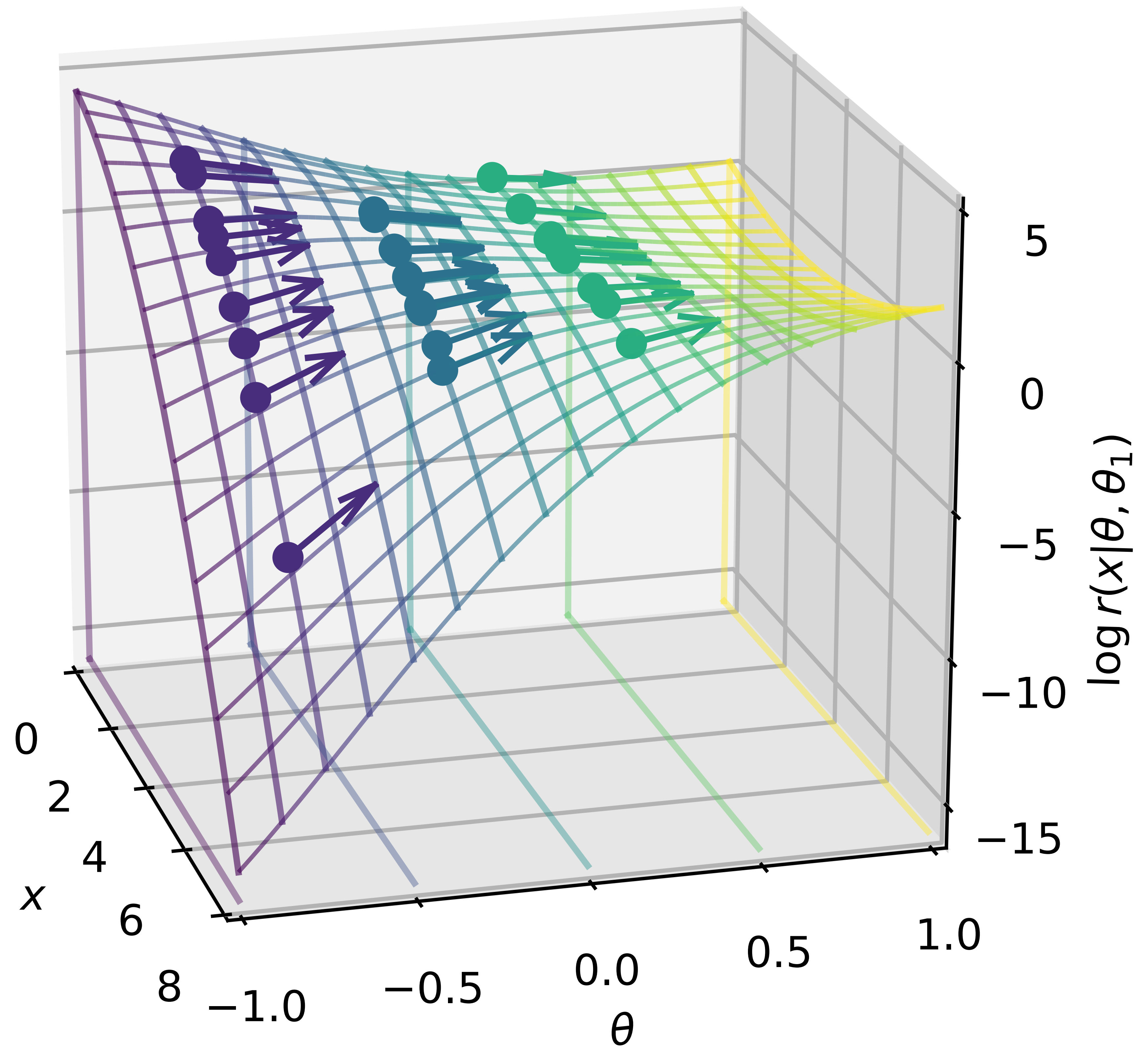
- improvement in training efficiency





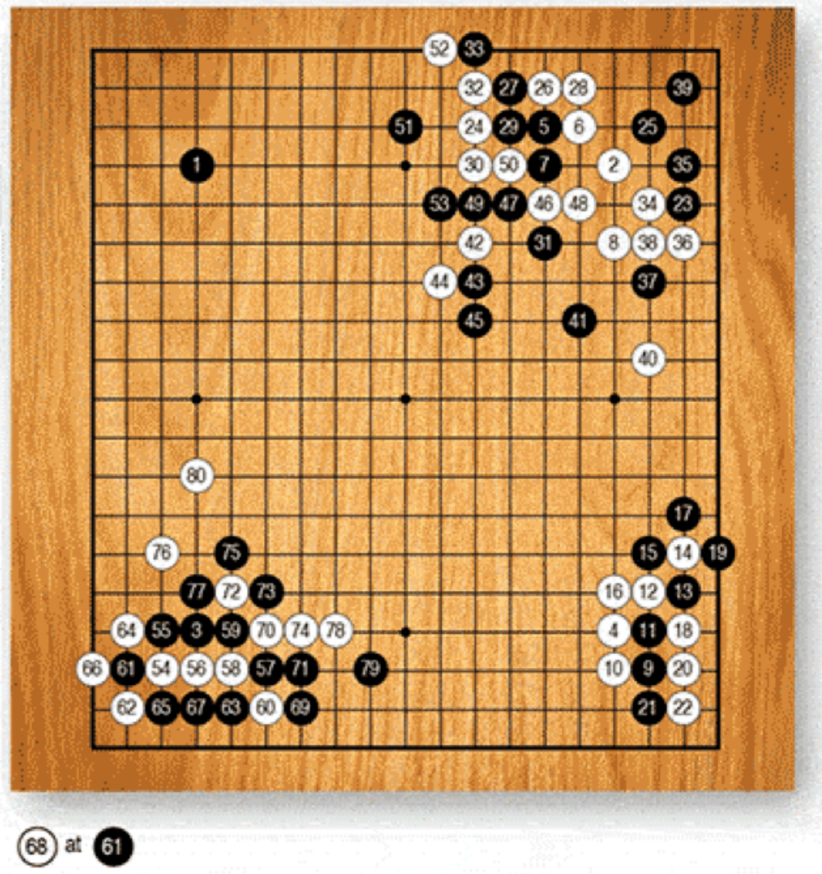
New techniques require less data than without augmented data

Traditional Approach no NN



From inference to design

Reinforcement Learning & Scientific Method



Captured Stones

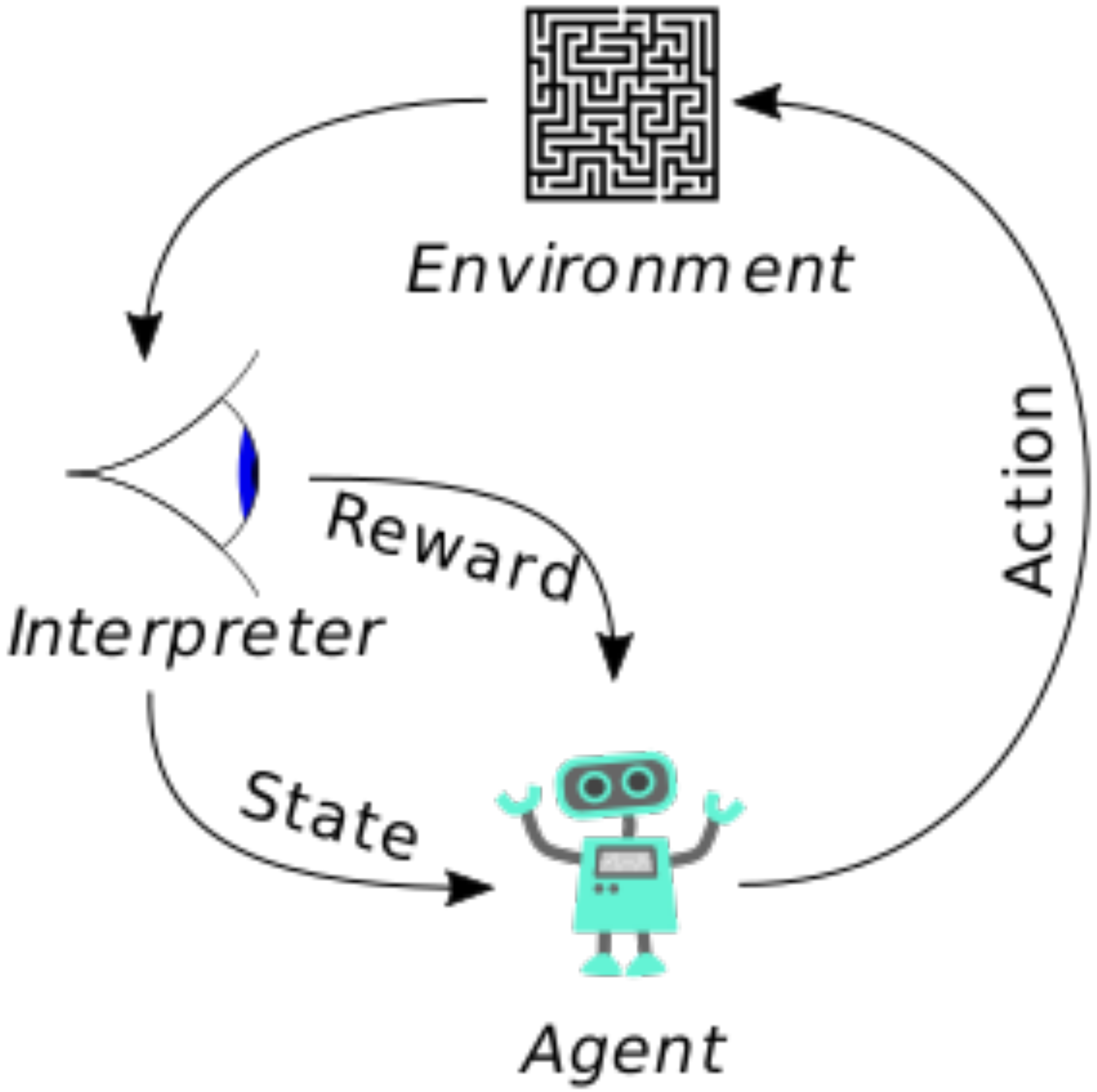
70 hours

AlphaGo Zero plays at super-human level. The game is disciplined and involves multiple challenges across the board.

40 days

AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.

Reinforcement Learning & Scientific Method

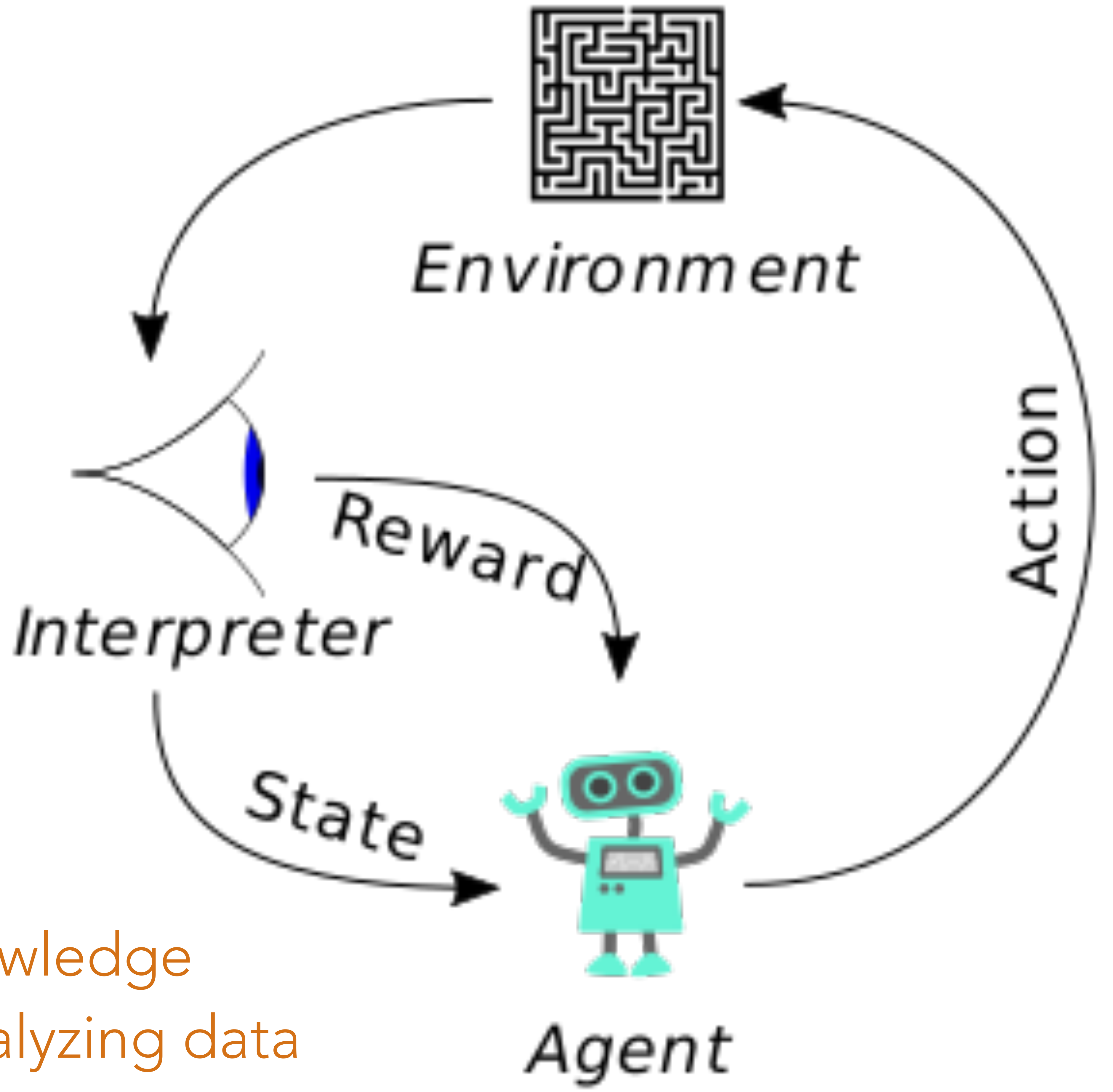


Reinforcement Learning & Scientific Method

Scientist trying to decide what experiment to do next

perform experiment,
gather data

statistical analysis



decide which
experiment to
perform

updated knowledge
based on analyzing data

Statistical Decision theory in 1 slide

Θ - States of nature; X - possible observations; A - action to be taken

$p(x|\theta)$ - statistical model (likelihood); $\pi(\theta)$ - prior

$\delta: X \rightarrow A$ - **decision rule** (take some action based on observation)

$L: \Theta \times A \rightarrow \mathbb{R}$ - **loss function**, real-valued function true parameter and action

$R(\theta,\delta) = E_{p(x|\theta)}[L(\theta, \delta)]$ - **risk**

$r(\pi, \delta) = E_{\pi(\theta)}[R(\theta,\delta)]$ - **Bayes risk** (expectation over θ w.r.t. prior and possible observations)

$\rho(\pi, \delta \mid x) = E_{\pi(\theta|x)}[L(\theta,\delta(x))]$ - **expected loss** (expectation over θ w.r.t. posterior $\pi(\theta|x)$)

Expected Information Gain

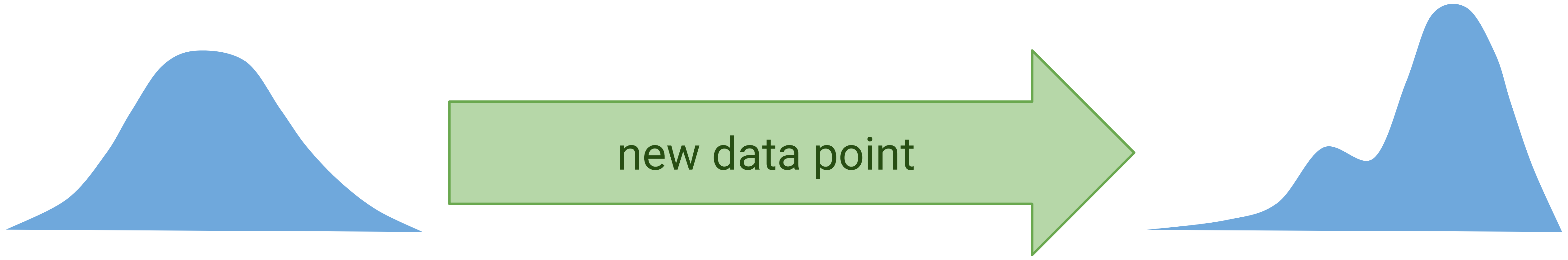
Choose experiment that will maximize expected information gain in prior → posterior

Active Learning & Control

Dustin Tran
Research Scientist at Google Brain
trandustin@google.com
Blog

Given data points $\{x, y\}$, how to select the next data point to fit the model?

Ex. Select data points which maximize expected information gain. [Lindley et al. 1956; Mackay 1992; Houthooft et al. 2016]



$$\arg\max_{\boldsymbol{x}} \mathrm{H}[\boldsymbol{\theta}|\mathcal{D}] - \mathbb{E}_{y\sim p(y|\boldsymbol{x}\mathcal{D})}[\mathrm{H}[\boldsymbol{\theta}|y,\boldsymbol{x},\mathcal{D}]]$$

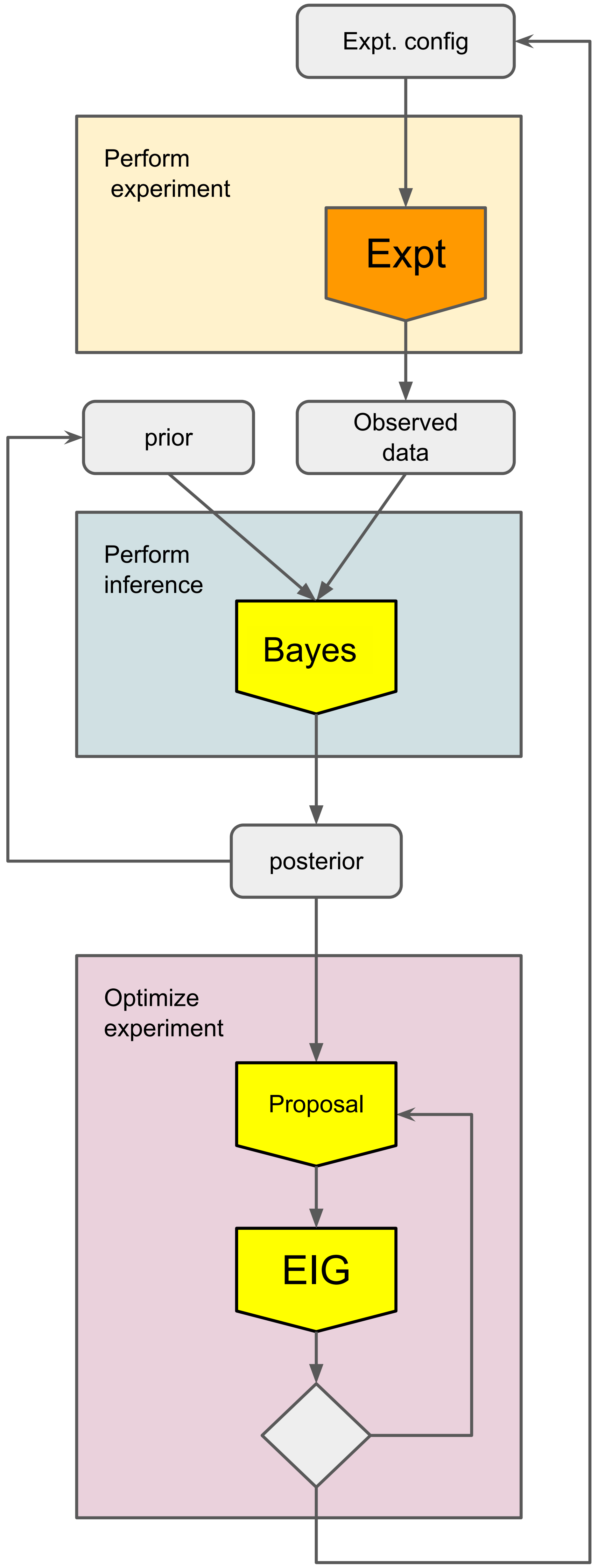
Uncertainty determines which **x** is most informative and, therefore, the model's success.

[Hafner et al., 2019]

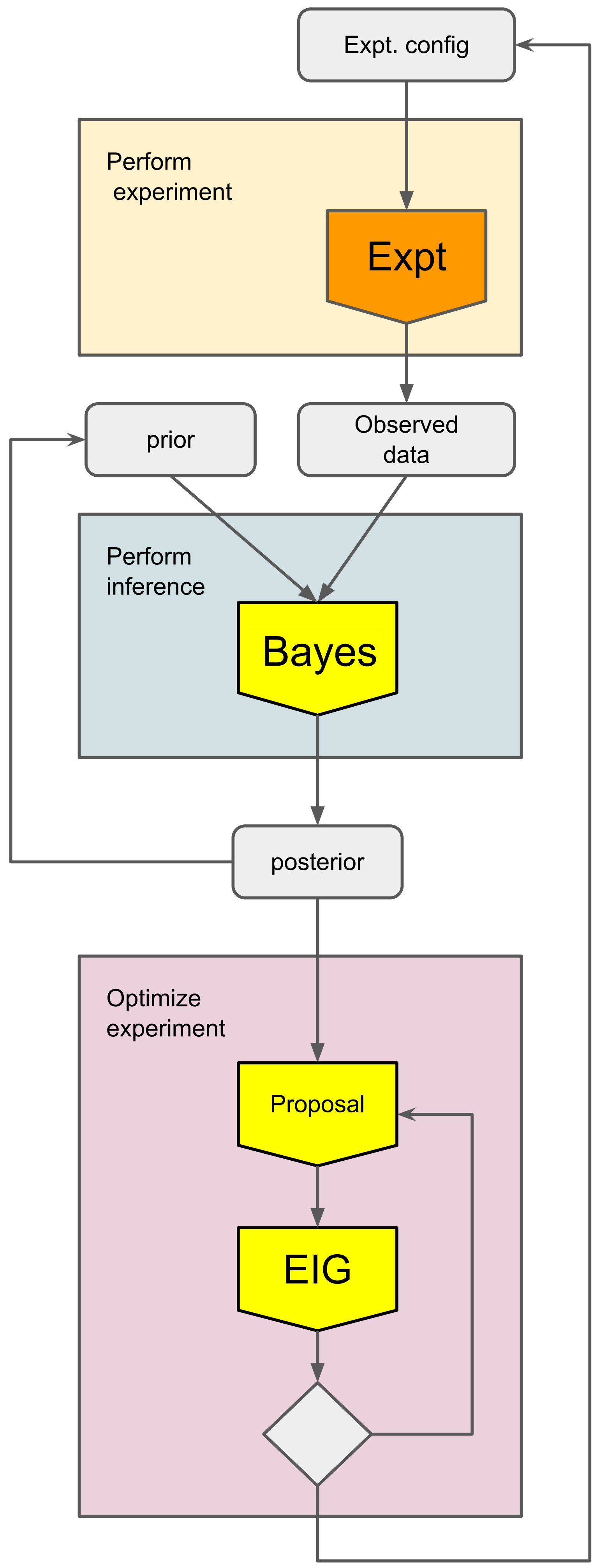
Slide from Dustin Tran at Hammers & Nails 2019

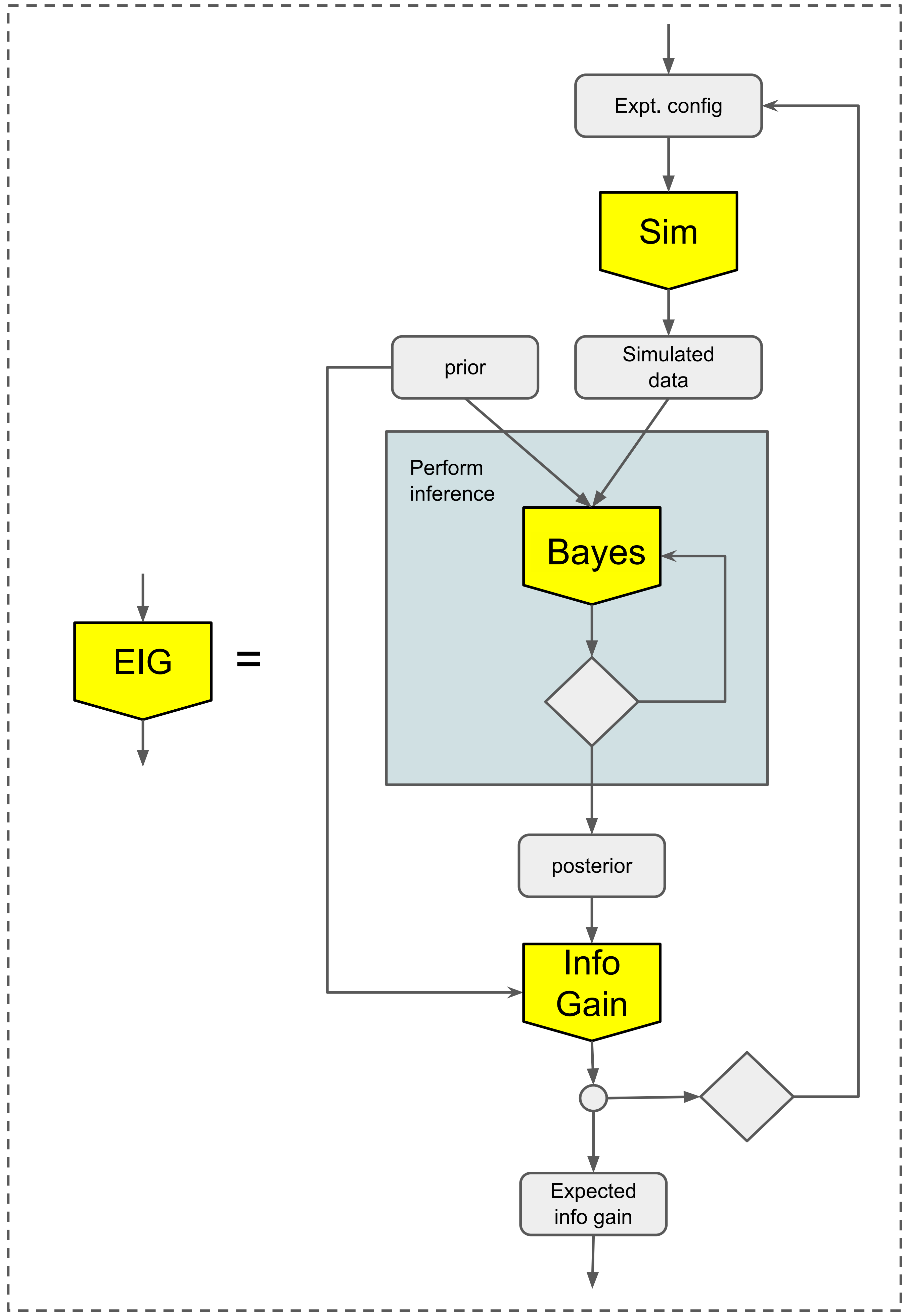
"Active Sciencing"

https://github.com/cranmer/active_sciencing



"Active Sciencing"

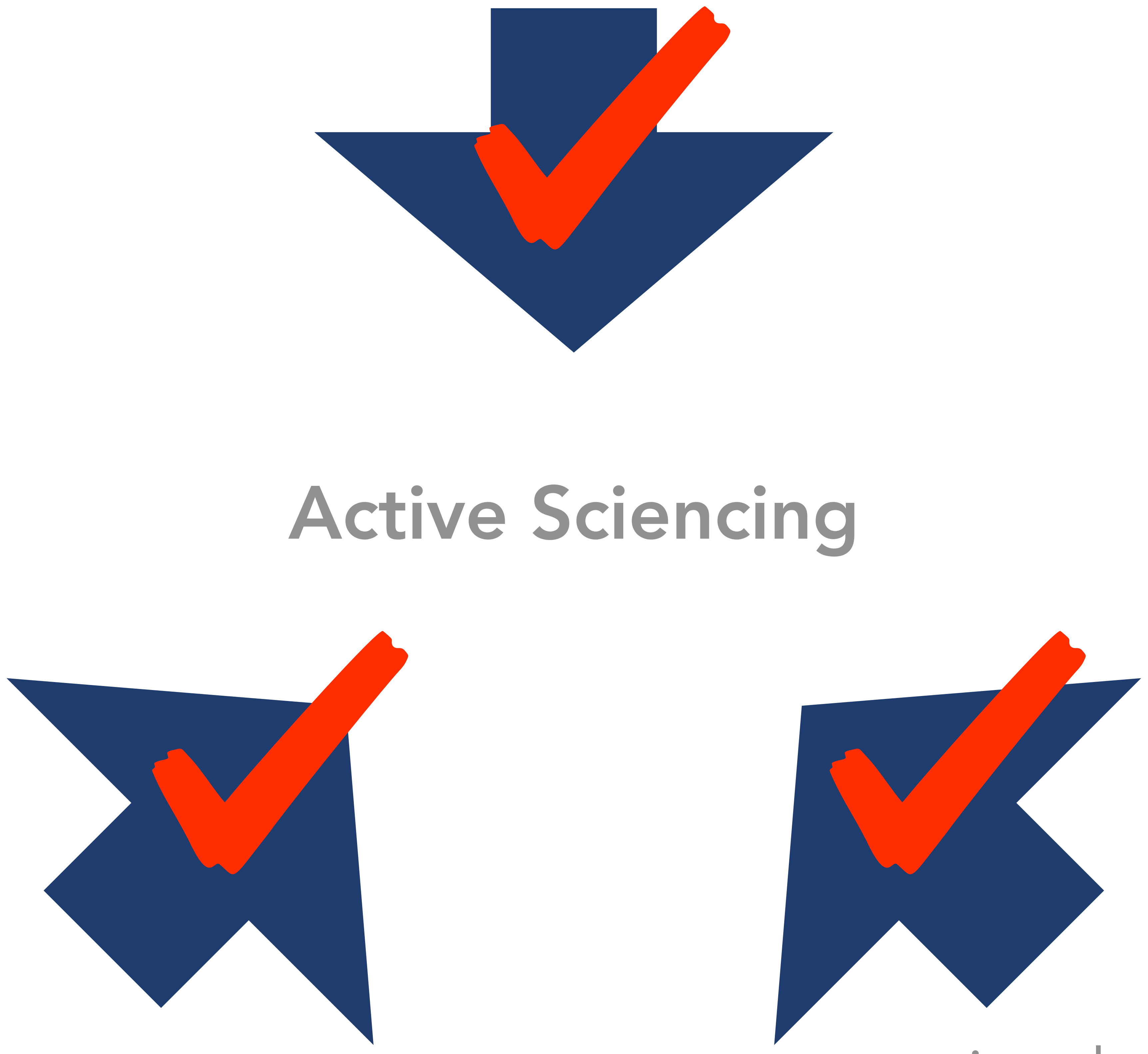




https://github.com/cranmer/active_sciencing

Synthesis

active learning / sequential design / black box optimization



Active Sciencing

reusable workflows

simulation-based /
likelihood-free
inference engines

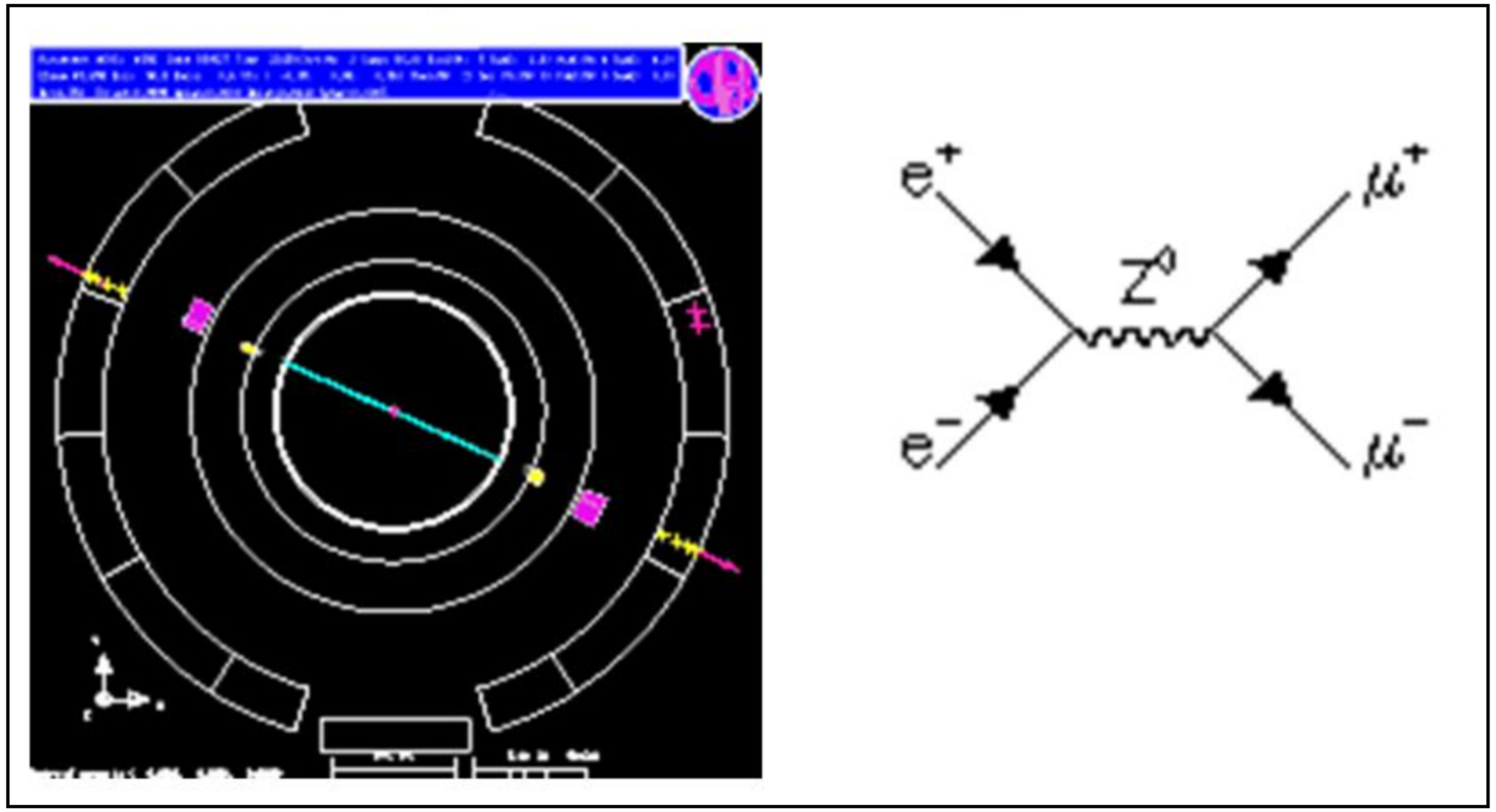
Experimental Design

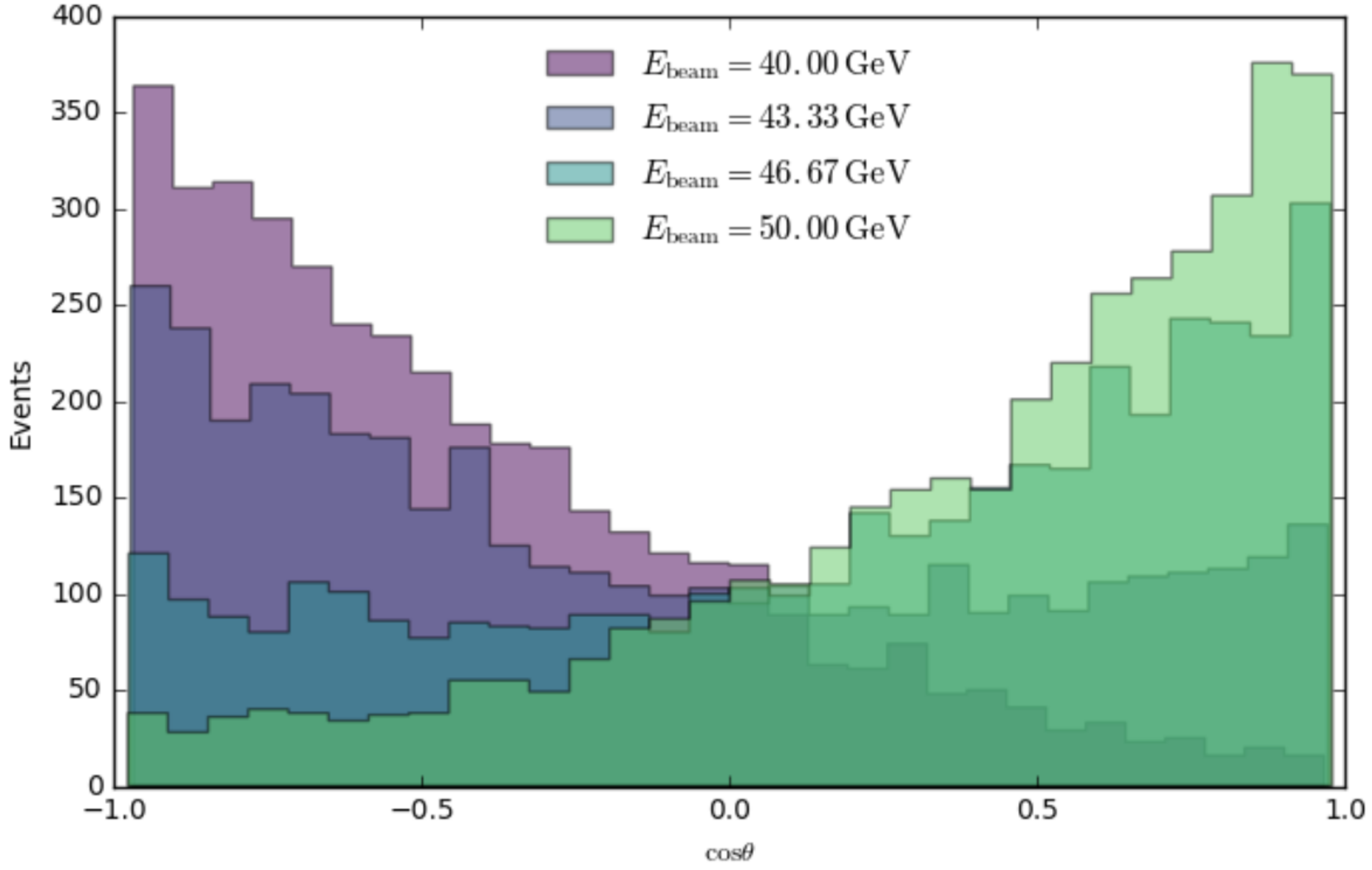
Goal:

- measure parameter of theory (eg. Weinberg angle in Standard Model of particle Physics) from raw data
- optimize experiment (eg. beam energy) for most sensitive measurement

$$
\begin{aligned}
\mathcal{L}_{SM} = \; & \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}}_{\text{kinetic energies and self-interactions of the gauge bosons}} \\
& + \underbrace{\bar{L}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})L + \bar{R}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g'YB_{\mu})R}_{\text{kinetic energies and electroweak interactions of fermions}} \\
& + \underbrace{\frac{1}{2}\left|(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})\,\phi\right|^{2} - V(\phi)}_{W^{\pm},Z,\gamma,\text{and Higgs masses and couplings}} \\
& + \underbrace{g''(\bar{q}\gamma^{\mu}T_{a}q)\,G^{a}_{\mu}}_{\text{interactions between quarks and gluons}} + \underbrace{(G_{1}\bar{L}\phi R + G_{2}\bar{L}\phi_{c}R + h.c.)}_{\text{fermion masses and couplings to Higgs}}
\end{aligned}
$$

$$
\begin{pmatrix} \gamma \\ Z^{0} \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} B^{0} \\ W^{0} \end{pmatrix}
$$

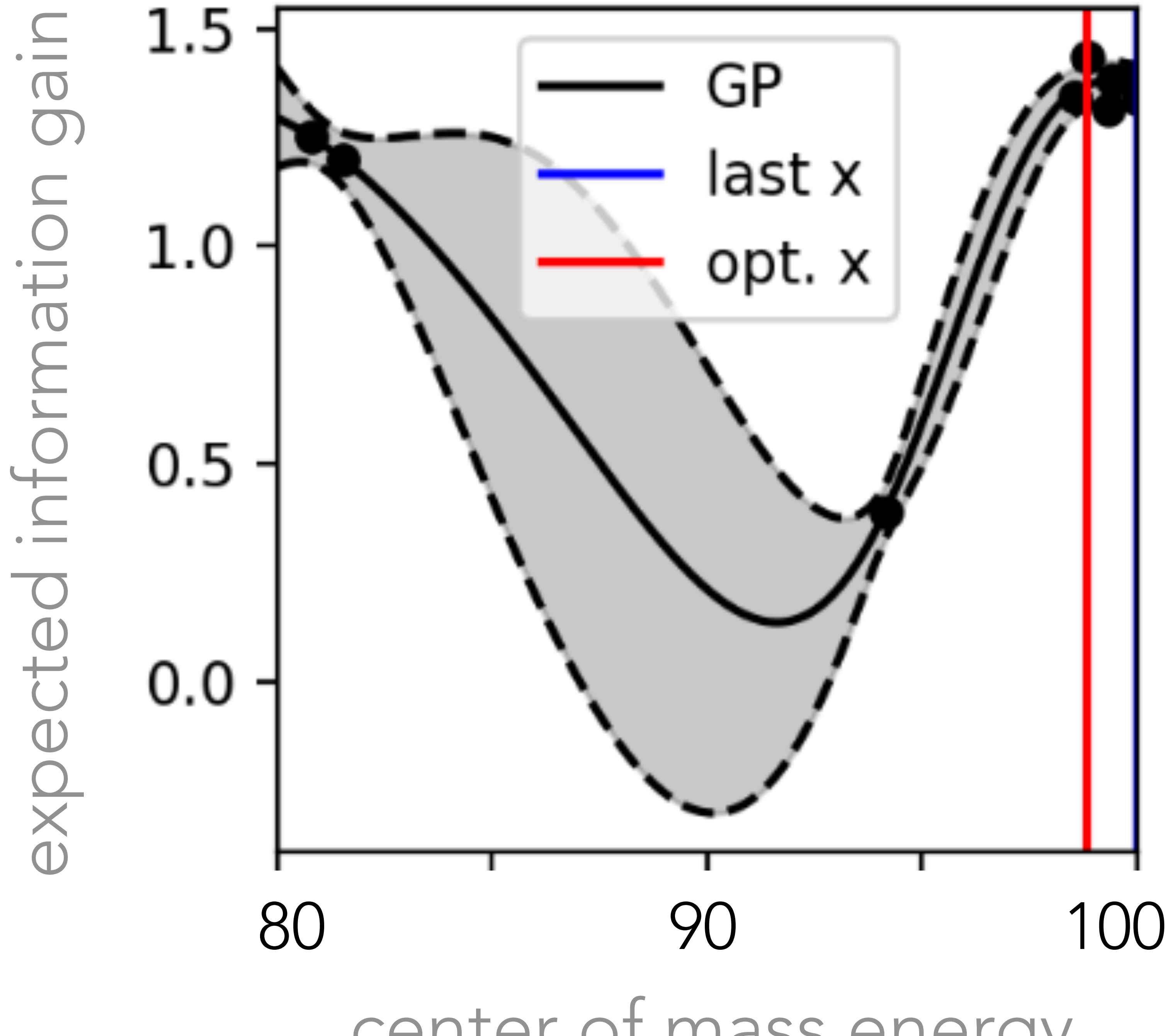




https://github.com/cranmer/active_sciencing

Proof-of-principle algorithm can:

- measure / infer the parameter of the theory (eg. Weinberg angle in Standard Model of particle Physics) from raw data using simulation-based inference
- optimize experiment (eg. beam energy) for most sensitive measurement



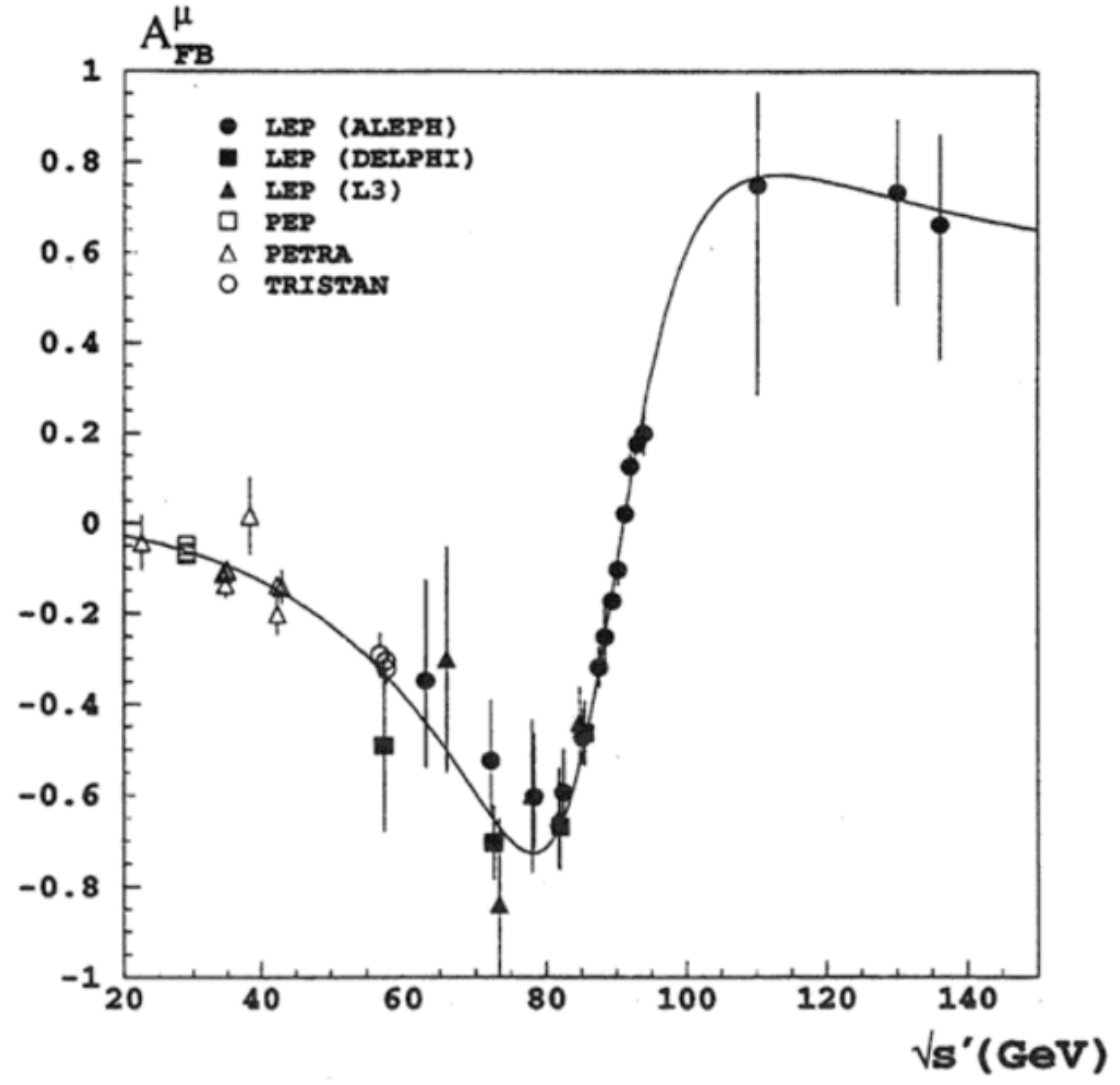
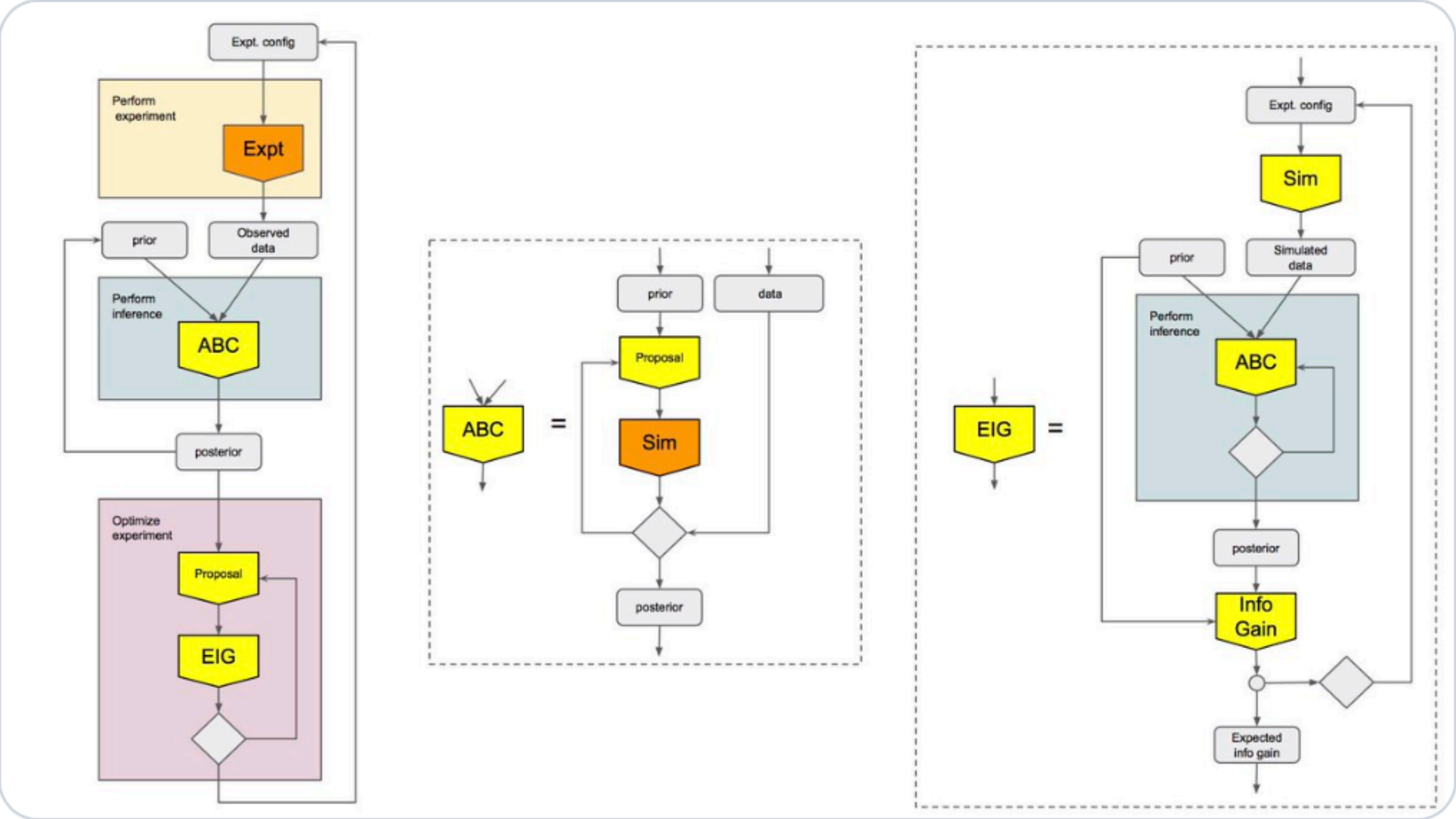


Figure 2: Measured forward-backward asymmetries of muon-pair production compared with the model independent fit results.

Kyle Cranmer @KyleCranmer · Jun 11, 2017

Demo for YComb research
active learning + workflows + implicit models = #ActiveSciencing
@lukasheinrich_ @glouppe
github.com/cranmer/active...



4 22 61

Danilo J. Rezende @DeepSpiker · Jul 19, 2017

This is great!

1 3

Kyle Cranmer @KyleCranmer · Jul 19, 2017

Thanks!!!

1 2

Danilo J. Rezende
@DeepSpiker

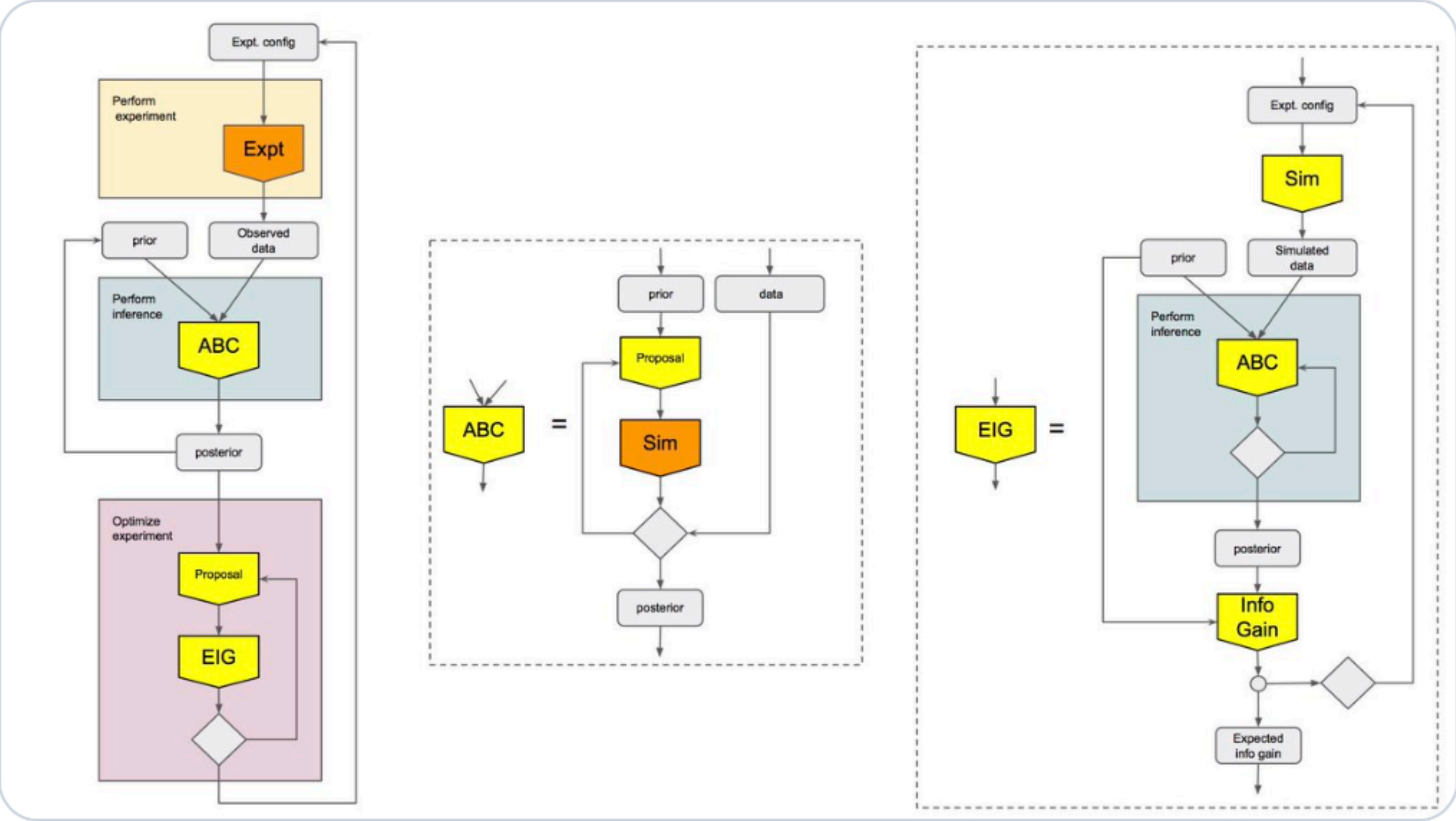
Replying to @KyleCranmer @lukasheinrich_ and @glouppe

You have the full loop of the scientific method in a python notebook :)

3:12 PM · Jul 19, 2017 · Twitter for iPhone

Kyle Cranmer @KyleCranmer · Jun 11, 2017

Demo for YComb research
active learning + workflows + implicit models = #ActiveSciencing
@lukasheinrich_ @glouppe
github.com/cranmer/active...



4 · 22 · 61

Danilo J. Rezende @DeepSpiker · Jul 19, 2017

This is great!

1 · 3

Kyle Cranmer @KyleCranmer · Jul 19, 2017

Thanks!!!

1 · 2

Danilo J. Rezende
@DeepSpiker

Replying to @KyleCranmer @lukasheinrich_ and @glouppe

You have the full loop of the scientific method in a python notebook :)

3:12 PM · Jul 19, 2017 · Twitter for iPhone

Reality check...

Keep in mind that

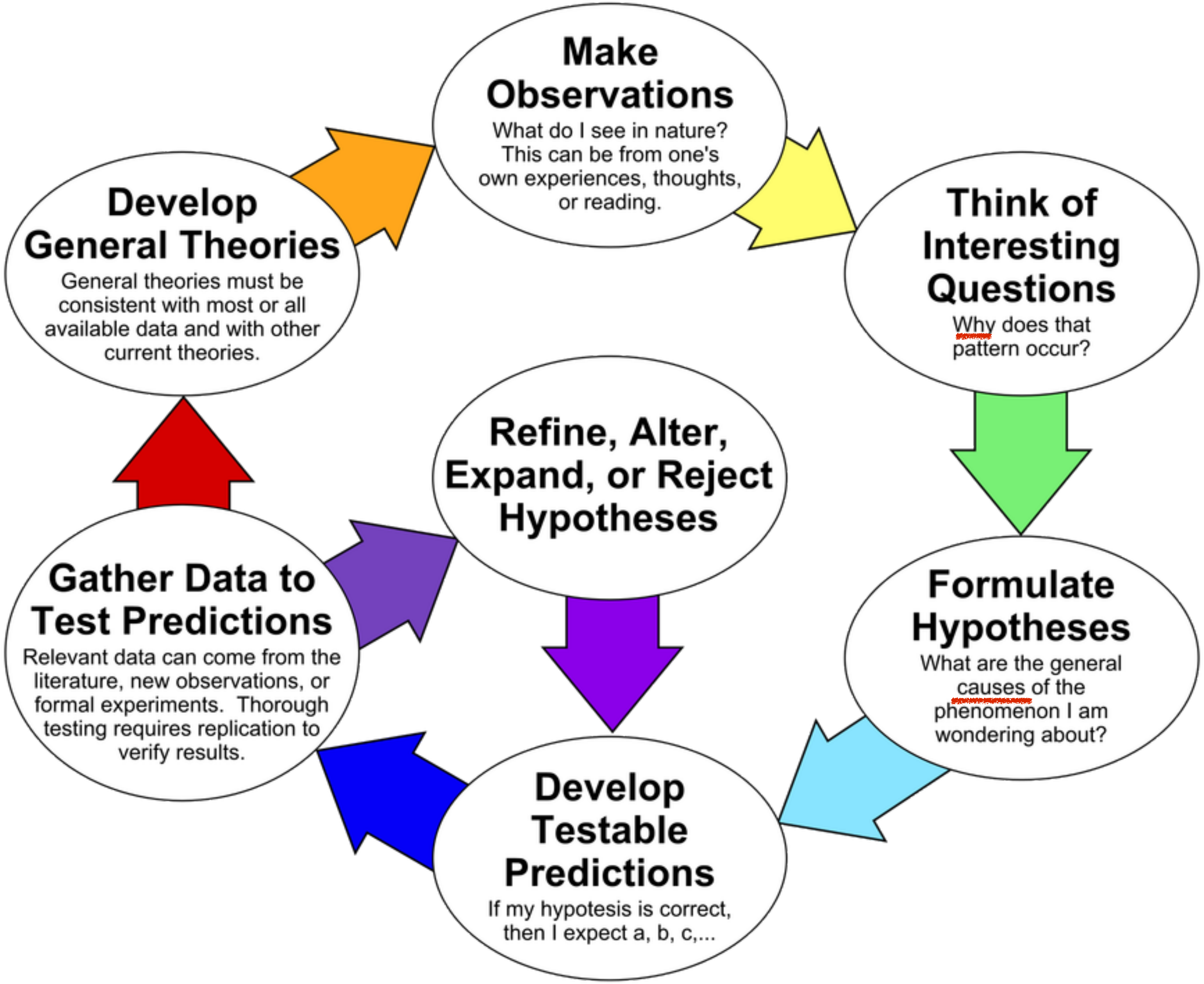
- the simulator model was specified
- the space of experimental configurations was well specified

Still it was hard enough!

Going to open world of experimental configurations and potential models much harder.

Hypothesis generation also hard.

The Scientific Method as an Ongoing Process

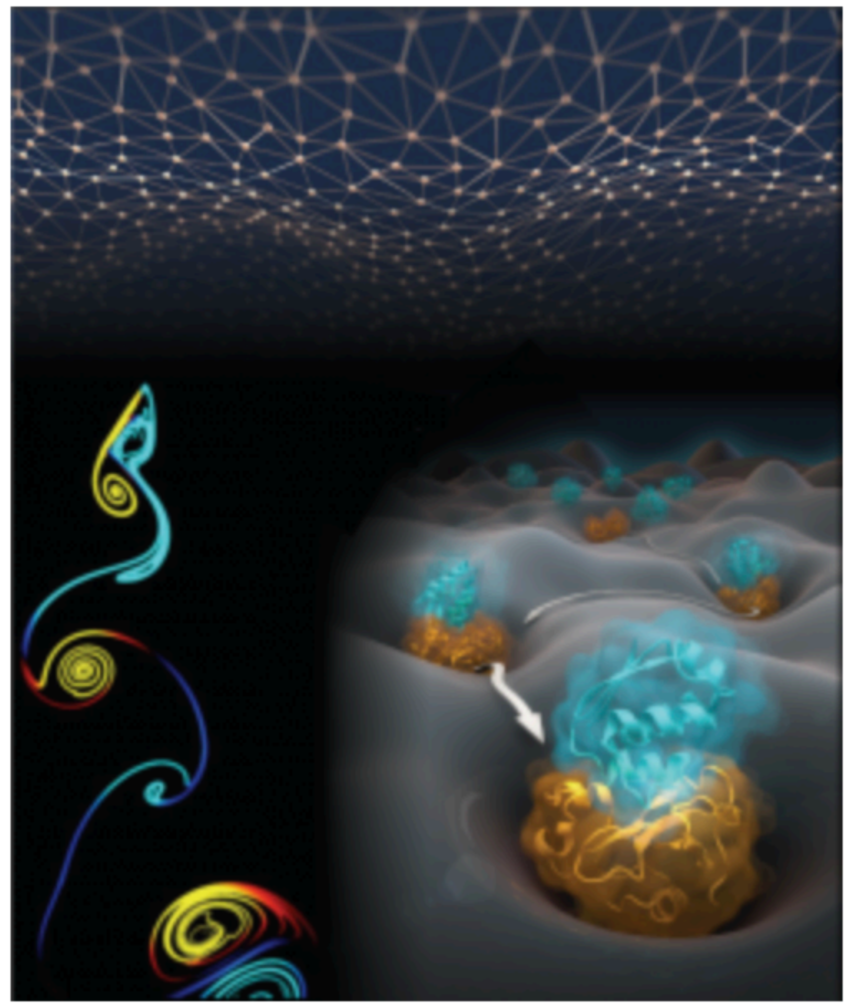


Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

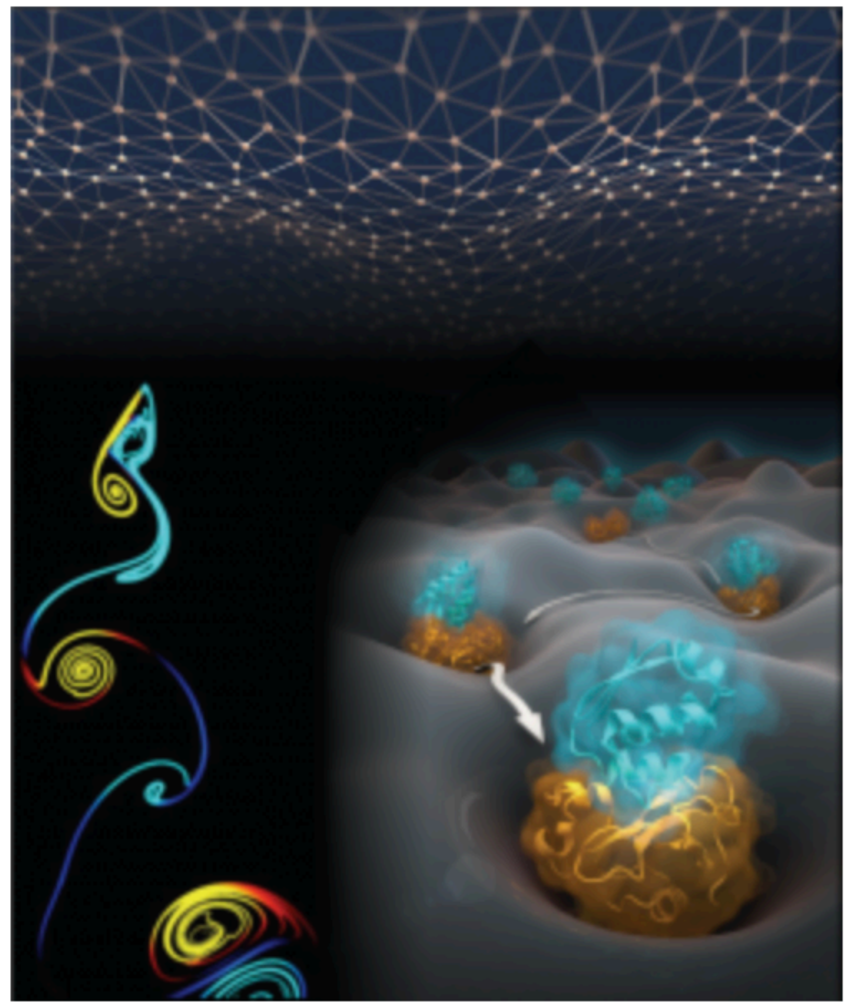
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

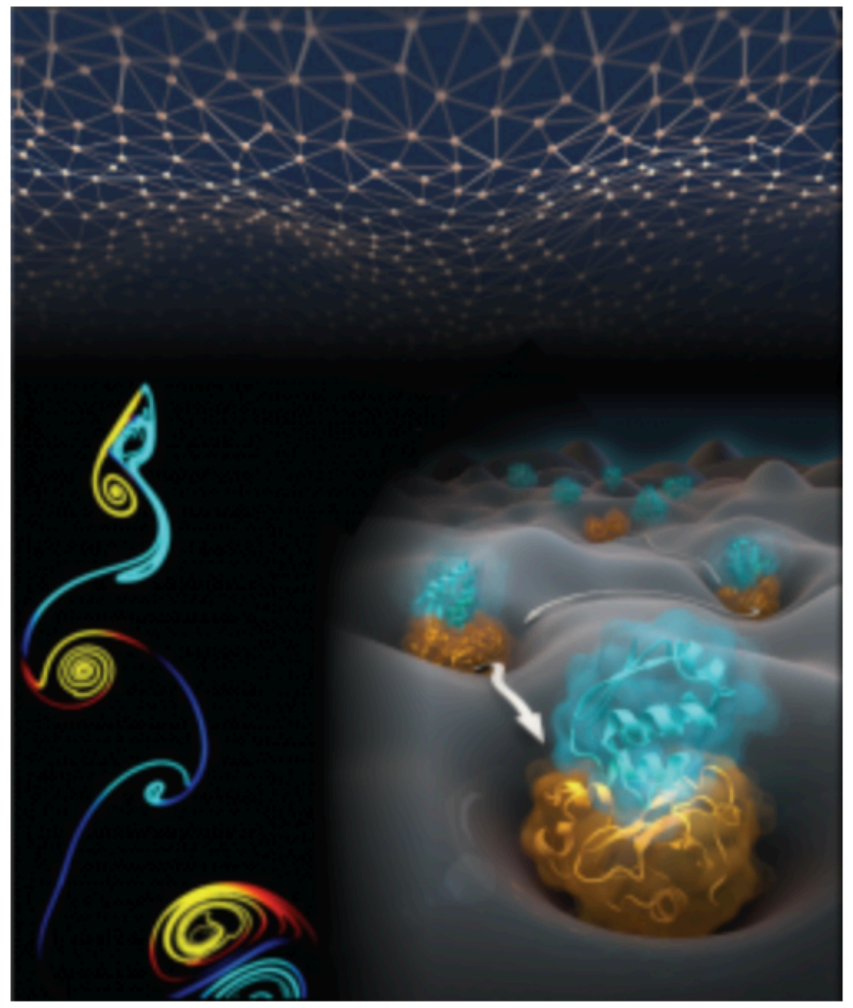
Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

Overview

Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.



As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

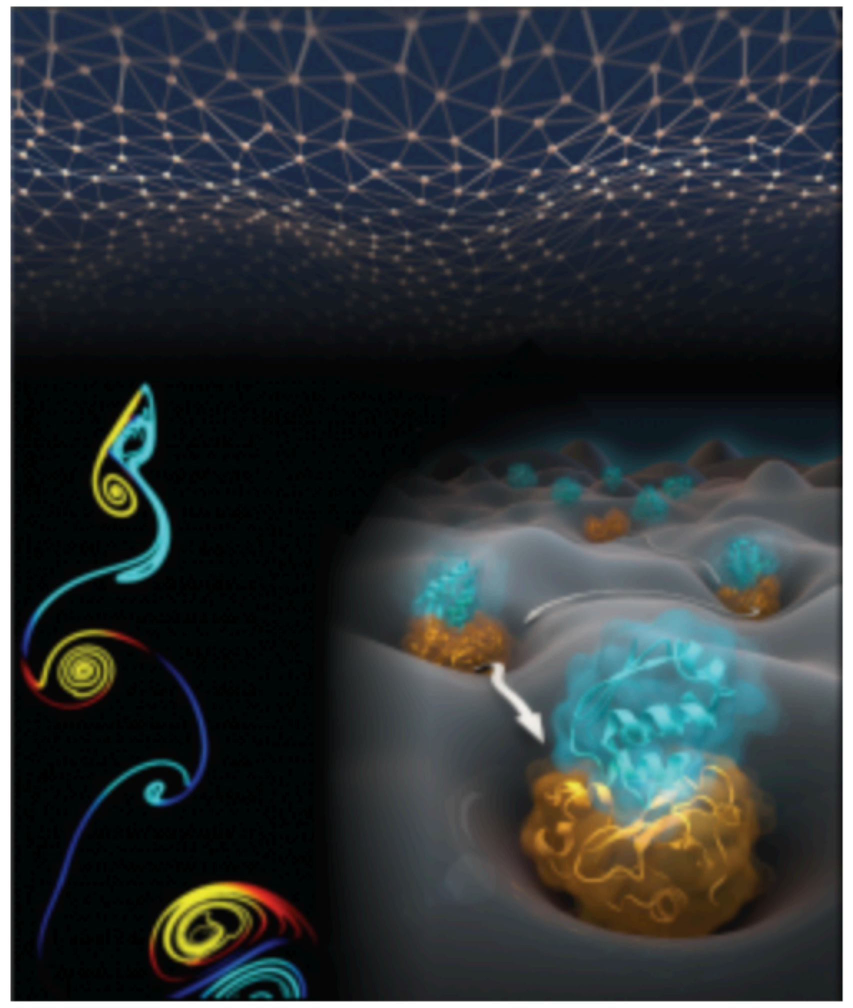
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

Machine Learning for Physics and the Physics of Learning

SEPTEMBER 4 - DECEMBER 8, 2019

OVERVIEW | PARTICIPANT LIST | ACTIVITIES | APPLICATION

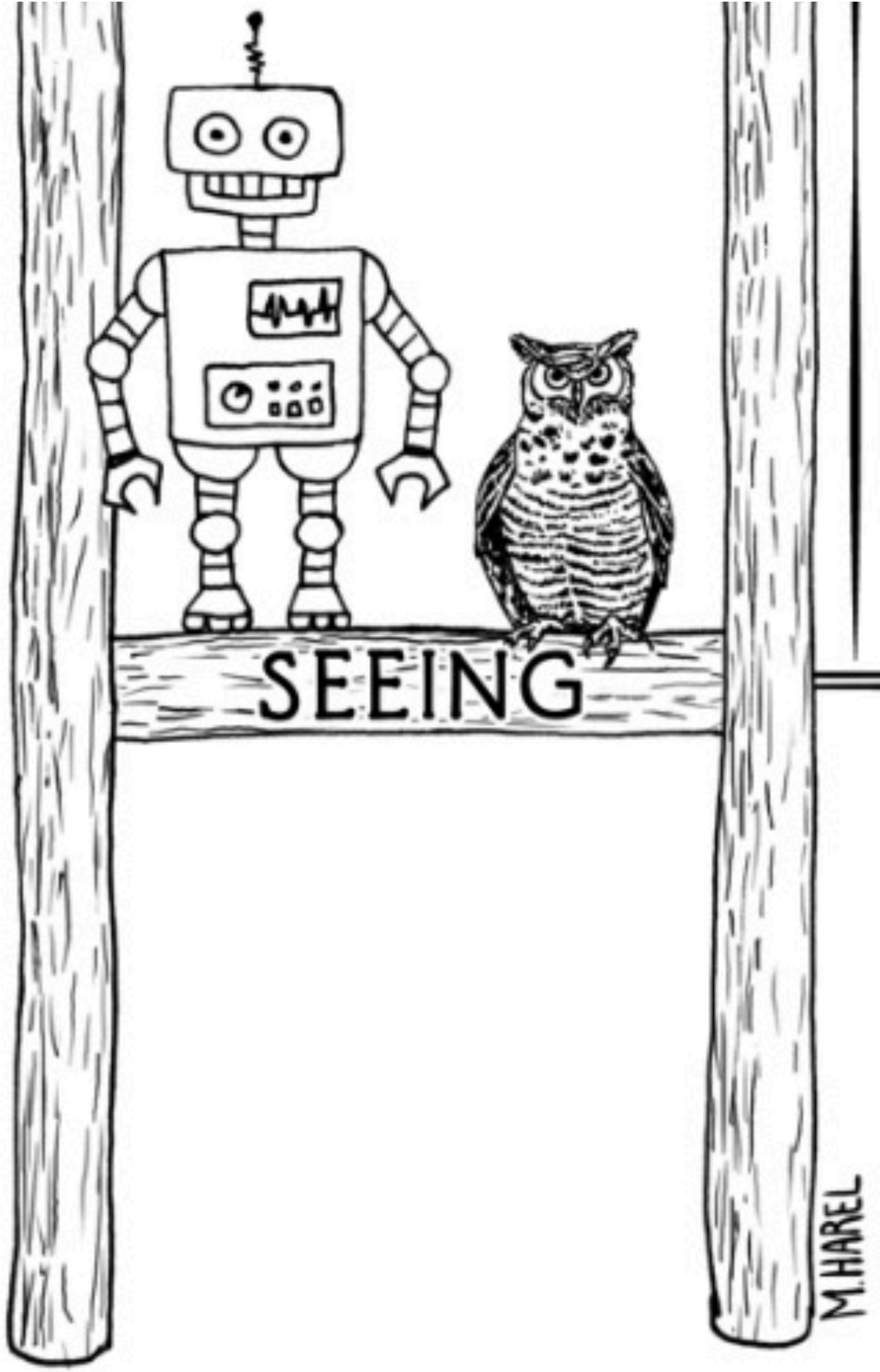
Overview



Machine Learning (ML) is quickly providing new powerful tools for physicists and chemists to extract essential information from large amounts of data, either from experiments or simulations. Significant steps forward in every branch of the physical sciences could be made by embracing, developing and applying the methods of machine learning to interrogate high-dimensional complex data in a way that has not been possible before.

As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity to learn the models themselves–that is, to learn the physical principles and structures underlying the data–and that with more realistic constraints, machine learning will also be able to generate and design complex and novel physical structures and objects. Finally, physicists would not just like to fit their data, but rather obtain models that are physically understandable; e.g., by maintaining relations of the predictions to the microscopic physical quantities used as an input, and by respecting physically meaningful constraints, such as conservation laws or symmetry relations.

The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information theory and statistics are intimately related in their goal to extract valid information from noisy data, and we want to push the cross-pollination further in the specific context of discovering physical principles from data.

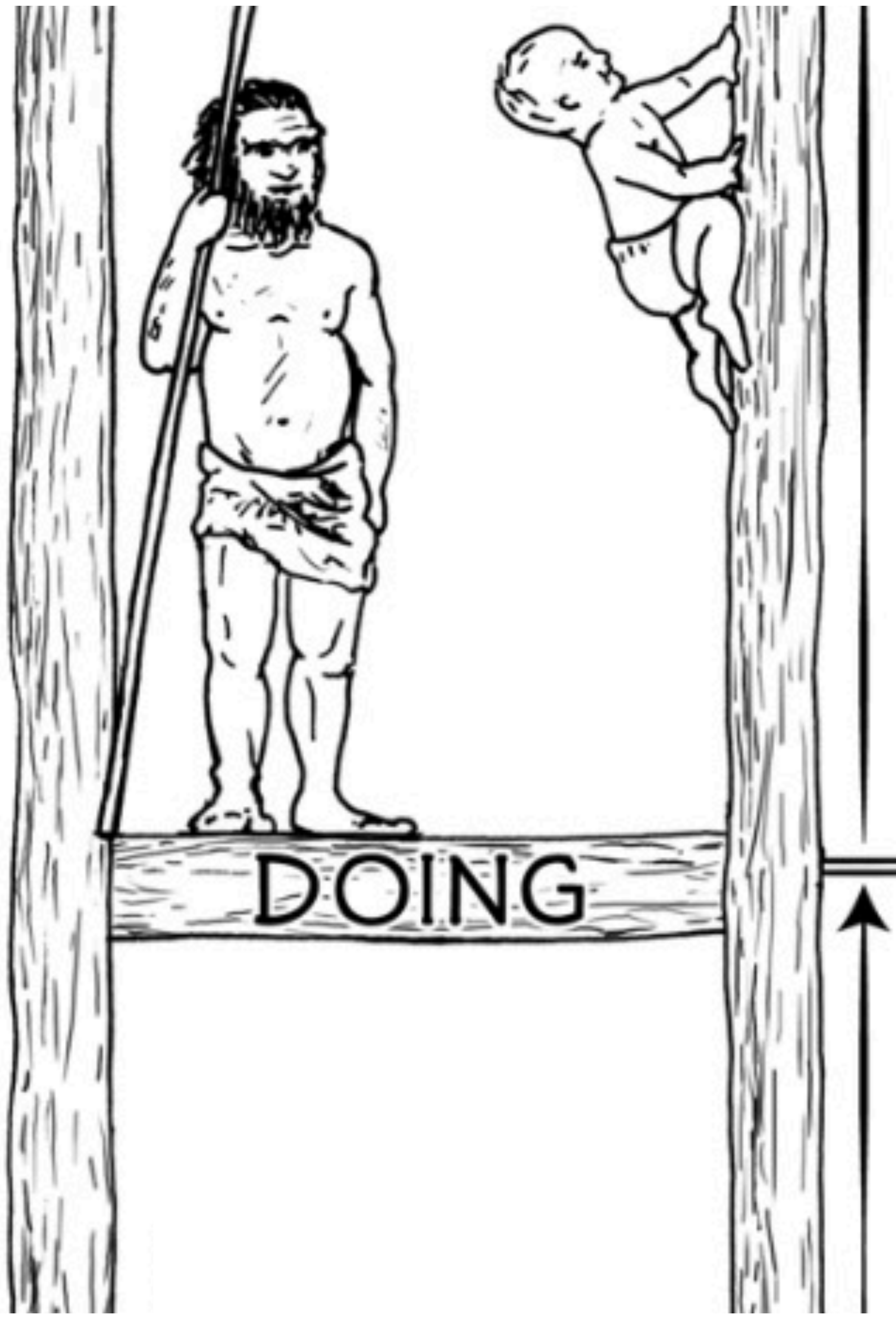


1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

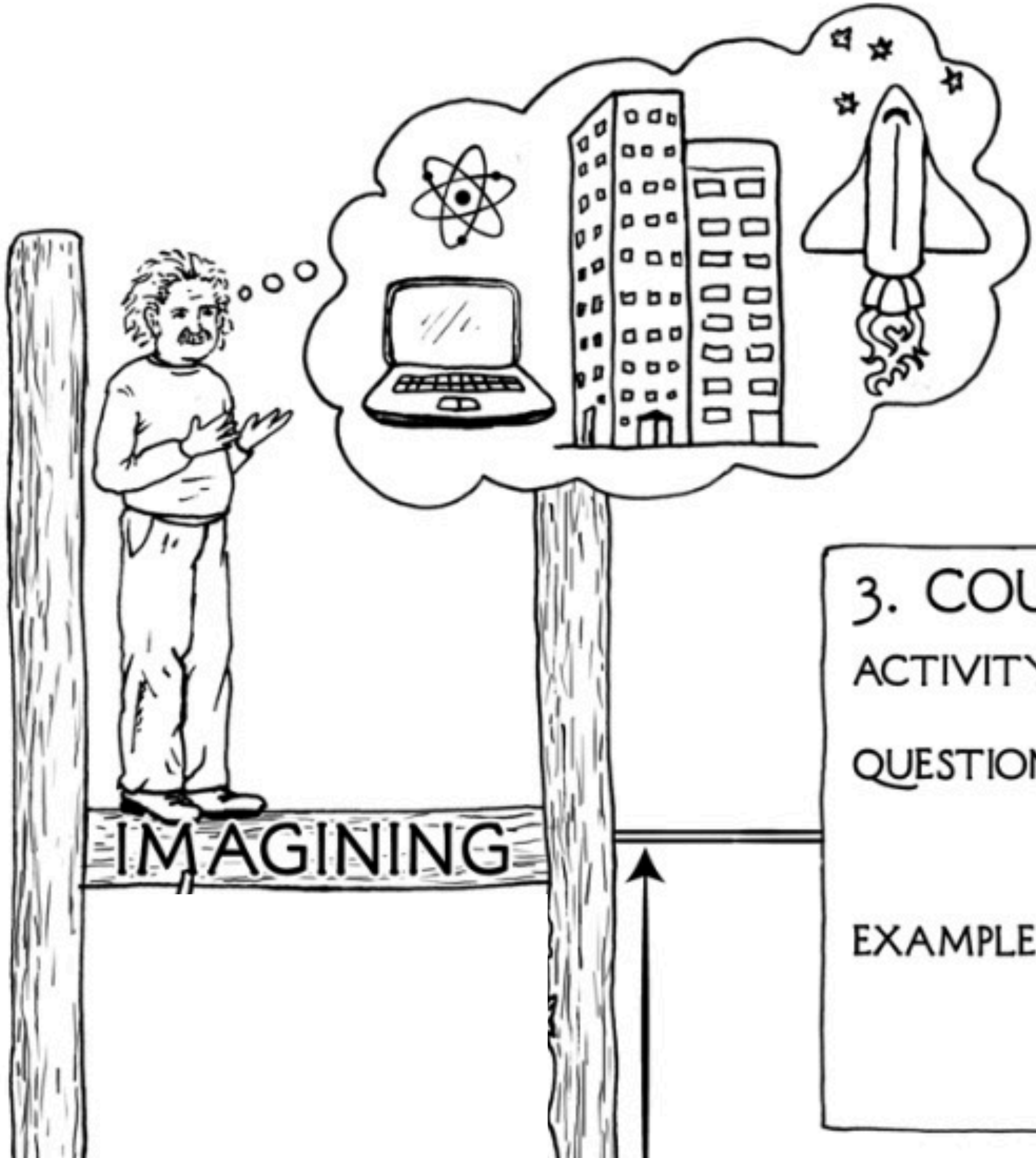


2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

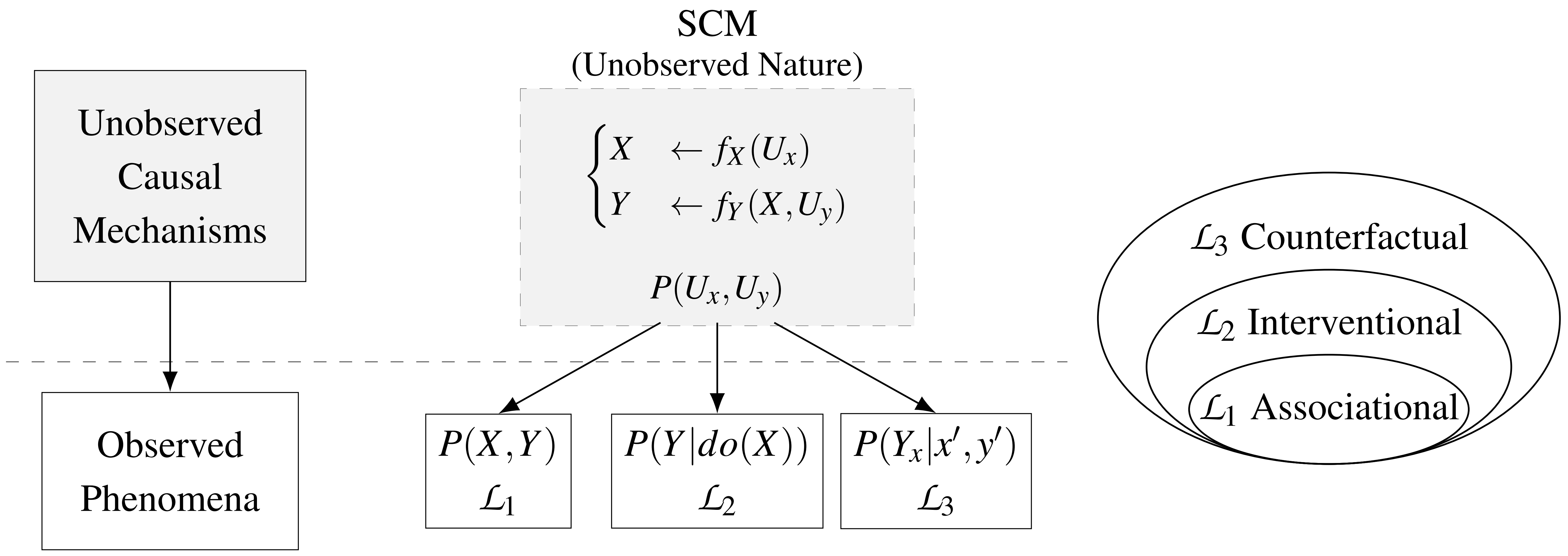
More formally

On Pearl's Hierarchy and the Foundations of Causal Inference

Elias Bareinboim[†] EB@CS.COLUMBIA.EDU
Juan D. Correa[†] JDCORREA@CS.COLUMBIA.EDU
[†] *Columbia University*
New York, NY 10027, USA

Duligur Ibeling[‡] DULIGUR@STANFORD.EDU
Thomas Icard[‡] ICARD@STANFORD.EDU
[‡] *Stanford University*
Stanford, CA 94305, USA

Abstract. Cause and effect relationships play a central role in how we perceive and make sense of the world around us, how we act upon it, and ultimately, how we understand ourselves. Almost two decades ago, computer scientist Judea Pearl made a breakthrough in understanding causality by discovering and systematically studying the "Ladder of Causation" [Pearl and Mackenzie 2018], a framework that highlights the distinct roles of *seeing*, *doing*, and *imagining*. In honor of this landmark discovery, we name this the *Pearl Causal Hierarchy* (PCH). In this chapter, we develop a novel and comprehensive treatment of the PCH through two complementary lenses, one logical-probabilistic and another inferential-graphical. Following Pearl's own presentation of the hierarchy, we begin by showing how the PCH organically emerges from a well-specified collection of causal mechanisms (a structural causal model, or SCM). We then turn to the logical lens. Our first result, the Causal Hierarchy Theorem (CHT), demonstrates that the three layers of the hierarchy almost always separate in a measure-theoretic sense. Roughly speaking, the CHT says that data at one layer virtually always underdetermines information at higher layers. Since in most practical settings the scientist does not have access to the precise form of the underlying causal mechanisms



(a) (b)

Figure 1.1: (a) Collection of causal mechanisms (or SCM) generating certain observed phenomena (qualitatively different probability distributions). (b) PCH's containment structure.

causal model, or SCM). We then turn to the logical lens. Our first result, the Causal Hierarchy Theorem (CHT), demonstrates that the three layers of the hierarchy almost always separate in a measure-theoretic sense. Roughly speaking, the CHT says that data at one layer virtually always underdetermines information at higher layers. Since in most practical settings

A toy example

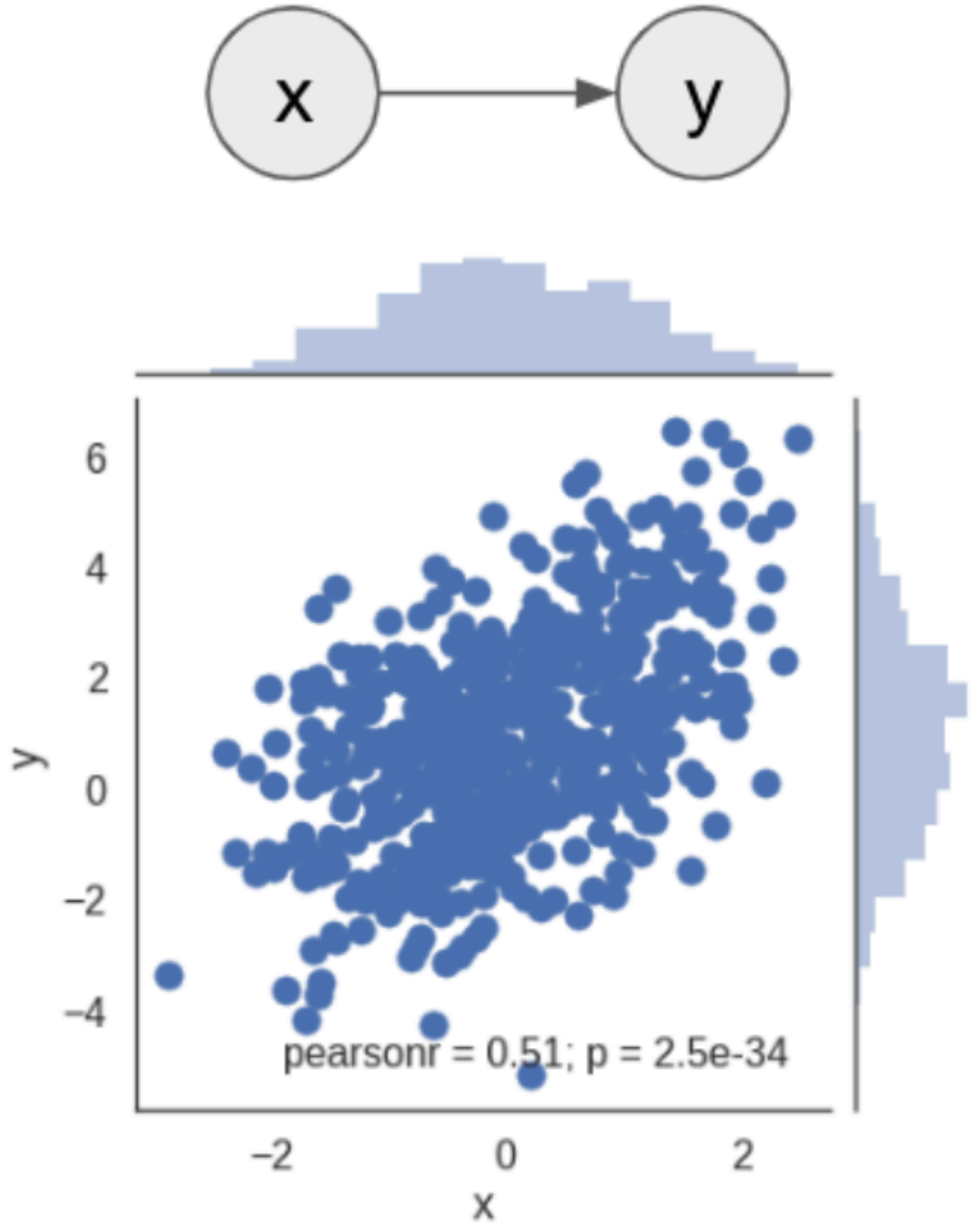
Ferenc Huszár

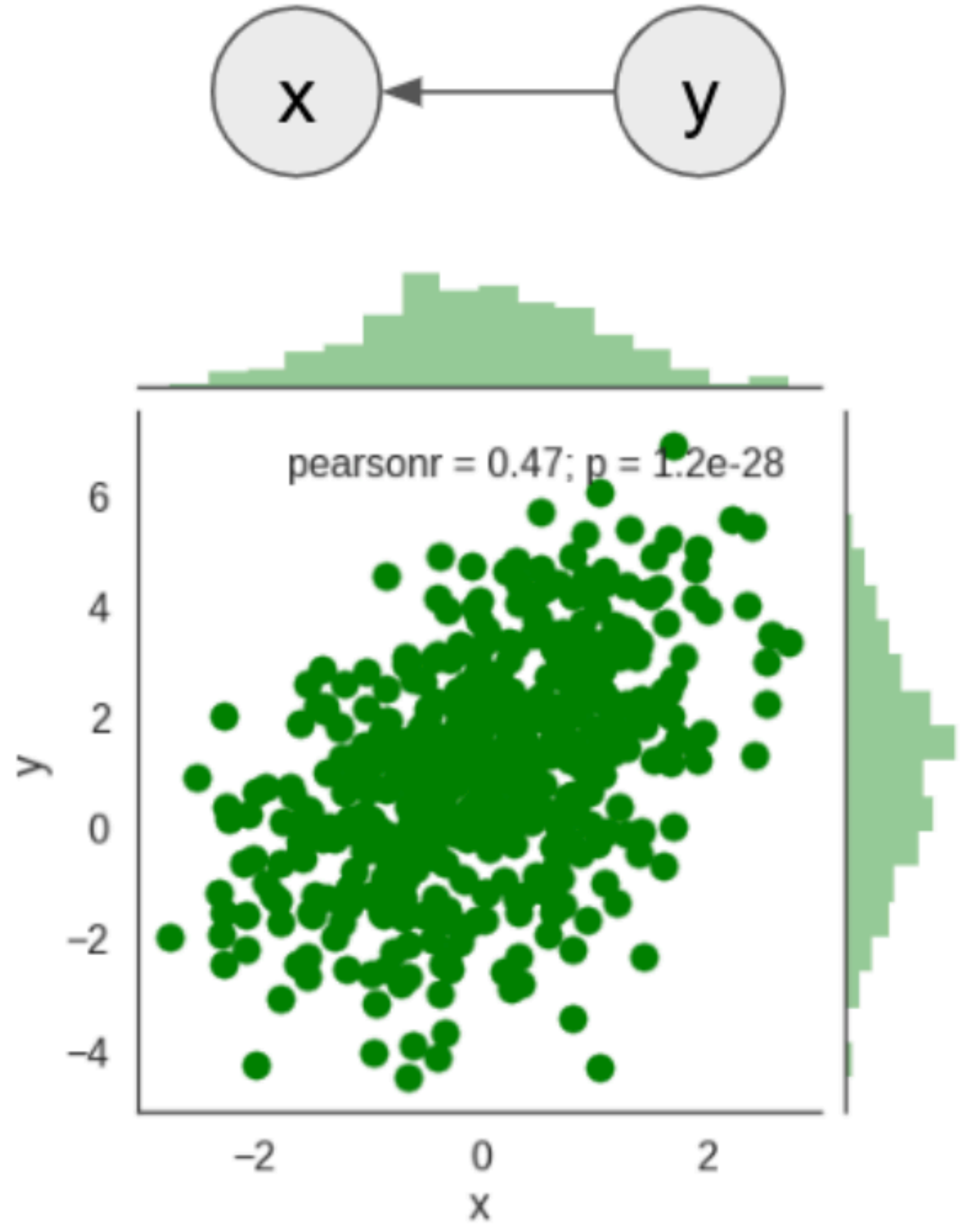
inFERENCe

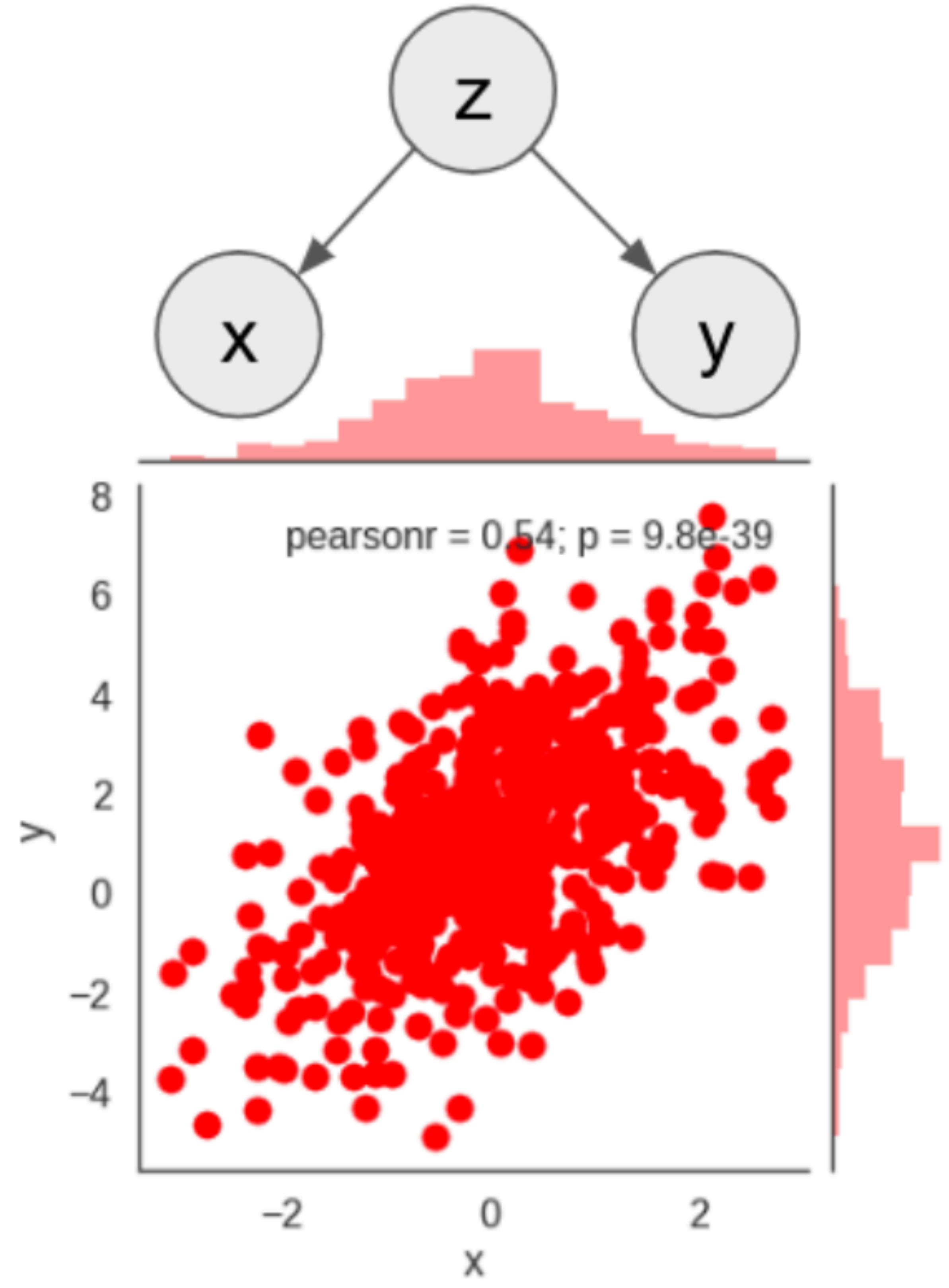
```
x = randn()
y = x + 1 + sqrt(3)*randn()
```

```
y = 1 + 2*randn()
x = (y-1)/4 + sqrt(3)*randn()/2
```

```
z = randn()
y = z + 1 + sqrt(3)*randn()
x = z
```







https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

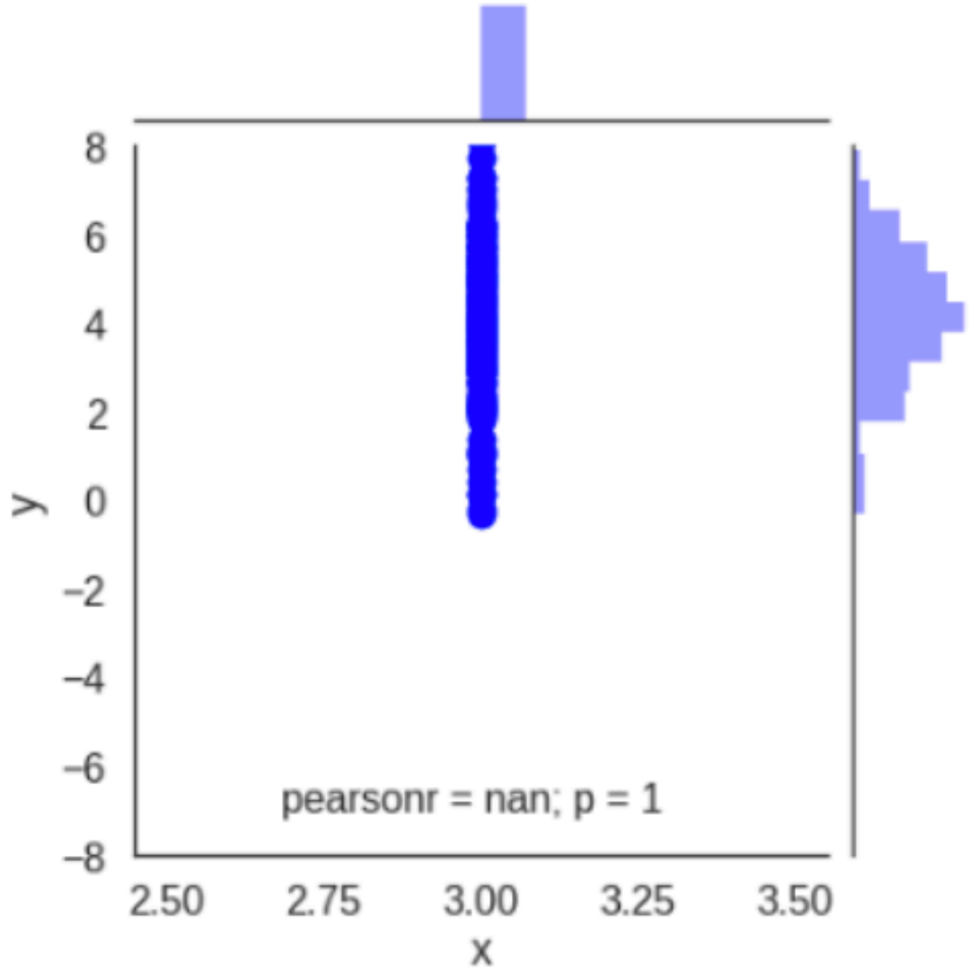
A toy example

Ferenc Huszár

inFERENCe

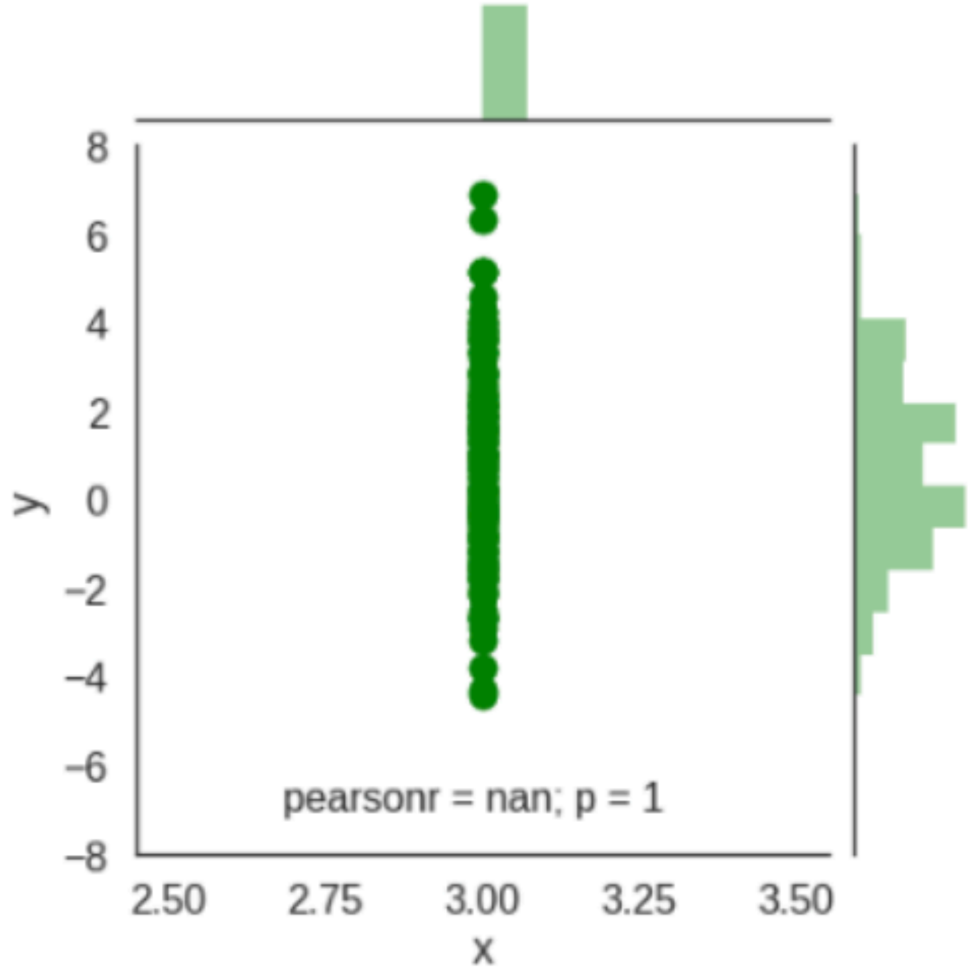
$$P(y|do(X)) = p(y|x)$$

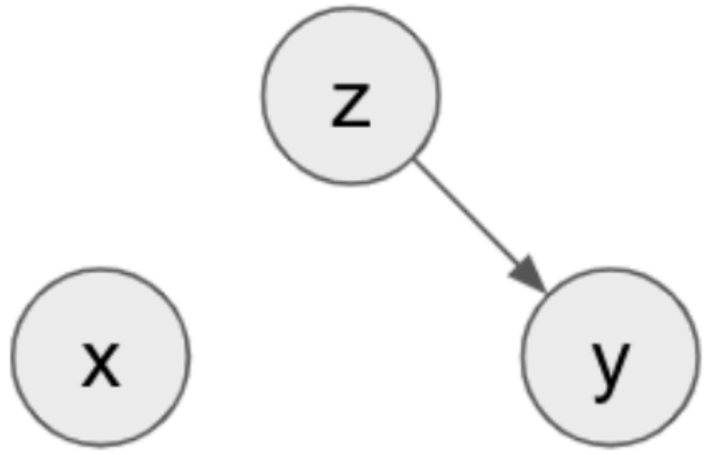
```
x = randn()
x = 3
y = x + 1 + sqrt(3)*randn()
x = 3
```



$$P(y|do(X)) = p(y)$$

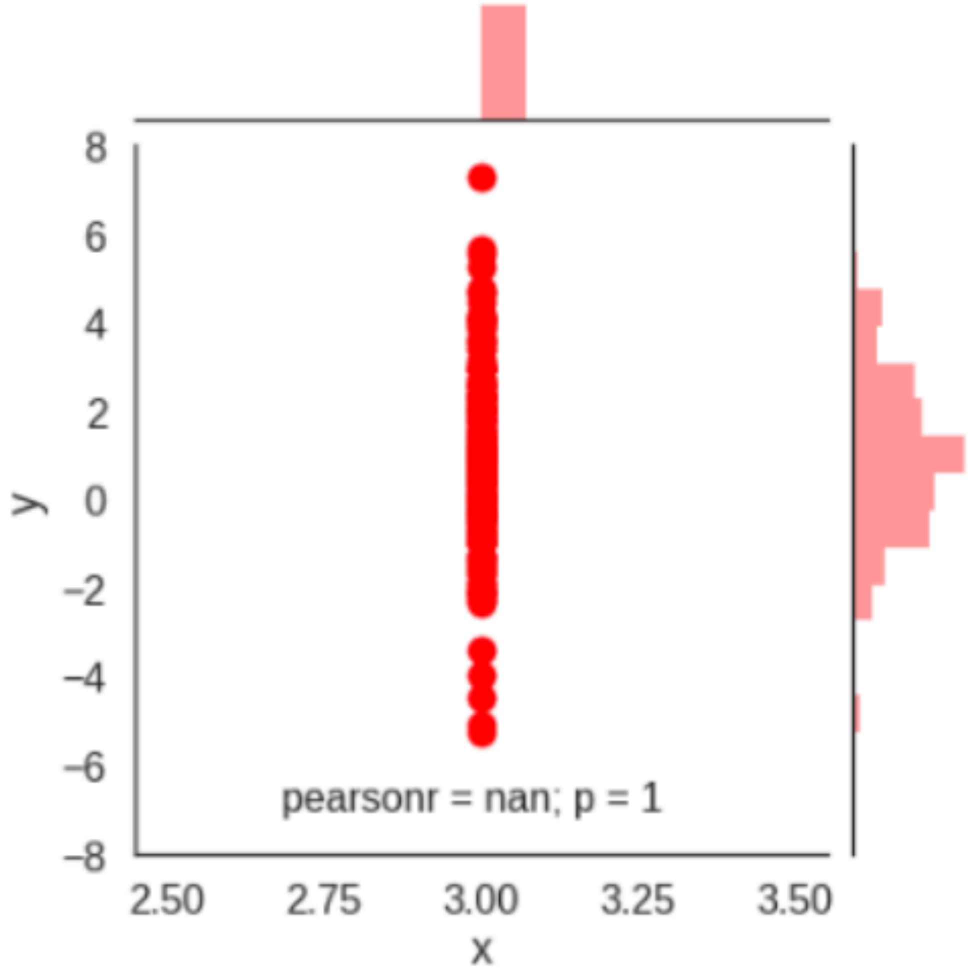
```
y = 1 + 2*randn()
x = 3
x = (y-1)/4 + sqrt(3)*randn()/2
x = 3
```



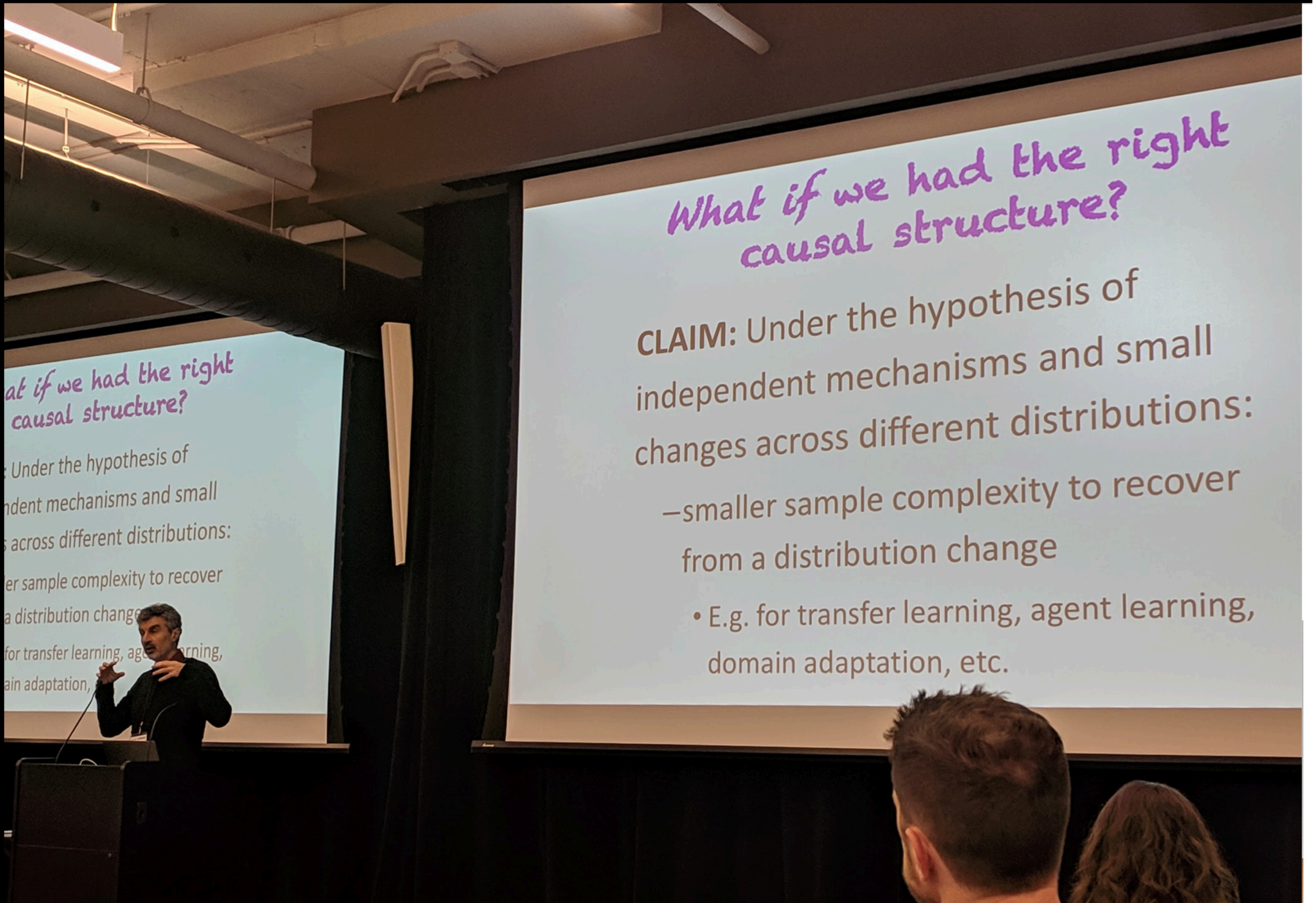


$$P(y|do(X)) = p(y)$$

```
z = randn()
x = 3
x = z
x = 3
y = z + 1 + sqrt(3)*randn()
x = 3
```



https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/



Max Welling Isn't this what Bernhard Schoelkopf has been saying for a while?

Like · Reply · 6w

Yann LeCun ...and Leon Bottou ?

Like · Reply · 6w

Leon Bottou Yoshua's paper says: if you observe a distribution change that comes from a causal effect, then you'll adapt faster if your generative model matches the causal model.

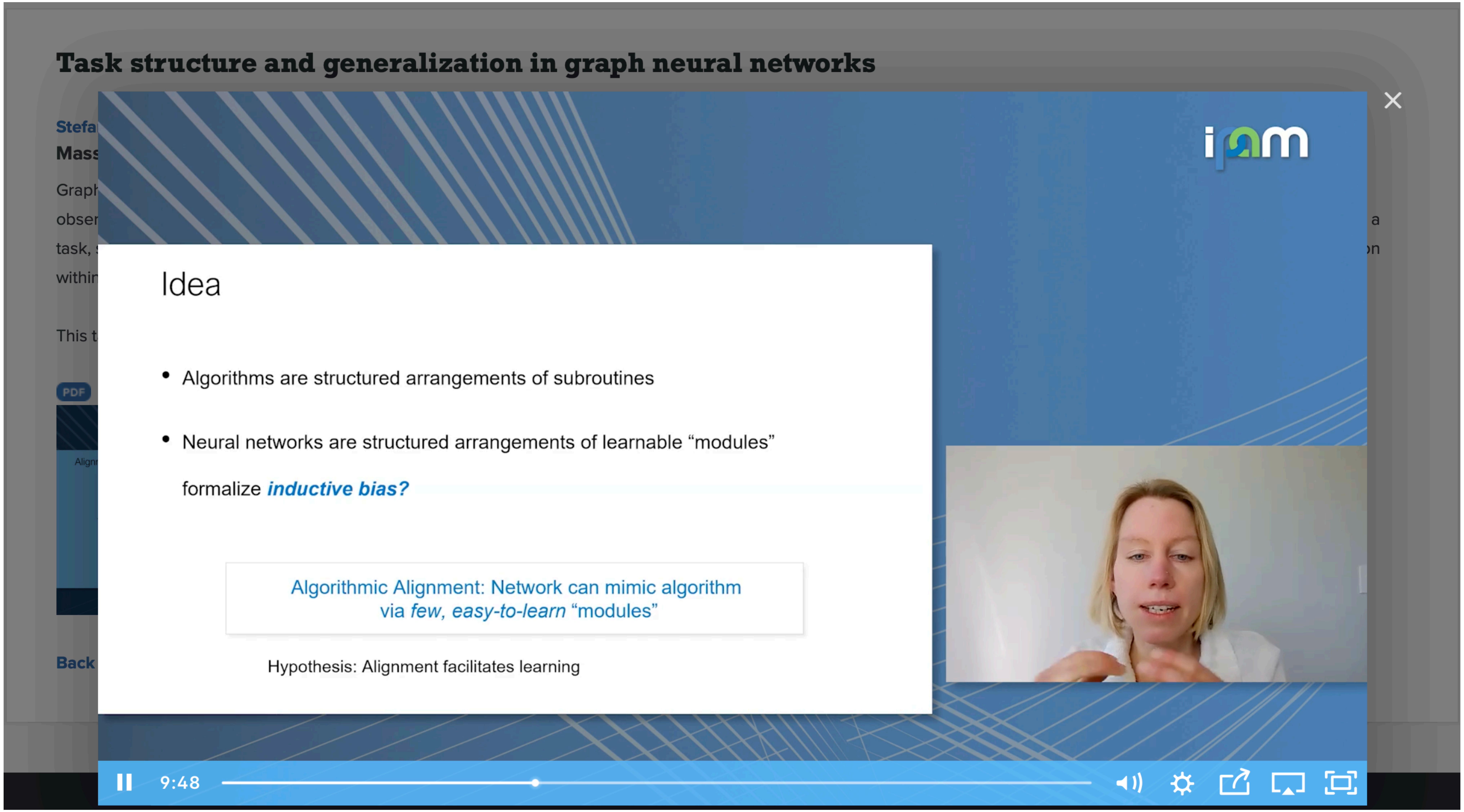
Another way of seeing it is : the right causal graph suggests a particular factorization of the joint distribution (a directed bayesian network). A causal intervention means that you only change one of these factors (or a few factors) while leaving the other ones unchanged. Therefore if your generative model is the right causal model, meaning that it factorizes the joint in the same way, it will be easy to adapt it to the change because only a few parameters need changing (those associated with the factors that actually changed).

Max Welling Dan Roy I am, and I think most of us, are keenly aware that Josh has been the big proponent of this view. And I think most people agree with him on this view. Integrating this view with deep learning for more narrowly defined tasks seems to me an interesting intellectual pursuit though. I think that's what's happening here but I was not at the talk 😉

Peter Battaglia

Algorithmic Alignment

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021



Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

Inductive Bias

Compositionality

Relationships

Symmetry

Causality

Separation

Learning Physics with Deep Neural Networks

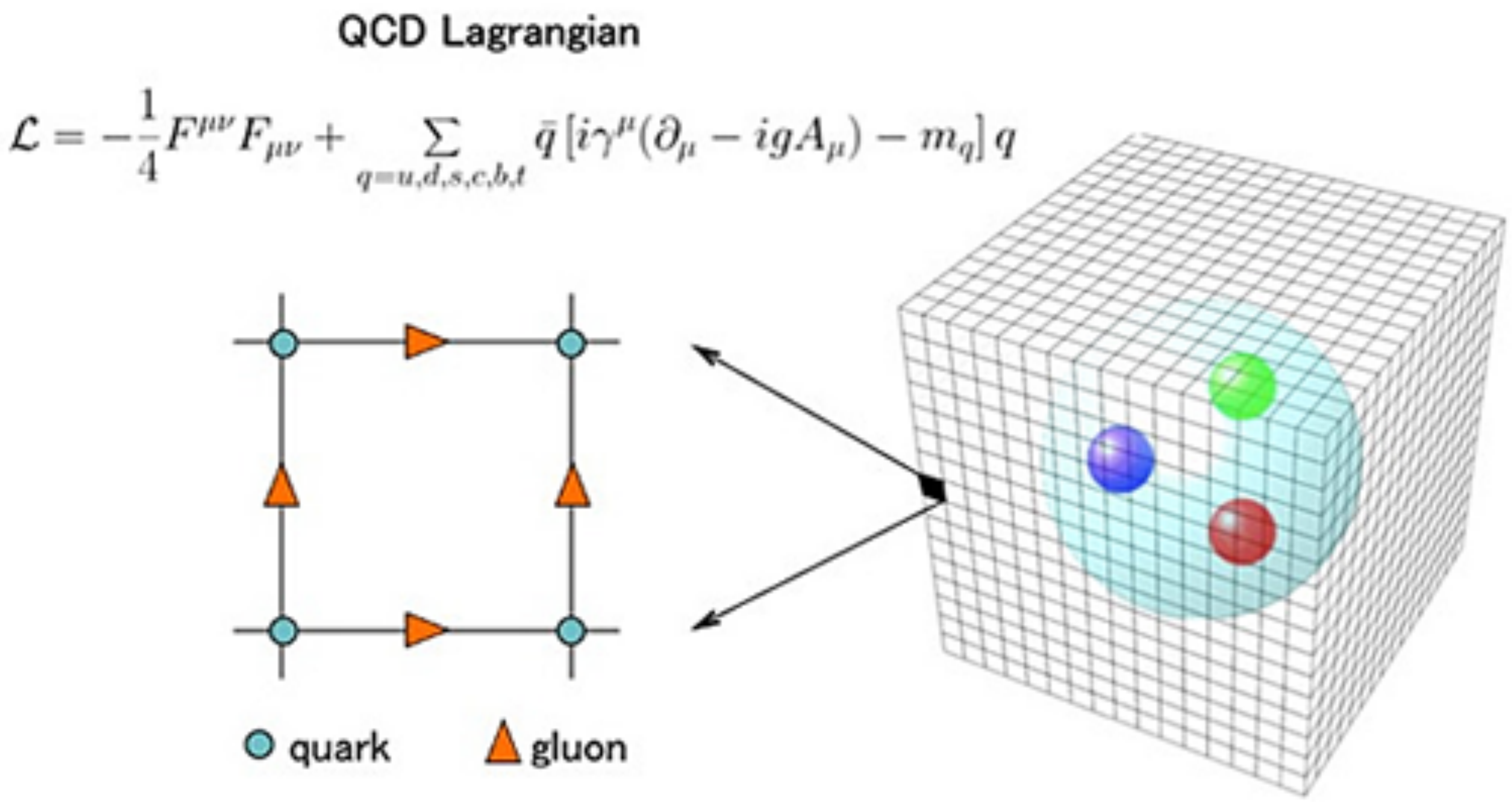
Stéphane Mallat, Ph.D.
École Normale Supérieure de Paris (ENS)

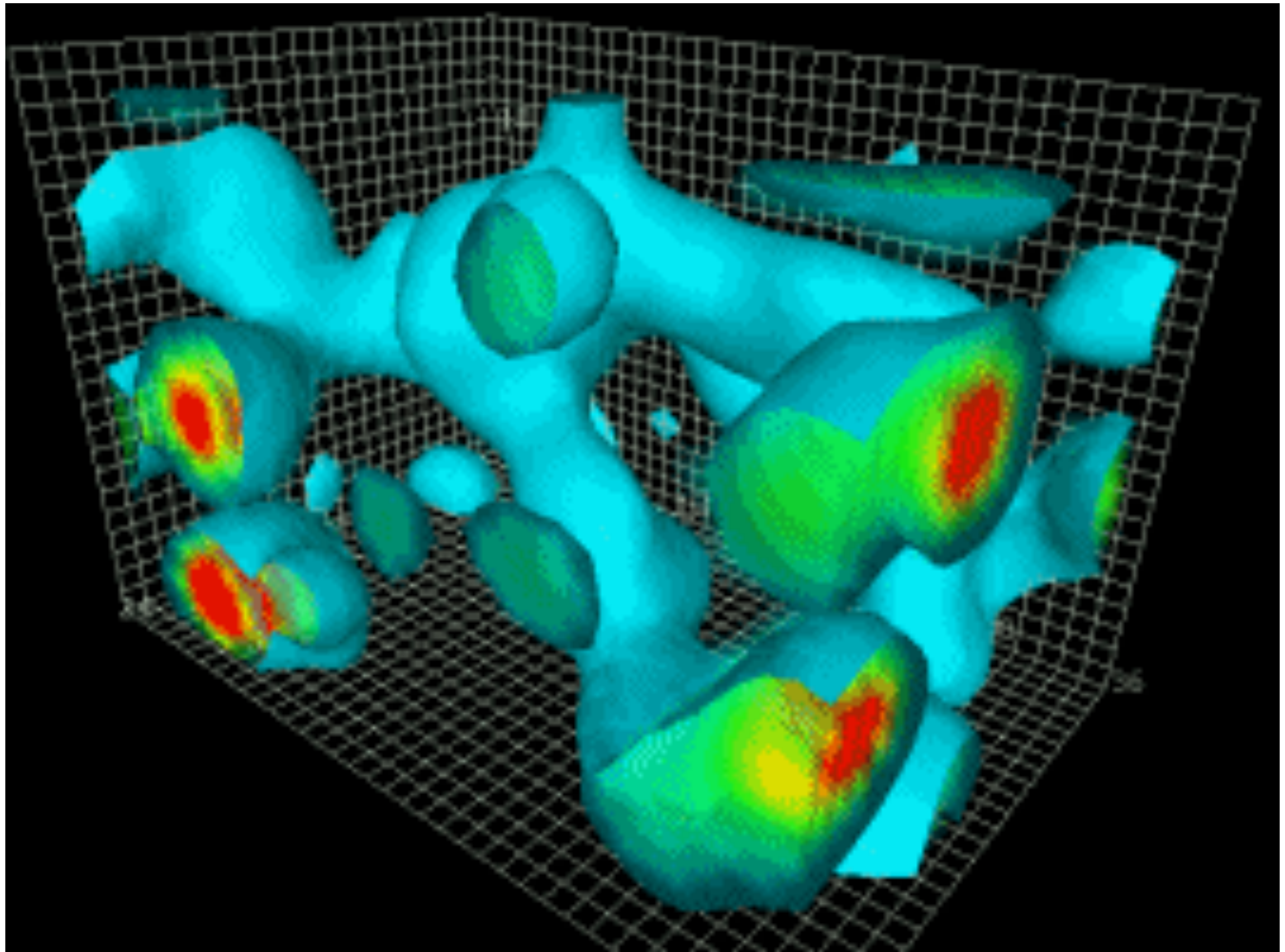
Can we learn physical properties from data? Machine learning offers a solution. It has many similarities with physics, requiring the approximation of functionals which depend on large numbers of variables, such as millions of pixels in images, letters in text, or particles in a physical system. Machine learning algorithms have considerably improved in the last 10 years through the processing of massive amounts of data. In particular, deep neural networks have spectacular applications, such as image classification and medical, industrial and physical data analysis.

In this lecture, Stéphane Mallat will show how machine learning can be applied to statistical physics, turbulent fluids and quantum chemistry. Beyond applications, he will highlight common mathematical approaches in physics and machine learning to overcome the issue of dimensionality. Two central pillars of such approaches are finding symmetries and separating phenomena at different scales. He will show that these pillars also govern the architecture and properties of deep convolutional neural networks.

Very expensive simulations with high dimensional data: eg. $x \in \mathbb{R}^{10^9}$

Use normalizing flow to approximate target distribution of configurations that is implied by the action $S(x)$ via the Boltzmann Equation $p(x) = e^{-S(x)}/Z$

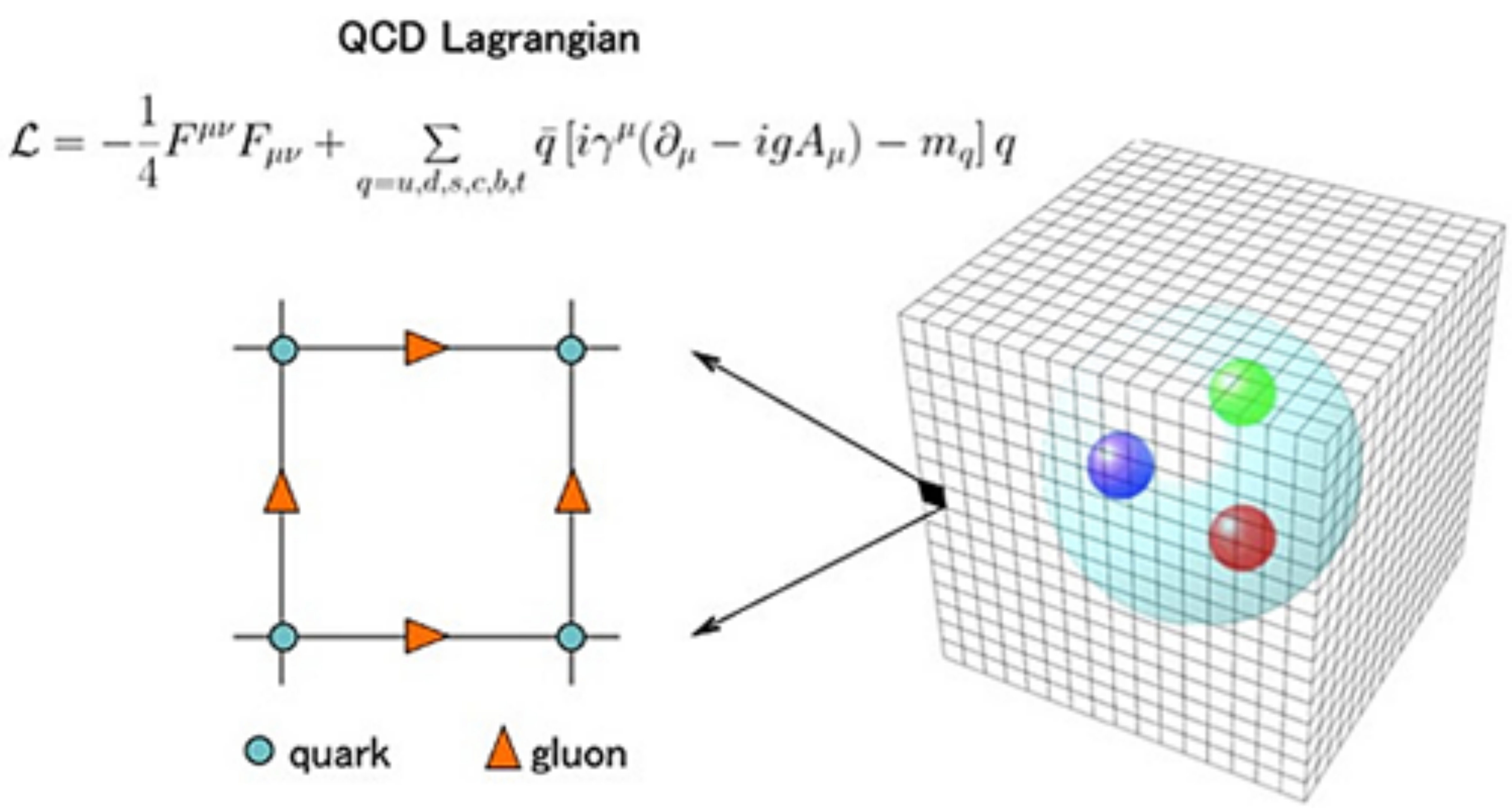


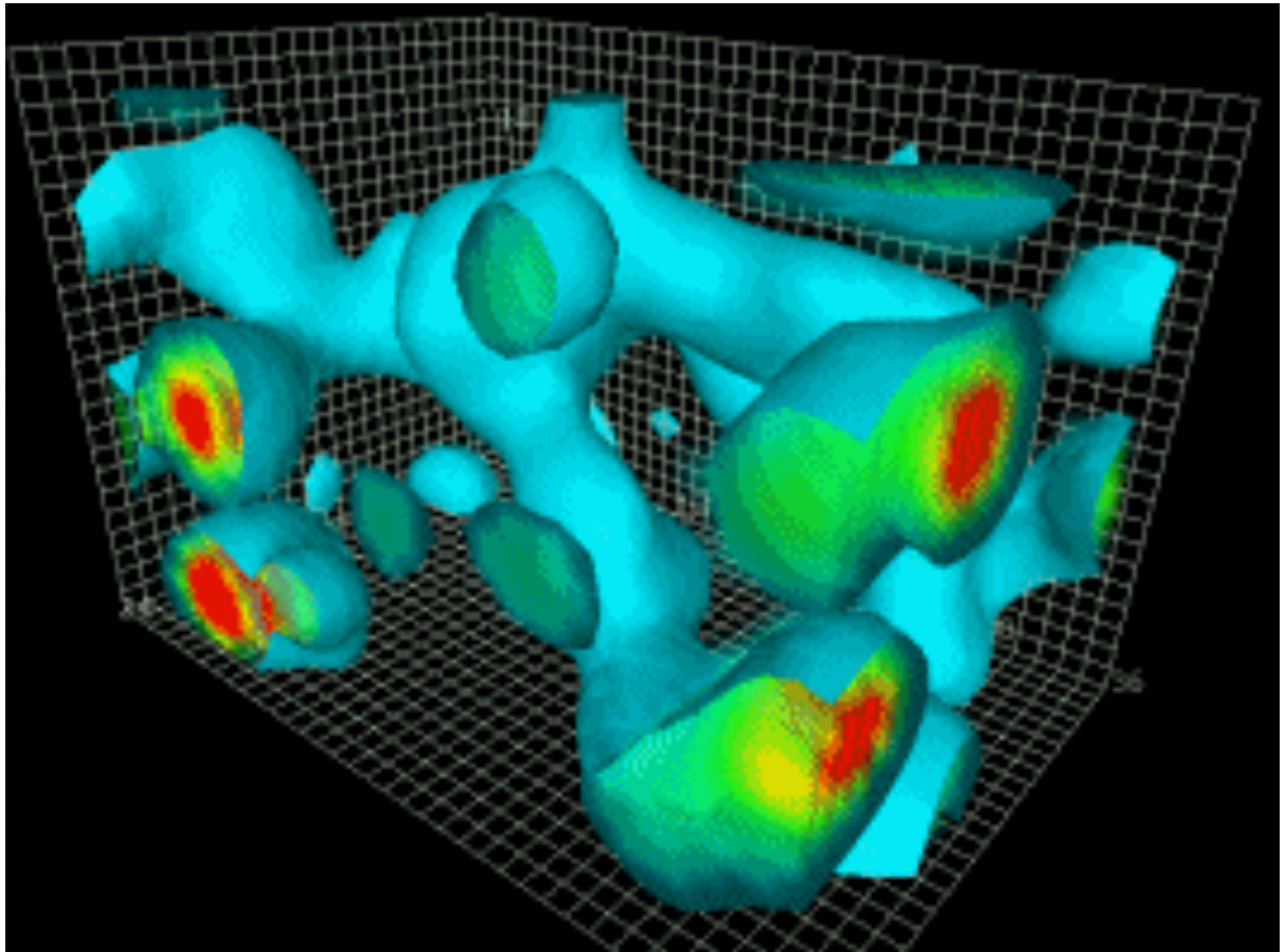


Lattice Field Theory

Very expensive simulations with high dimensional data: eg. x∈ $\mathbb{R}^{10^9}$

Use normalizing flow to approximate target distribution of configurations that is implied by the action $S(x)$ via the Boltzmann Equation $p(x) = e^{-S(x)}/Z$





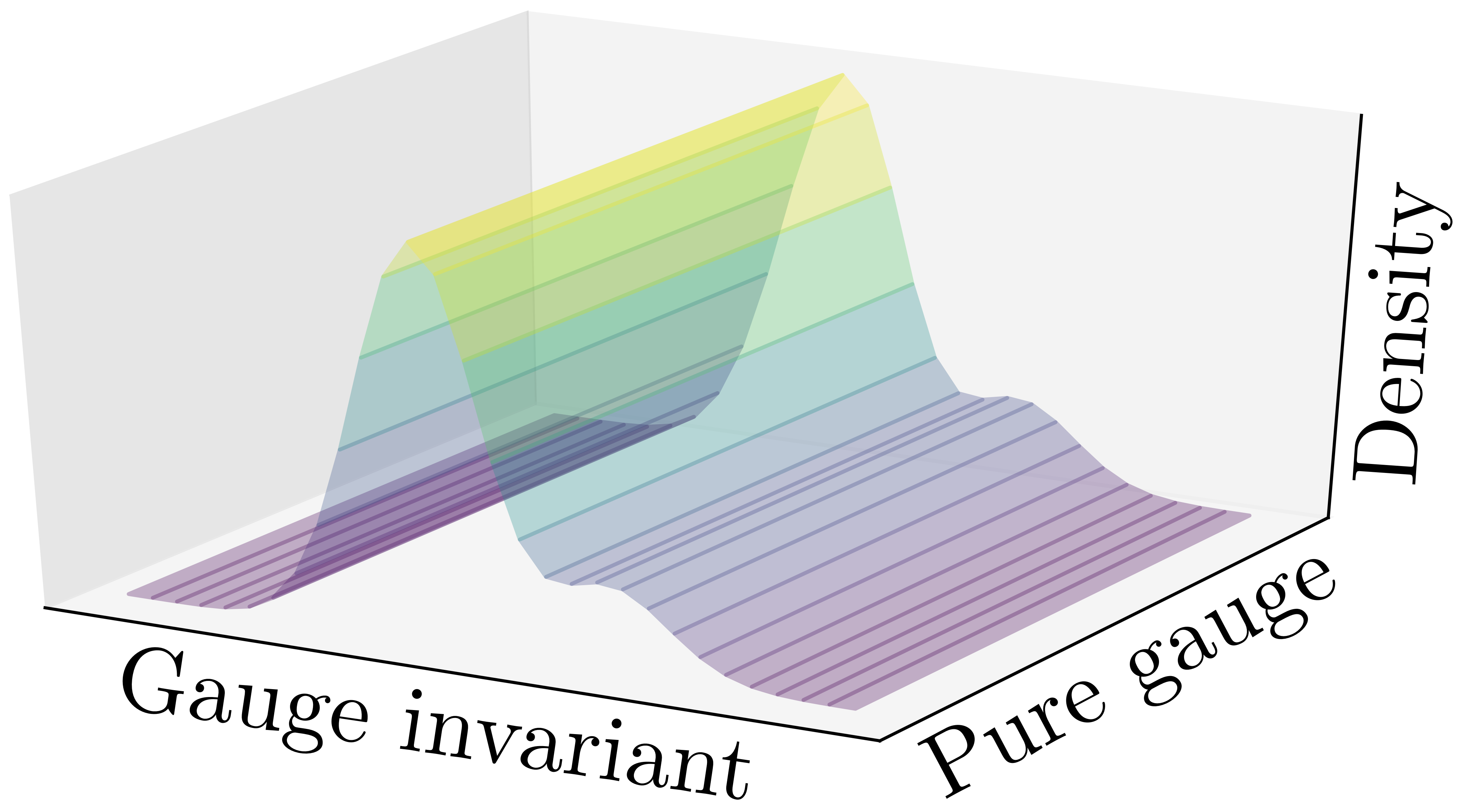
Space-time & Local Gauge Symmetry

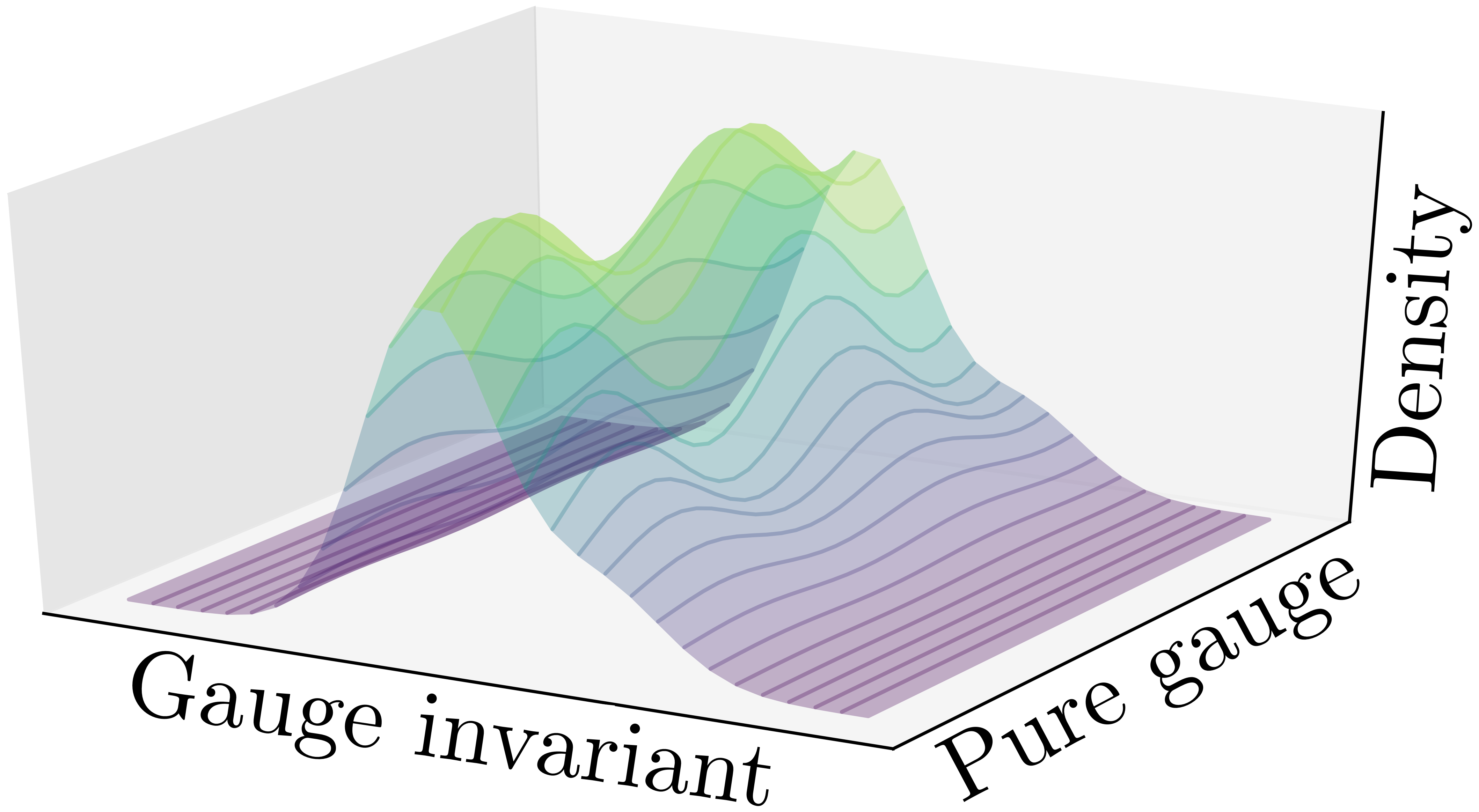
Abstract: I will discuss recent work to incorporate symmetries, in particular gauge symmetries (local symmetry transformations that form Lie groups), into generative flow models. This work is motivated by the applications of generative models for physics simulation, in particular for lattice field theory.

The action is invariant to gauge transformations, so the distribution is constant in those directions

Many more pure gauge degrees of freedom than physical ones

We would like to enforce this symmetry in the network, and not have to learn it.





Space-time & Local Gauge Symmetry

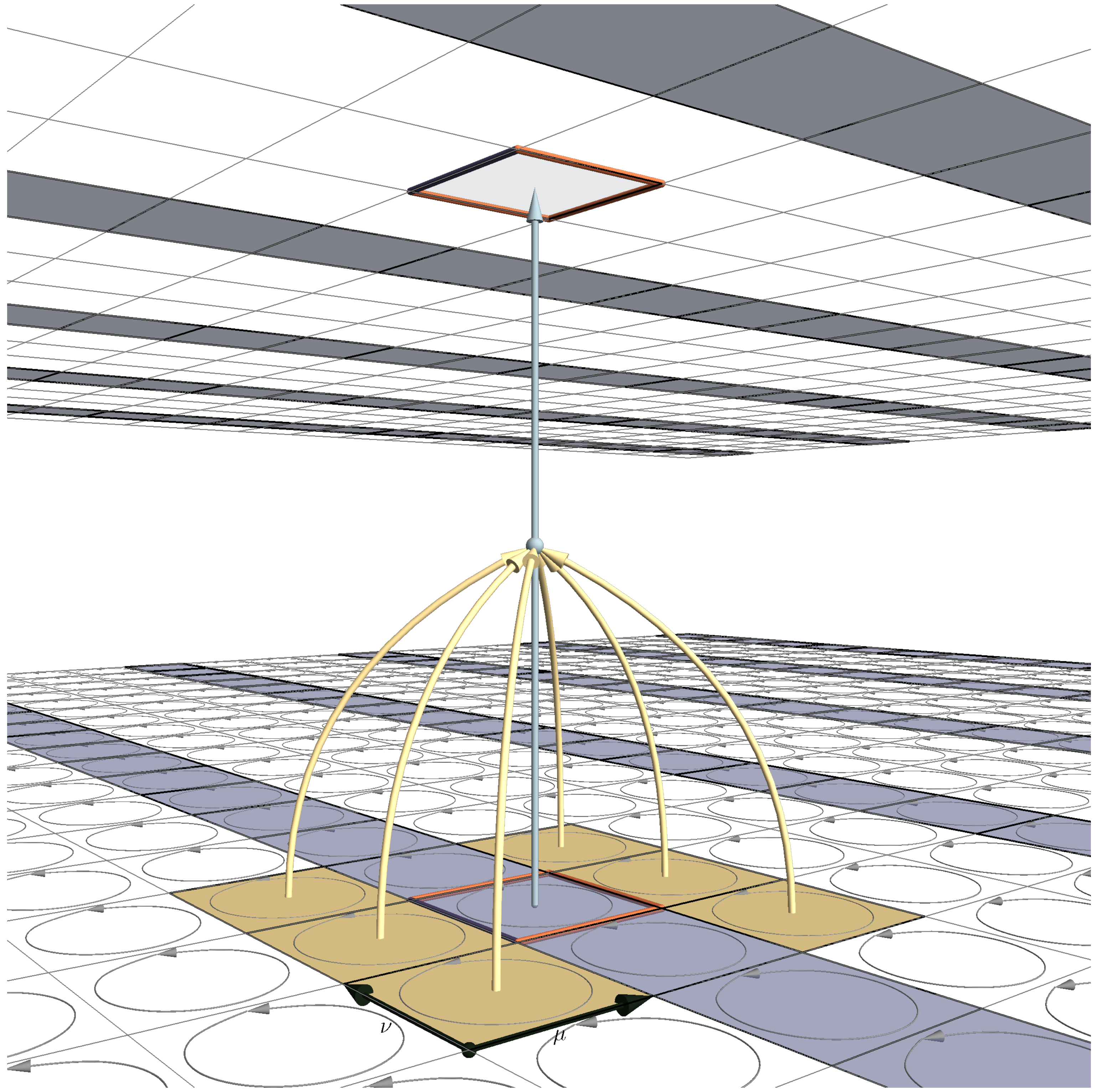
Kanwar, Albergo, Boyda, Cranmer, Hackett, Racaniere, Rezende, Shanahan arXiv:2002.02428 & arXiv:2003:06413

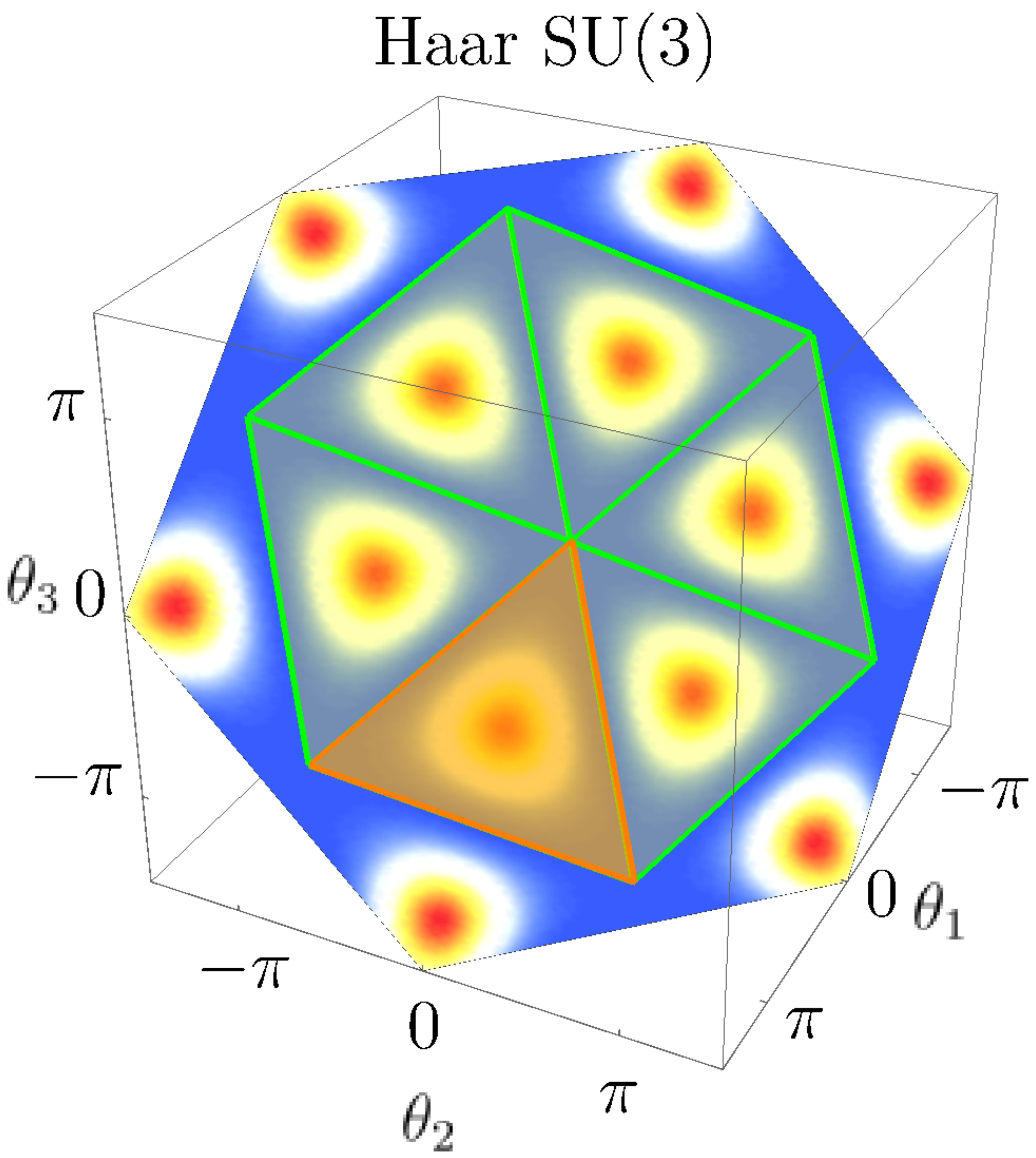
20 May 2020

Building symmetries into generative flow models

Phiala Shanahan, MIT, 12:00 EDT

Abstract: I will discuss recent work to incorporate symmetries, in particular gauge symmetries (local symmetry transformations that form Lie groups), into generative flow models. This work is motivated by the applications of generative models for physics simulation, in particular for lattice field theory.





Permutation symmetry

Physics ∩ ML

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

16
Dec 2020

Two talks from string_data 2020

Haggai Maron (NVIDIA Research) and Sergei Gukov (CalTech), 12:00 EDT

Haggai Maron

Title: Leveraging Permutation Group Symmetries for the design of Equivariant Neural Networks

Abstract: Learning of irregular data, such as sets and graphs, is a prominent research direction that has received considerable attention in the last few years. The main challenge that arises is which architectures should be used for such data types. I will present a general framework for designing network architectures for irregular data types that adhere to permutation group symmetries. In the first part of the talk, we will see that these architectures can be implemented using a simple parameter-sharing scheme. We will then demonstrate the applicability of the framework by devising neural architectures for two widely used irregular data types: (i) Graphs and hyper-graphs and (ii) Sets of structured elements.

Set2Graph: Learning Graphs from Sets

Hadar Serviansky[1] Nimrod Segol[1] Jonathan Shlomi[1] Kyle Cranmer[2]
Eilam Gross[1] Haggai Maron[3] Yaron Lipman[1]

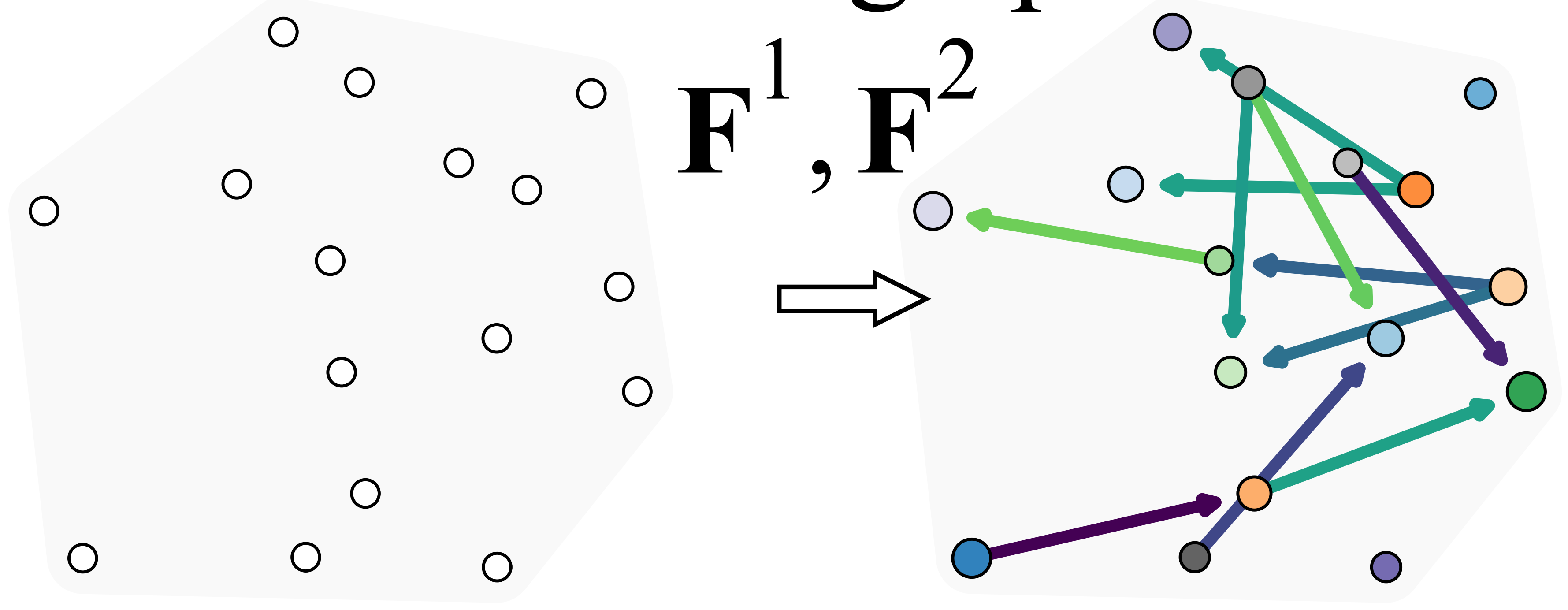
https://arxiv.org/abs/2002.08772

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE
1

NYU
2

NVIDIA
3

Set ⟶ graph



Hybrid “physics-aware” approaches to AI / ML

- inject knowledge of data generating process into inductive bias of ML models
- use “black box” style ML components to model the most uncertain aspects of problem where traditional approaches make overly restrictive assumptions
- align architectural components of ML models with causal mechanism

Insight of data generating process informs inductive bias on architecture

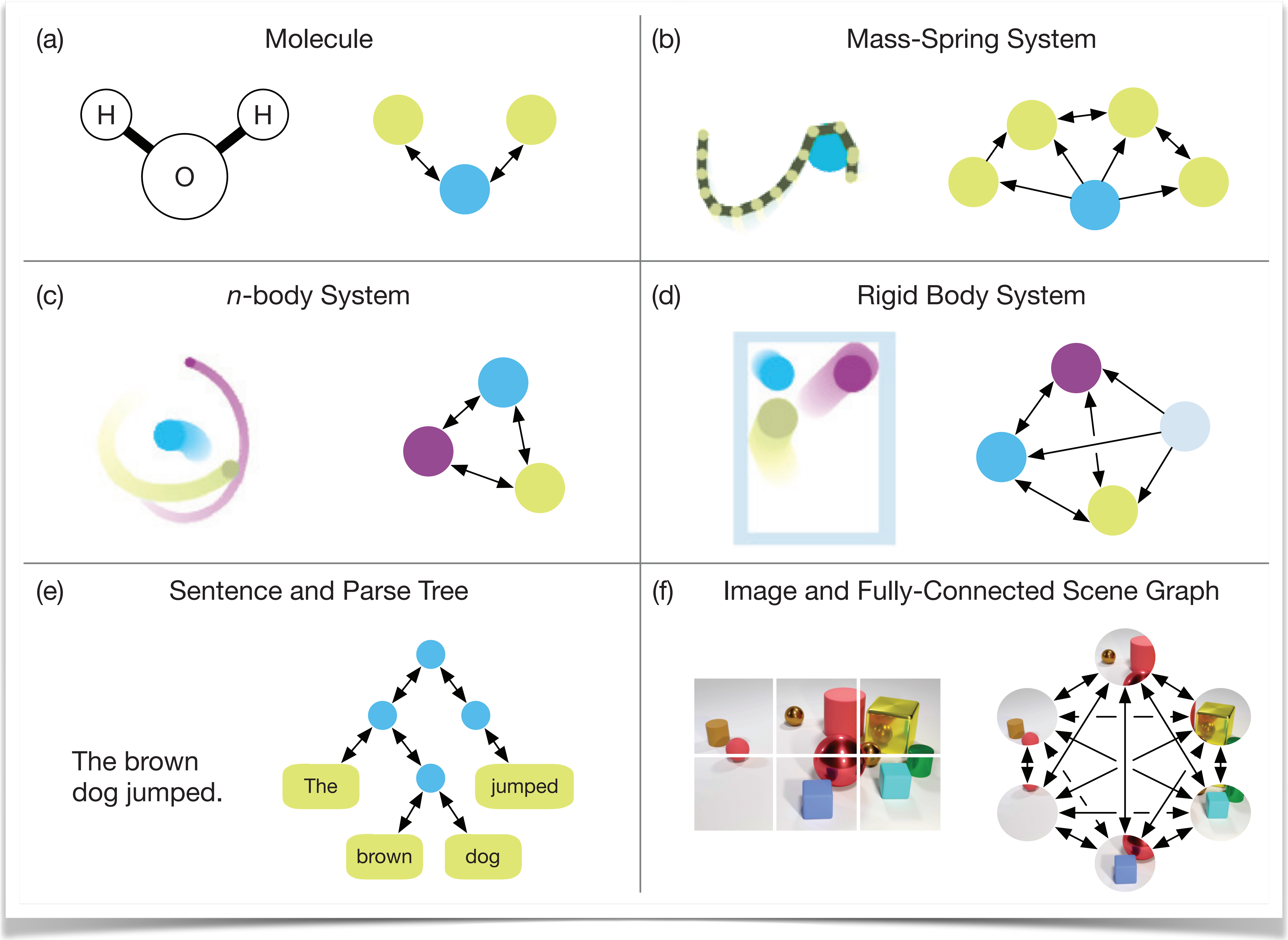
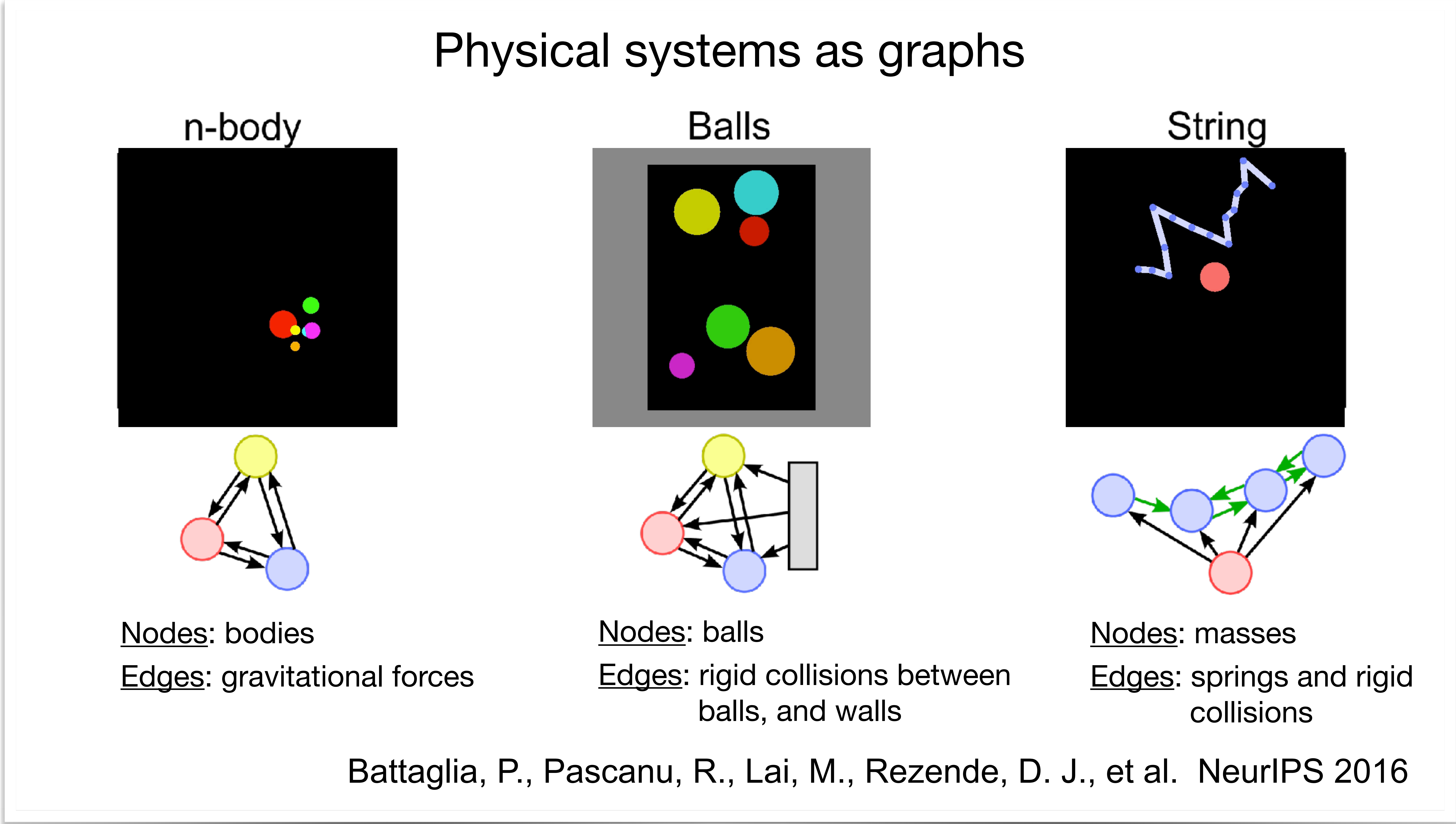


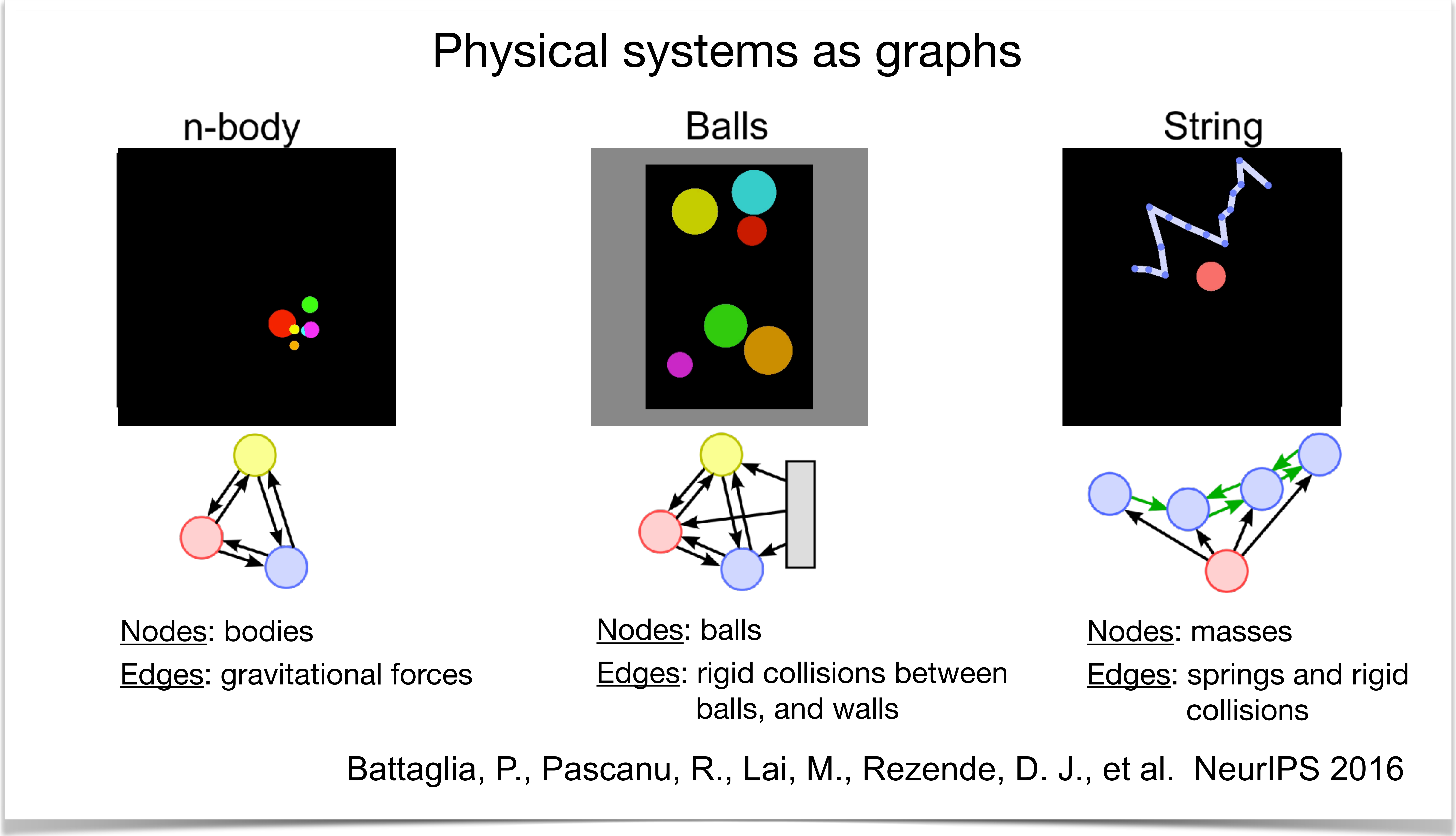
Image credit: Battaglia, et. al. arXiv:1806.01261

Dynamical systems



Peter Battaglia

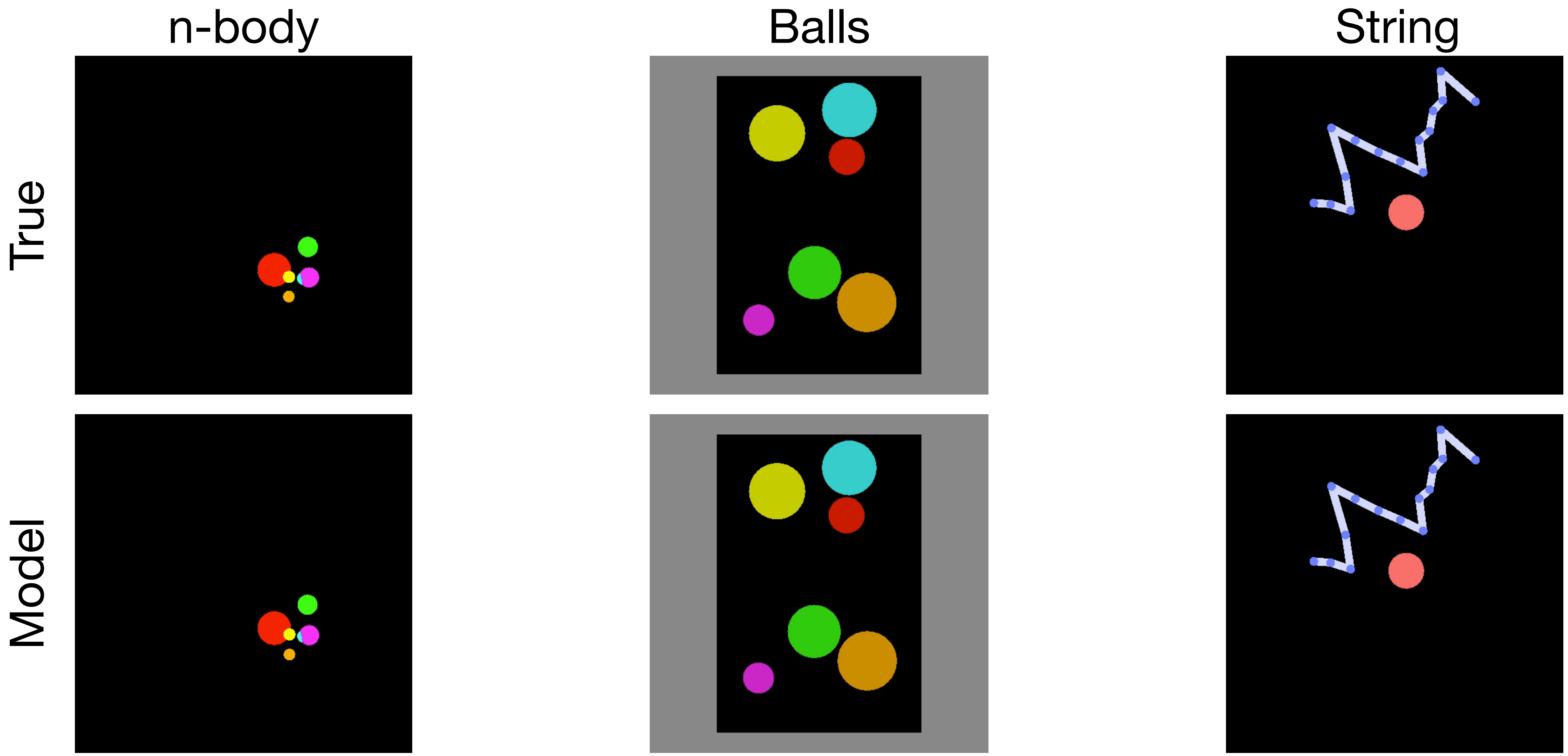
Dynamical systems




Peter Battaglia

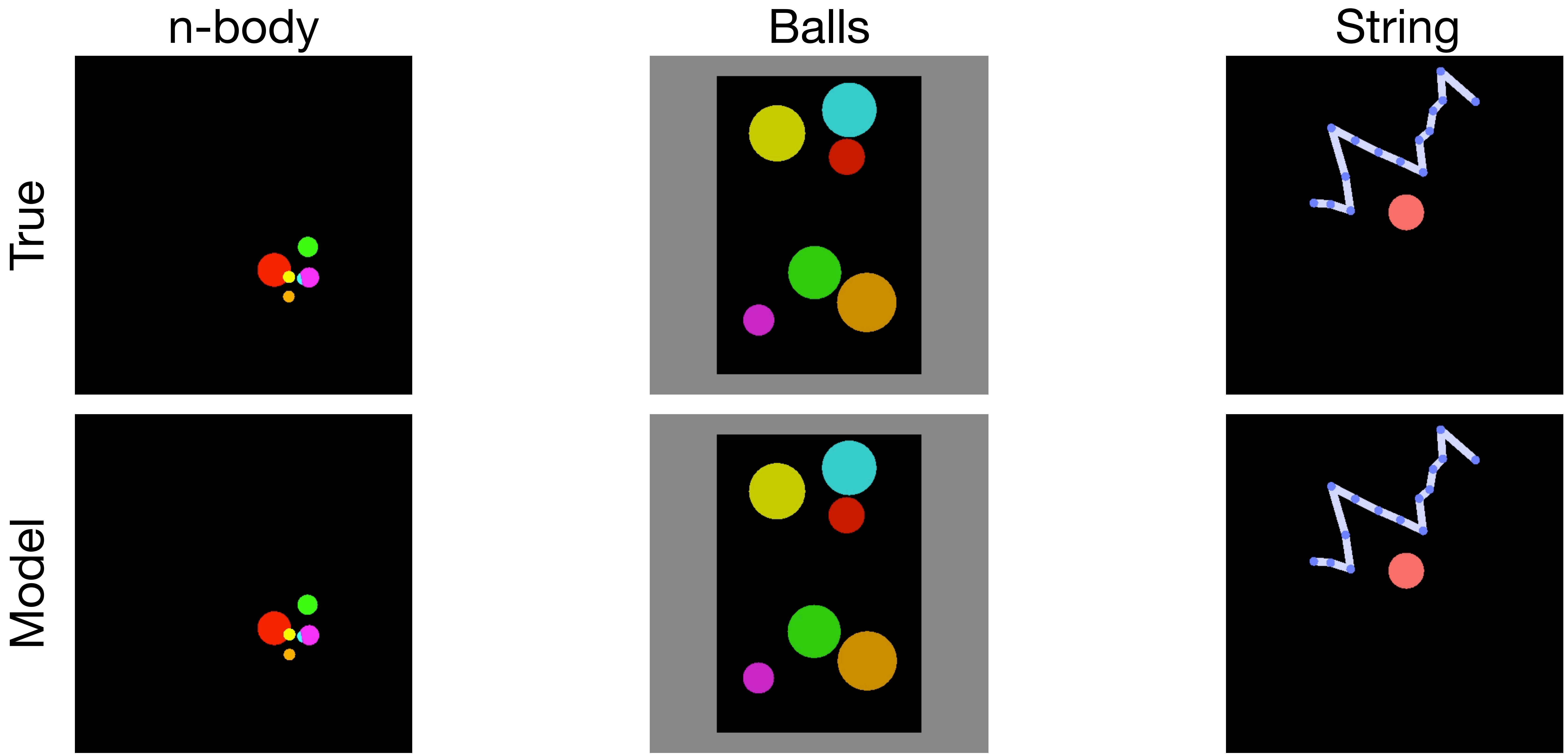
Slides borrowed with permission from Peter Battaglia based on Battaglia et al., 2016, NeurIPS

1000-step rollouts of true (top row) vs predicted (bottom row)



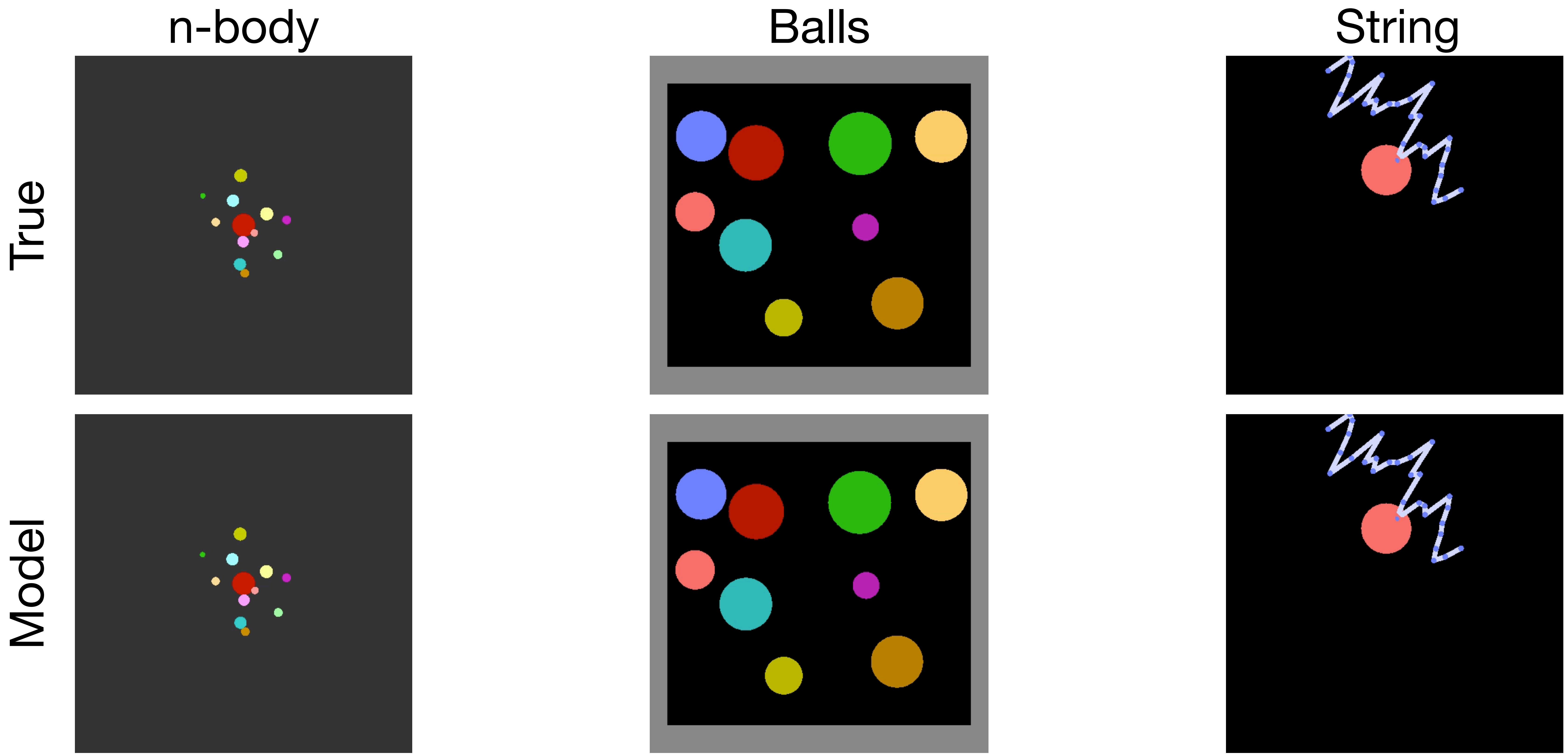
Slides borrowed with permission from Peter Battaglia based on Battaglia et al., 2016, NeurIPS

1000-step rollouts of true (top row) vs predicted (bottom row)



Slides borrowed with permission from Peter Battaglia based on Battaglia et al., 2016, NeurIPS

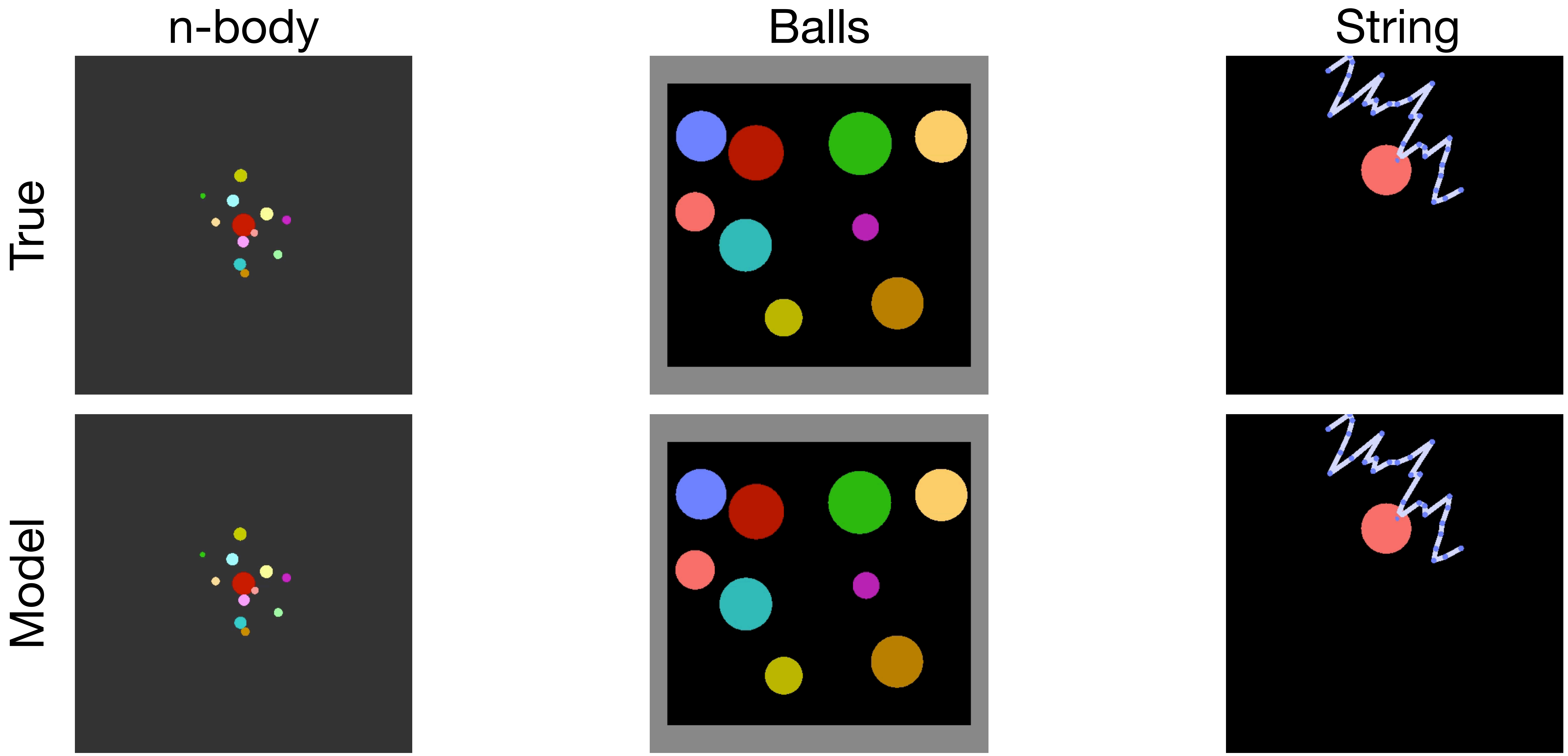
Zero shot generalisation to larger systems



Peter Battaglia

Slides borrowed with permission from Peter Battaglia based on Battaglia et al., 2016, NeurIPS

Zero shot generalisation to larger systems



Hamiltonian graph Networks

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics

Alvaro Sanchez Gonzalez

Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

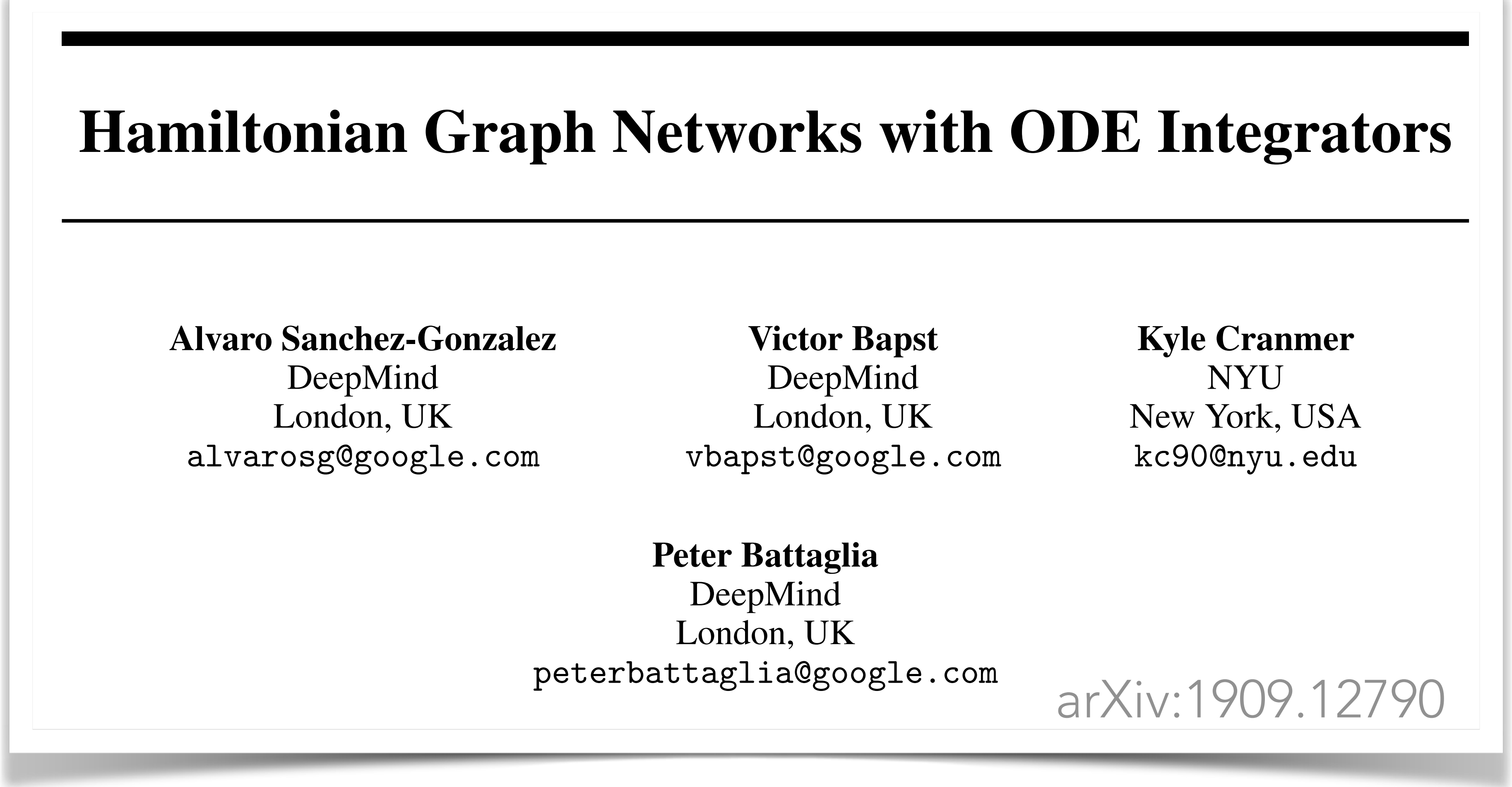
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

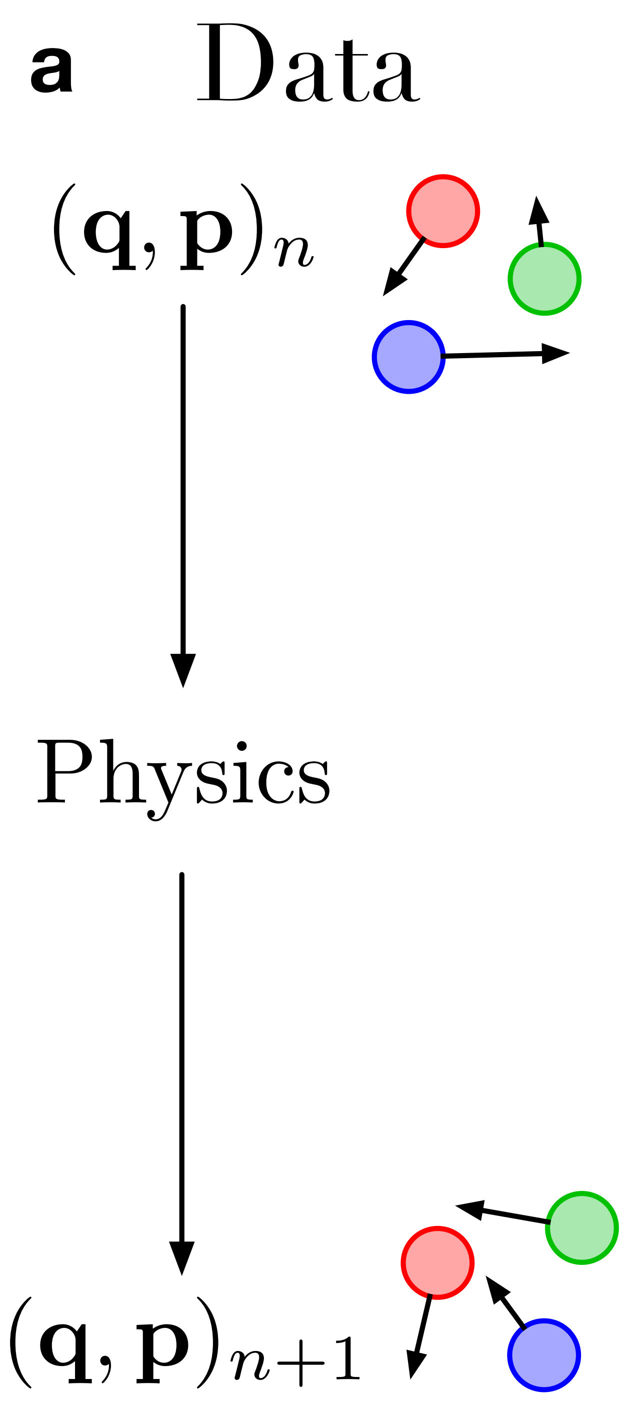
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

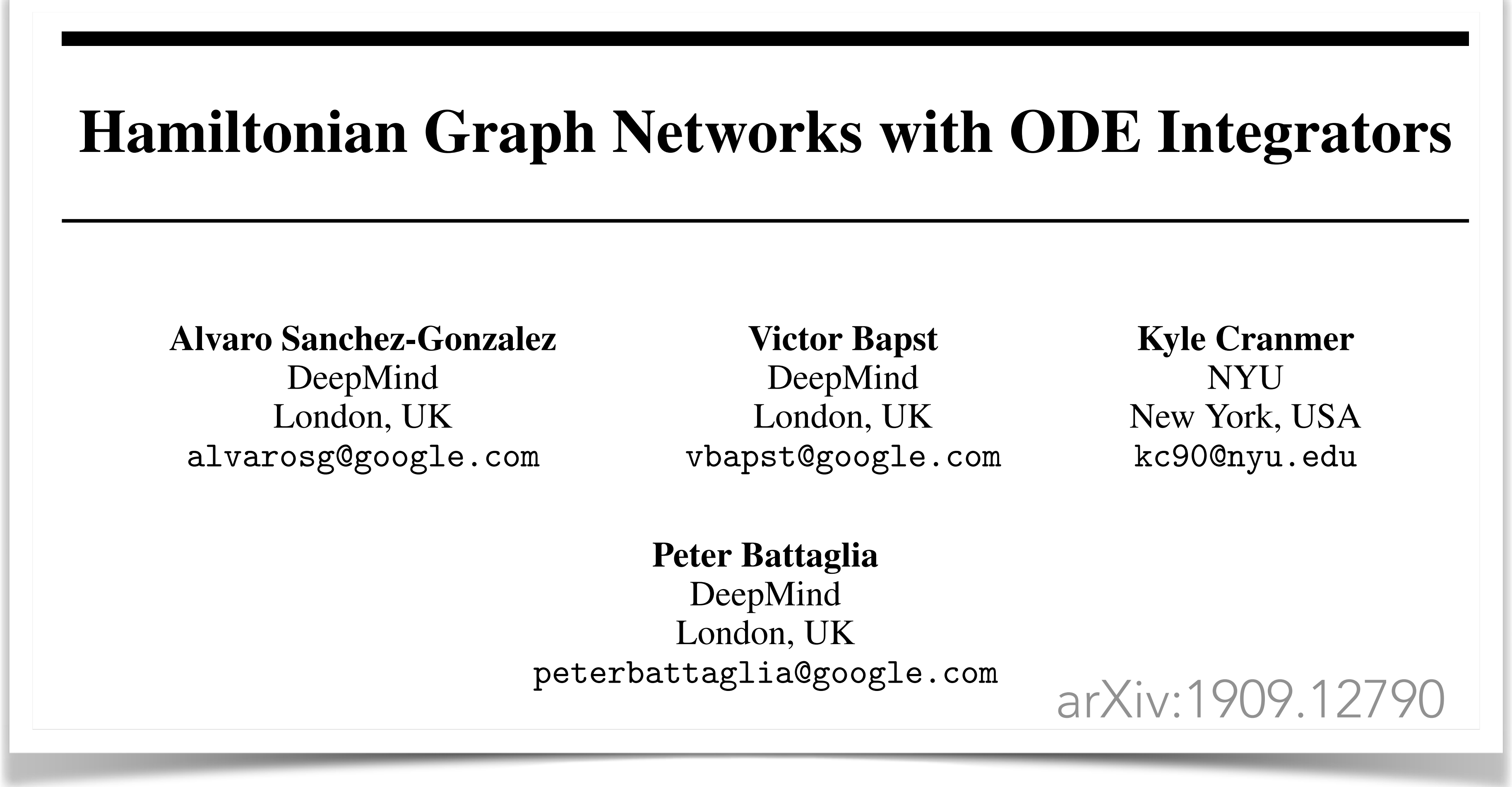
We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics



Hamiltonian graph Networks

Alvaro Sanchez Gonzalez


Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

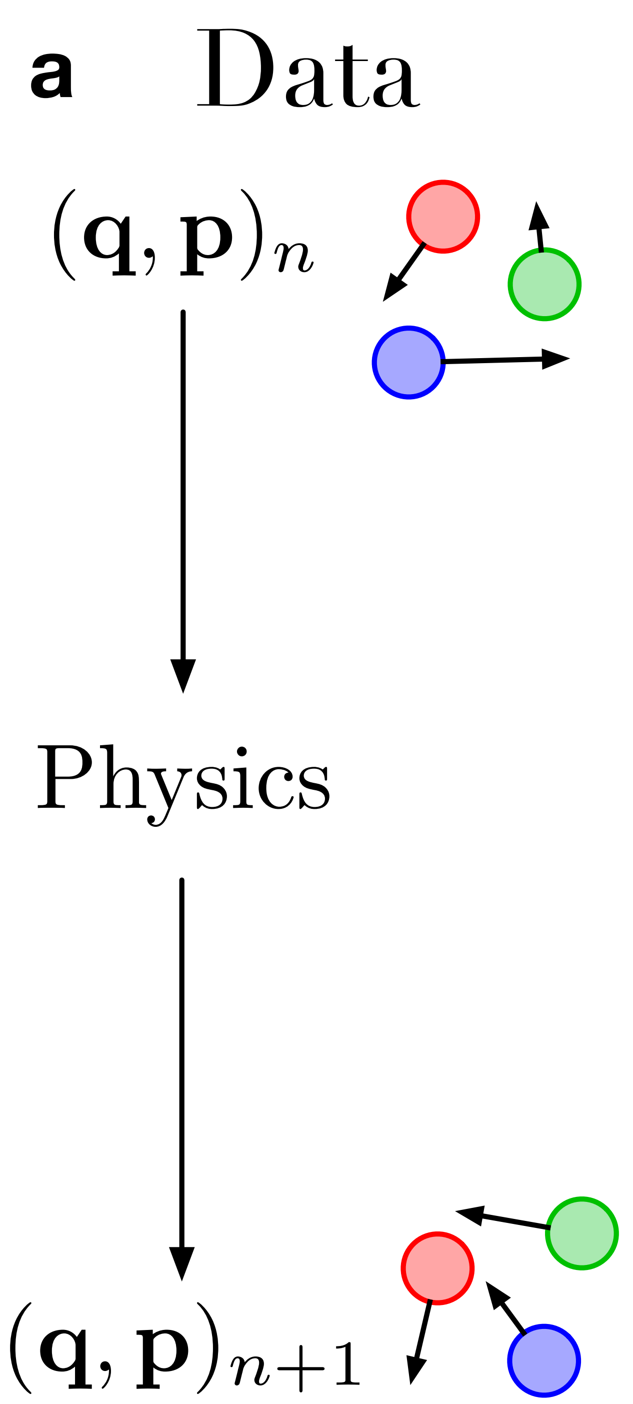
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

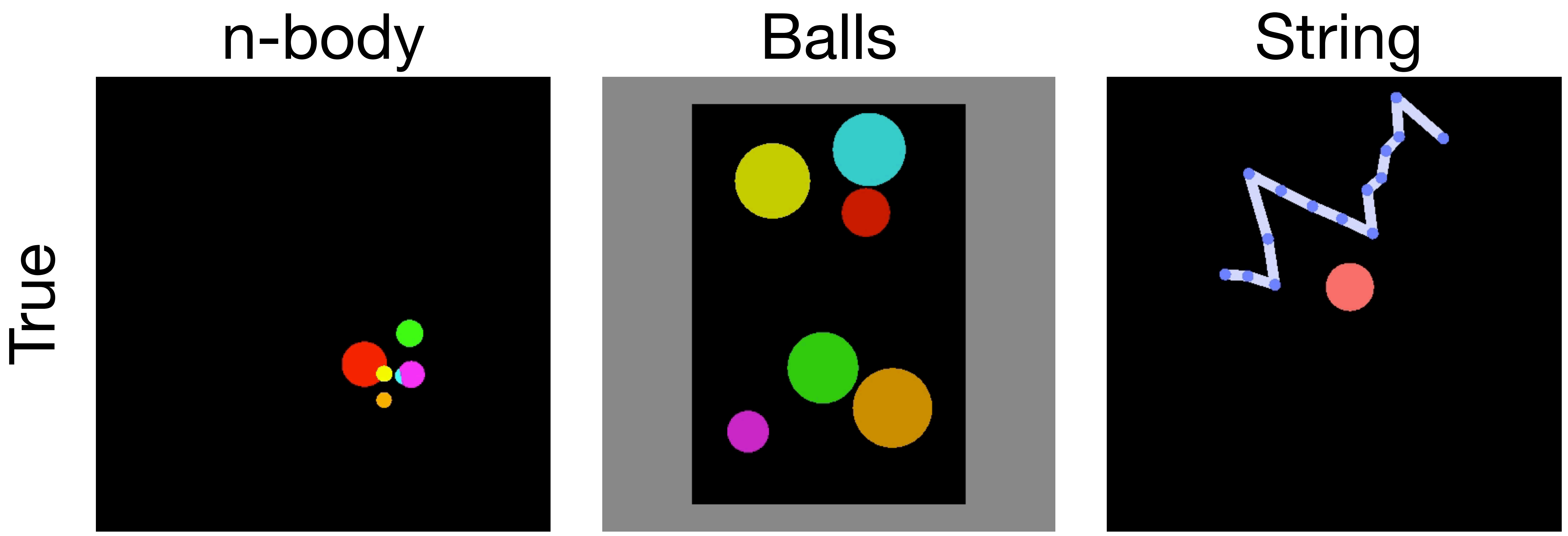
Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

We incorporated two physically-informed inductive biases

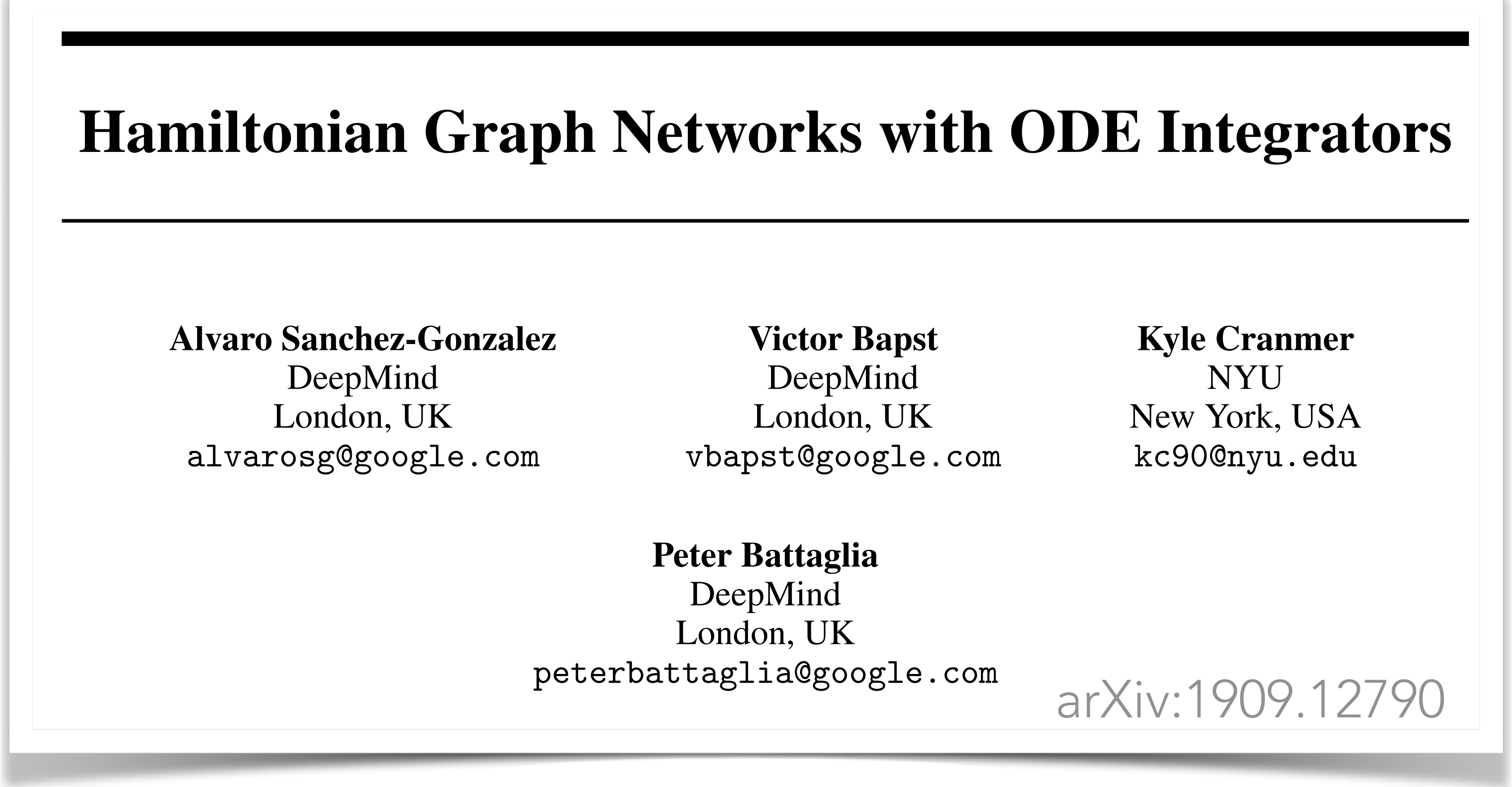
- ODE integrators
- Hamiltonian mechanics





Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

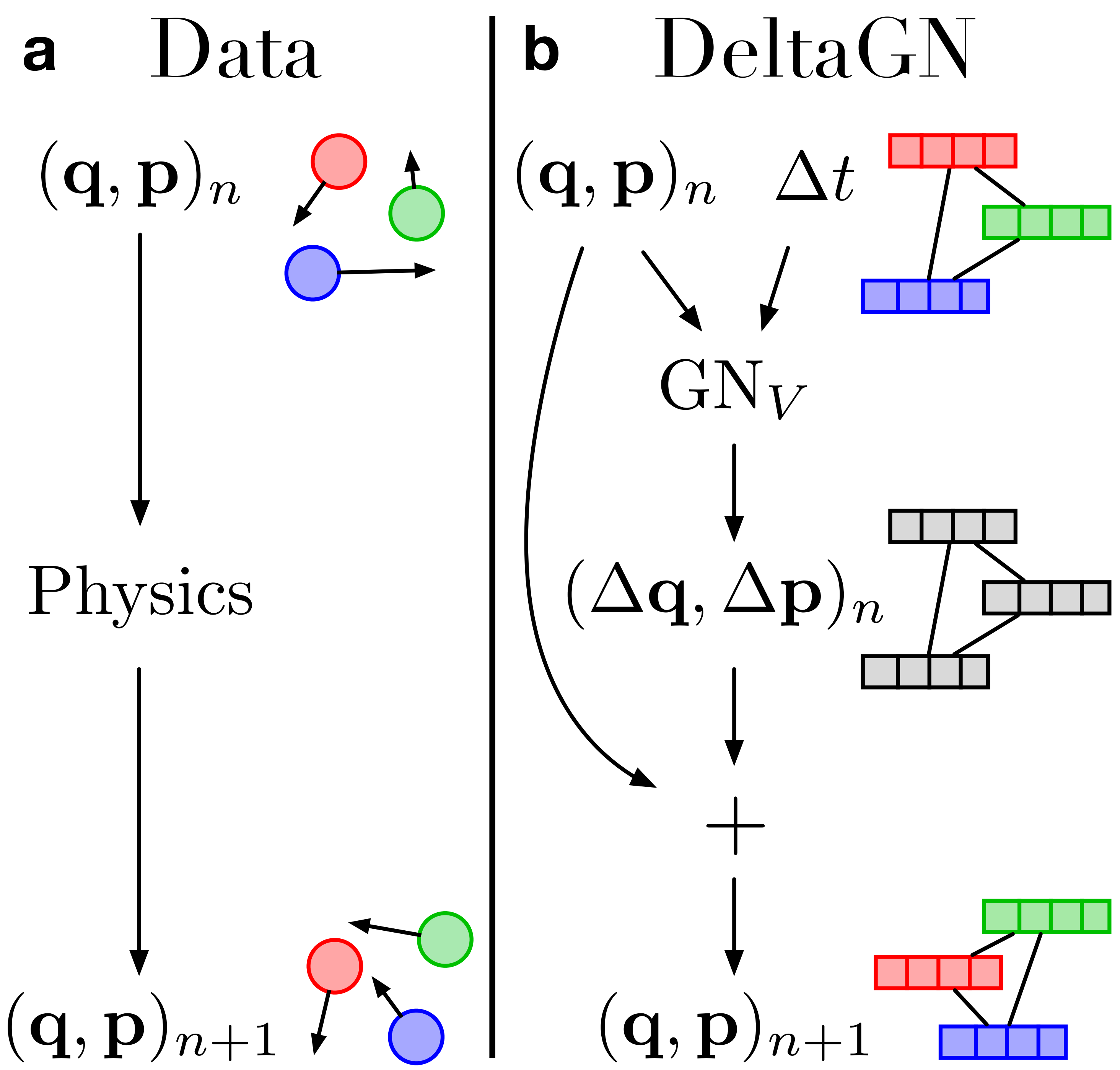
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

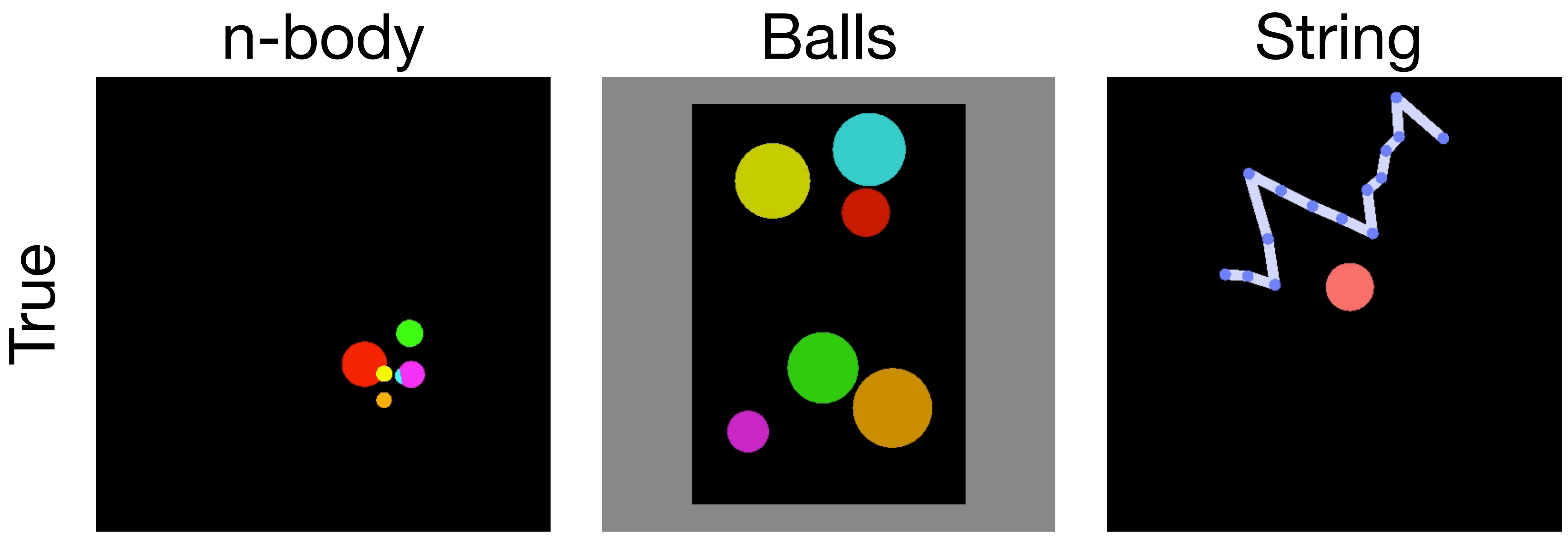
Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

We incorporated two physically-informed inductive biases

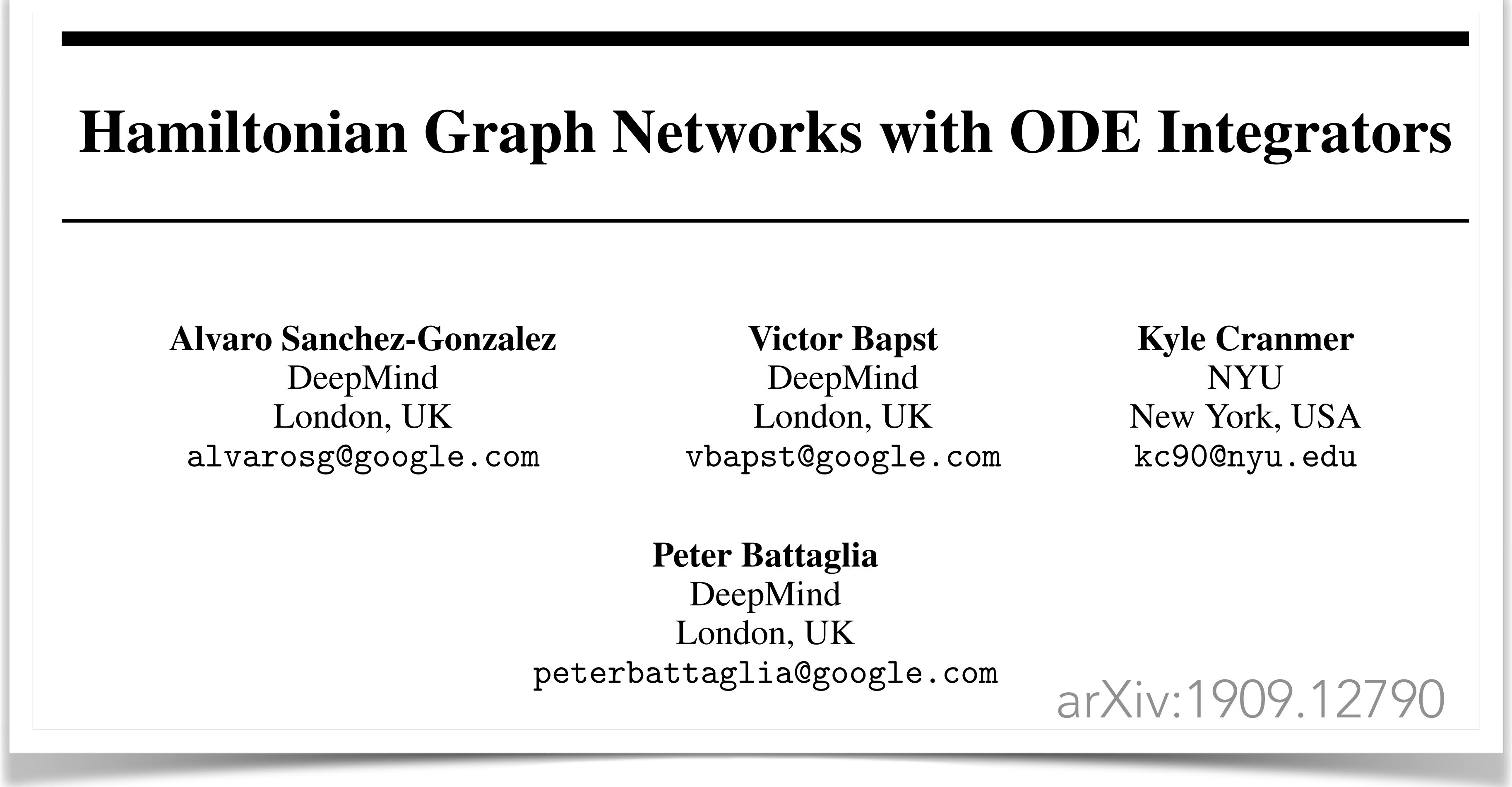
- ODE integrators
- Hamiltonian mechanics





Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

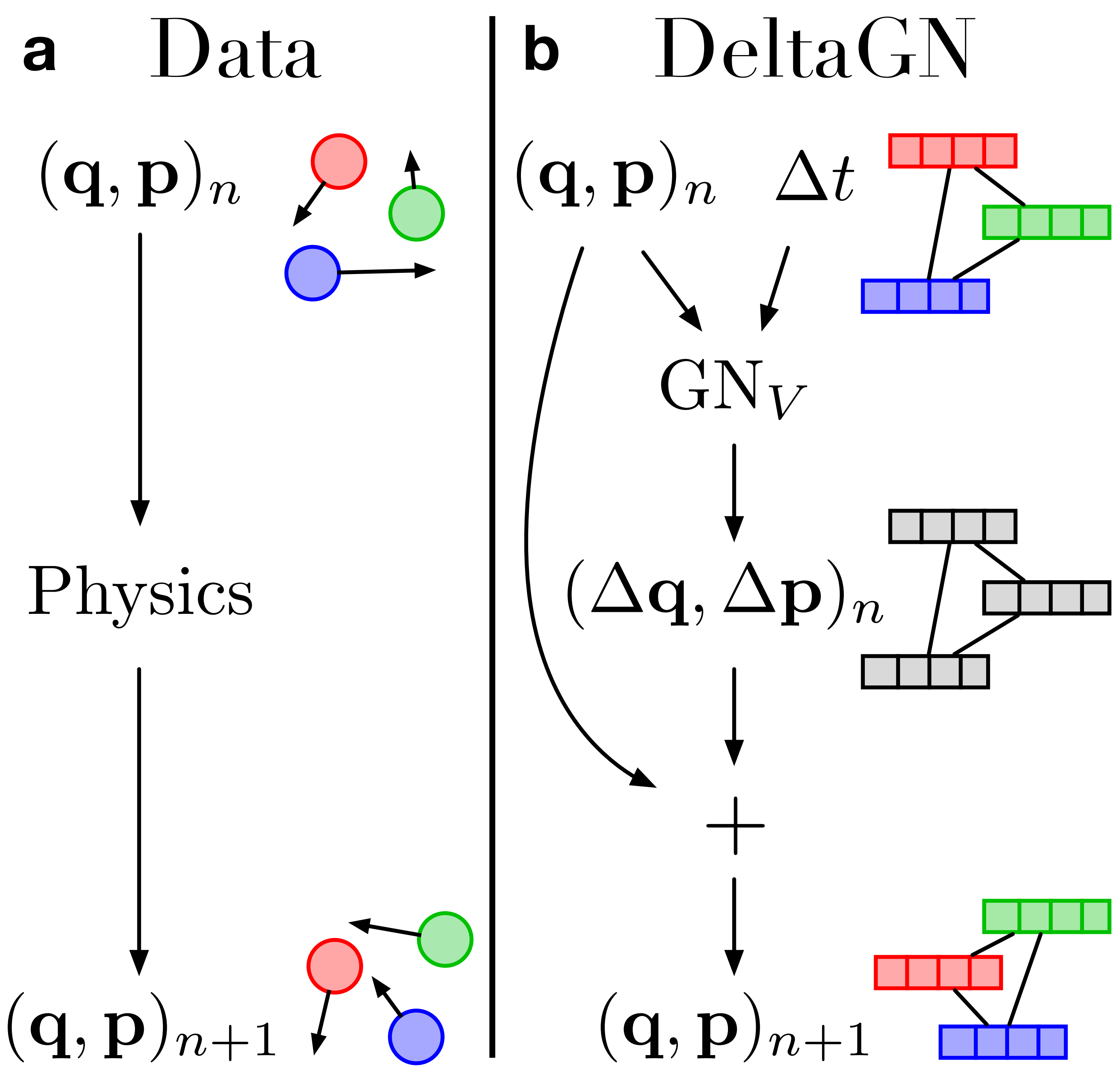
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

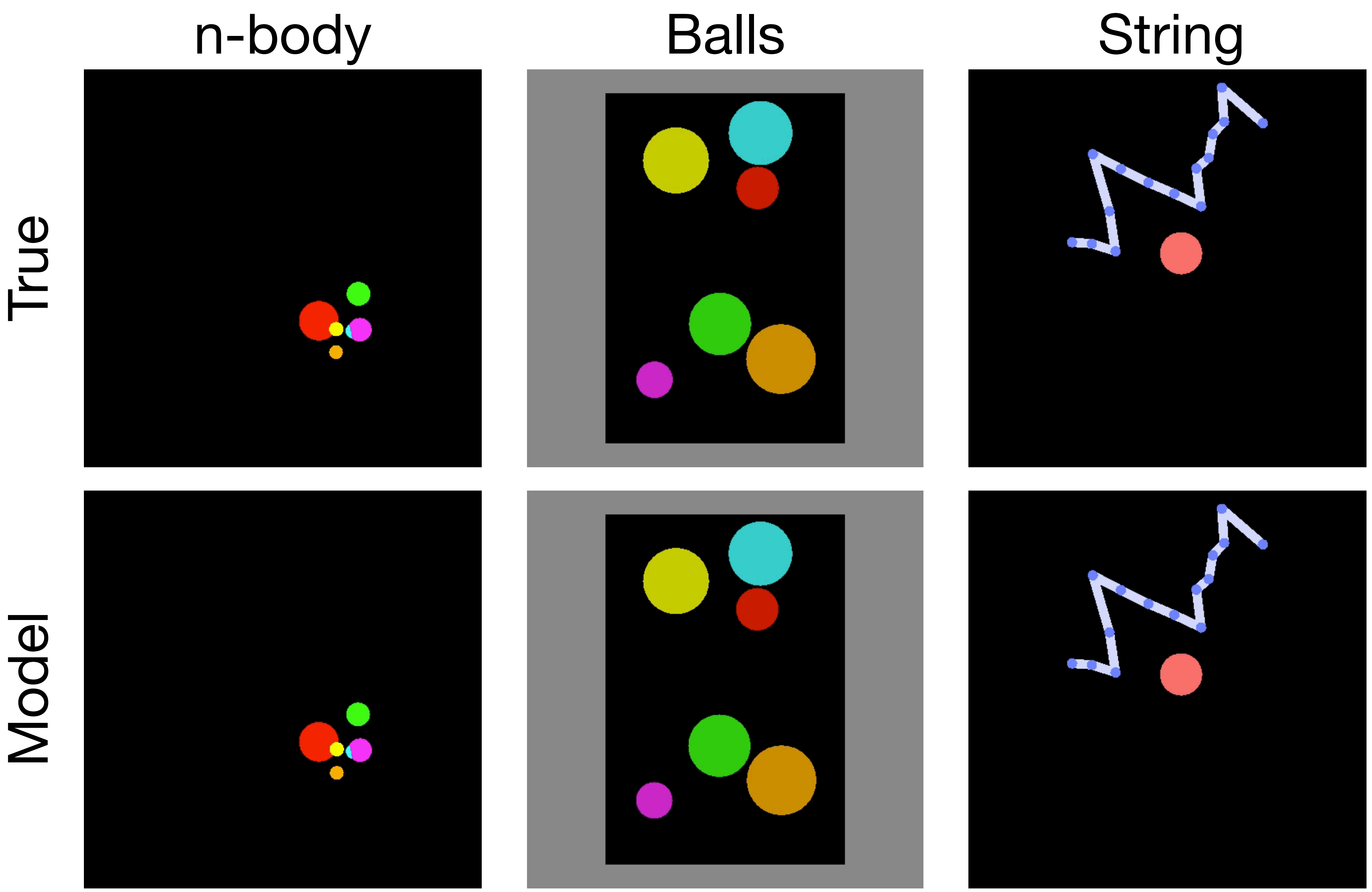
Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics

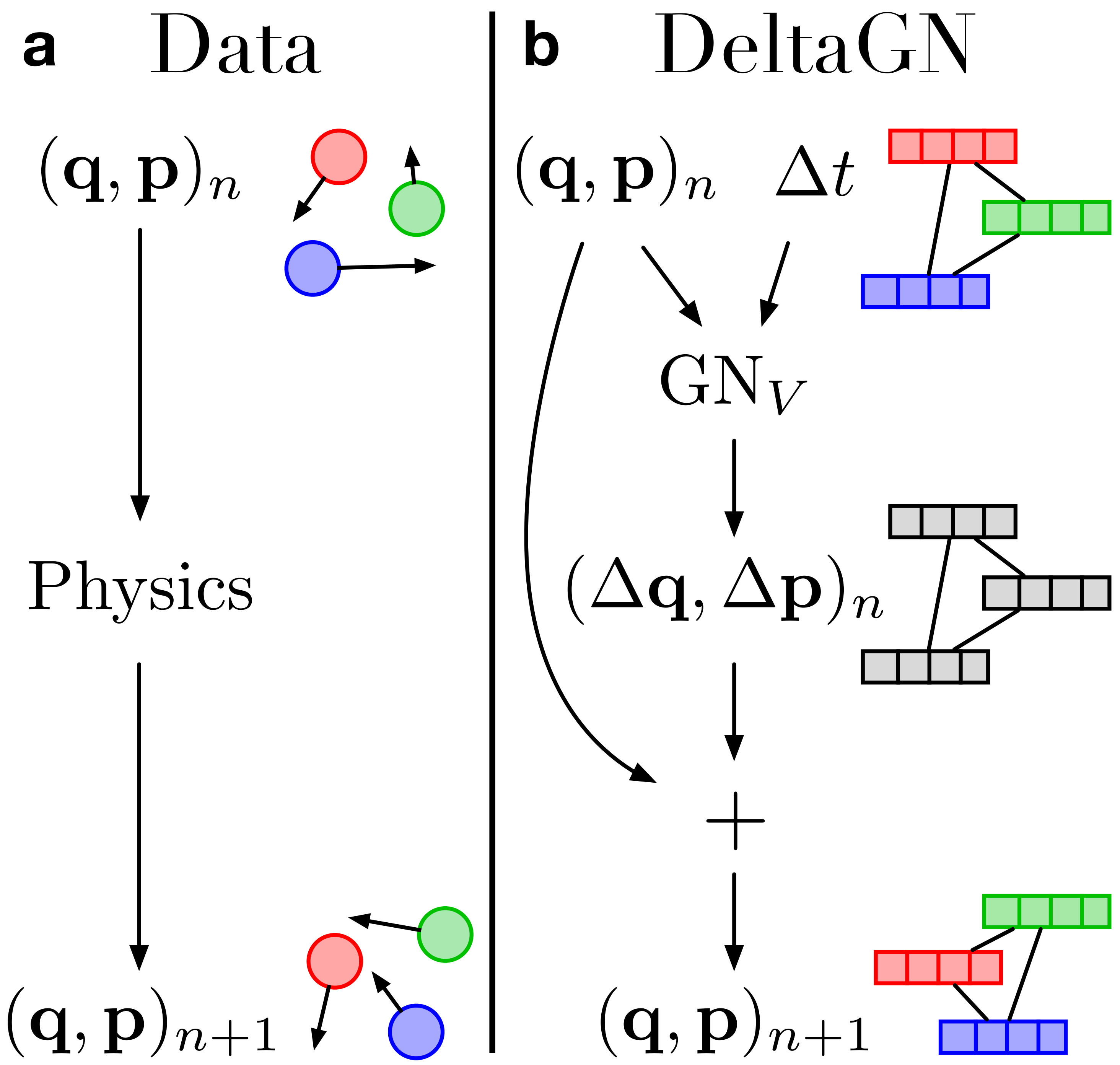




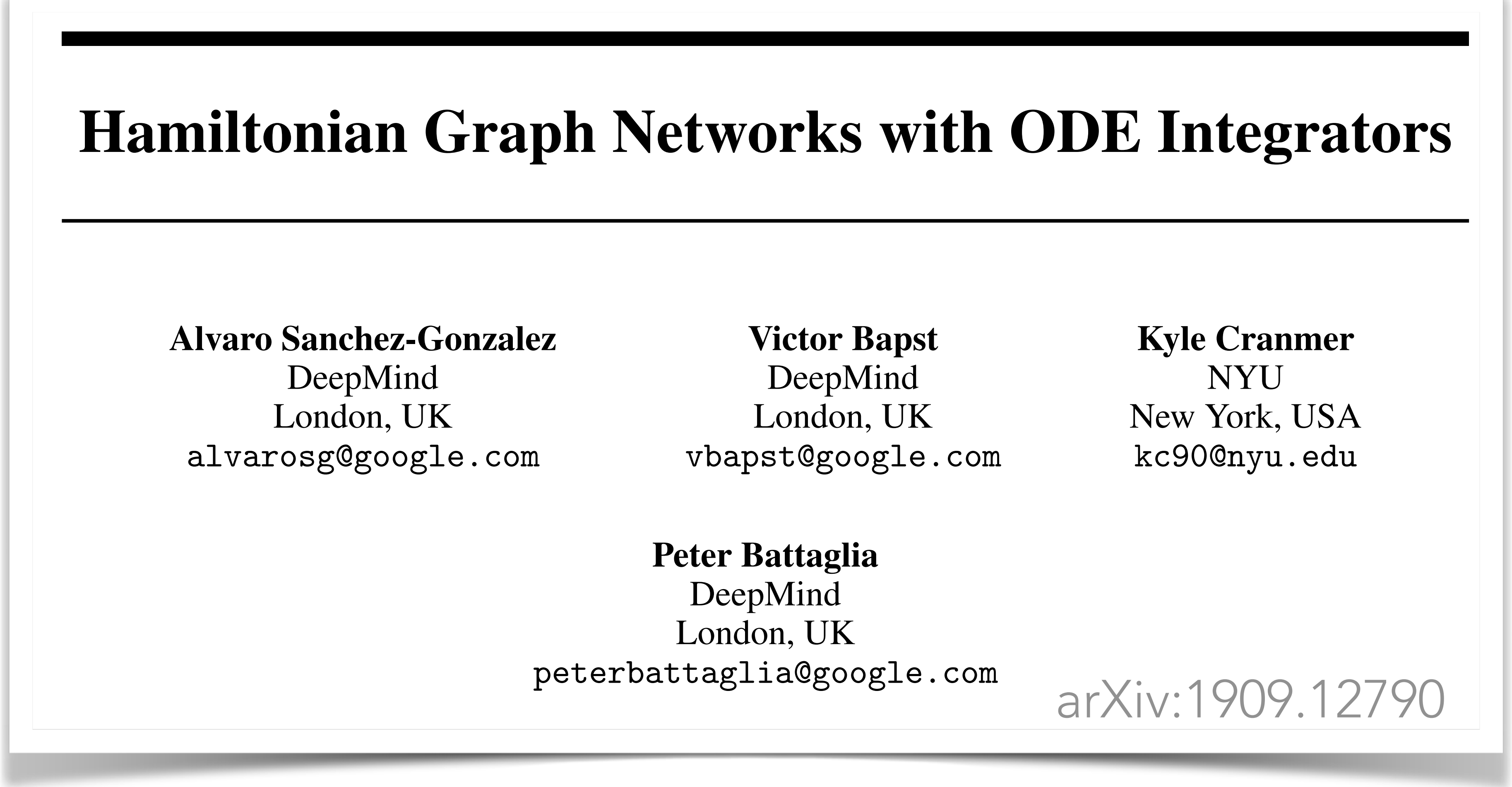
Hamiltonian graph Networks

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics




Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

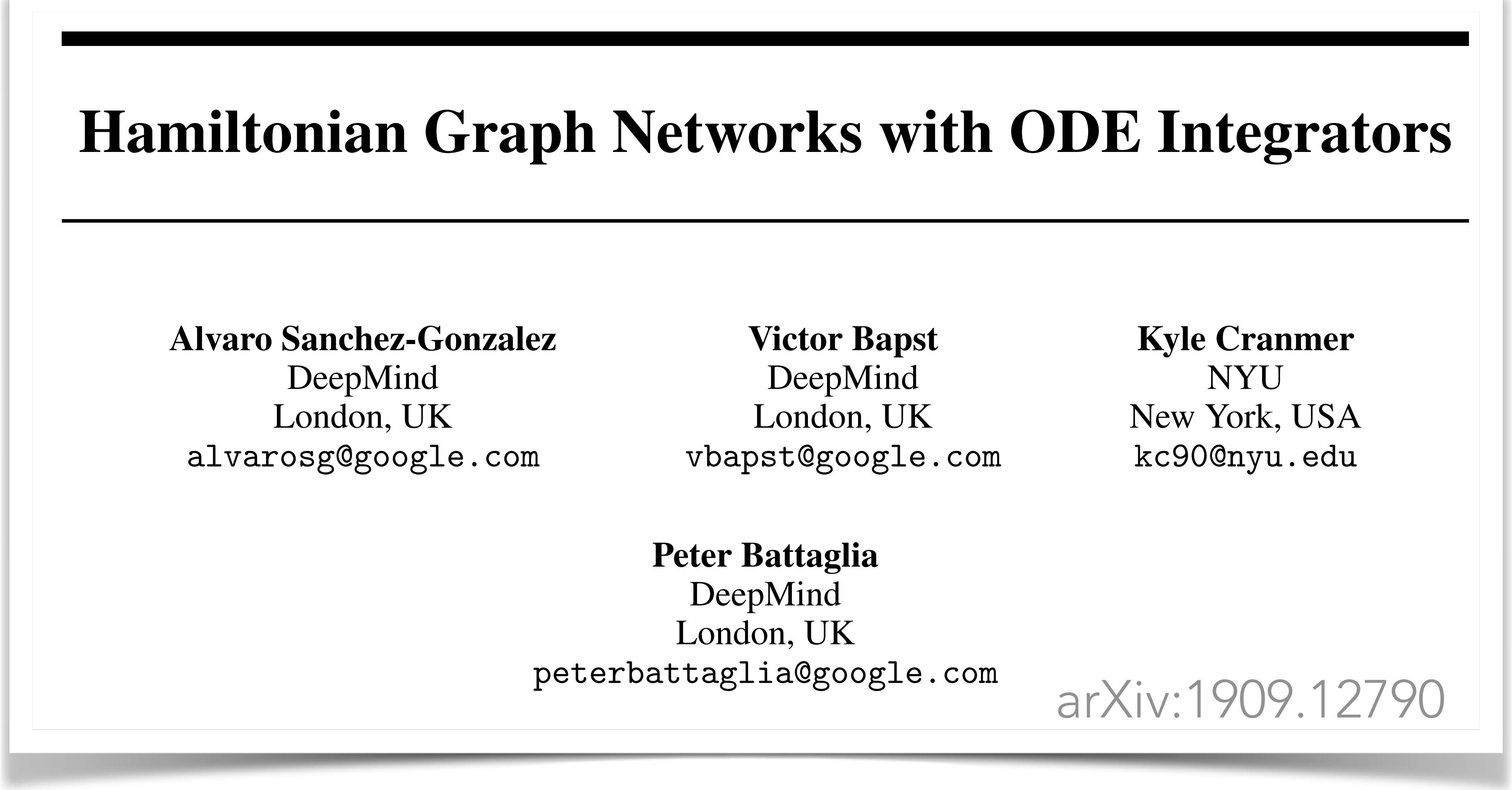
arXiv:1909.12790

Hamiltonian graph Networks

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

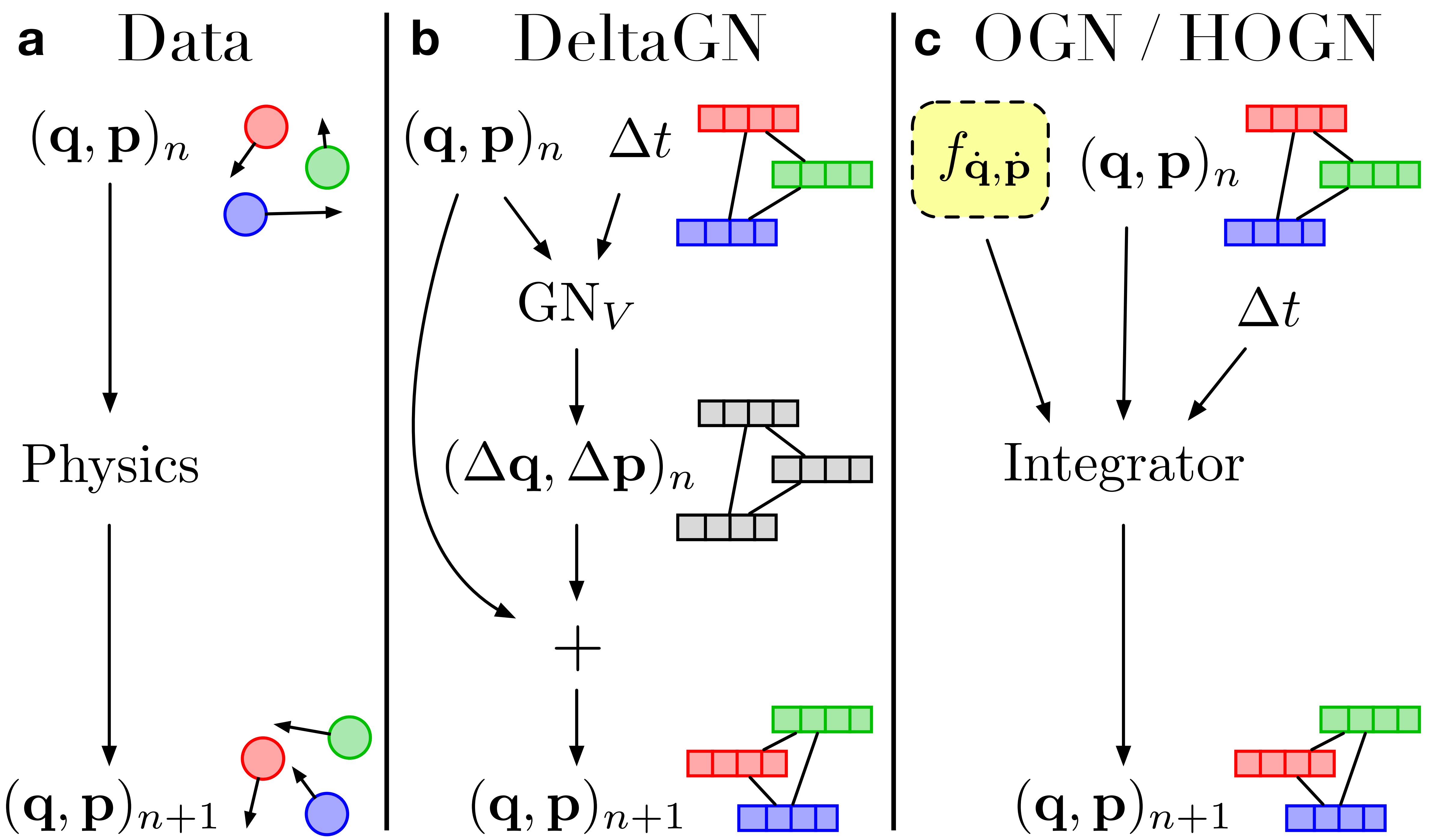
Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

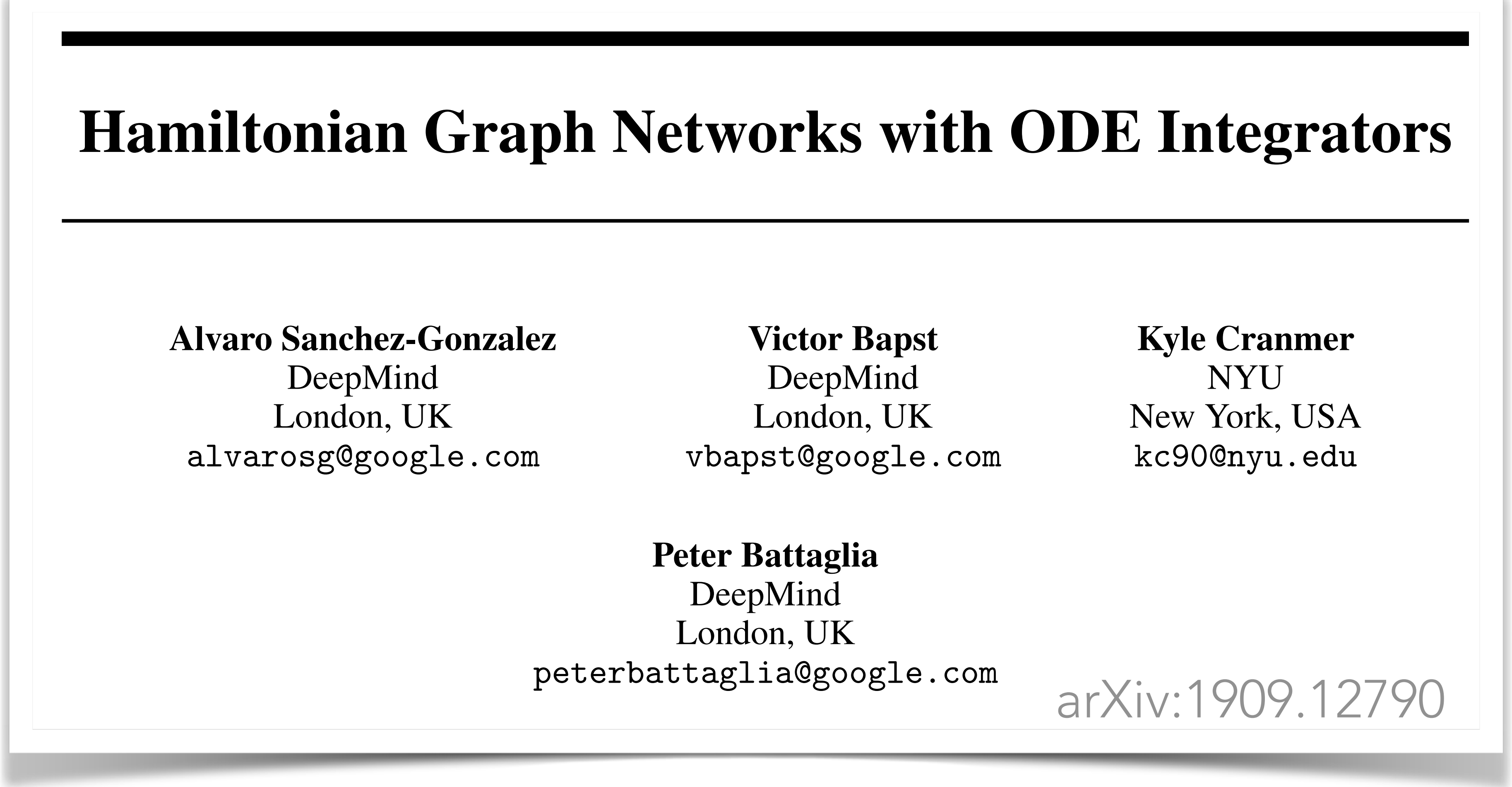
Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790



Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

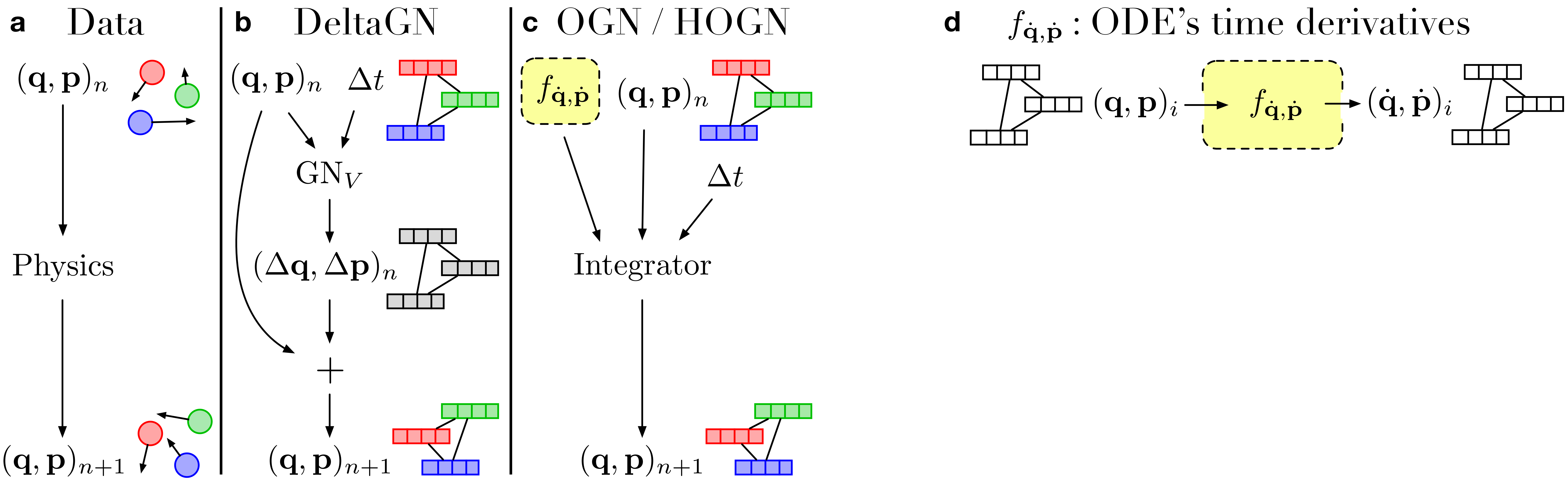
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics

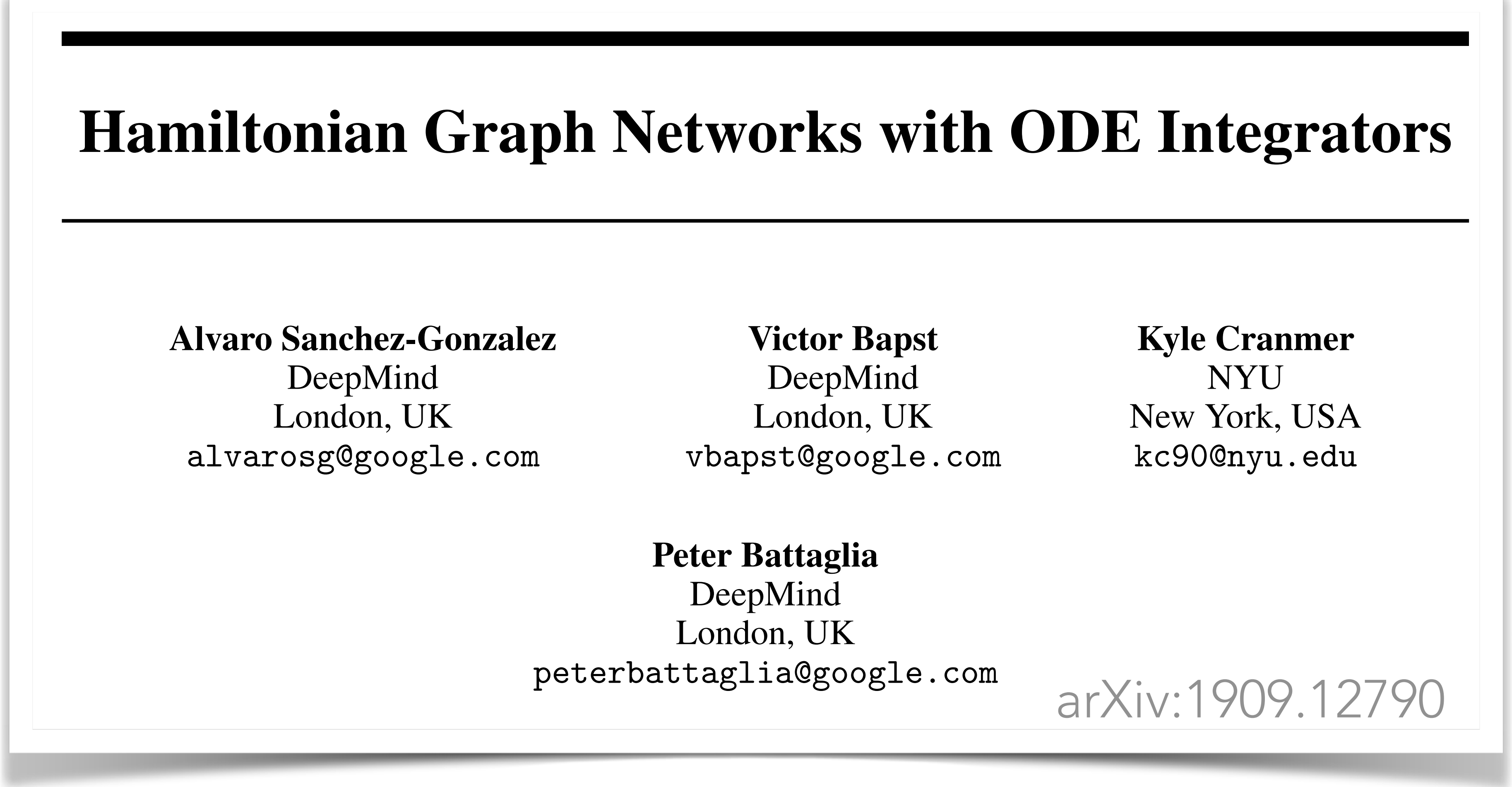


Hamiltonian graph Networks

We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics


Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

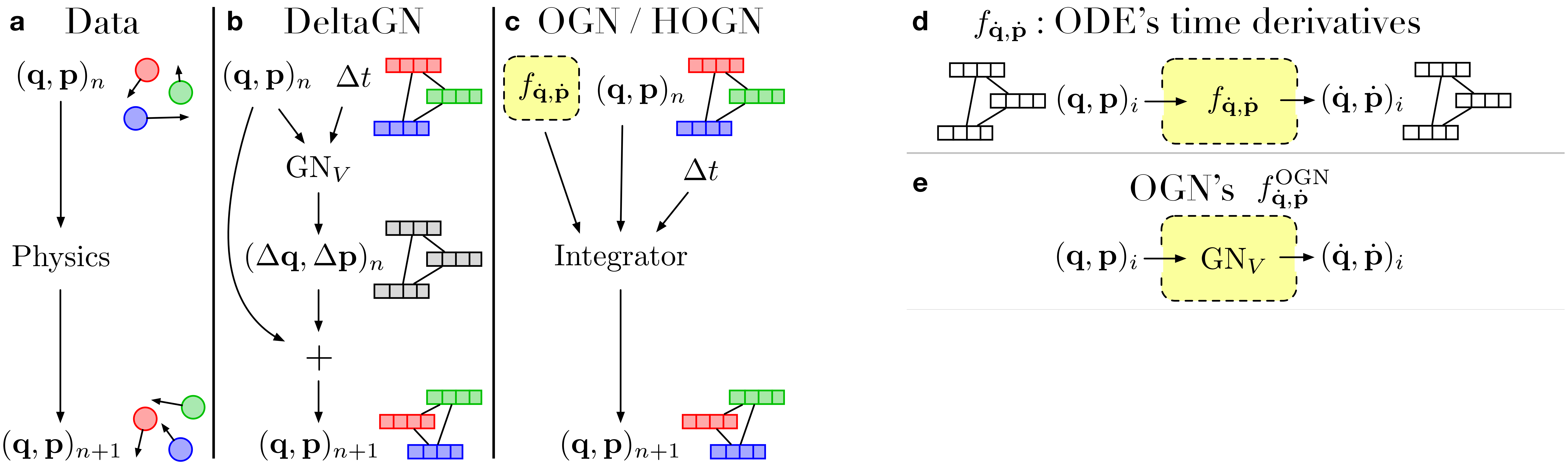
Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790



Hamiltonian graph Networks

Alvaro Sanchez Gonzalez

Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

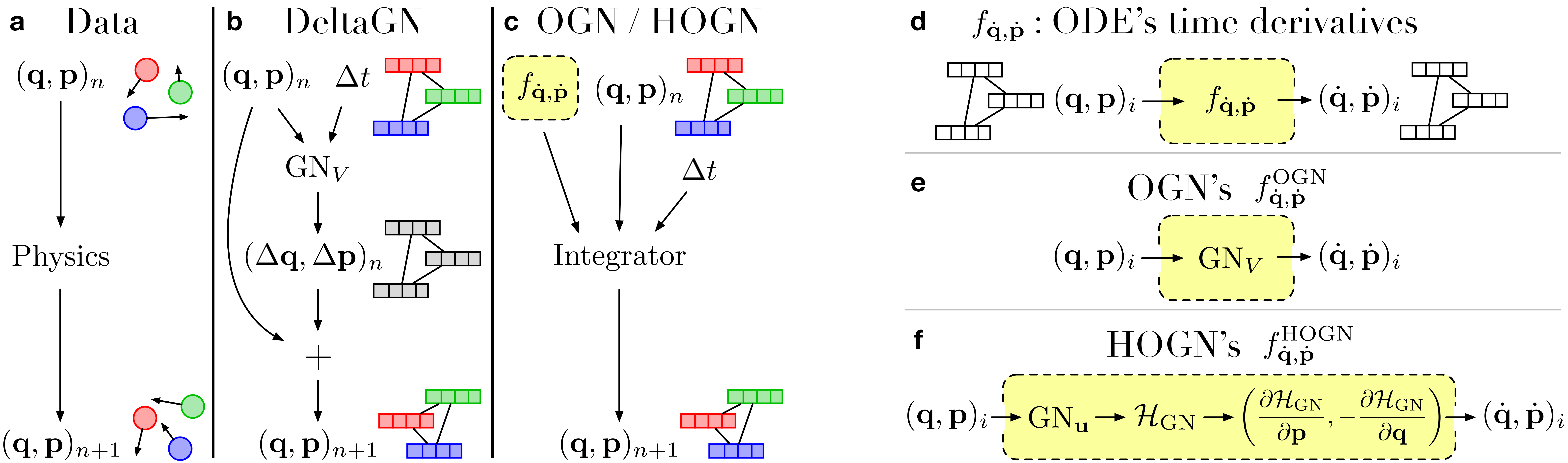
Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

arXiv:1909.12790

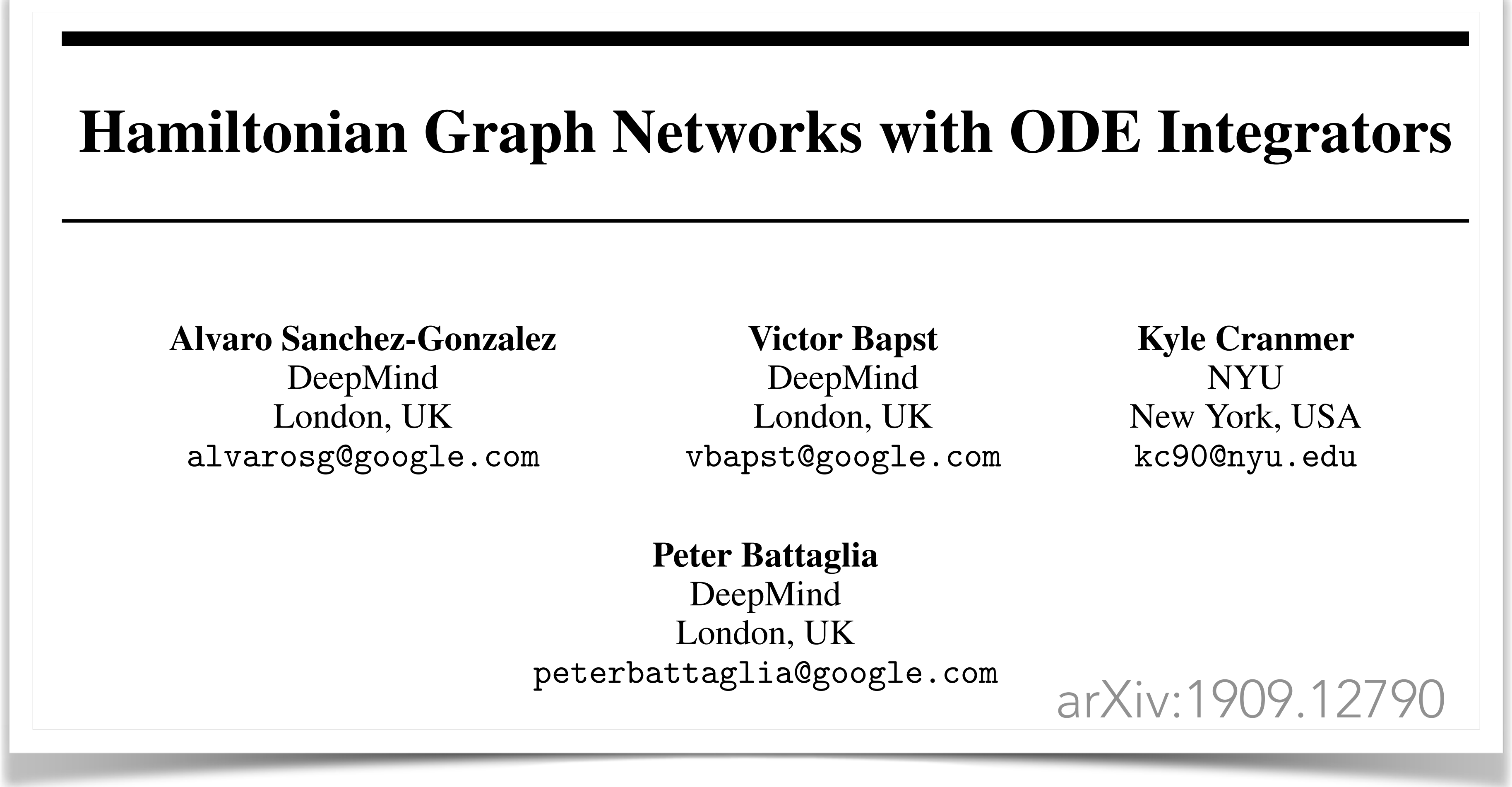
We incorporated two physically-informed inductive biases

- ODE integrators
- Hamiltonian mechanics



Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

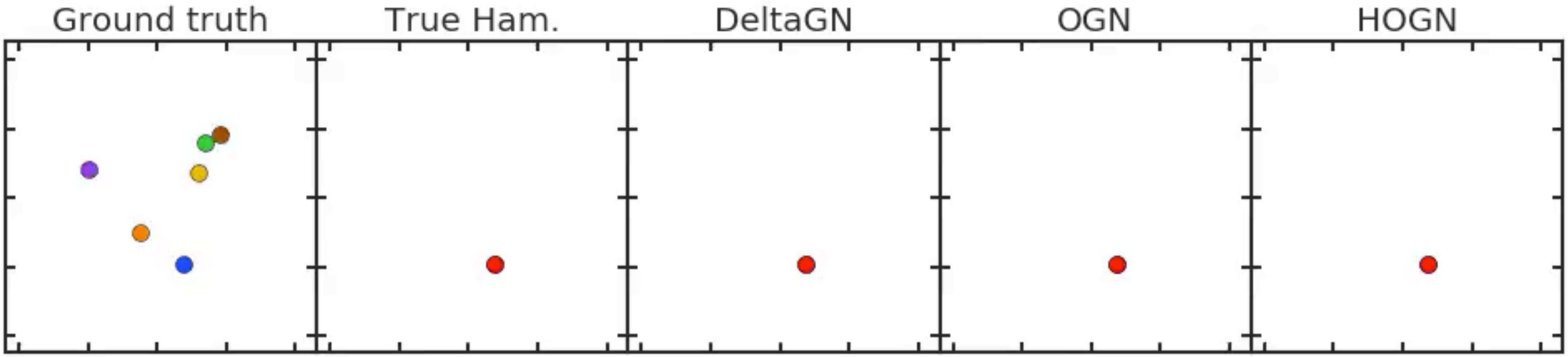
arXiv:1909.12790

We incorporated two physically-informed inductive biases

- ODE integrators , Hamiltonian mechanics

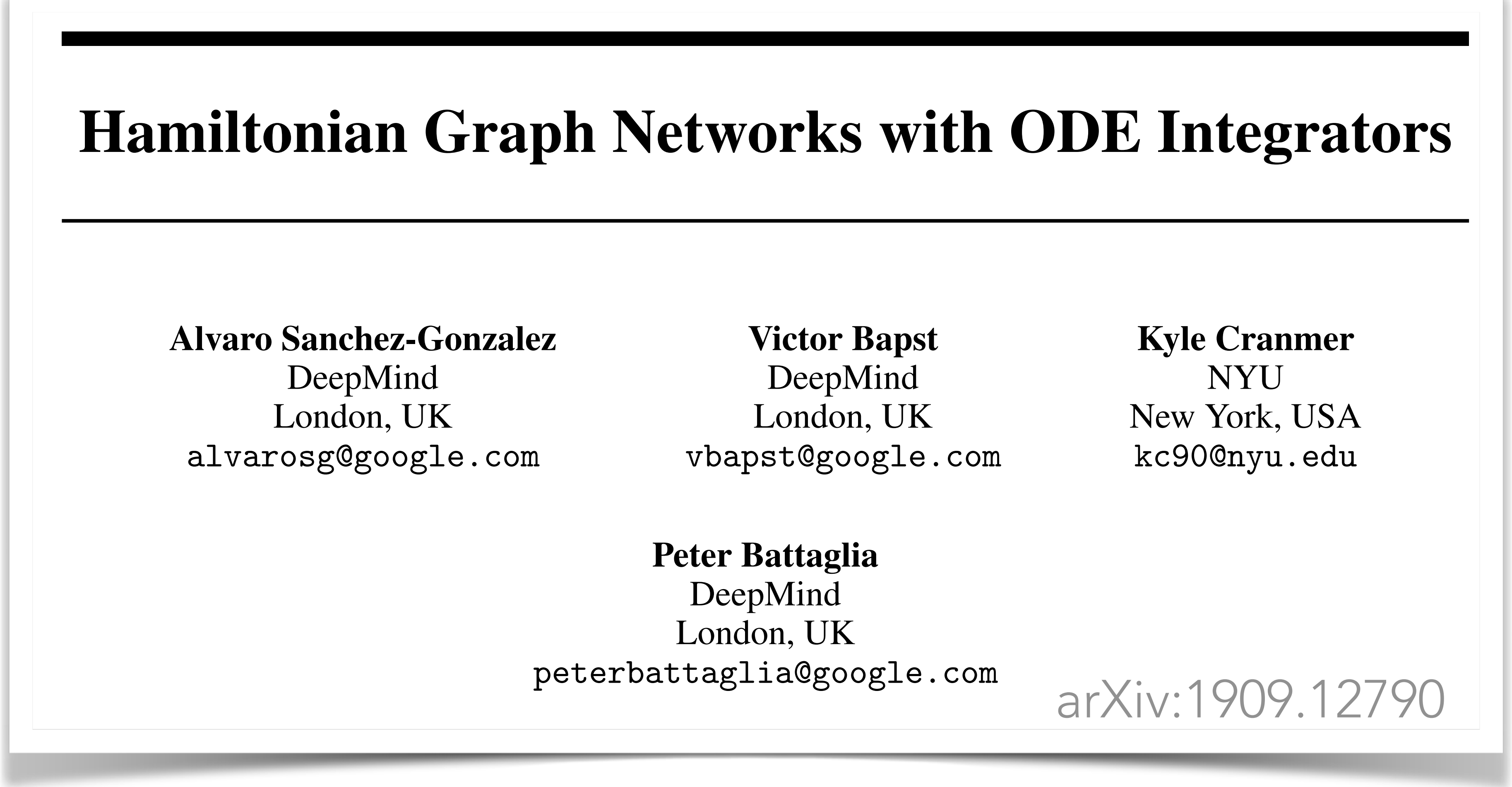
We found they could improve performance, energy, and zero-shot time-step generalization.

- HOGN generalizes better across integrators and time steps



Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

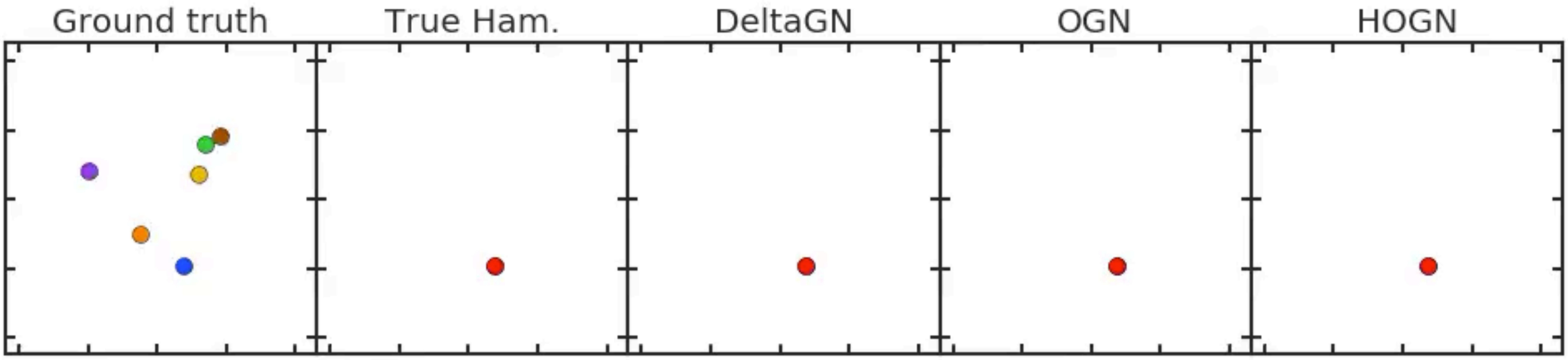
arXiv:1909.12790

We incorporated two physically-informed inductive biases

- ODE integrators , Hamiltonian mechanics

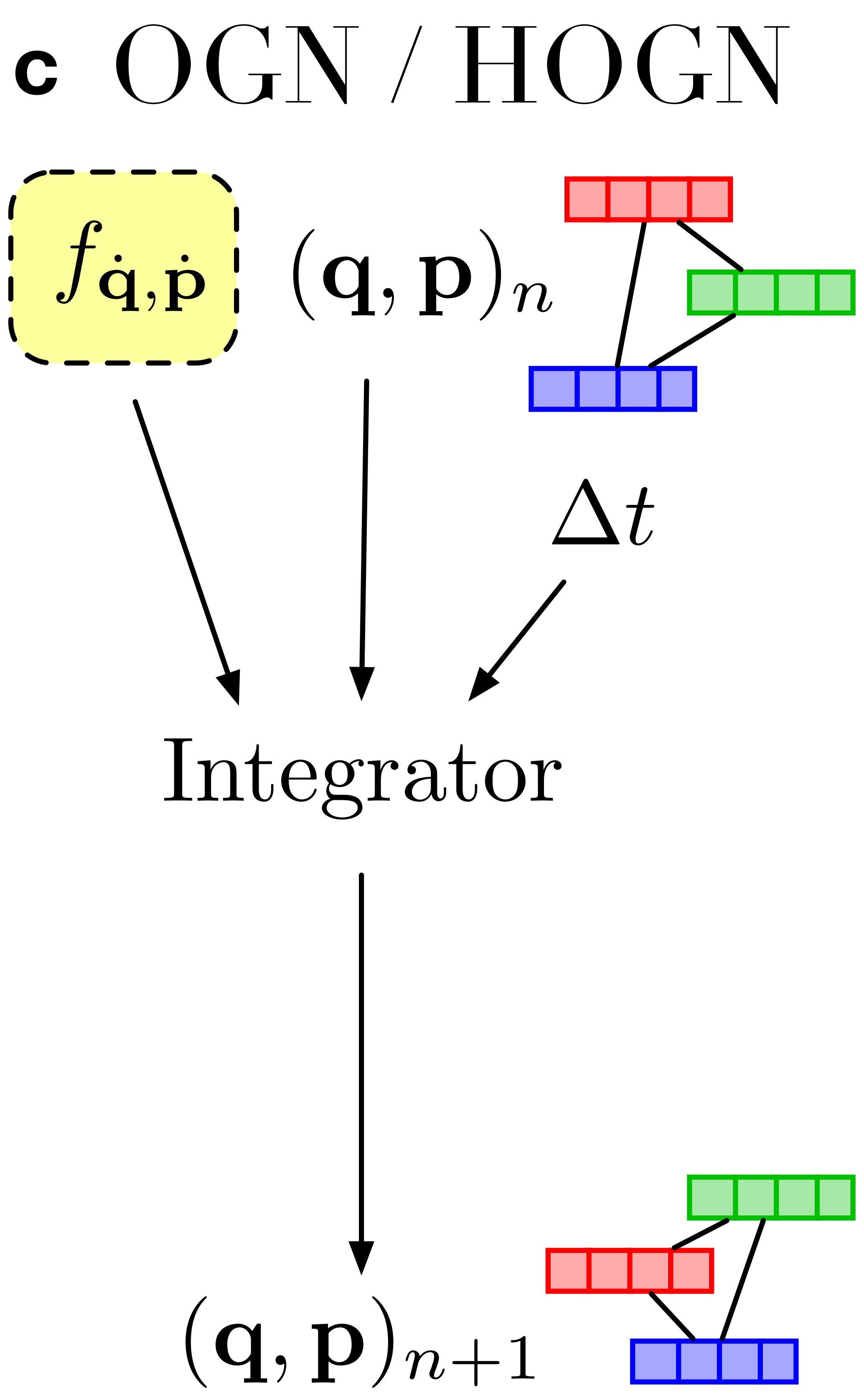
We found they could improve performance, energy, and zero-shot time-step generalization.

- HOGN generalizes better across integrators and time steps

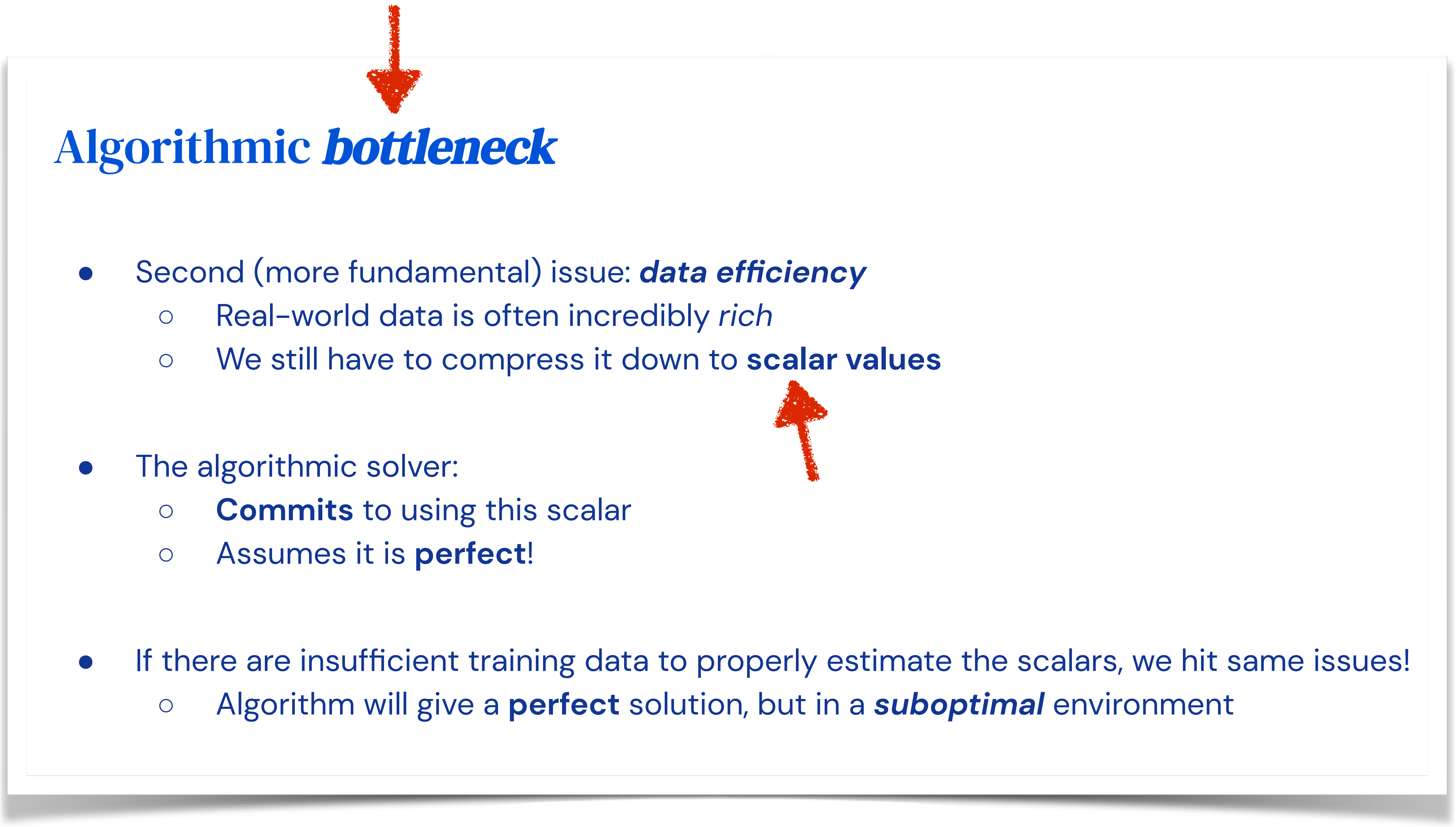


Bottlenecks

Reminder, in the HOGN, the global hidden state goes through a severe **bottleneck**… the Hamiltonian is a scalar!

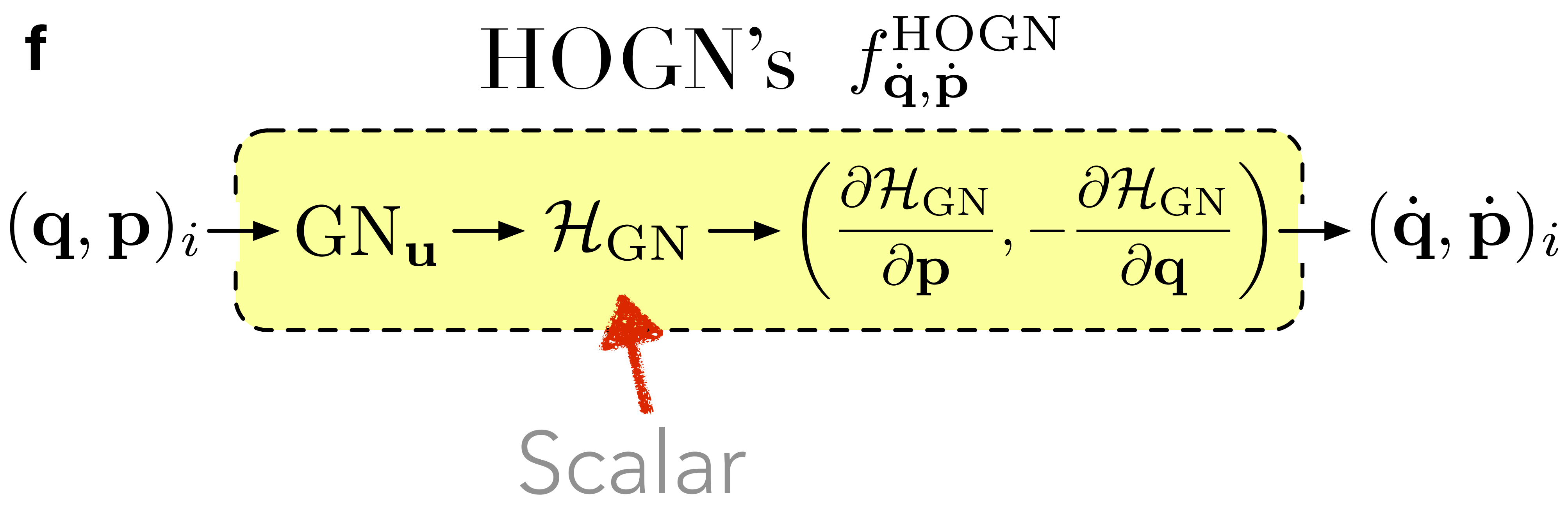


Again similarities with algorithmic reasoning

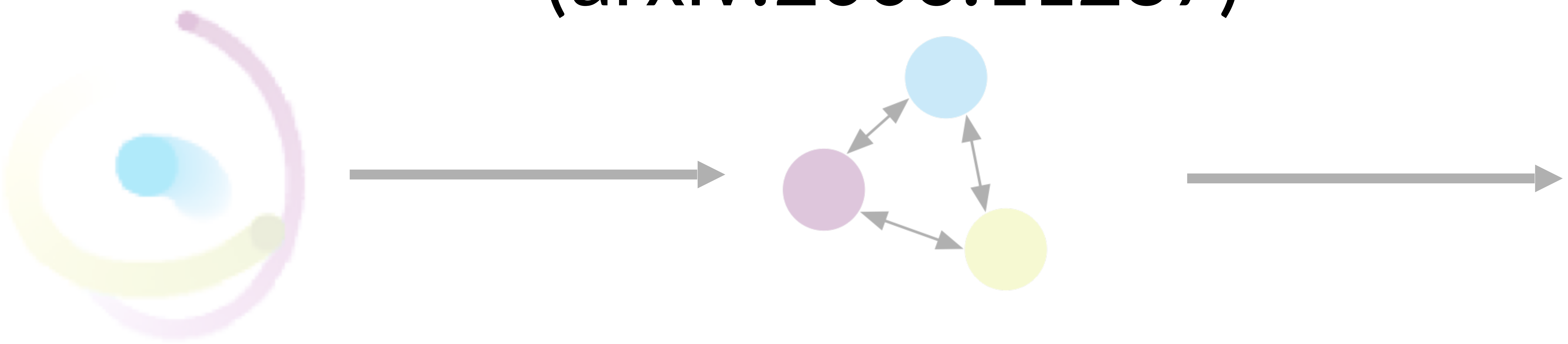


Algorithmic ***bottleneck***

- Second (more fundamental) issue: ***data efficiency***
  - Real-world data is often incredibly *rich*
  - We still have to compress it down to **scalar values**
- The algorithmic solver:
  - **Commits** to using this scalar
  - Assumes it is **perfect**!
- If there are insufficient training data to properly estimate the scalars, we hit same issues!
  - Algorithm will give a **perfect** solution, but in a ***suboptimal*** environment



Discovering Symbolic Models from Deep Learning with Inductive Biases (arxiv:2006.11287)



$$\frac{m_1 m_2}{r^2}\hat{r}$$

Miles Cranmer, Dept. of Astrophysics, Princeton University

@MilesCranmer

Miles Cranmer

with: Alvaro Sanchez-Gonzalez (DeepMind), Peter Battaglia (DeepMind), Rui Xu (Princeton), Kyle Cranmer (NYU), David Spergel (Flatiron/Princeton), Shirley Ho (Flatiron)

Physics ∩ ML

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

29
Jul 2020

Discovering Symbolic Models in Physical Systems using Deep Learning

Shirley Ho, Flatiron Institute, 12:00 EDT

Abstract: We develop a general approach to distill symbolic representations of a learned deep model by introducing strong inductive biases. We focus on Graph Neural Networks (GNNs). The technique works as follows: we first encourage sparse latent representations when we train a GNN in a supervised setting, then we apply symbolic regression to components of the learned model to extract explicit physical relations. We find the correct known equations, including force laws and Hamiltonians, can be extracted from the neural network. We then apply our method to a non-trivial cosmology example—a detailed dark matter simulation—and discover a new analytic formula that can predict the concentration of dark matter from the mass distribution of nearby cosmic structures. The symbolic expressions extracted from the GNN using our technique also generalized to out-of-distribution-data better than the GNN itself. Our approach offers alternative directions for interpreting neural networks and discovering novel physical principles from the representations they learn.

M

with:

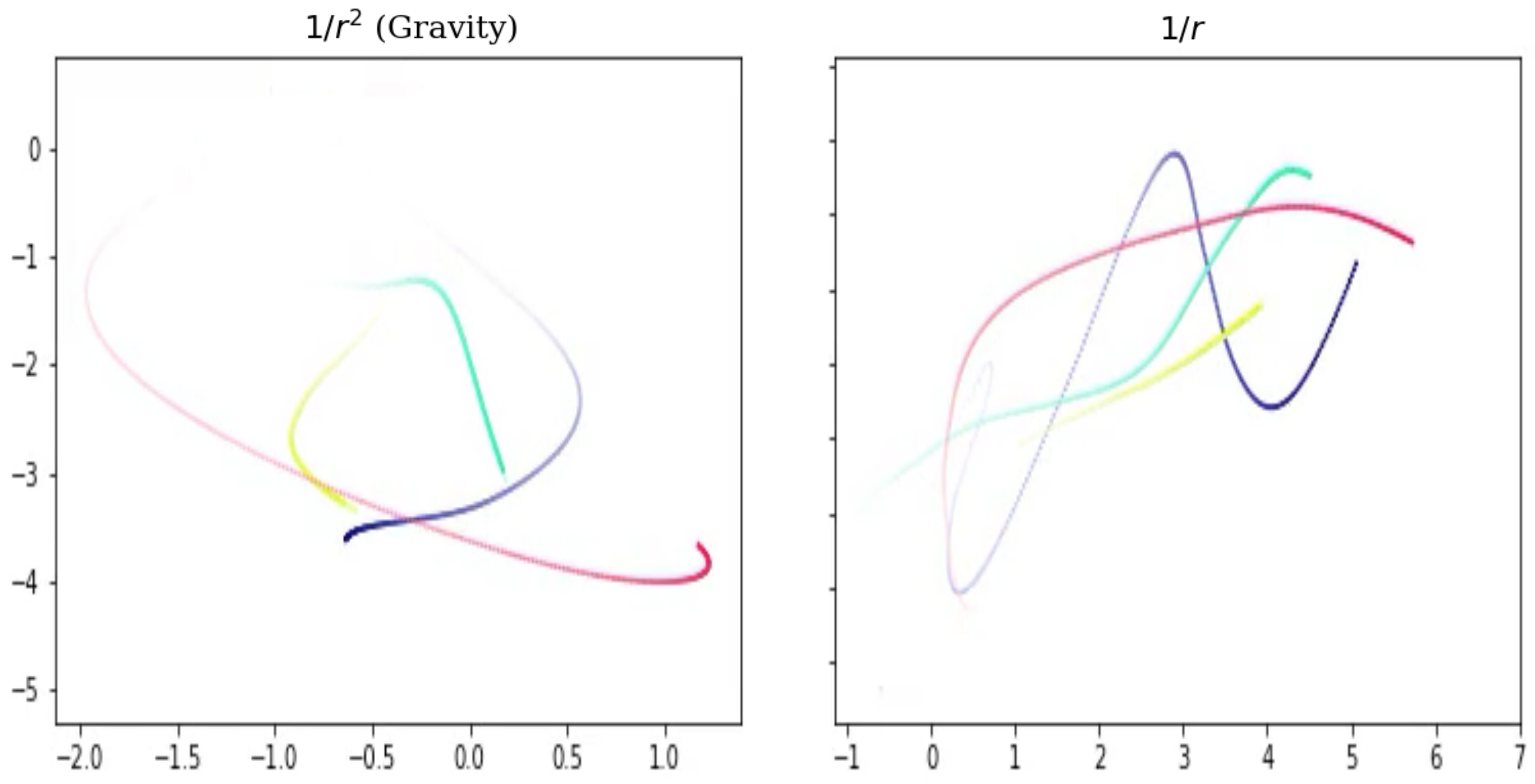
Miles Cranmer

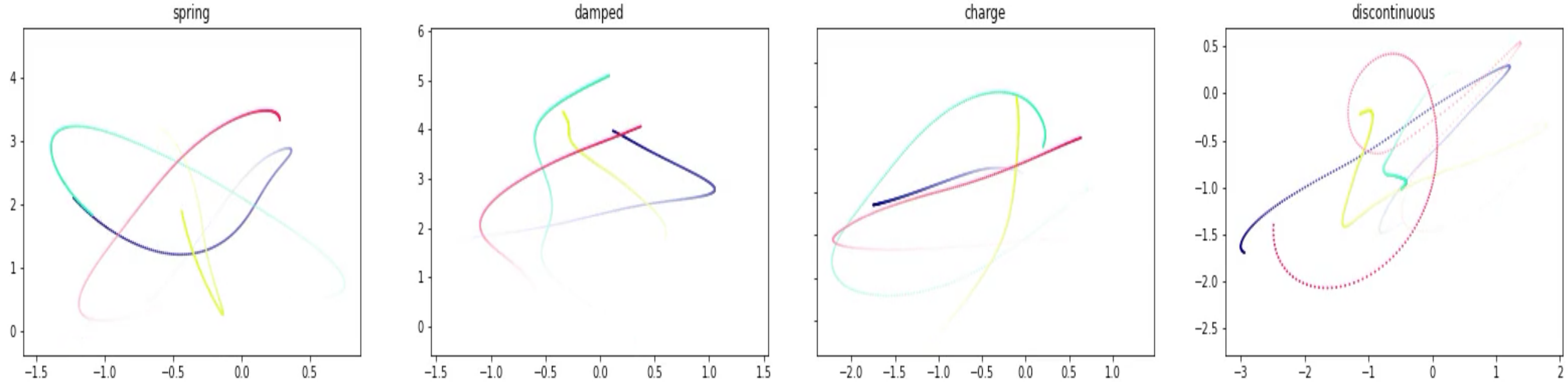
Experiments

M. Cranmer, et. al. arxiv:2006.11287

Newtonian systems (force laws)

$$
\begin{aligned}
1/r^2 &: \; U_{12} = -m_1 m_2 / r'_{12} \\
1/r &: \; U_{12} = m_1 m_2 \log(r'_{12}) \\
\text{Spring} &: \; U_{12} = (r'_{12} - 1)^2 \\
\text{Damped} &: \; U_{12} = (r'_{12} - 1)^2 + \mathbf{r}_1 \cdot \dot{\mathbf{r}}_1 / n \\
\text{Charge} &: \; U_{12} = q_1 q_2 / r'_{12} \\
\text{Dicontinuous} &: \; U_{12} = \begin{cases} 0, & r'_{12} < 2 \\ (r'_{12} - 1)^2, & r'_{12} \geq 2 \end{cases}
\end{aligned}
$$

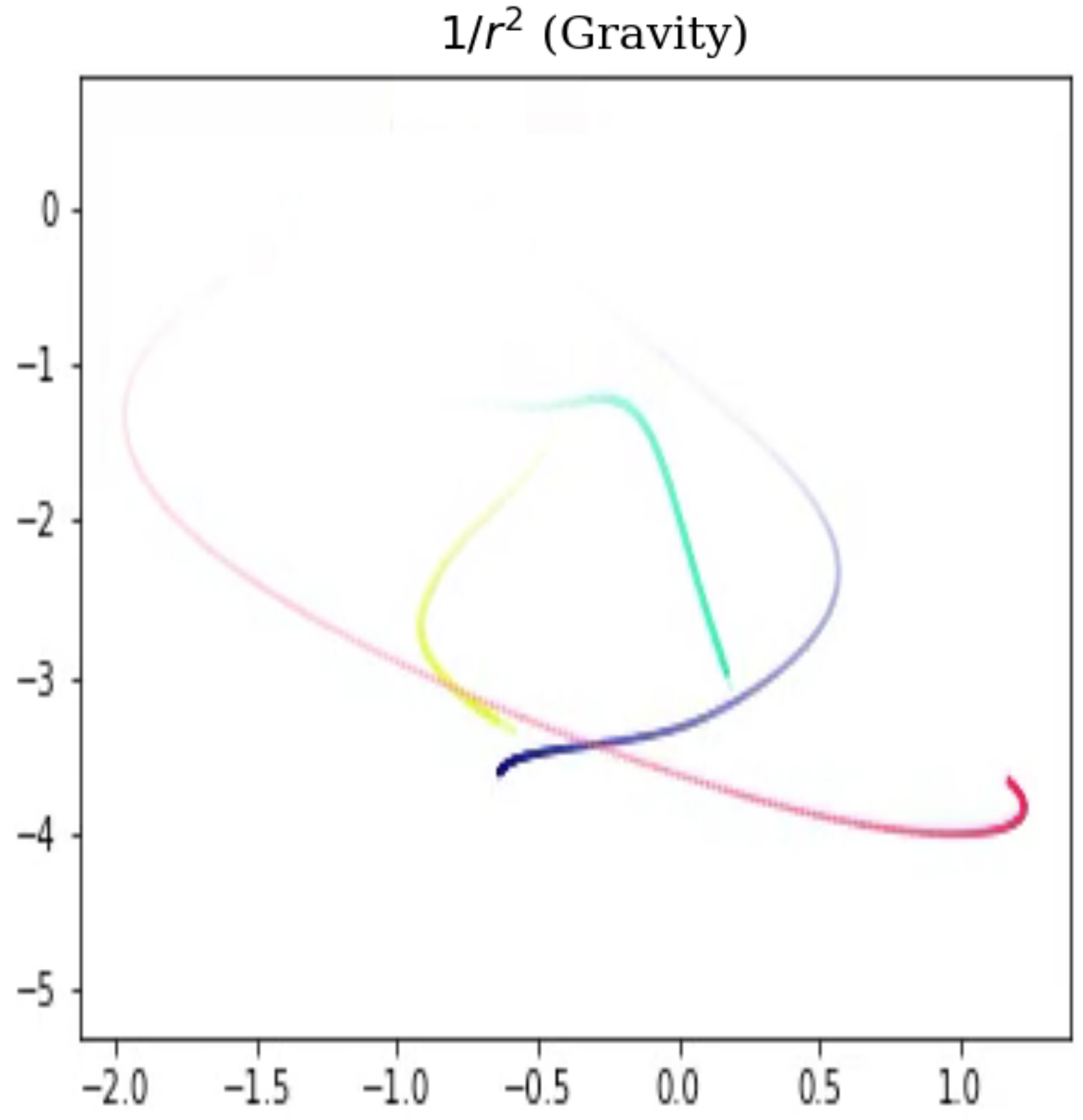


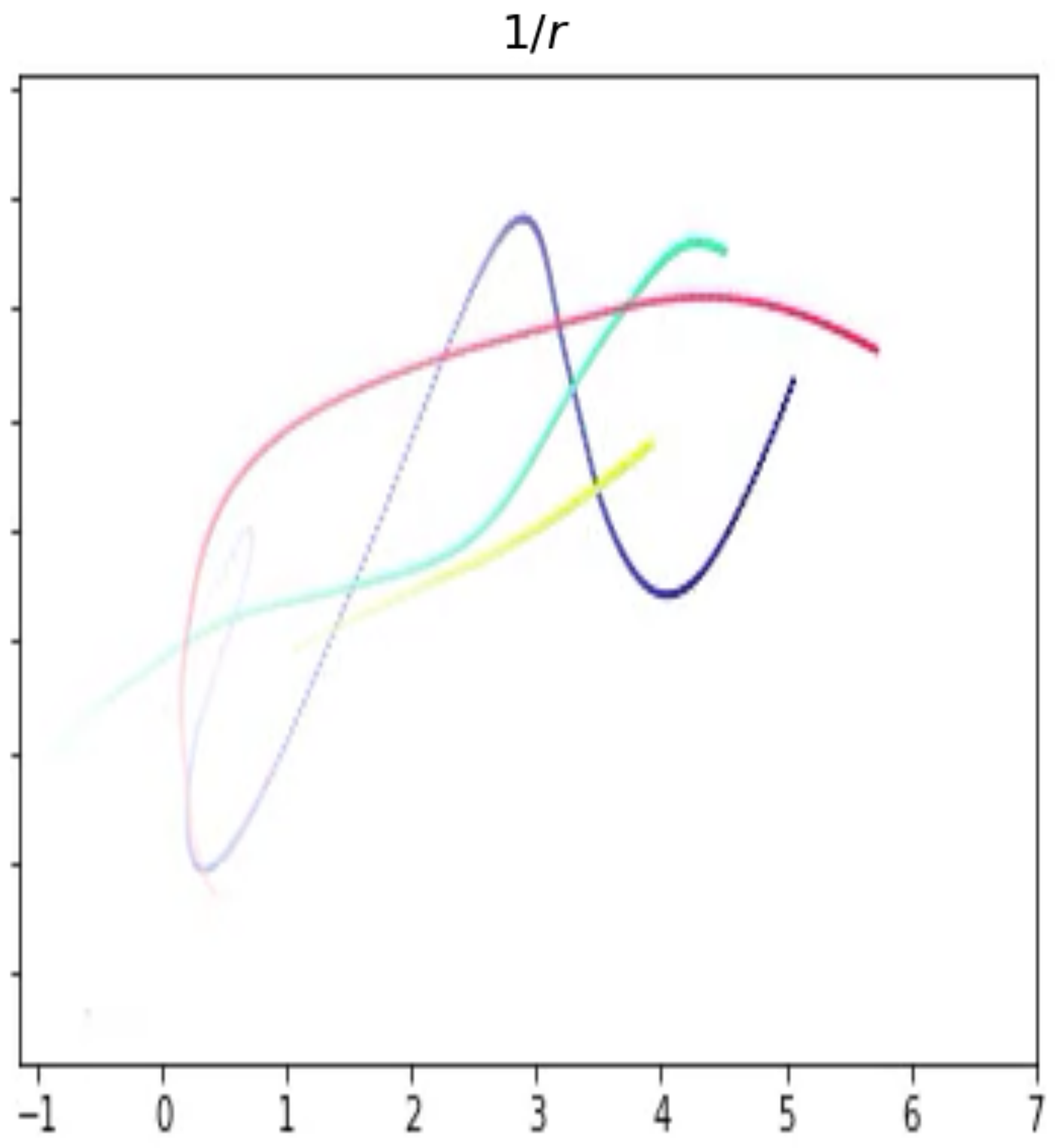


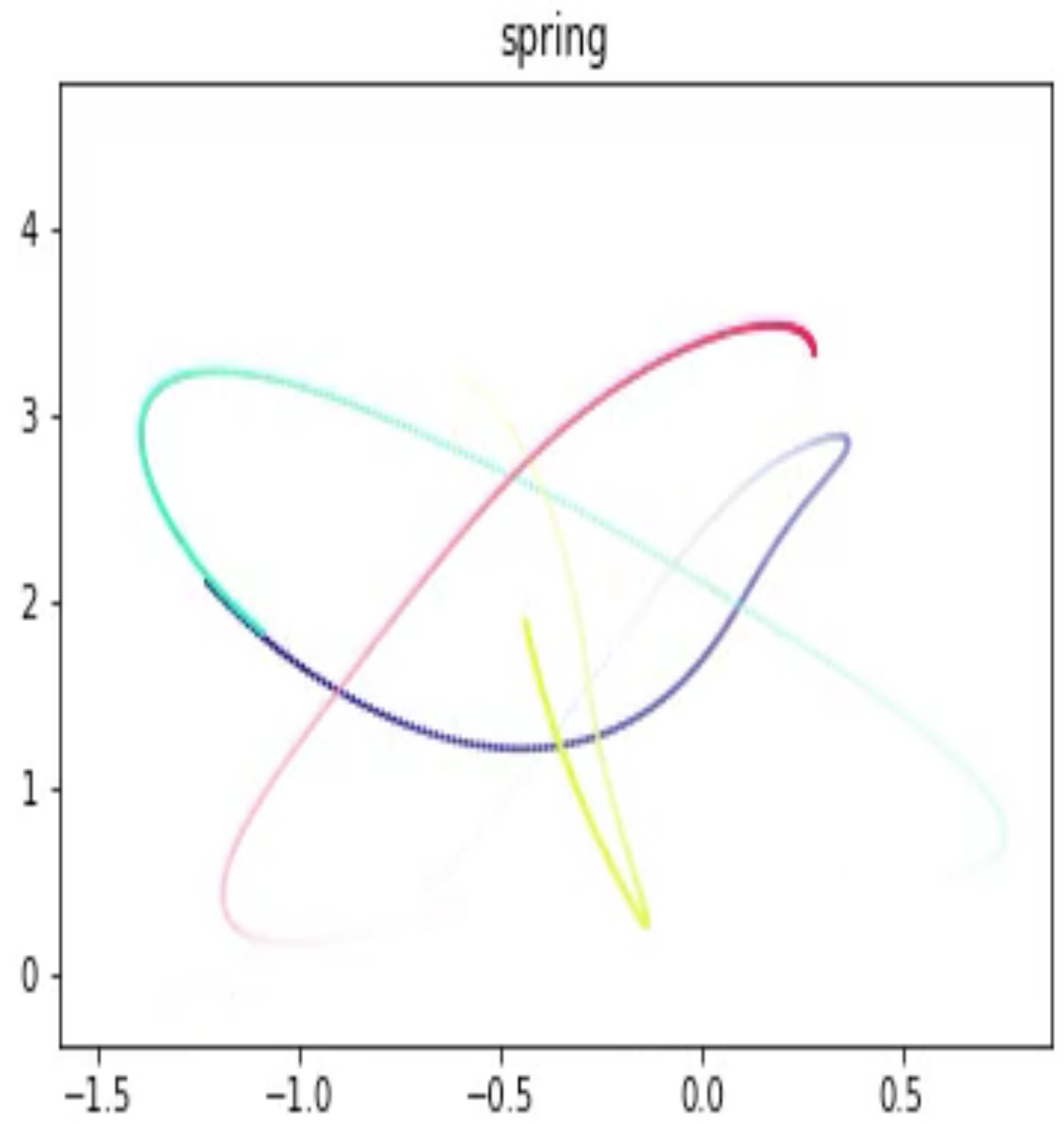
Experiments

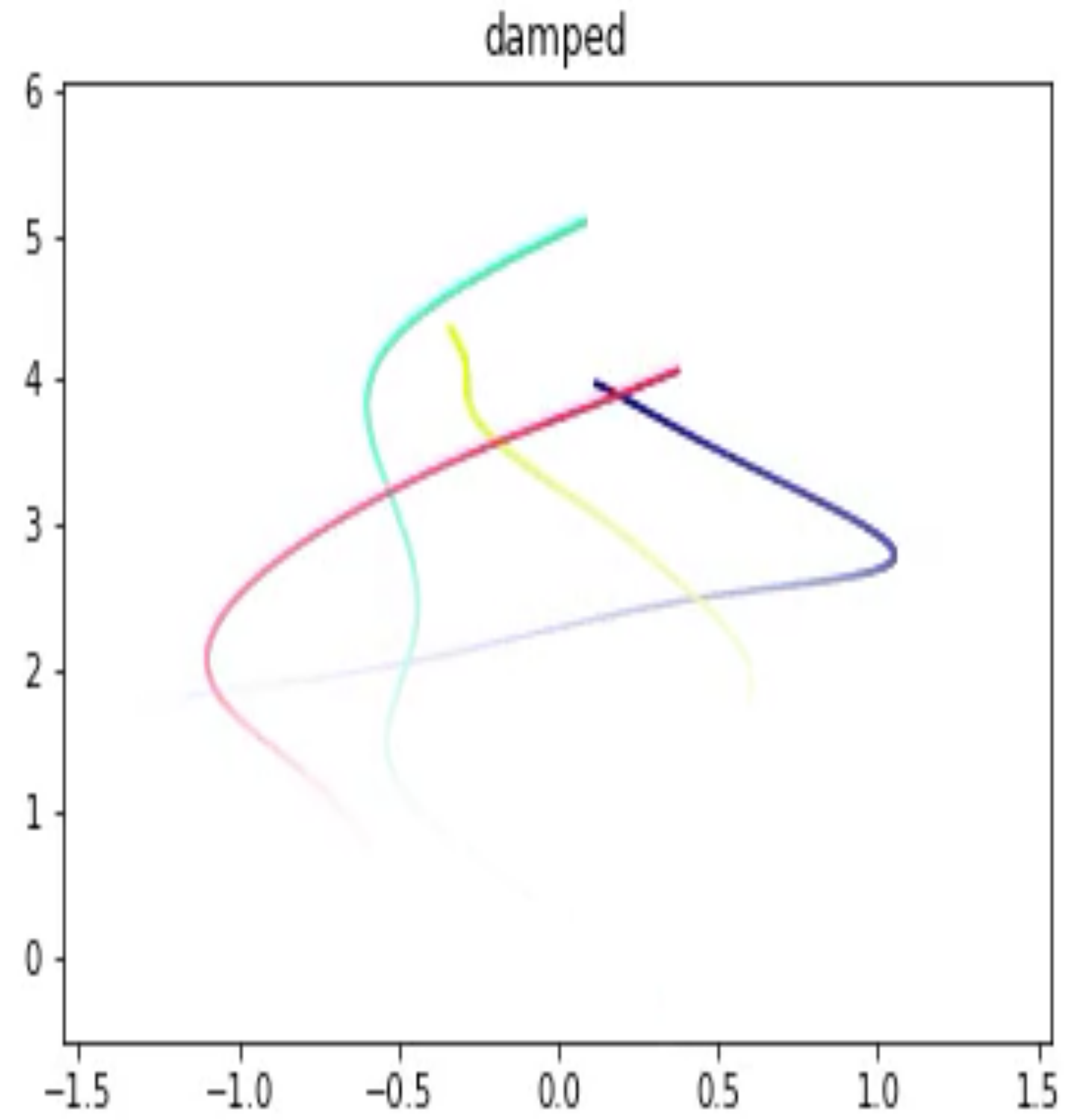
Newtonian systems (force laws)

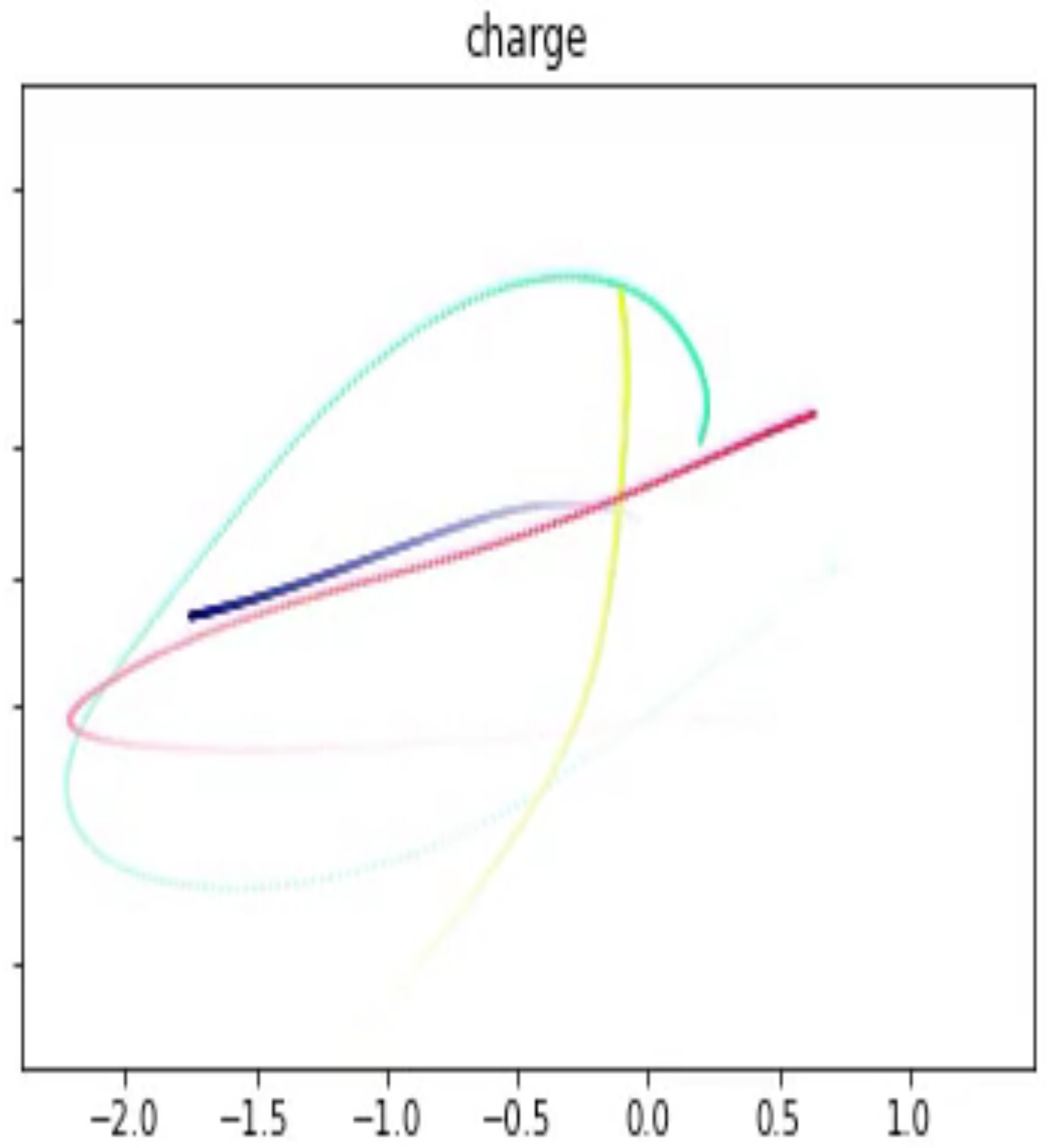
$$
\begin{aligned}
1/r^2 &: U_{12} = -m_1 m_2 / r'_{12} \\
1/r &: U_{12} = m_1 m_2 \log(r'_{12}) \\
\text{Spring} &: U_{12} = (r'_{12} - 1)^2 \\
\text{Damped} &: U_{12} = (r'_{12} - 1)^2 + \mathbf{r}_1 \cdot \dot{\mathbf{r}}_1 / n \\
\text{Charge} &: U_{12} = q_1 q_2 / r'_{12} \\
\text{Dicontinuous} &: U_{12} = \begin{cases} 0, & r'_{12} < 2 \\ (r'_{12} - 1)^2, & r'_{12} \geq 2 \end{cases}
\end{aligned}
$$

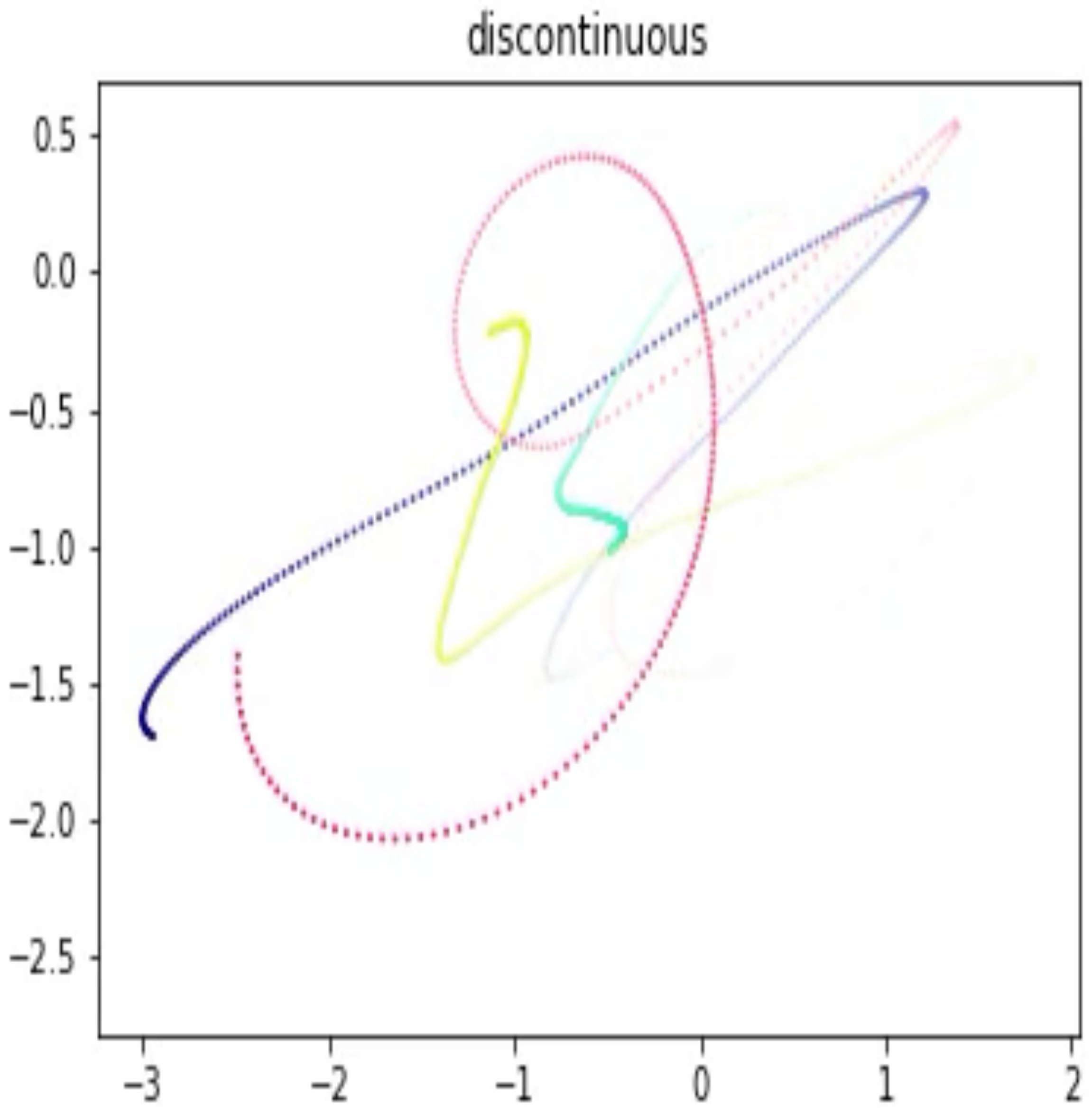






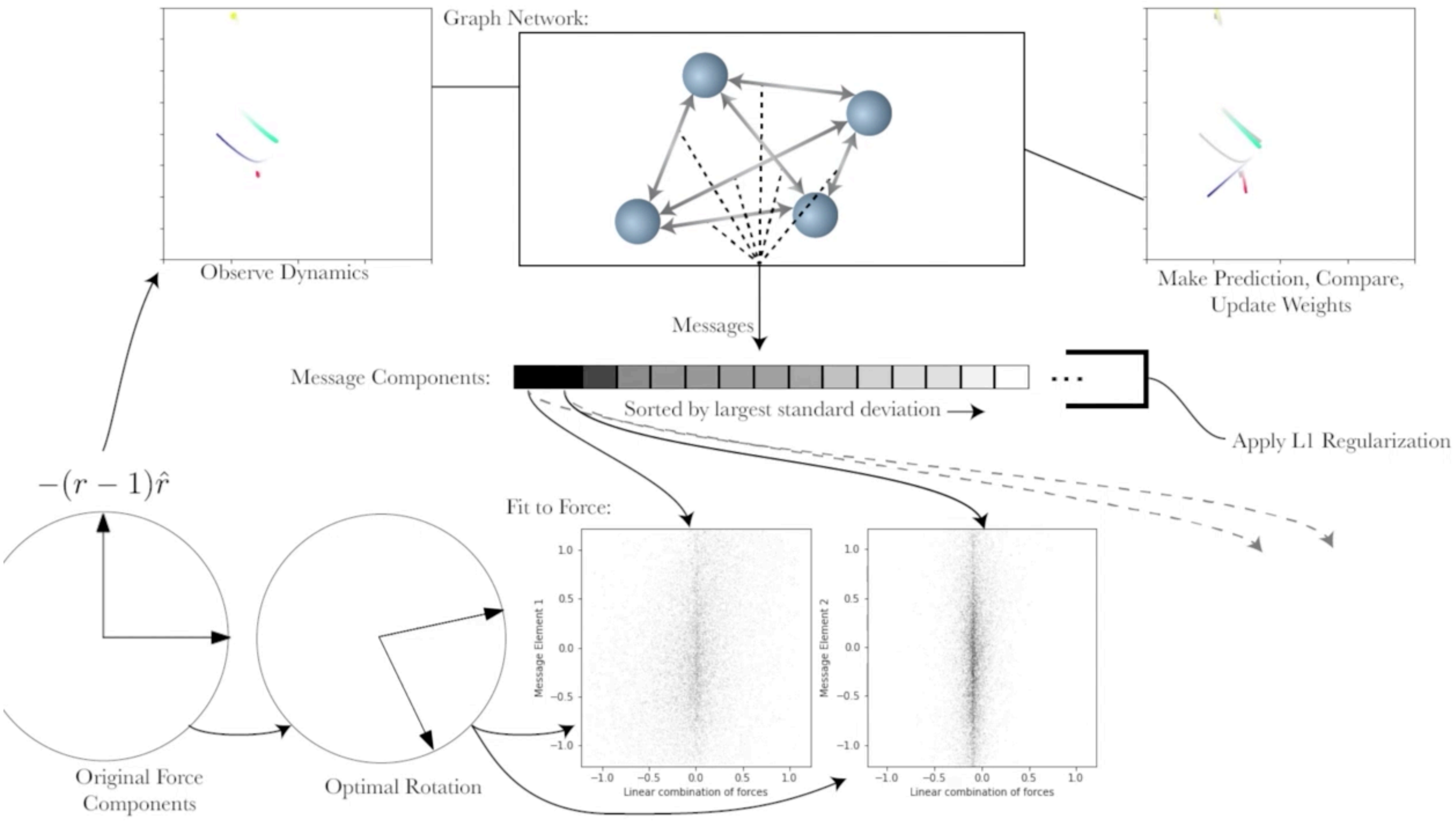




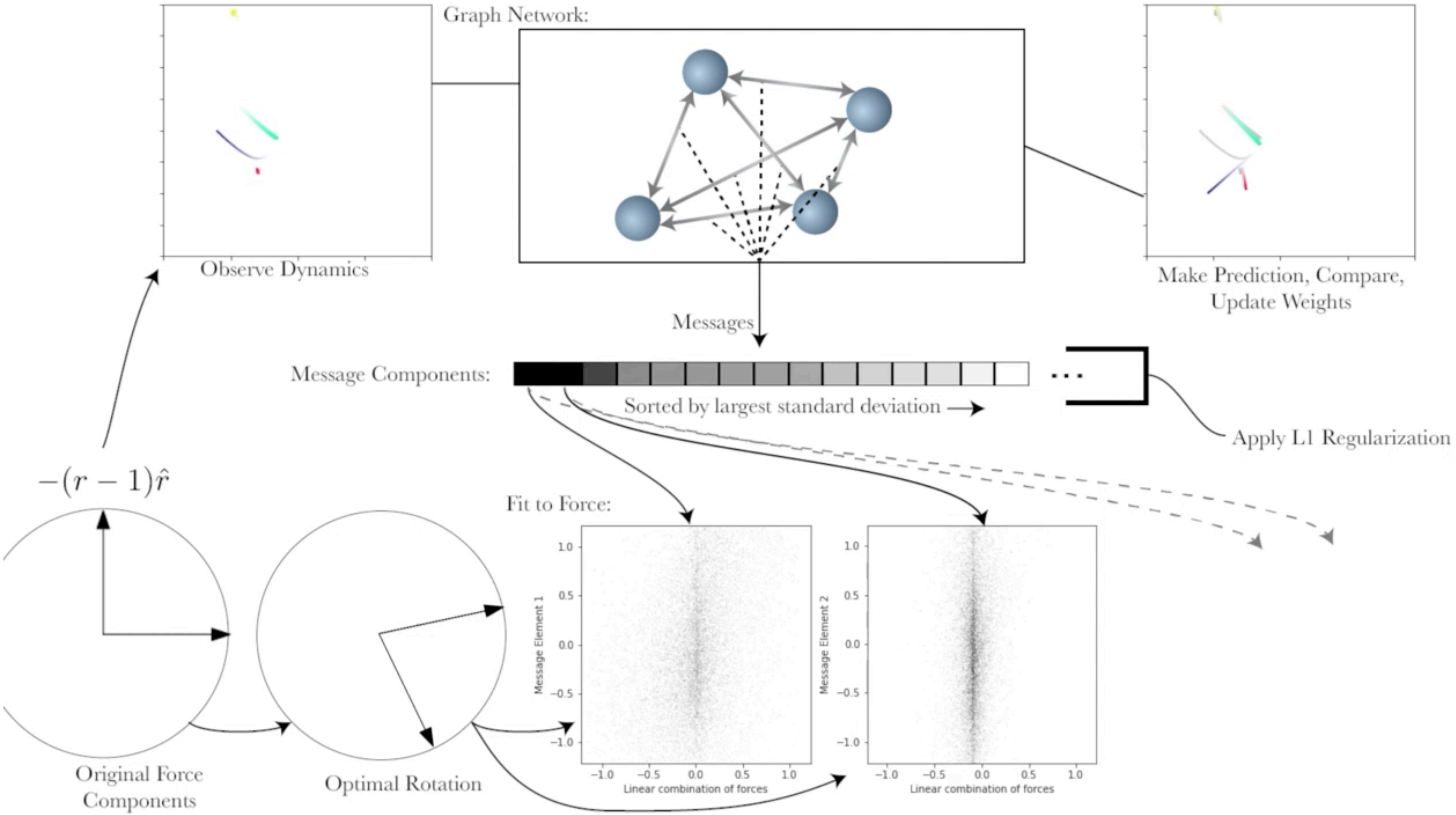


Symbolic Regression

M. Cranmer, et. al. arxiv:2006.11287

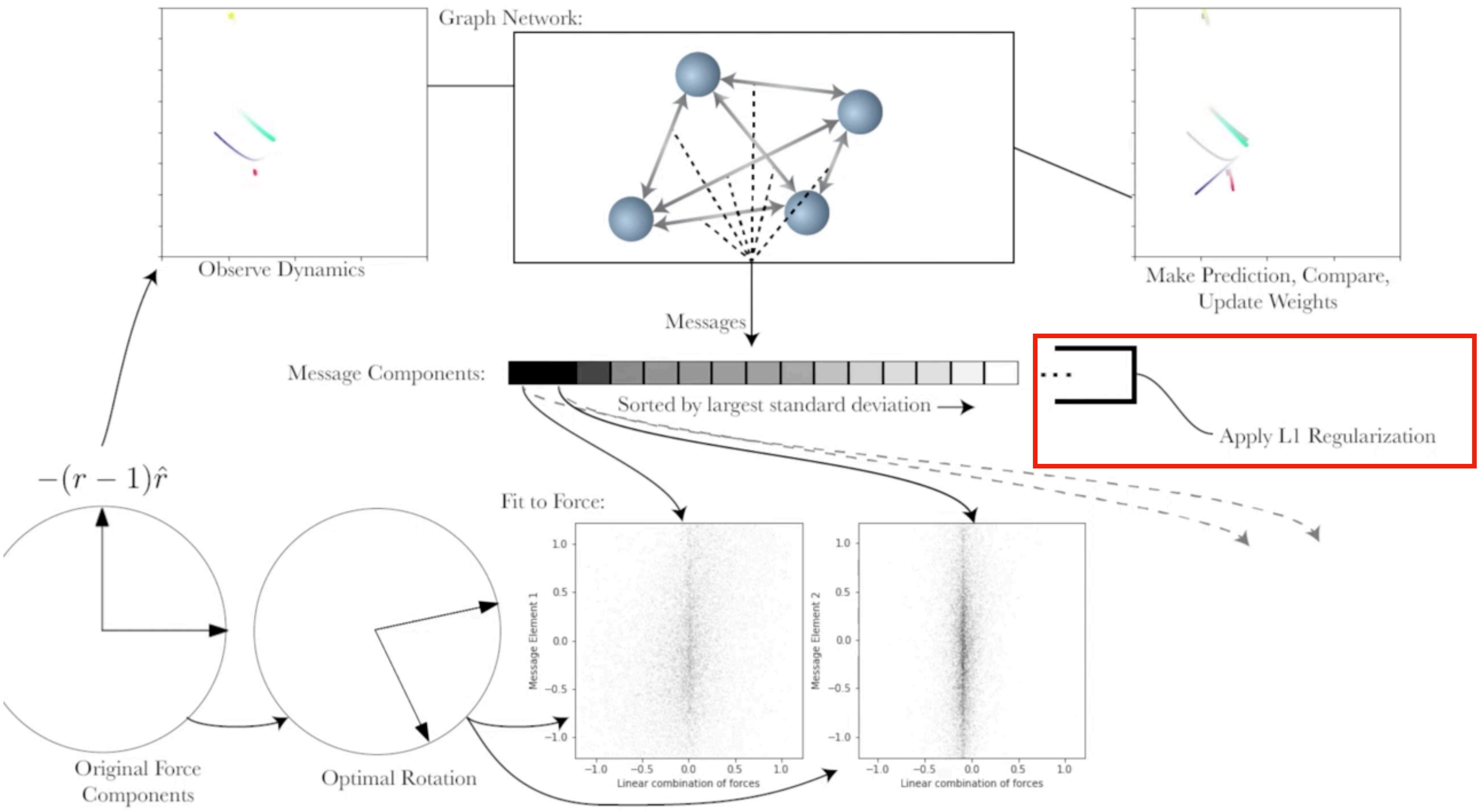


Symbolic Regression



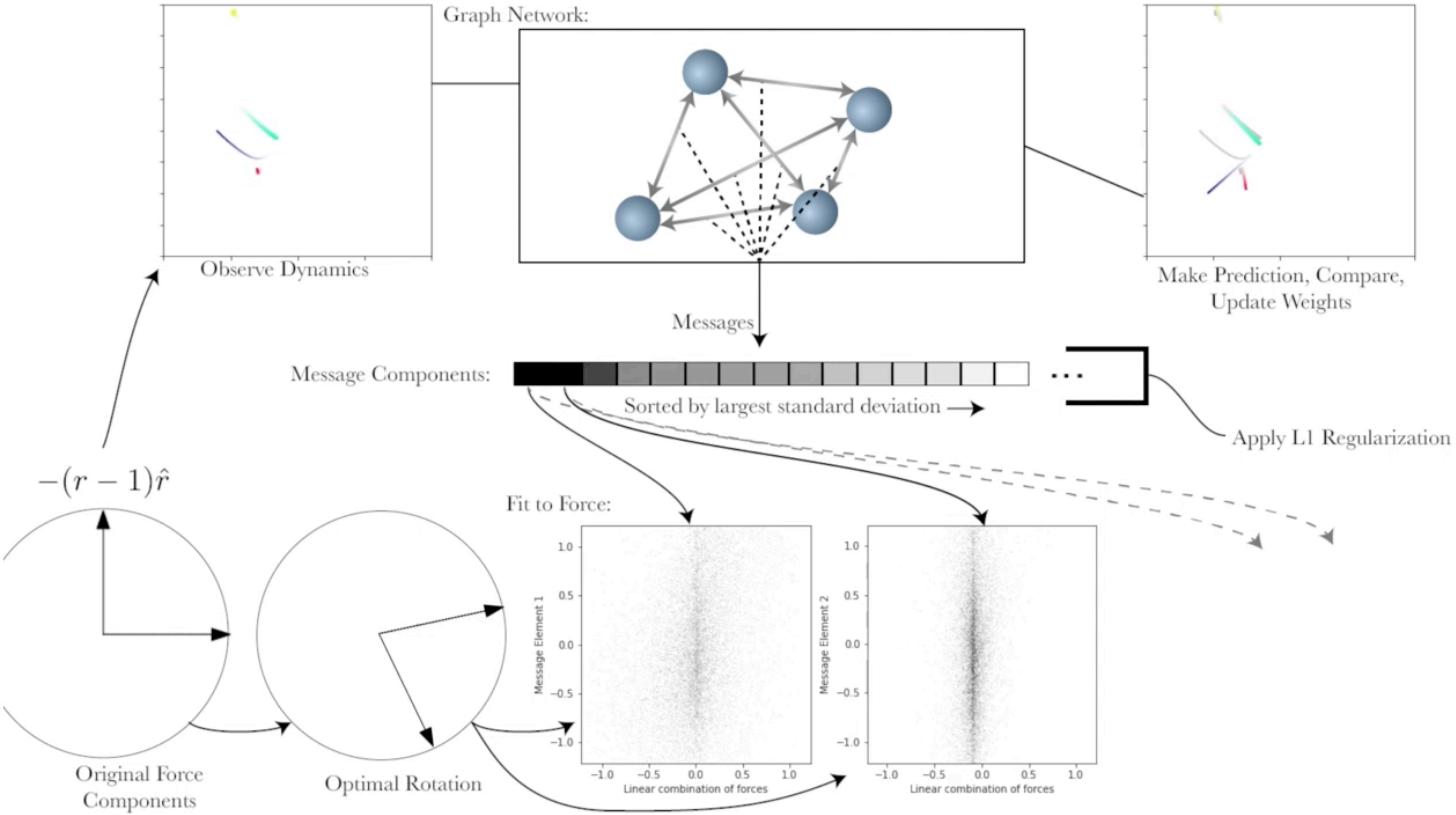
Symbolic Regression

M. Cranmer, et. al. arxiv:2006.11287



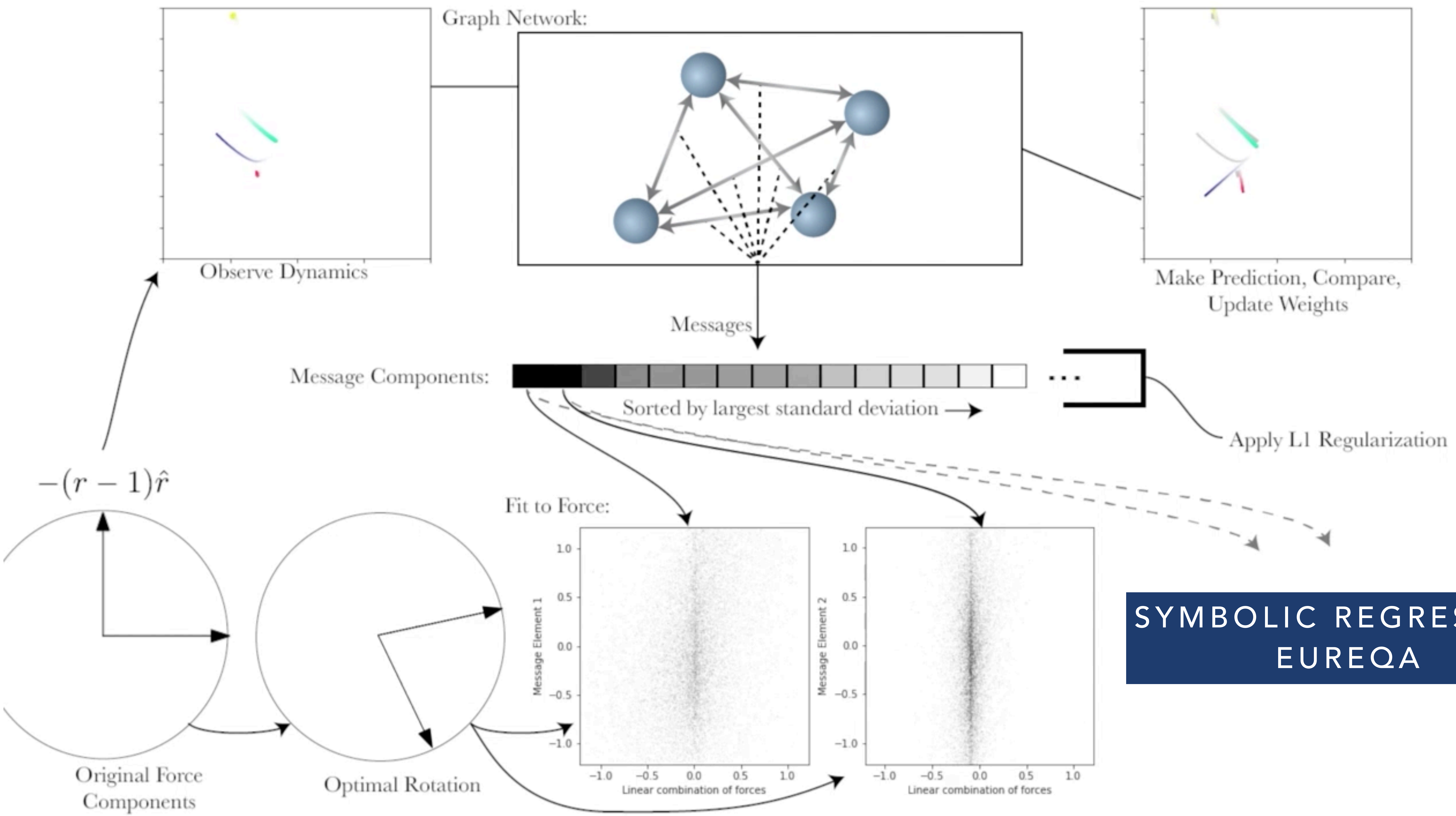
Symbolic Regression

M. Cranmer, et. al. arxiv:2006.11287



Symbolic Regression

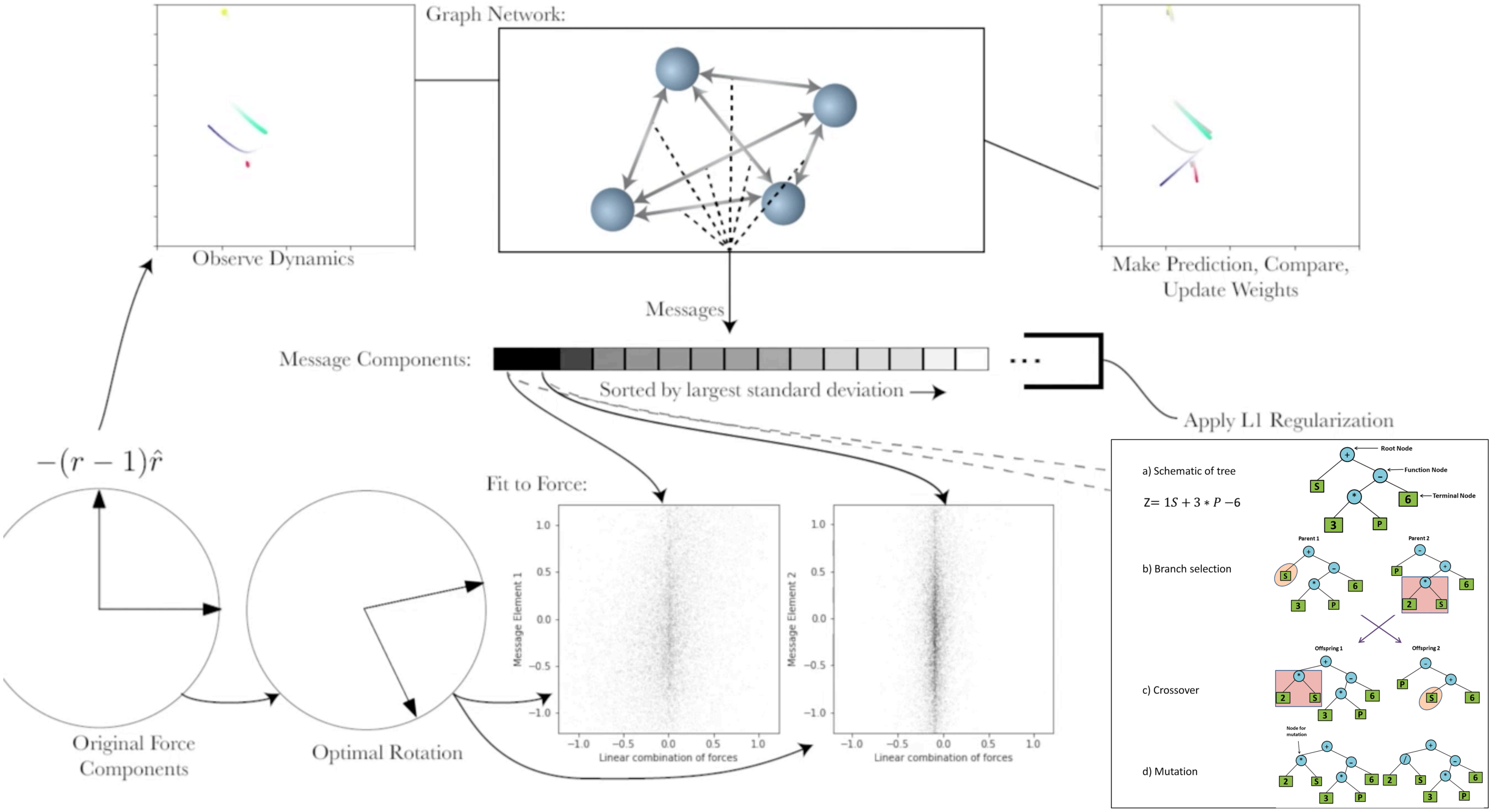
M. Cranmer, et. al. arxiv:2006.11287



SYMBOLIC REGRESSION EUREQA

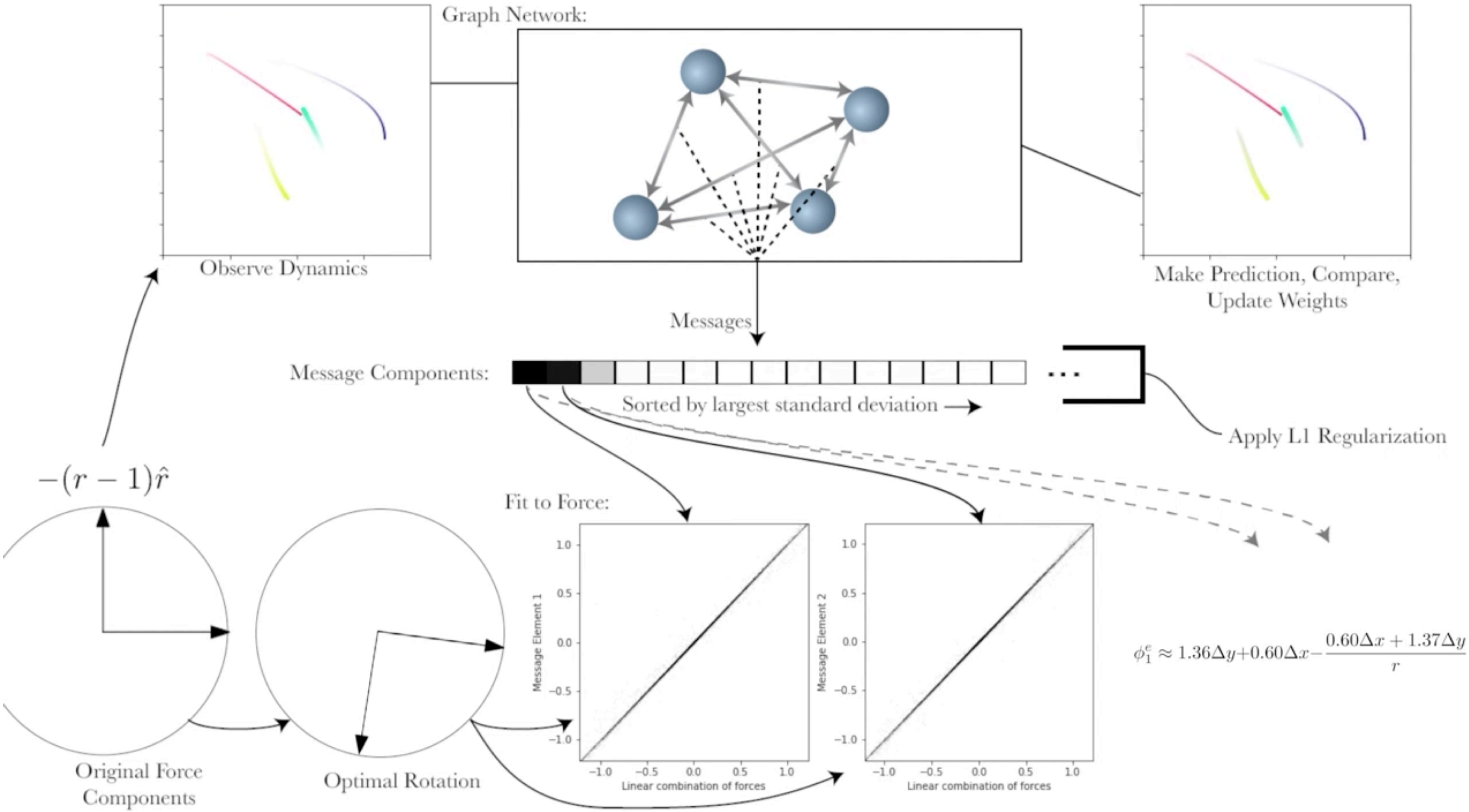
Symbolic Regression

M. Cranmer, et. al. arxiv:2006.11287



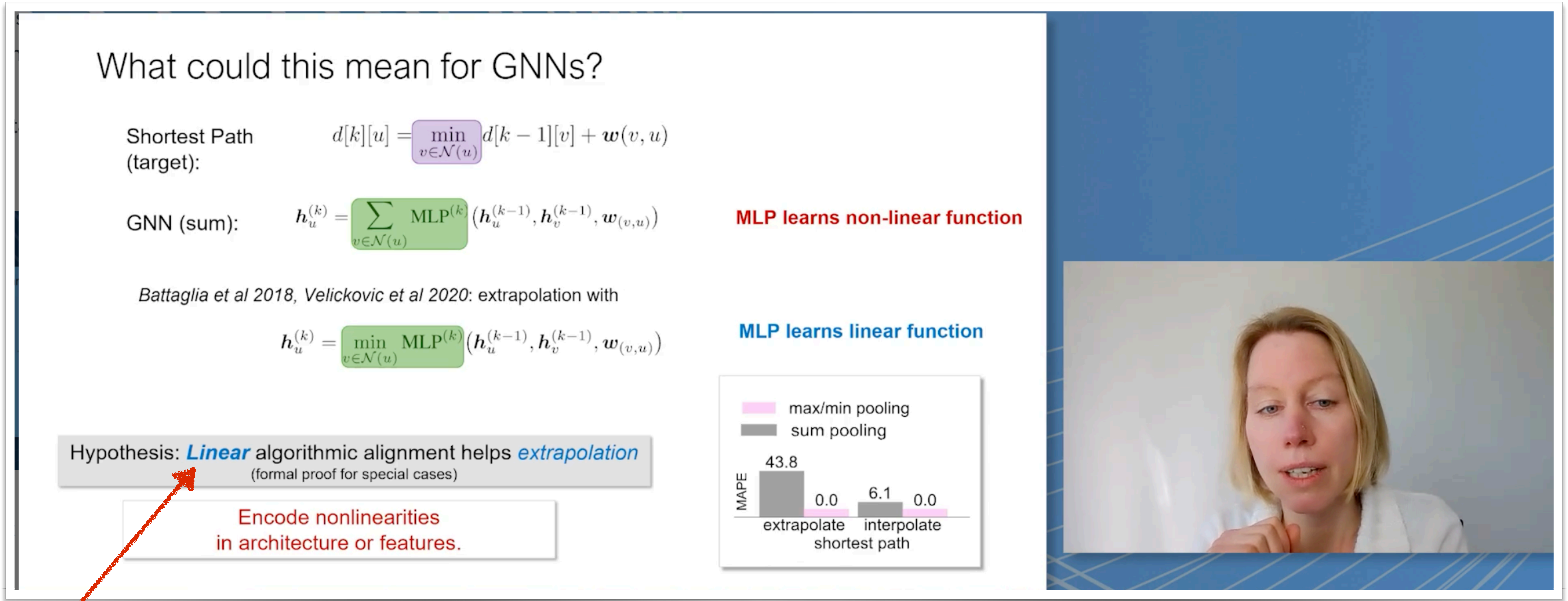
Symbolic Regression

M. Cranmer, et. al. arxiv:2006.11287



Algorithmic Alignment

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021

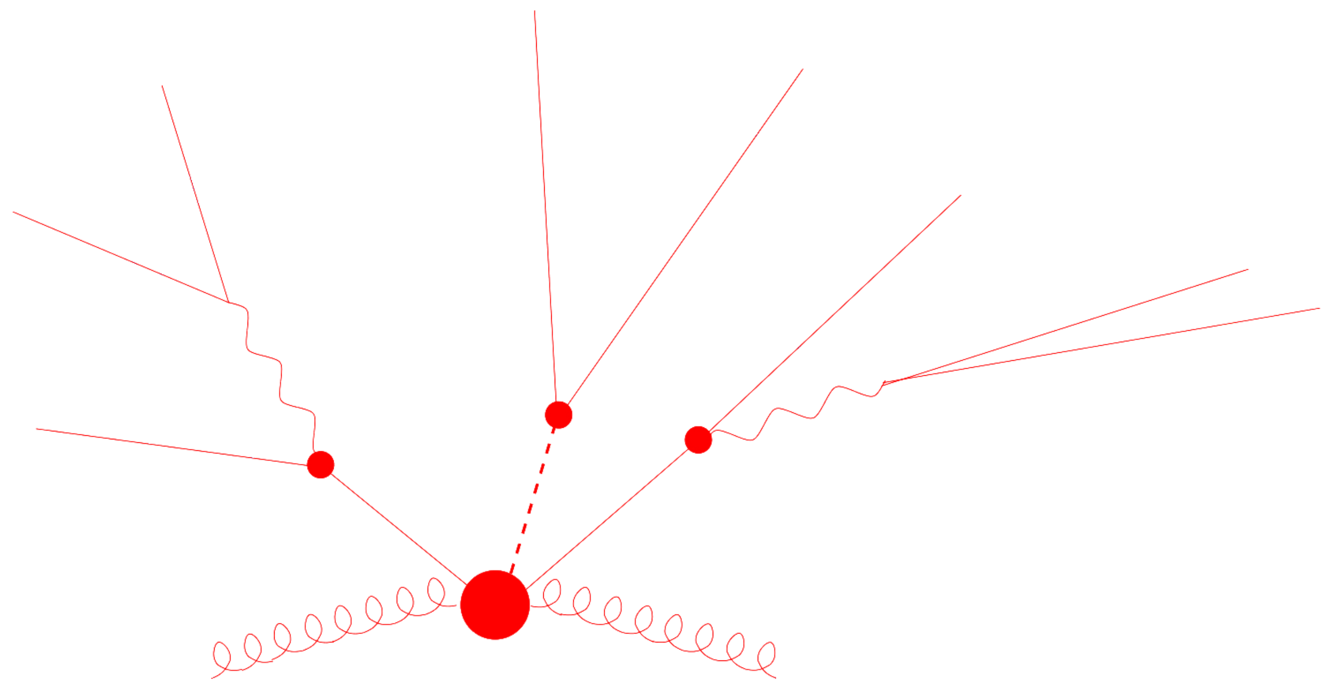


Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

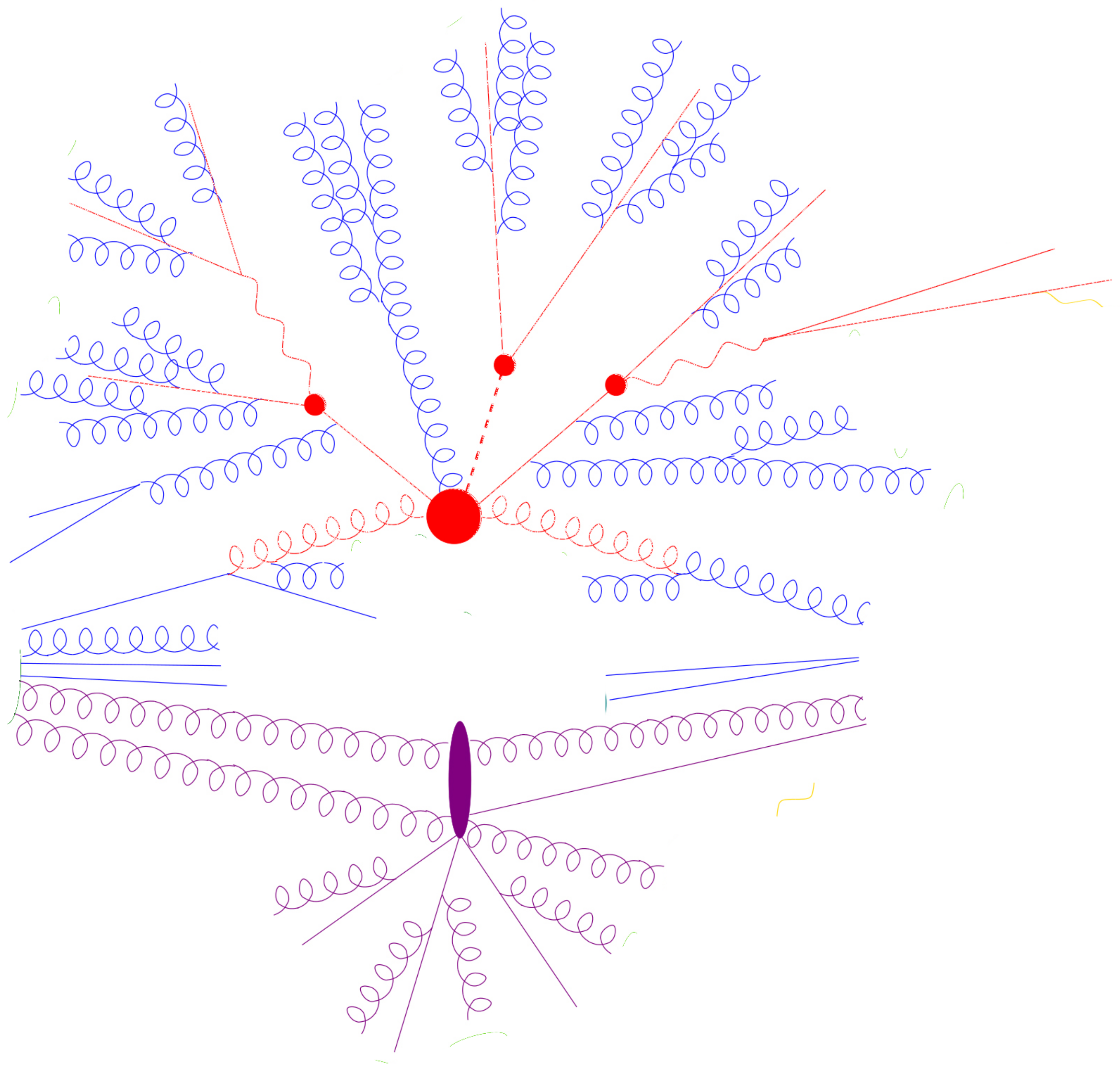

JETS
Run: 329716
Event: 857582452
2017-07-14 10:48:51 CEST
ATLAS
EXPERIMENT

Physicists' Model For Jets

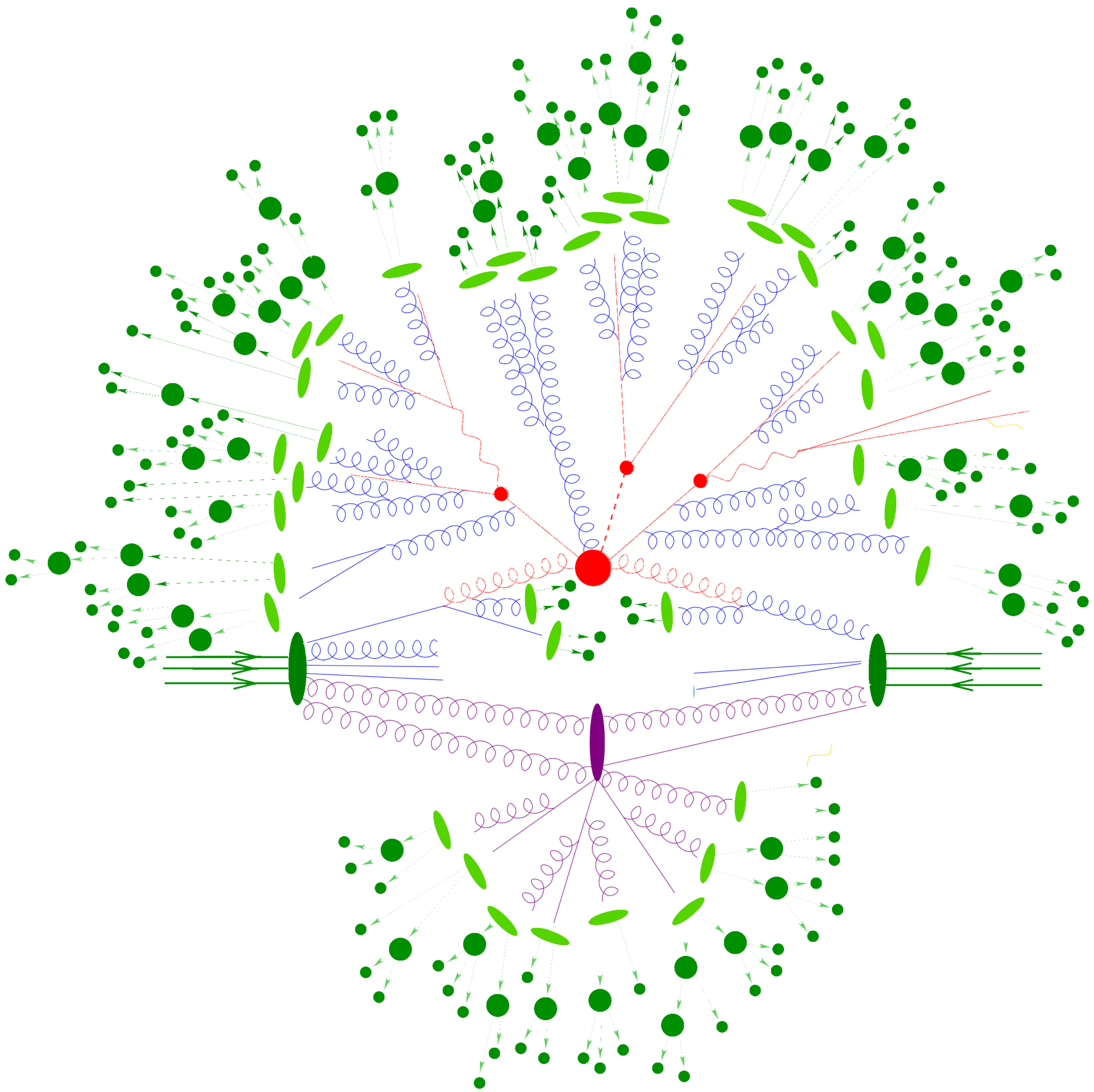
Physicists' Model For Jets



Physicists' Model For Jets

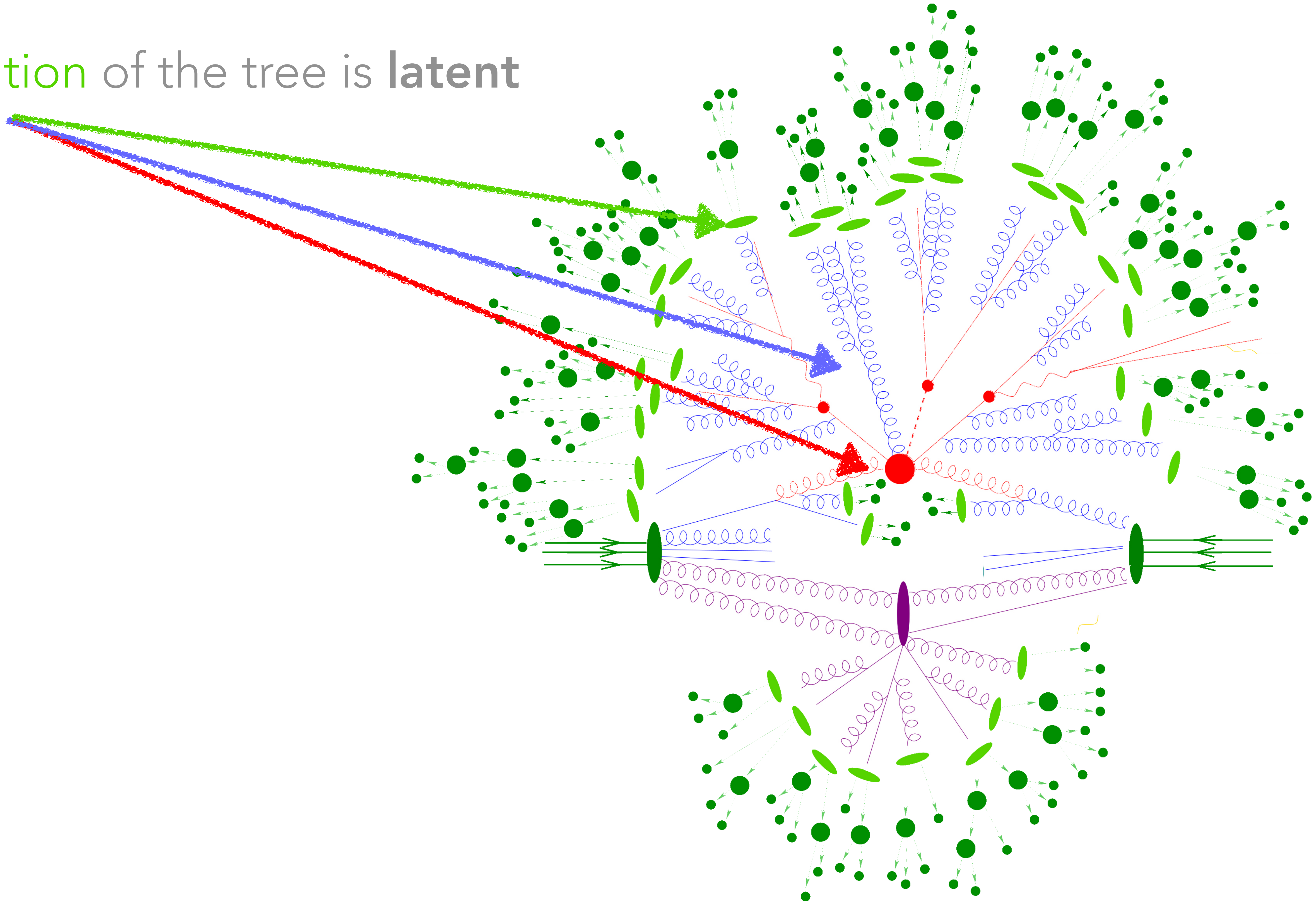


Physicists' Model For Jets



Physicists' Model For Jets

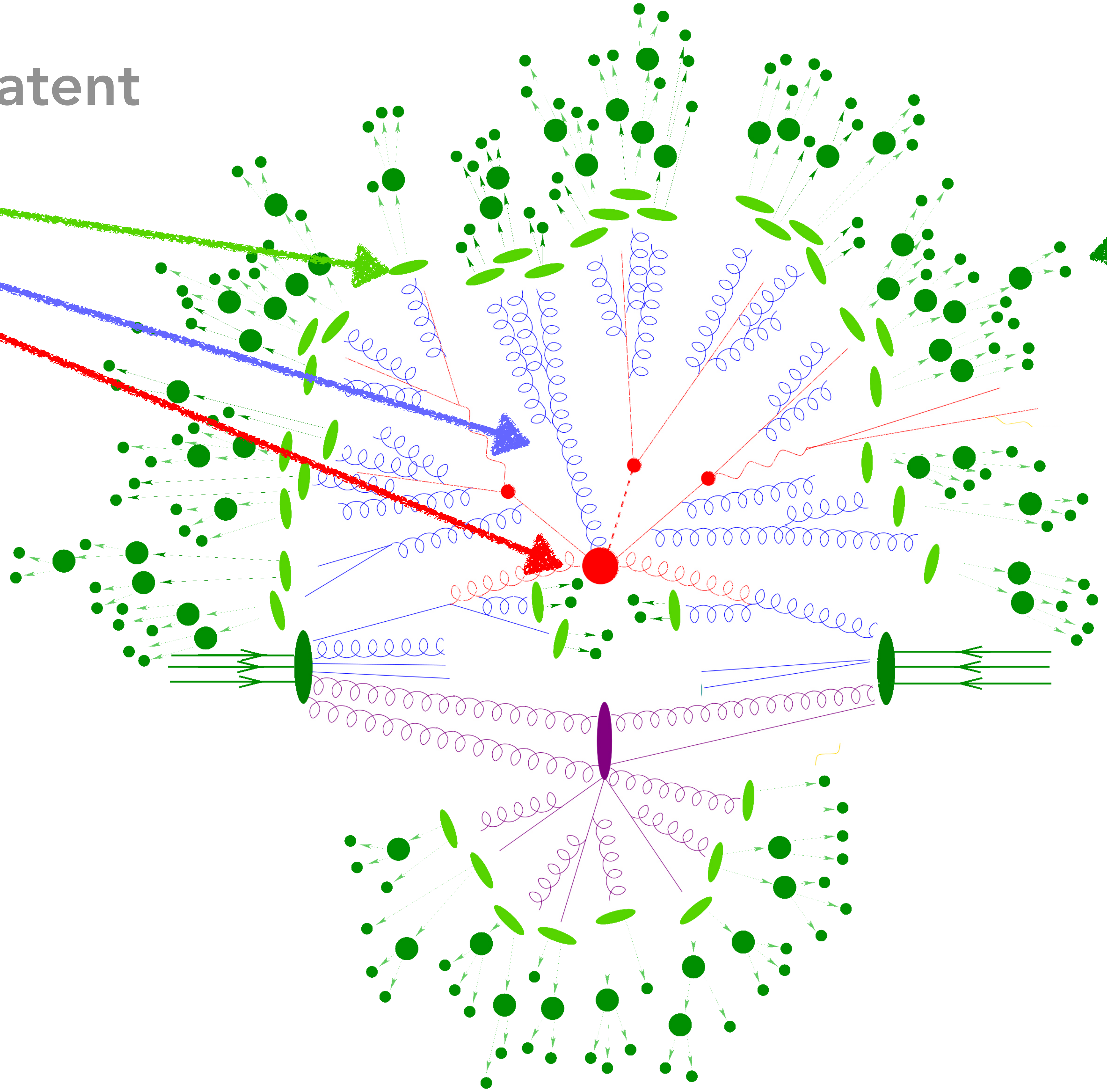
Evolution of the tree is **latent**



Physicists' Model For Jets

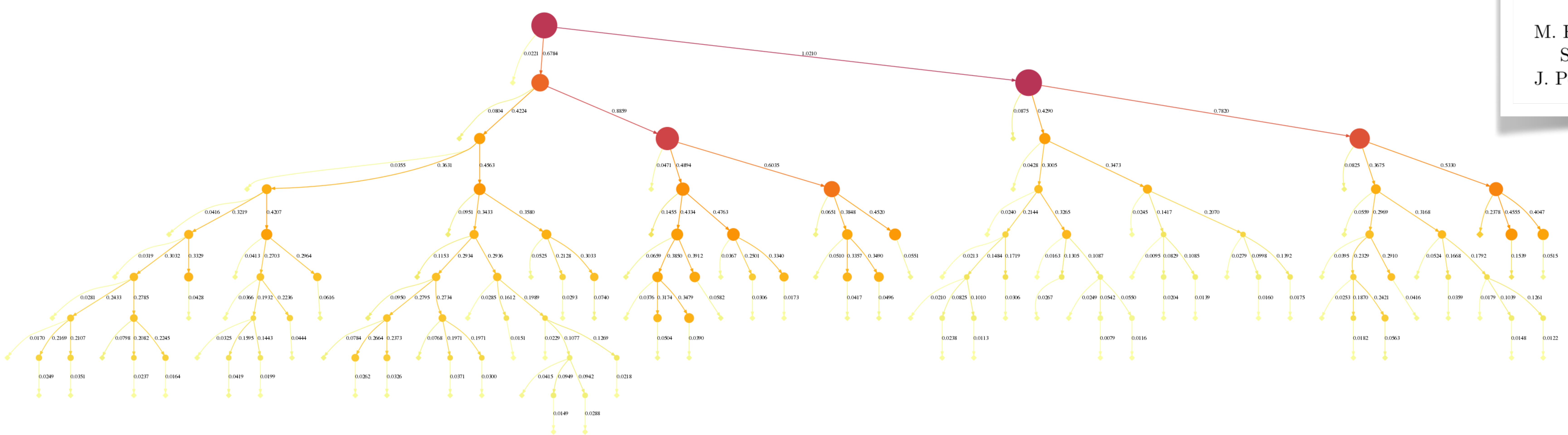
Evolution of the tree is **latent**

We only **observe** the **leaves**



QCD-inspired Recursive Neural Networks

arXiv:1902.09914



The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

- Generative process approximately stationary Markov Process producing tree of decays
- algorithms exist to estimate the latent tree
- TreeRNN performs well on binary classification task, has many fewer parameters, and needs much less data to train!

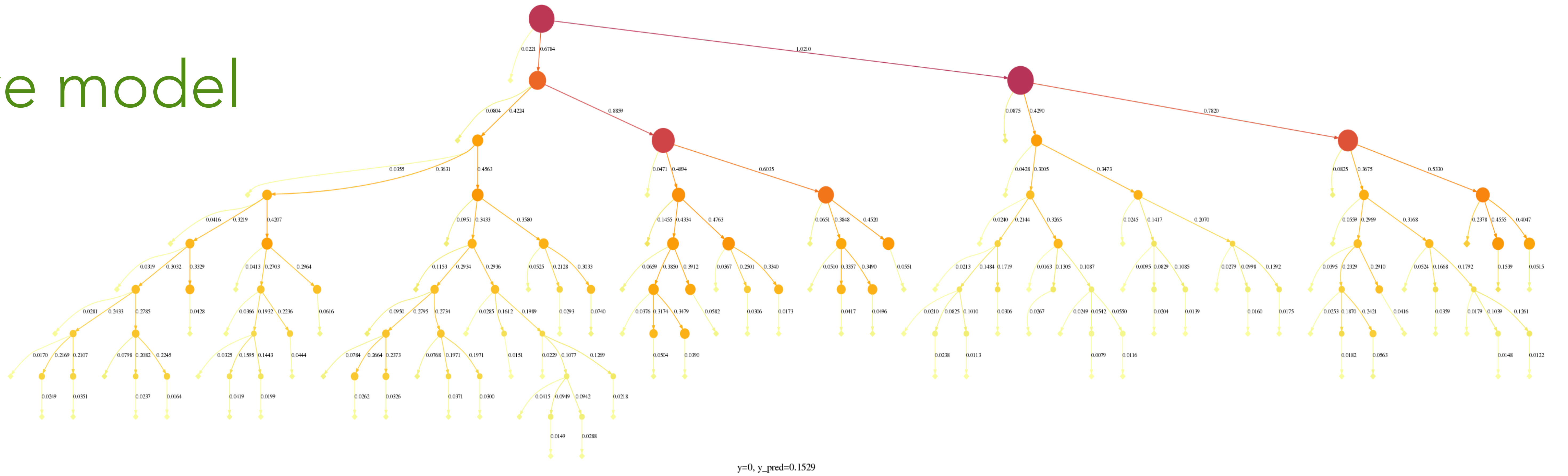
	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	802±12	783±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	926±20	886±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1209±23	1167±24	34k
P-CNN	0.980	0.930	732±24	838±13	841±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1383±45	1374±41	498k
LBN [19]	0.981	0.931	836±17	852±67	971±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	751±11	127k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	734±13	729±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1005±21	1005±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

Louppe, Cho, Becot, KC [arXiv:1702.00748]

Alignment for generative model

Andreassen, Feige, Frye, Schwartz arXiv:1804.09720

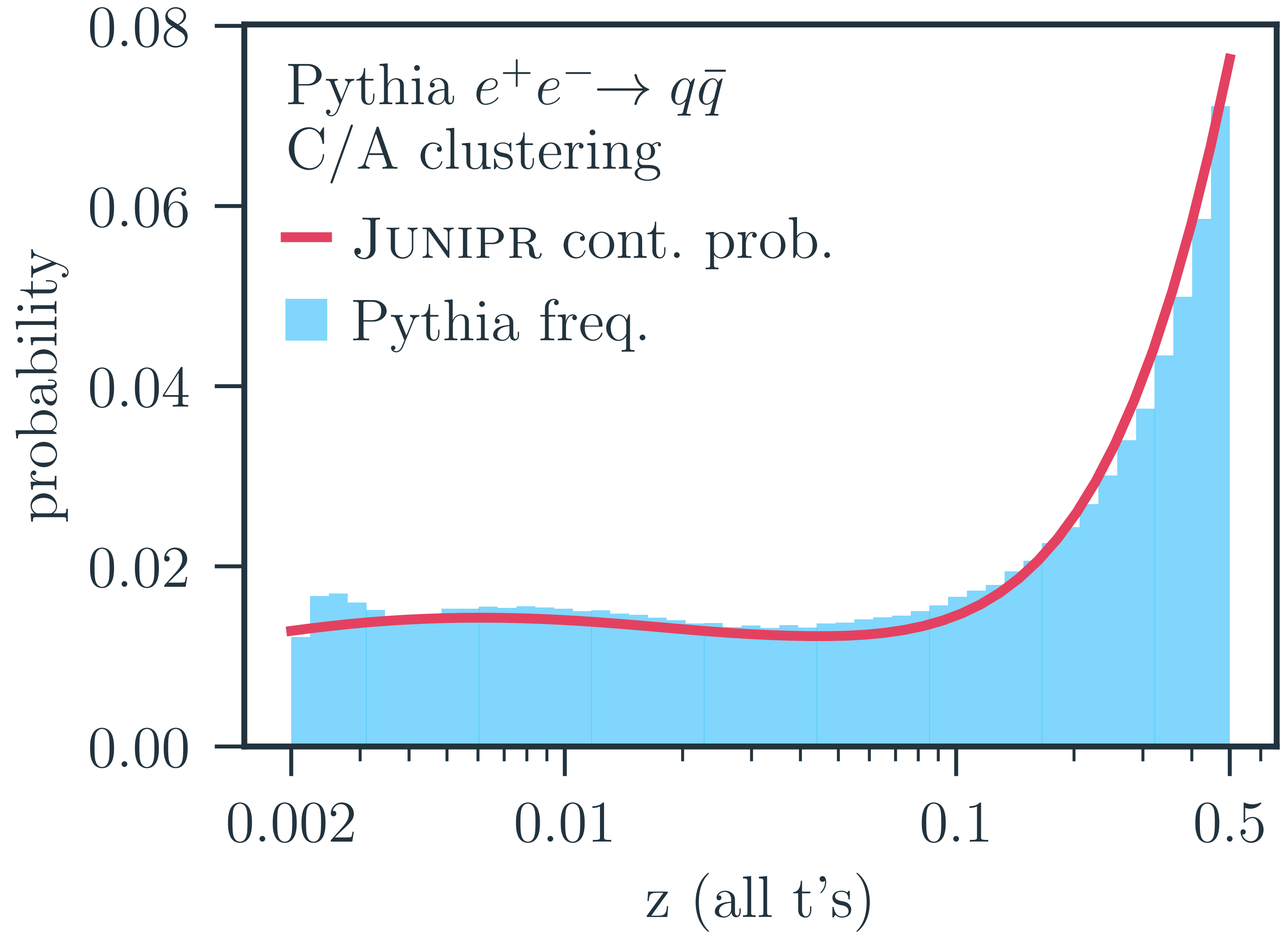
See also GInkgo by Sebastian Macaluso and Duccio Pappadopulo, and KC
https://github.com/SebastianMacaluso/ToyJetsShower



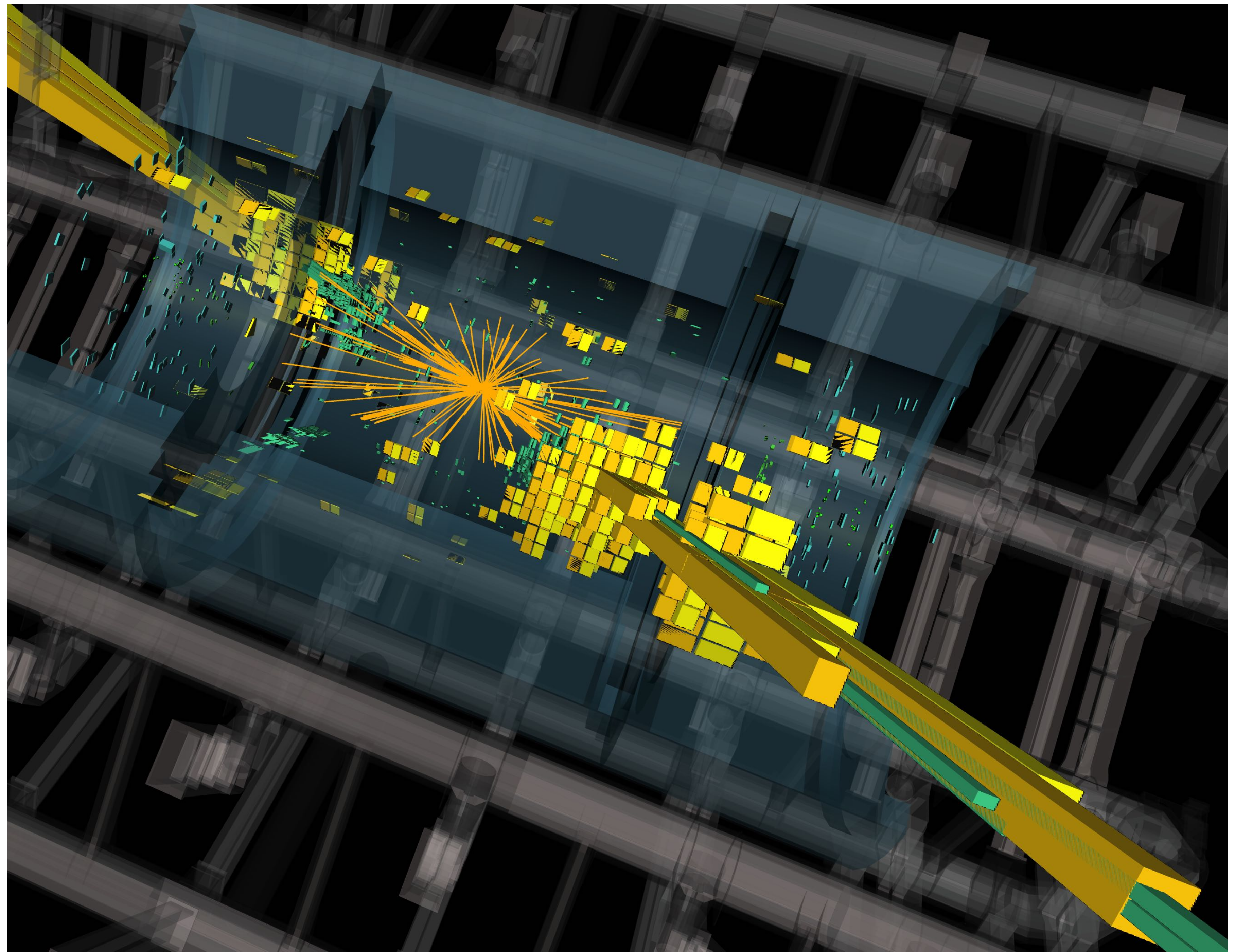
JUNIPR is an autoregressive generative model for jets with a tractable likelihood

$$P_{\text{jet}}(\{p_1,\dots p_n\}) = \left[\prod_{t=1}^{n-1} P_t\big(k_1^{(t+1)},\dots,k_{t+1}^{(t+1)}\big|k_1^{(t)},\dots,k_t^{(t)}\big)\right] \times P_n\big(\text{end}\big|k_1^{(n)},\dots,k_n^{(n)}\big).$$

Autoregressive structure matches causal structure in traditional physics simulators



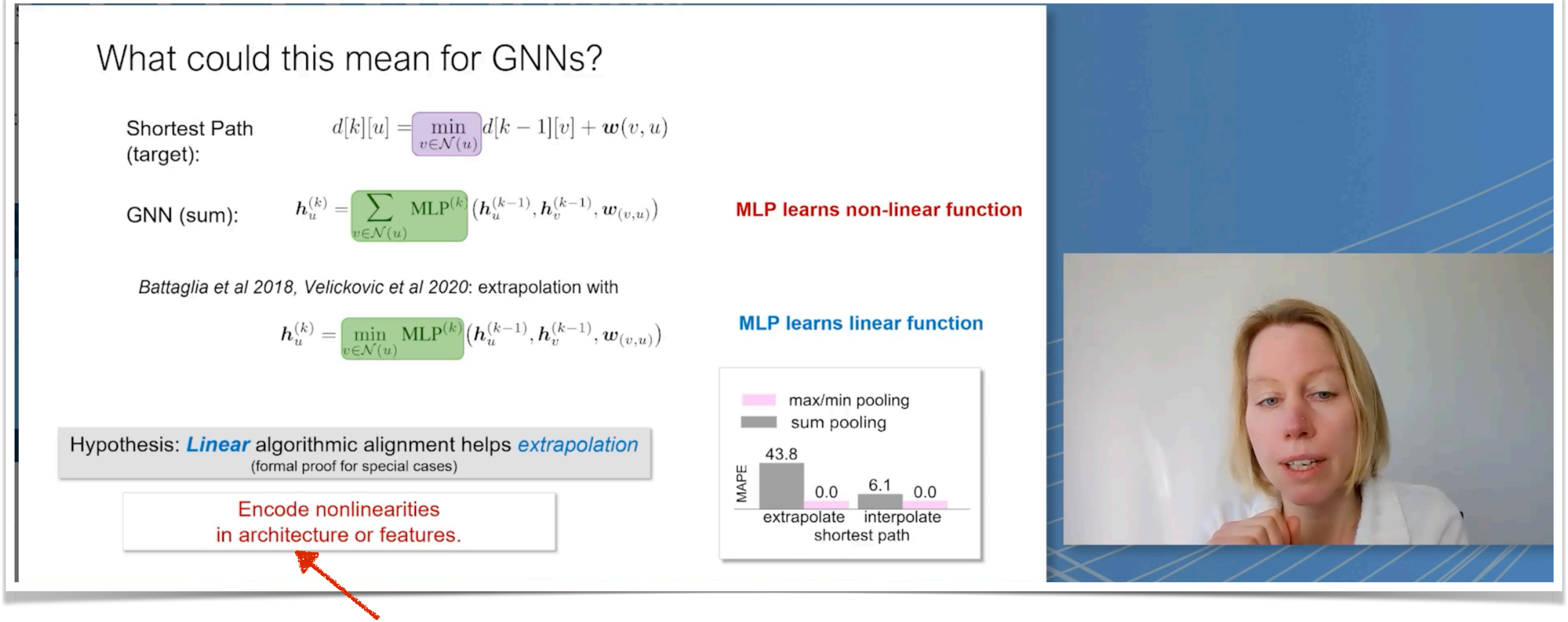
← alignment!



Latent variables are interpretable

Algorithmic Alignment

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021



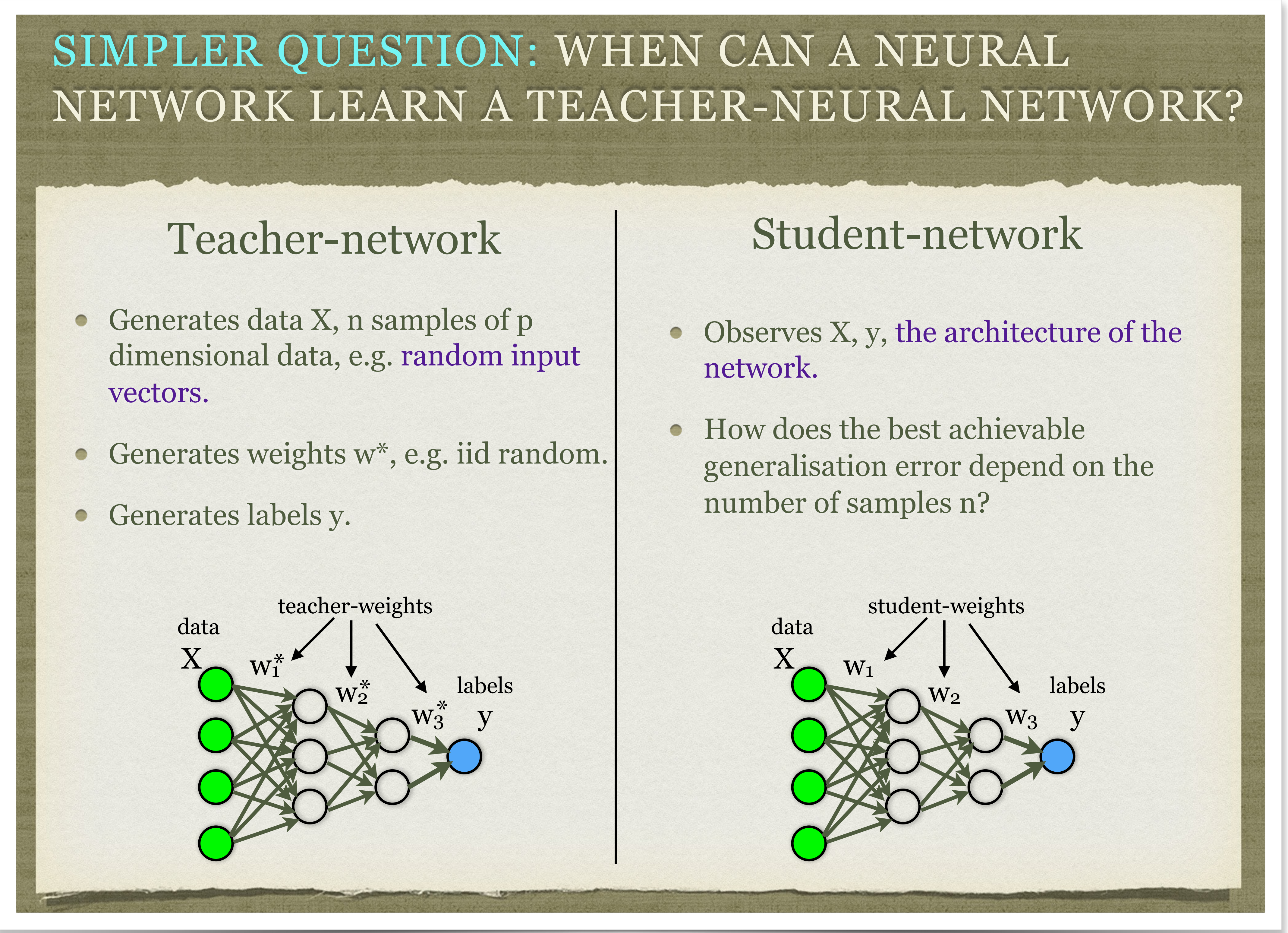
Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

Physical systems are a good testbed for AI research

Interplay of key ingredients of Deep Learning

Hard to analyze the effect of data structure for real-world data sources.

Toy models are useful!



* Lenka Zdeborova:
Talk: https://ml4physicalsciences.github.io
Position piece: https://rdcu.be/b4p1m

Interplay of key ingredients of Deep Learning

Hard to analyze the effect of data structure for real-world data sources.

Toy models are useful!

Physics ∩ ML

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

09 Sep 2020

Insights on gradient-based algorithms in high-dimensional learning

Lenka Zdeborova, EPFL, 12:00 EDT

Abstract: Gradient descent algorithms and their noisy variants, such as the Langevin dynamics or multi-pass SGD, are at the center of attention in machine learning. Yet their behaviour remains perplexing, in particular in the high-dimensional non-convex setting. In this talk, I will present several high-dimensional and (mostly) non-convex statistical learning problems in which the performance of gradient-based algorithms can be analysed down to a constant. The common point of these settings is that the data come from a probabilistic generative model leading to problems for which, in the high-dimensional limit, statistical physics provides exact closed solutions for the performance of the gradient-based algorithms. The covered settings include the spiked mixed matrix-tensor model, the perceptron or phase retrieval.

Slides and video of the talk are both available.

* Lenka Zdeborova:
Talk: https://ml4physicalsciences.github.io
Position piece: https://rdcu.be/b4p1m

Teacher → Causal, Generative Model (Simulator)

Richer set of problems can be investigated.

A profound shift

Scientific simulators are based on well-motivated mechanistic models

- However, the aggregate effect of many interactions between their low-level components leads to intractable inverse problems

The developments in machine learning have the potential to effectively bridge the microscopic - macroscopic divide & aid in these inverse problems

- They can provide effective models for macroscopic (emergent) phenomena that are tied back to the low-level microscopic (reductionist) model

Hybrid approaches align architectural components with causal mechanism

- A path to AI that can learn causal mechanism, generate hypotheses, and design future experiments

Conclusion

Physical models are highly-structured causal models:

- Insight of data generating process informs inductive bias on architecture
- There is growing empirical evidence that this inductive bias is helpful in terms of sample complexity and generalization
- Identification of semantics of physical system with network components is easy to take for granted, but has many important consequences. (vis-à-vis "alignment")

Simulators for physical systems provide families of problems that can be scaled to a complexity beyond the reach of current ML systems.

- They provide controlled experiments
- Some of problems have known solutions or strong baselines that aid theoretical and experimental analysis
- Ideal to probe interplay between architecture, data, & algorithms.

An IPAM workshop

Highly recommended:

http://www.ipam.ucla.edu/programs/workshops/deep-learning-and-combinatorial-optimization/

Deep Learning and Combinatorial Optimization

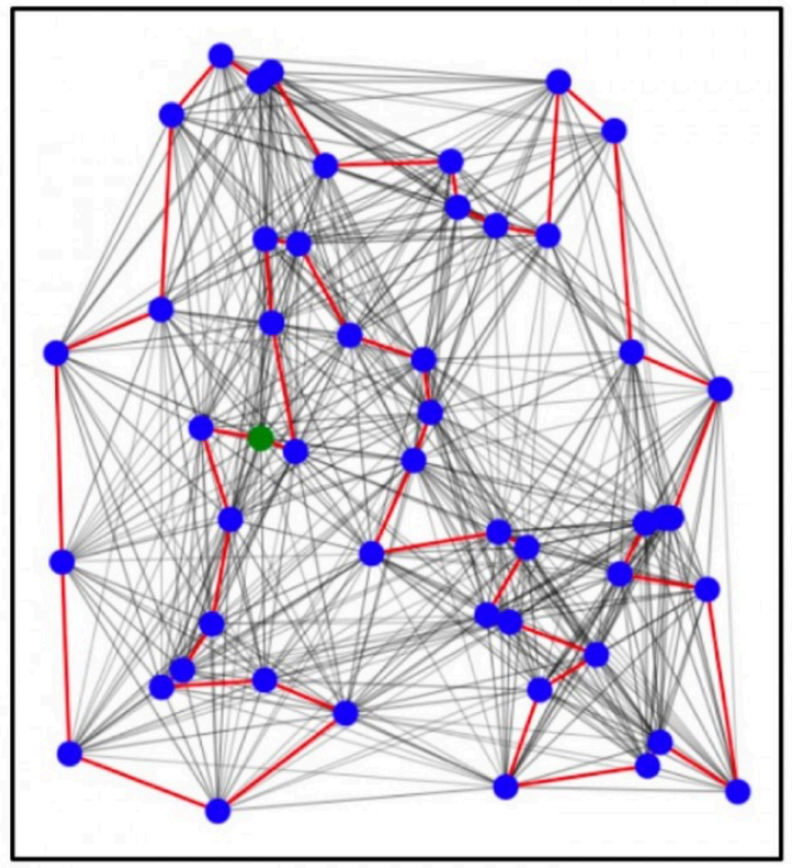
FEBRUARY 22 - 25, 2021

OVERVIEW | SPEAKER LIST | SCHEDULE

Overview

Virtual Workshop: In response to COVID-19, all participants will attend this workshop virtually via Zoom. Workshop registrants will receive the Zoom link a few days prior to the workshop, along with instructions on how to participate. The video of the recorded sessions will be made available on IPAM **website**.

Workshop Overview: In recent years, deep learning has significantly improved the fields of computer vision, natural language processing and speech recognition. Beyond these traditional fields, deep learning has been expended to quantum chemistry, physics, neuroscience, and more recently to combinatorial optimization (CO). Well-known CO problems are Travelling Salesman Problem, assignment problems, routing, planning, Bayesian search, and scheduling. CO is basically used every day in finance and revenue management, transportation, manufacturing, supply chain, public policy, hardware design, computing and information technology.



Most combinatorial problems are difficult to solve, often leading to heuristic solutions which require years of research work and significant specialized knowledge. For example, the famous TSP problem has been studied for more than 80 years, and the best solver leverages 30 years of theoretical developments, data structures and heuristics from computer science. In the last few years, deep learning has developed some preliminary but promising approaches to deal with classical CO problems such as TSP, MaxCut, Minimum Vertex Cover, Knapsack, Quadratic Assignment Problem and Vehicle Routing Problems. DL is particularly attractive to address CO problems given its high flexibility, approximate nature, and self-learning paradigm. In other words, DL has the potential to learn universal

Further insight

If you would like to know more details about constructing good processor networks:

https://www.youtube.com/watch?v=IPQ6CPoluok

https://drive.google.com/file/d/1_EQ9Yu7VEkvrHaVHl_WbT5ABvxrSNY-s/view?usp=sharing

Want to know more?

Combinatorial optimization and reasoning with graph neural networks

Quentin Cappart[1], Didier Chételat[2], Elias Khalil[3], Andrea Lodi[2], Christopher Morris[2], and Petar Veličković[*4]

[1]Department of Computer Engineering and Software Engineering, Polytechnique Montréal
[2]CERC in Data Science for Real-Time Decision-Making, Polytechnique Montréal
[3]Department of Mechanical & Industrial Engineering, University of Toronto
[4]DeepMind

Combinatorial optimization is a well-established area in operations research and computer science. Until recently, its methods have focused on solving problem instances in isolation, ignoring the fact that they often stem from related data distributions in practice. However, recent years have seen a surge of interest in using machine learning, especially graph neural networks (GNNs), as a key building block for combinatorial tasks, either as solvers or as helper functions. GNNs are an inductive bias that effectively encodes combinatorial and relational input due to their permutation-invariance and sparsity awareness. This paper presents a conceptual review of recent key advancements in this emerging field, aiming at both the optimization and machine learning researcher.

Our *43-page* survey on GNNs for CO!

https://arxiv.org/abs/2102.09544

Section 3.3. details algorithmic reasoning, with comprehensive references.