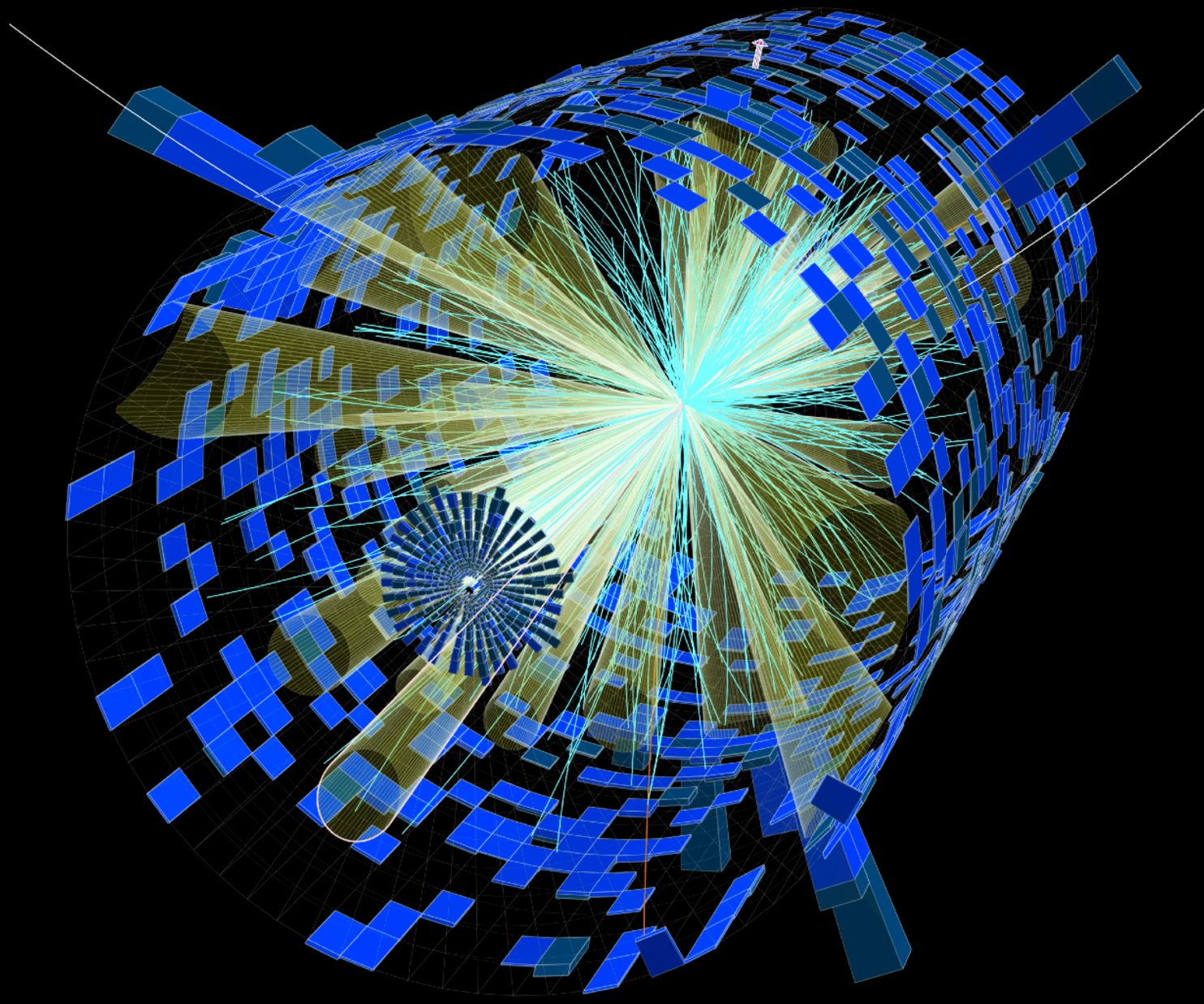
NYU Center for Data Science

Center for Cosmology and Particle Physics

THE INTERPLAY BETWEEN

PHYSICAL SIMULATIONS AND MACHINE LEARNING

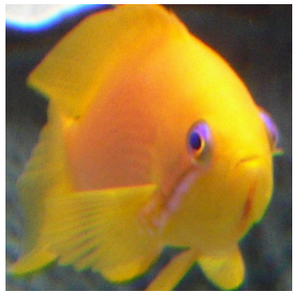


@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

SUPPORT

The SCAILFIN Project

scailfin.github.io



COLLABORATORS (+ MANY MORE)

Gilles Louppe
U. Liège

Kyunghyun Cho

Joan Bruna

Brenden Lake

Meghan Frate

Juan Pavez

Tilman Plehn

Johann Brehmer

Felix Kling

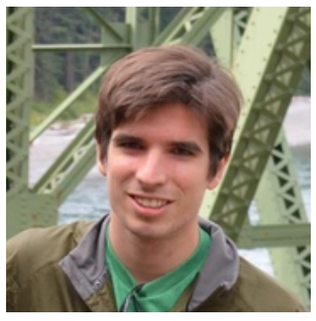
Lukas Heinrich

Markus Store

Tim Head

Michael Kagan

Irina Espejo



Peter Sadowski

Daniel Whiteson

Pierre Baldi

Lezcano Casado

Atılım Güneş Baydin
University of Oxford

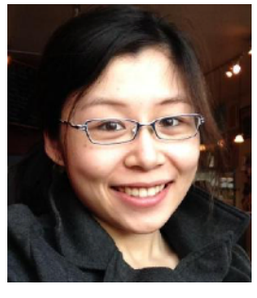
Prabhat
NERSC, Berkeley Lab

Wahid Bhimji
NERSC, Berkeley Lab

Frank Wood
University of Oxford

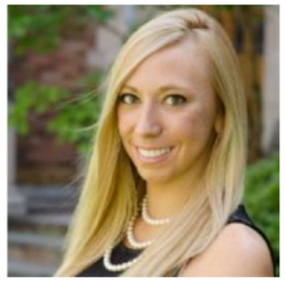
Phiala Shanahan

William Detmold



Karen Ng

Tuan Anh Le



Michela Paganini
Yale University

Daniela Huppenkothen
New York University

Savannah Thais
Yale University

Ruth Angus
Columbia University

Machine Learning and the Physical Sciences

Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS)

December 13 or 14, 2019

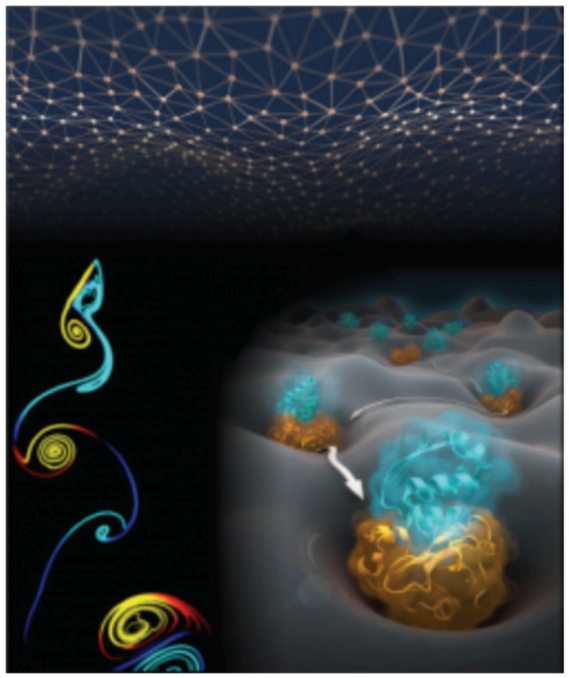
ml4physicalsciences.github.io

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.

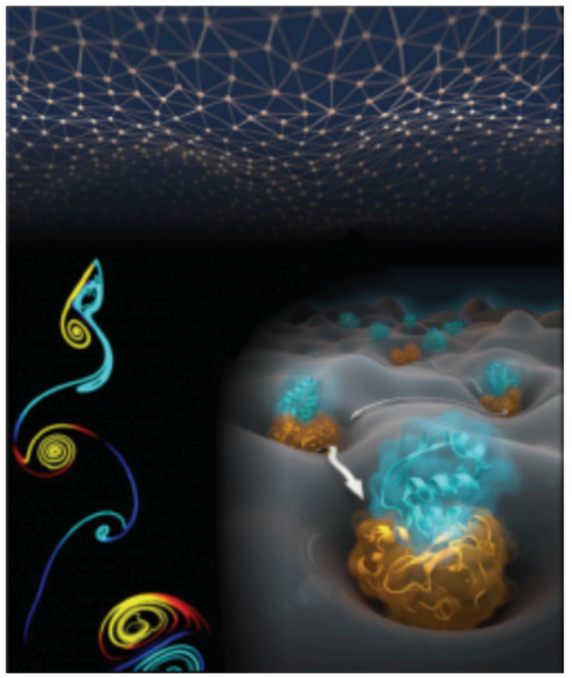
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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

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.

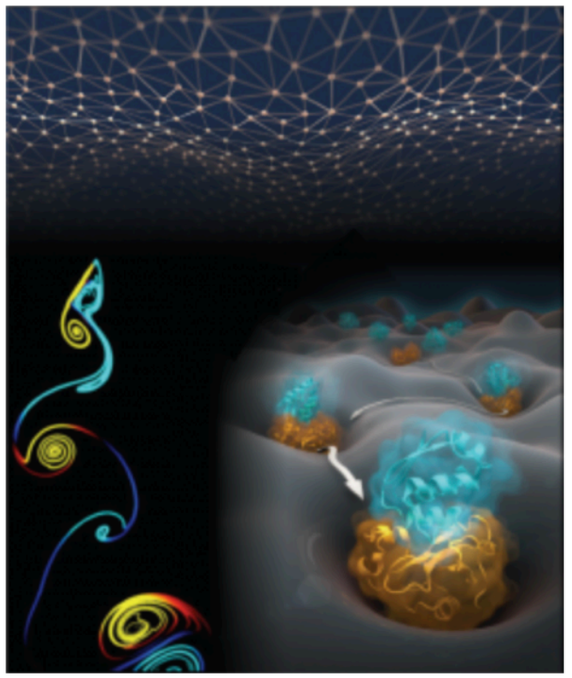
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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.

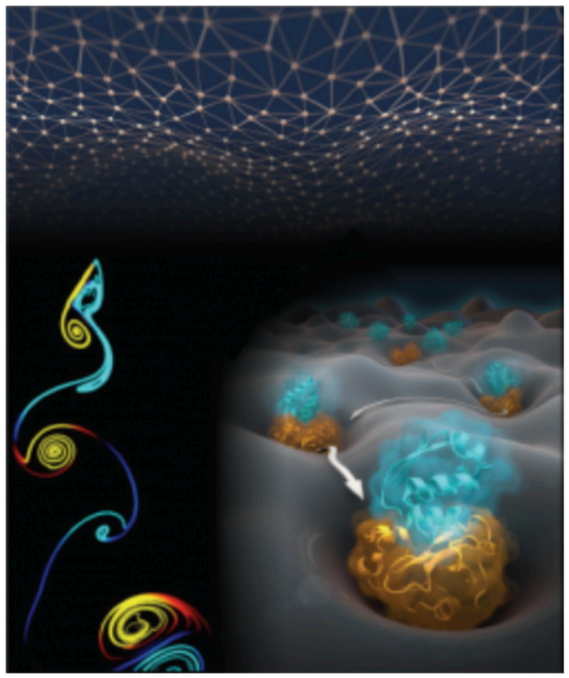
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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

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.

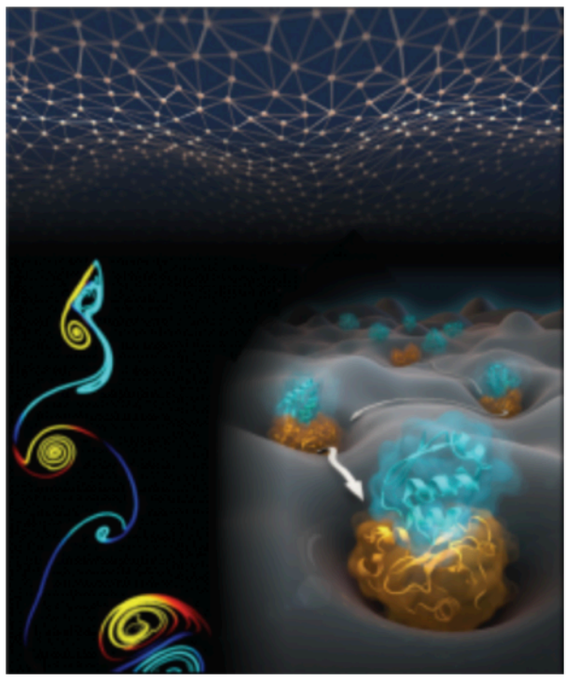
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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.

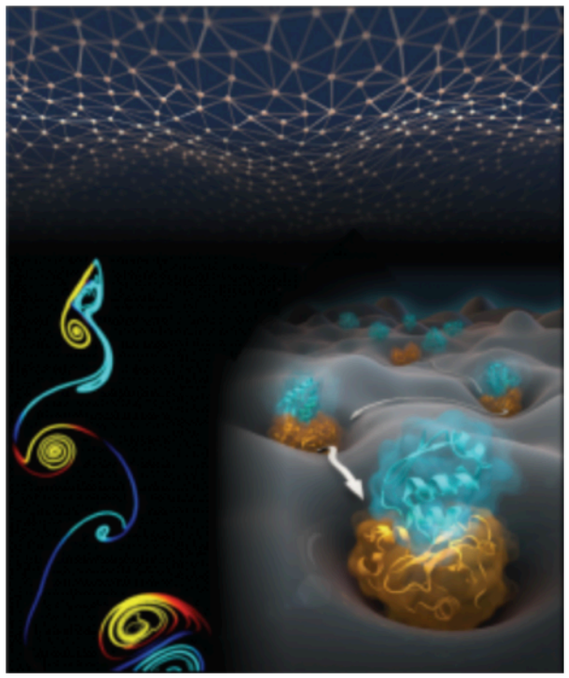
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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

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.



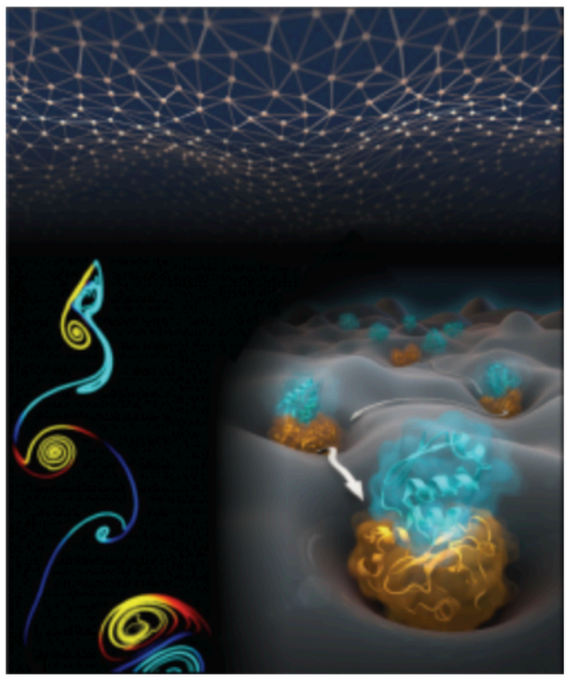
As yet, most applications of machine learning to physical sciences have been limited to the “low-hanging fruits,” as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity 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.

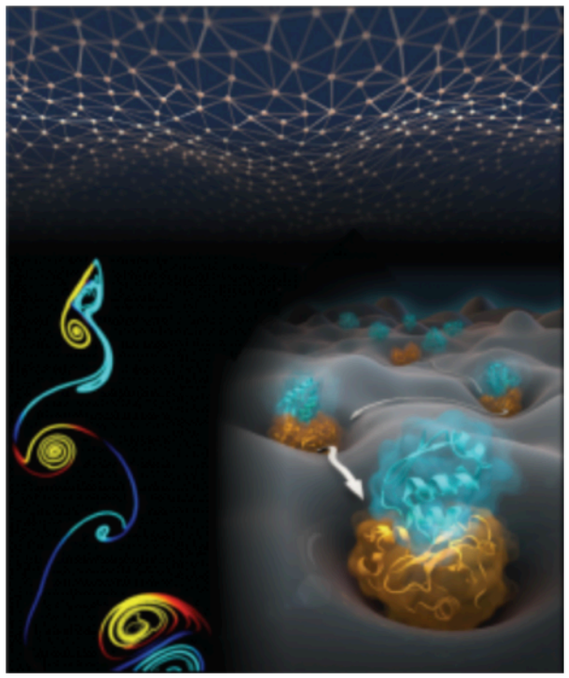
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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

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.

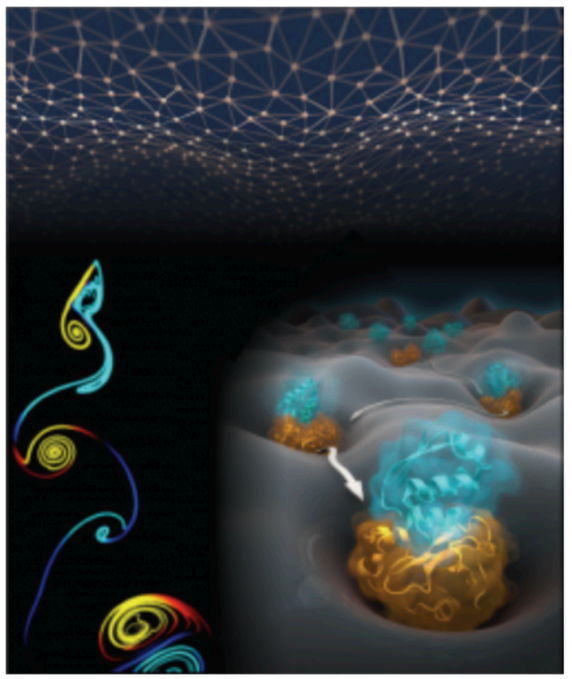
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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.

NOTATION / TERMINOLOGY

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

p(x, z | θ, ν)

x
observed data
simulated data

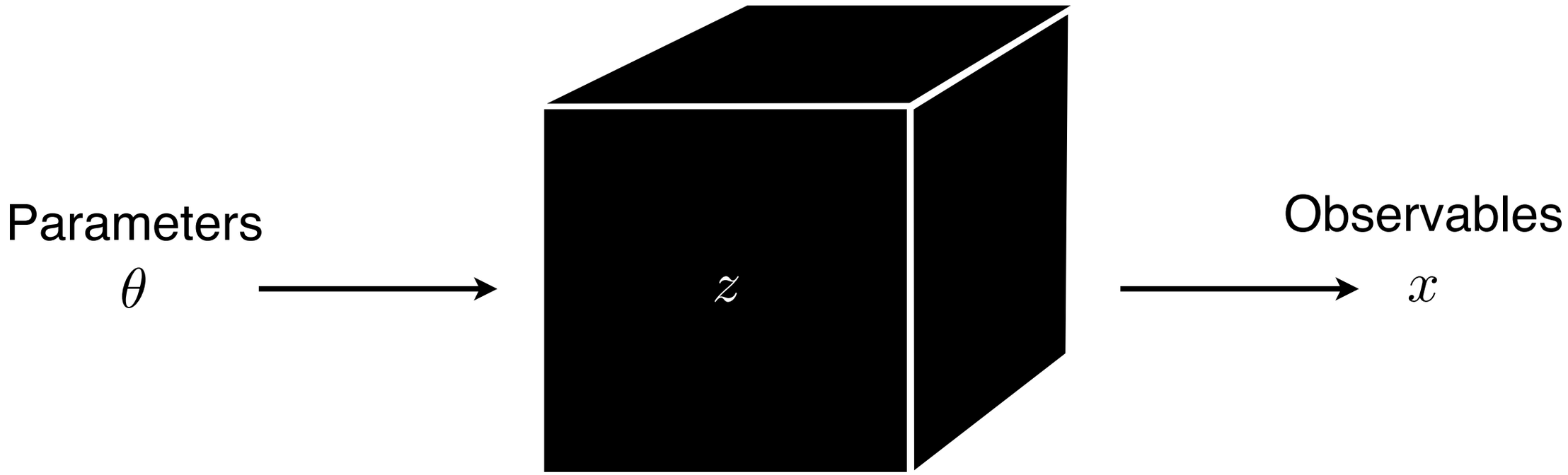
z
latent variables

ν
nuisance parameters

INFERENCE

inverse problem
measurement
parameter estimation

SIMULATION WITH A MECHANISTIC MODEL



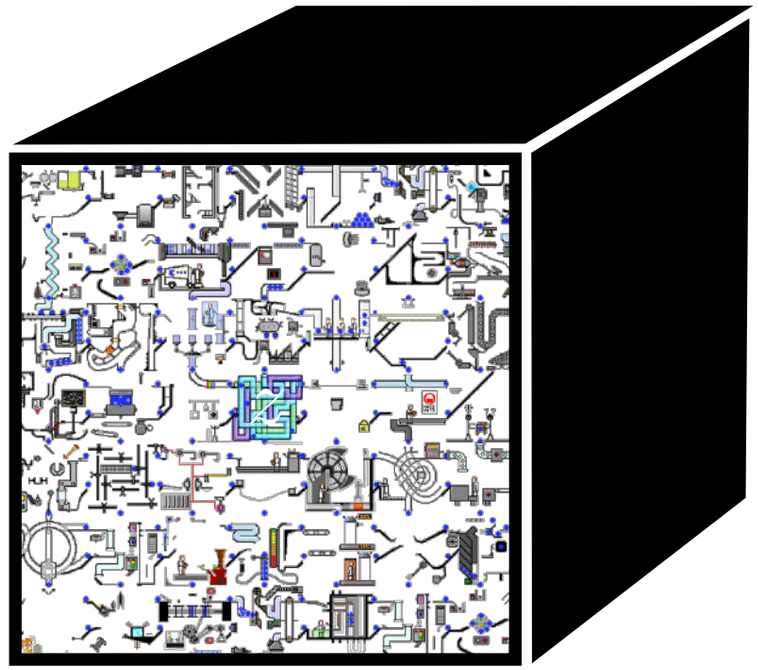
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

SIMULATION WITH A MECHANISTIC MODEL

Parameters

θ



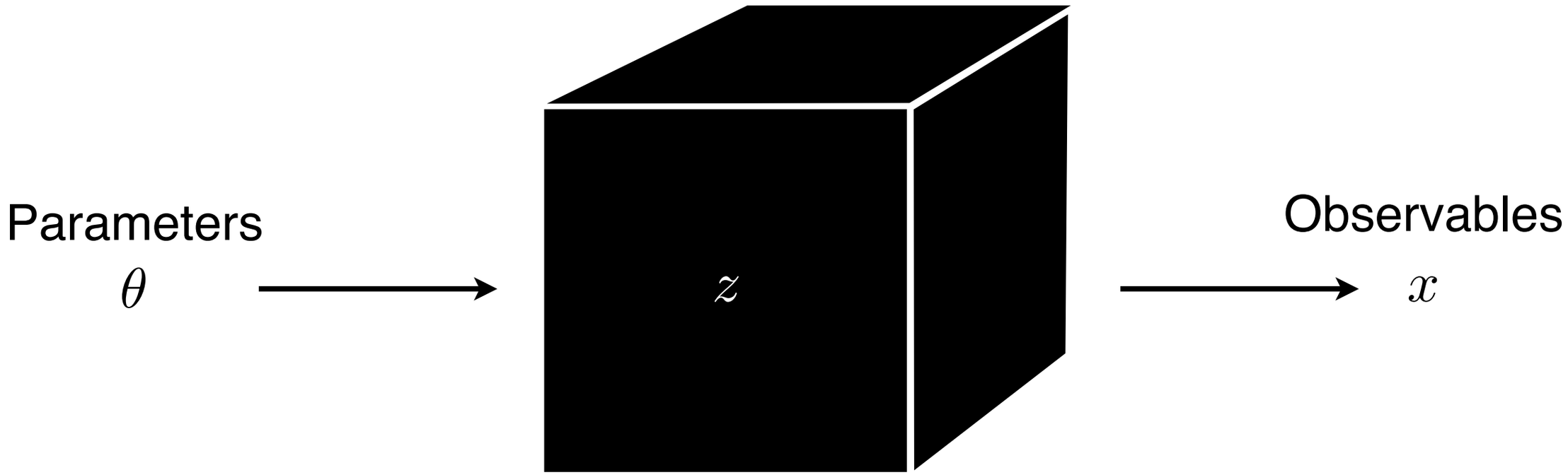
Observables

x

Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

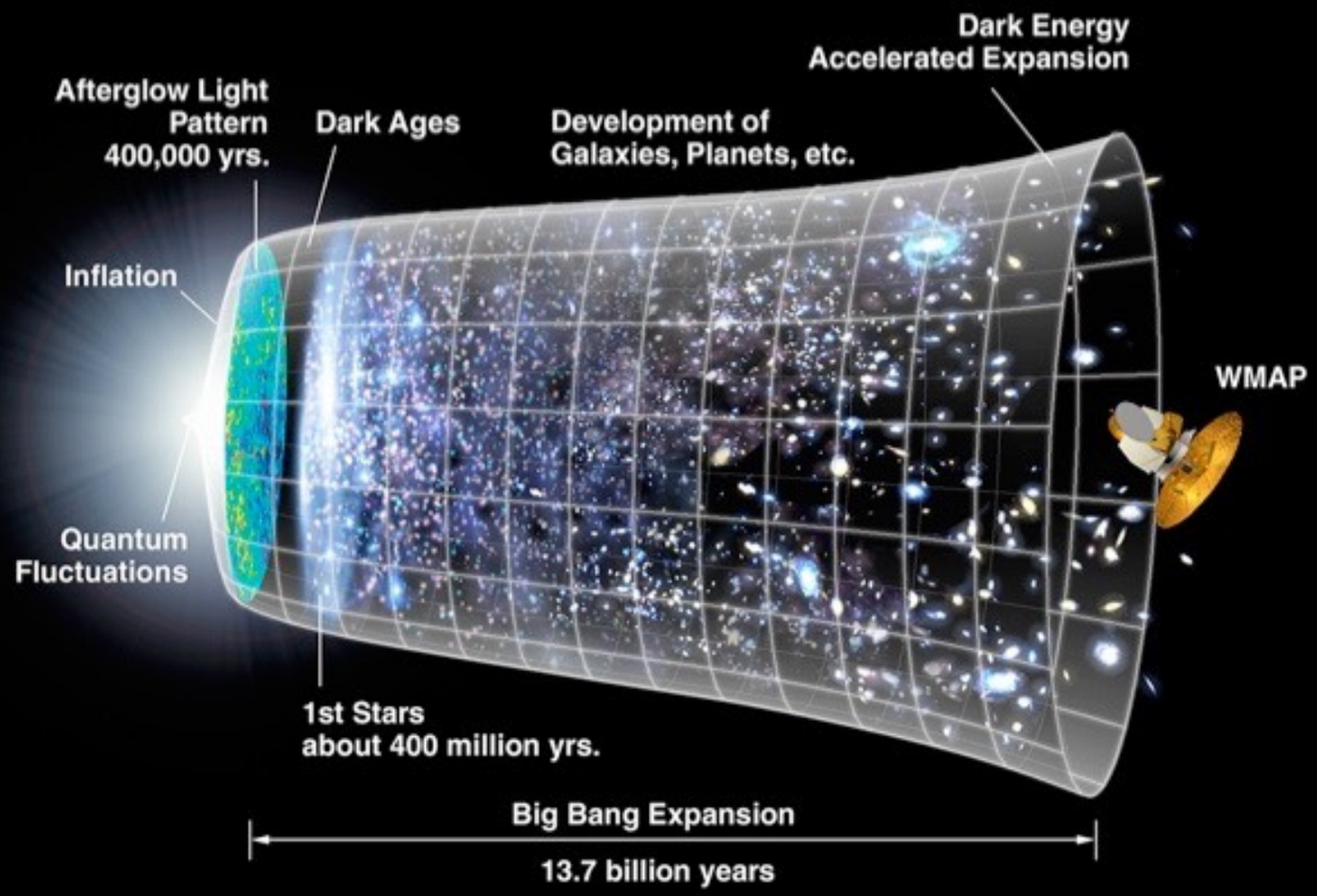
SIMULATION WITH A MECHANISTIC MODEL



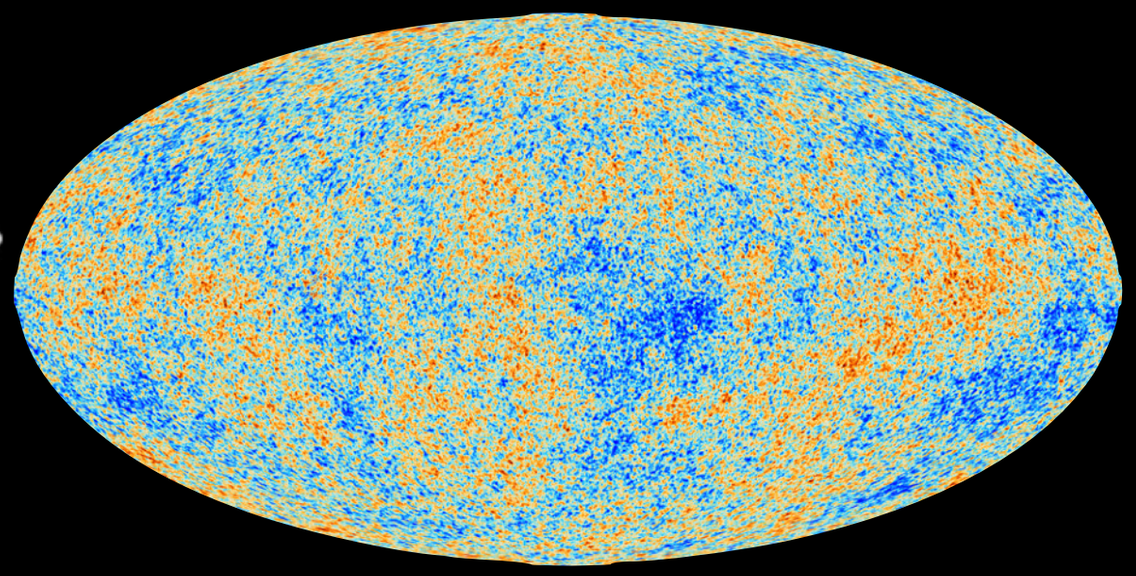
Prediction (simulation):

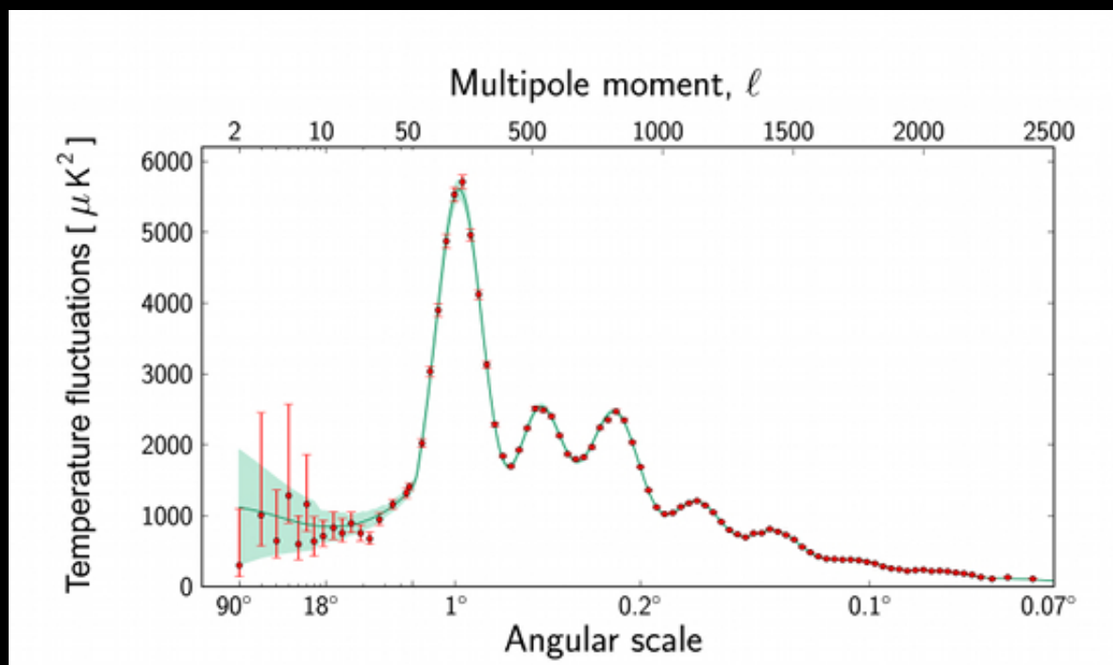
- Well-understood mechanistic model
- Simulator can generate samples

COSMOLOGY: 6 PARAMETERS



The Cosmic Microwave Background
A Gaussian Process in the Sky



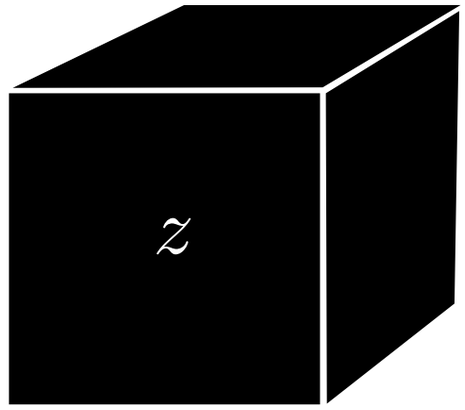


Symbol	Description	Value
$\Omega_B H^2$	Physical Baryon Density Parameter	0.02230 ± 0.00014
$\Omega_C H^2$	Physical Dark Matter Density Parameter	0.1188 ± 0.0010
T_0	Age Of The Universe	13.799 ± 0.021 × 10^9 Years
N_S	Scalar Spectral Index	0.9667 ± 0.0040
Δ_2	Curvature Fluctuation Amplitude	2.441 ± 0.09 × 10^{-9}
T	Reionization Optical Depth	0.066 ± 0.012

COSMOLOGICAL N-BODY SIMULATIONS

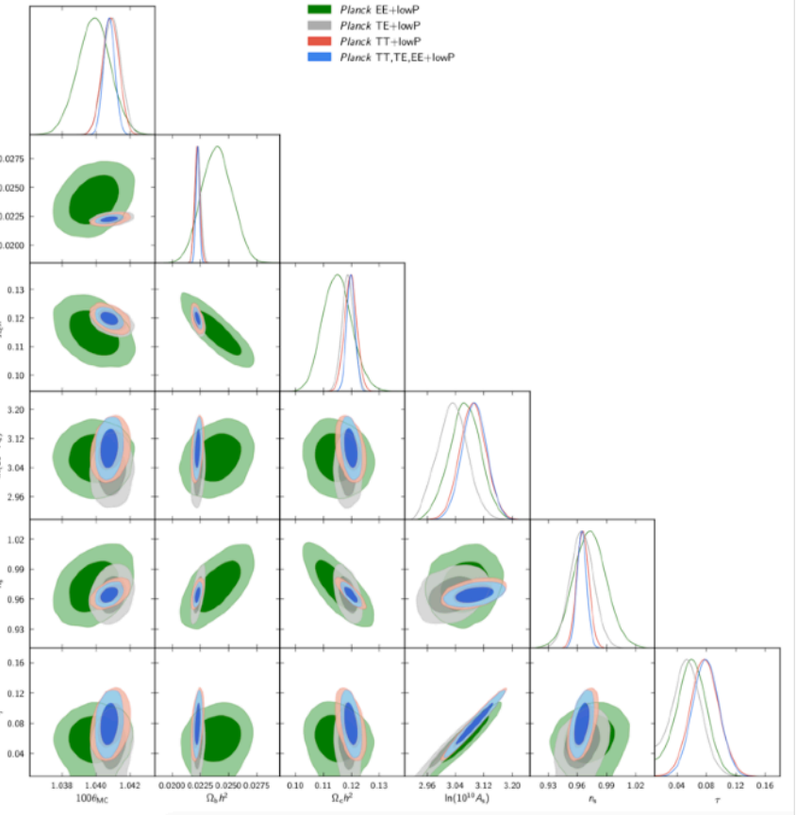
Parameters

θ

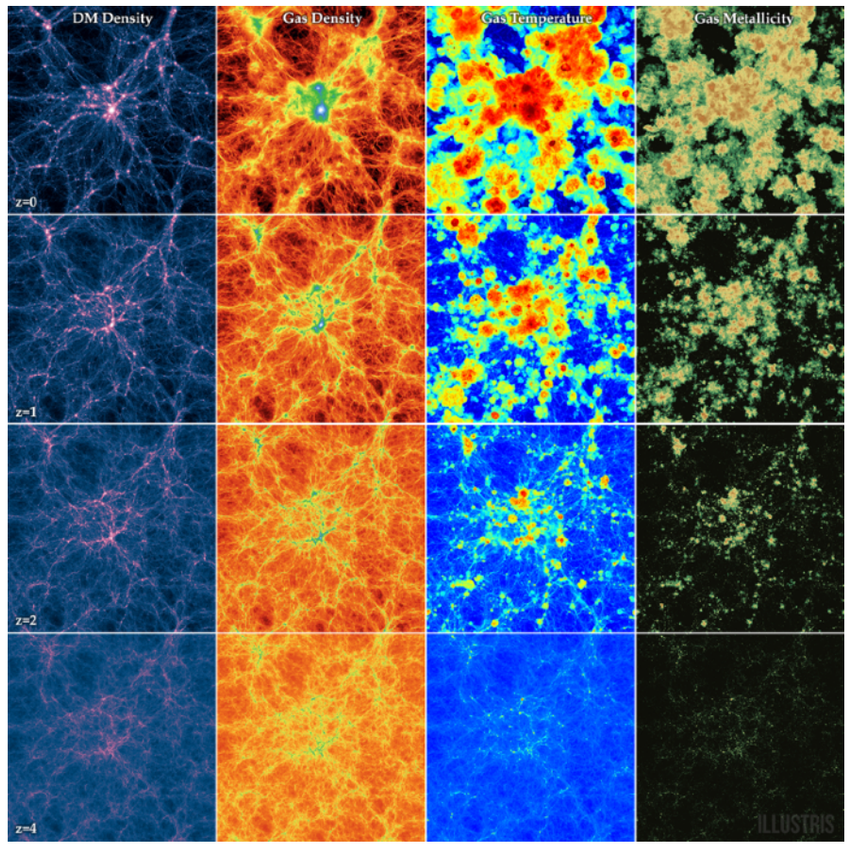


Observables

x



[Source: Planck 1502.01589]



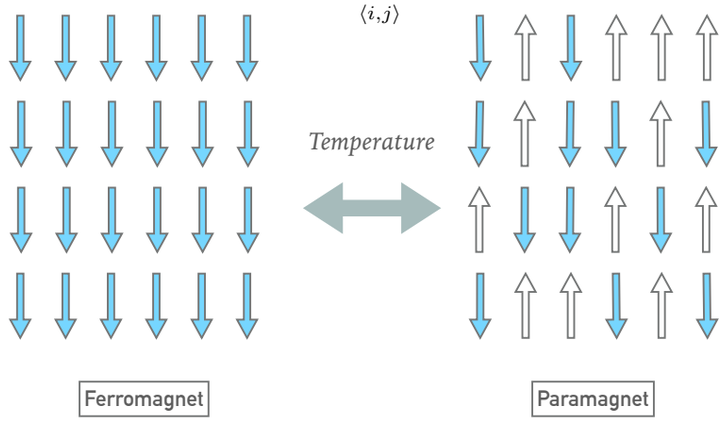
[Source: Illustris 1405.2921]

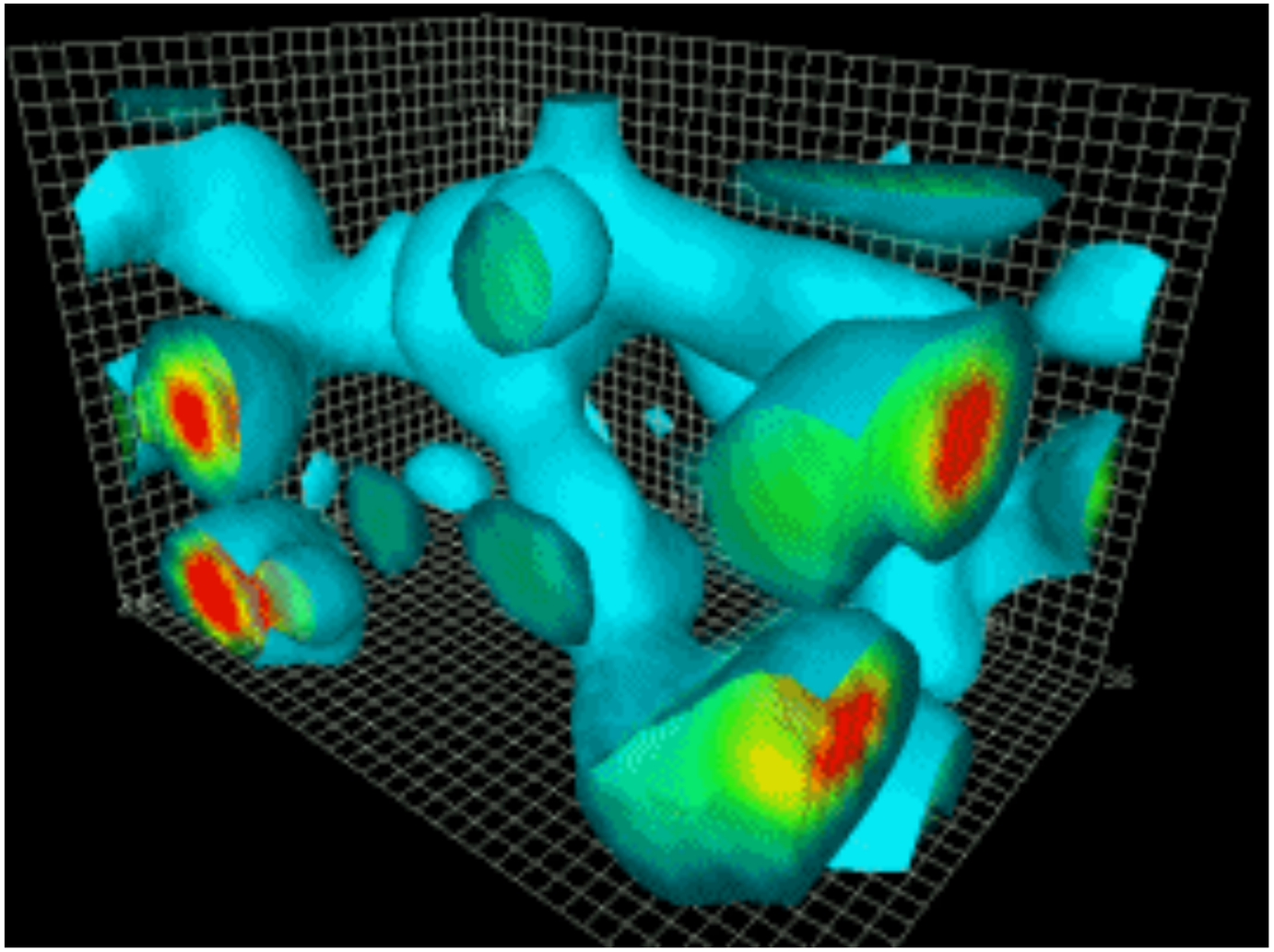
LATTICE FIELD THEORY

PHASES, PHASE TRANSITIONS, AND THE ORDER PARAMETER

Ising ferromagnet in two dimensions

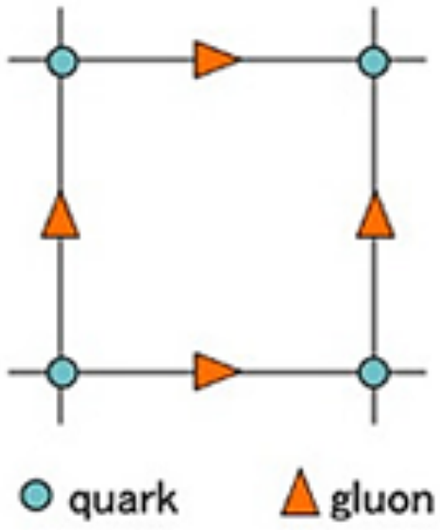
$$E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

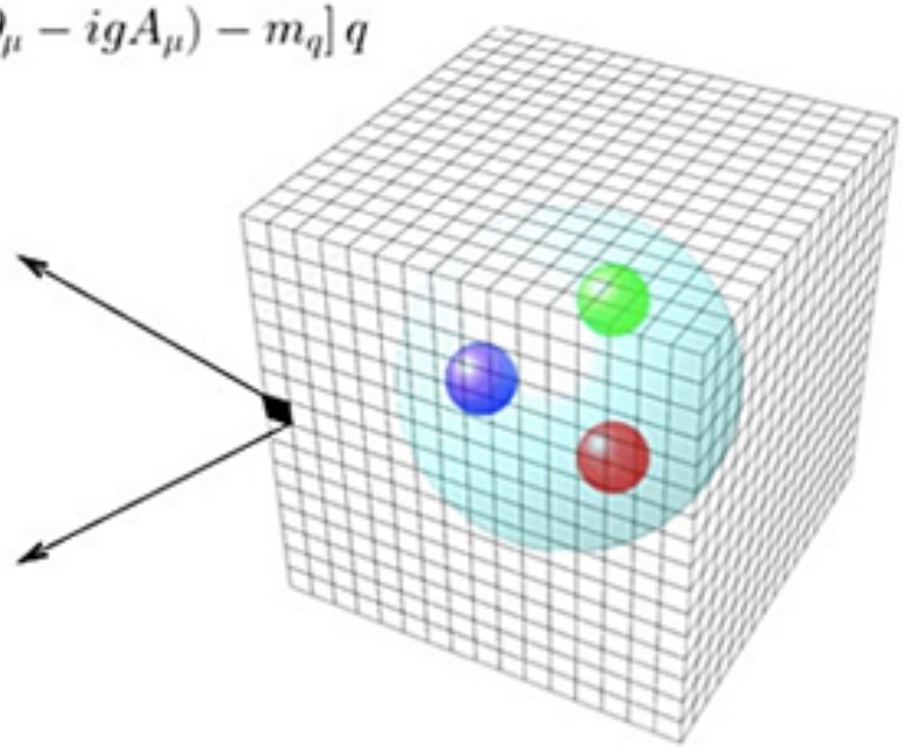




QCD Lagrangian

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \sum_{q=u,d,s,c,b,t} \bar{q} \left[i\gamma^\mu (\partial_\mu - igA_\mu) - m_q \right] q$$



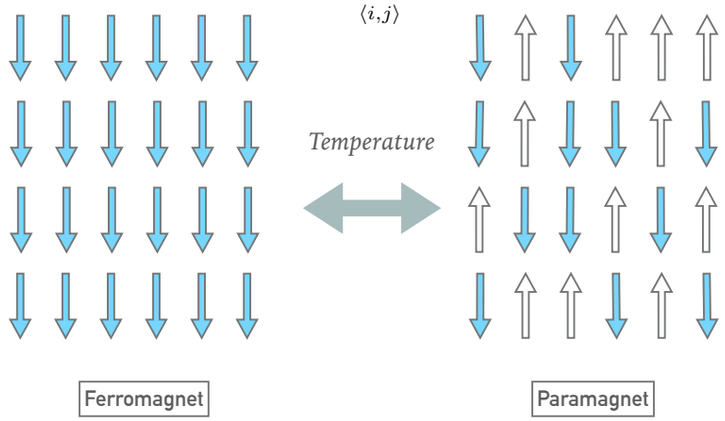


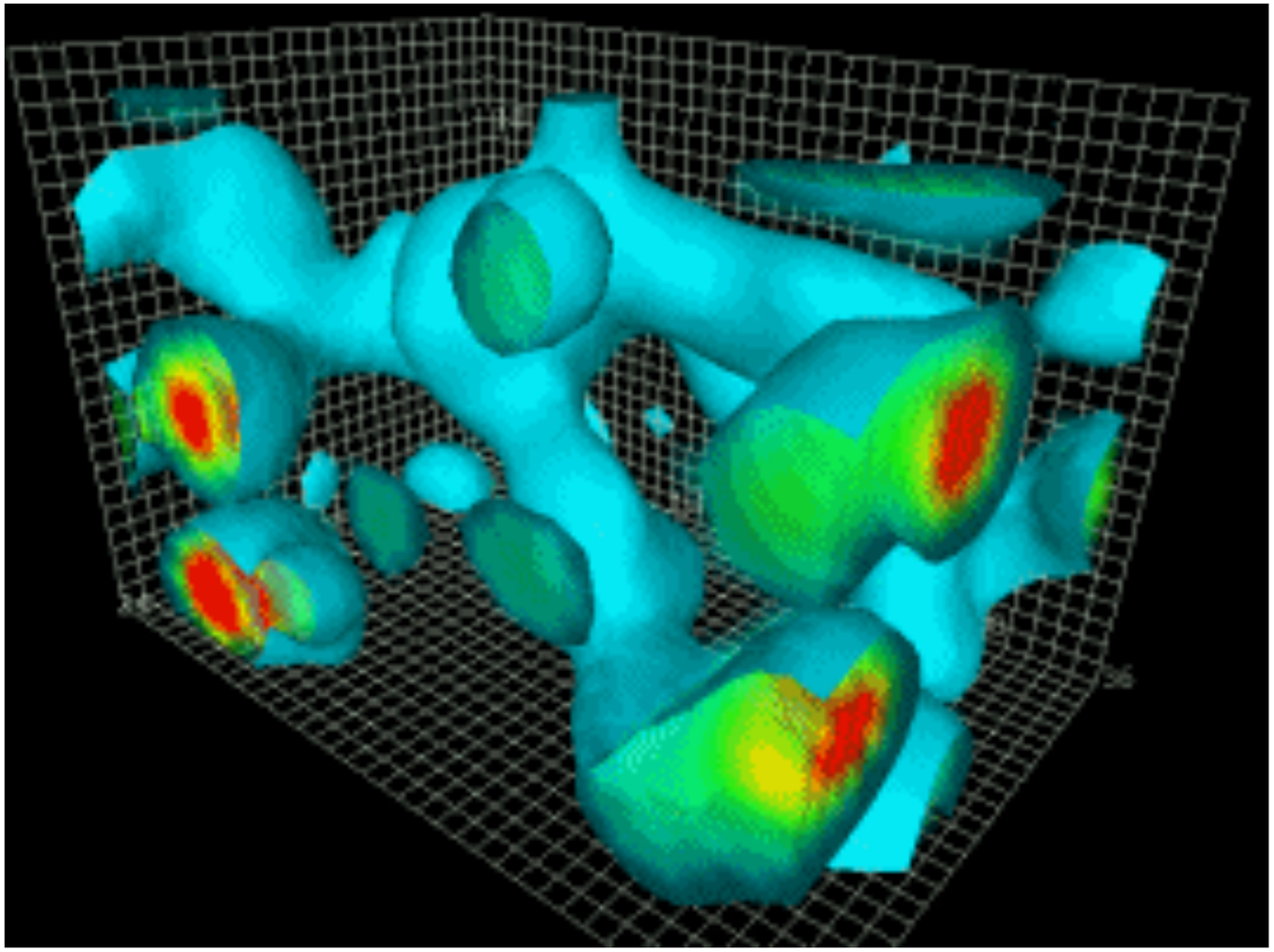
LATTICE FIELD THEORY

PHASES, PHASE TRANSITIONS, AND THE ORDER PARAMETER

Ising ferromagnet in two dimensions

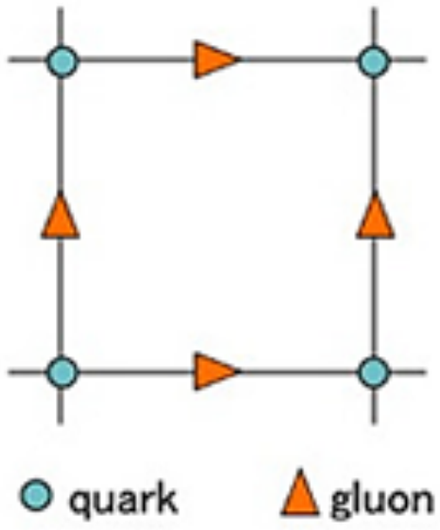
$$E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

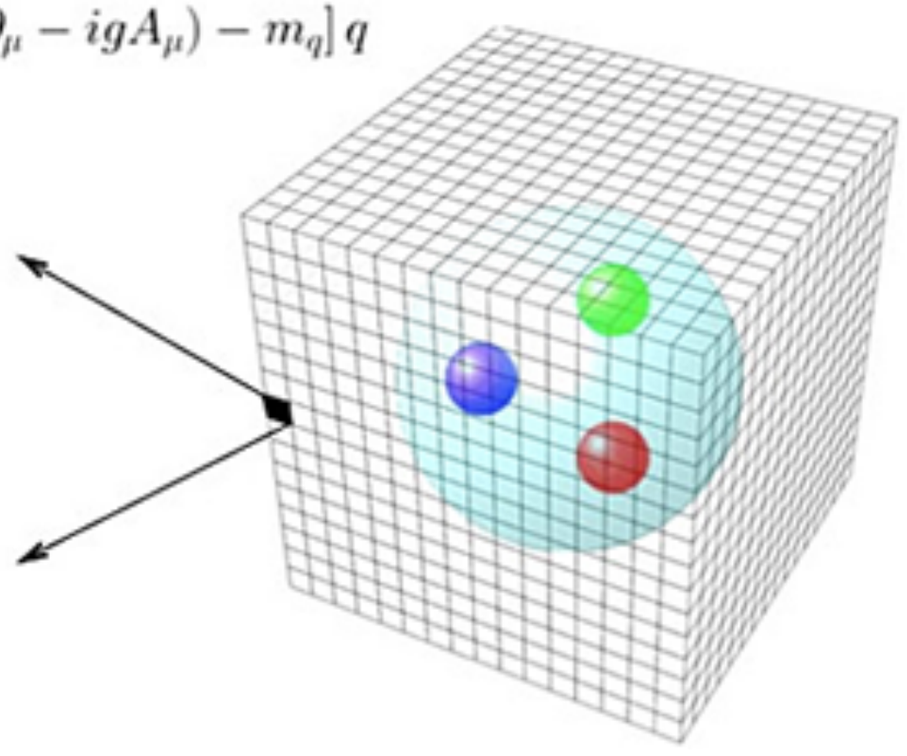




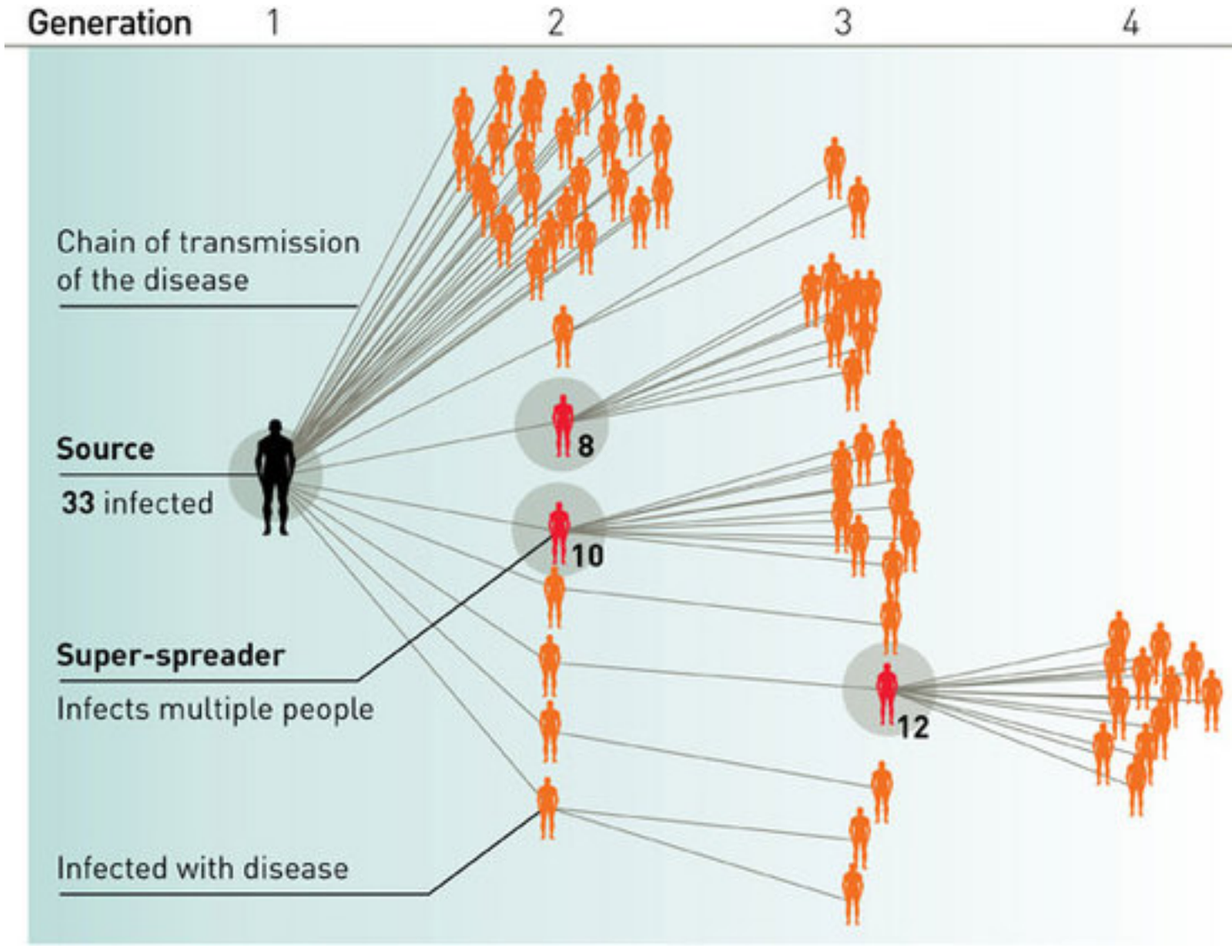
QCD Lagrangian

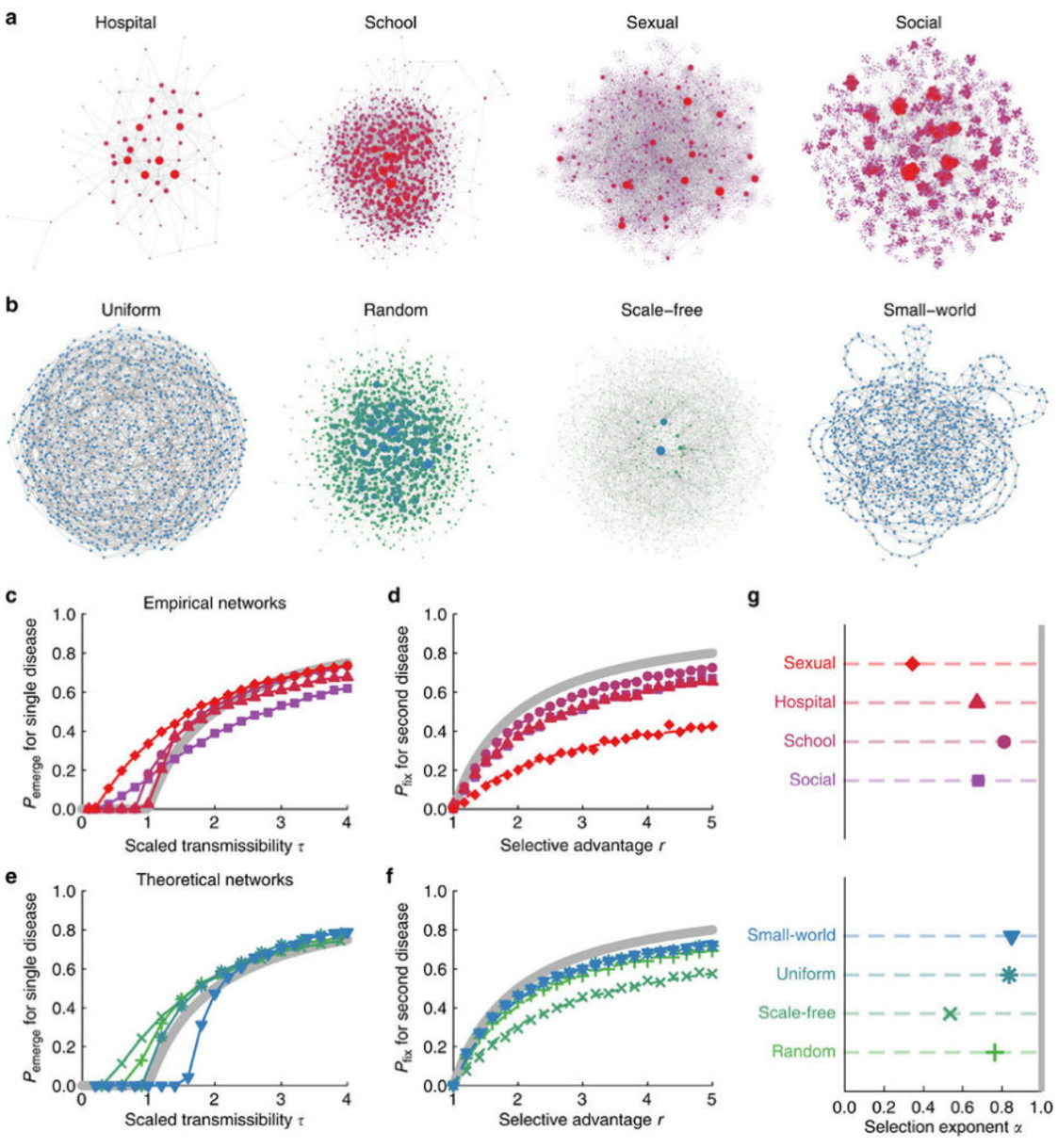
$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \sum_{q=u,d,s,c,b,t} \bar{q} \left[i\gamma^\mu (\partial_\mu - igA_\mu) - m_q \right] q$$

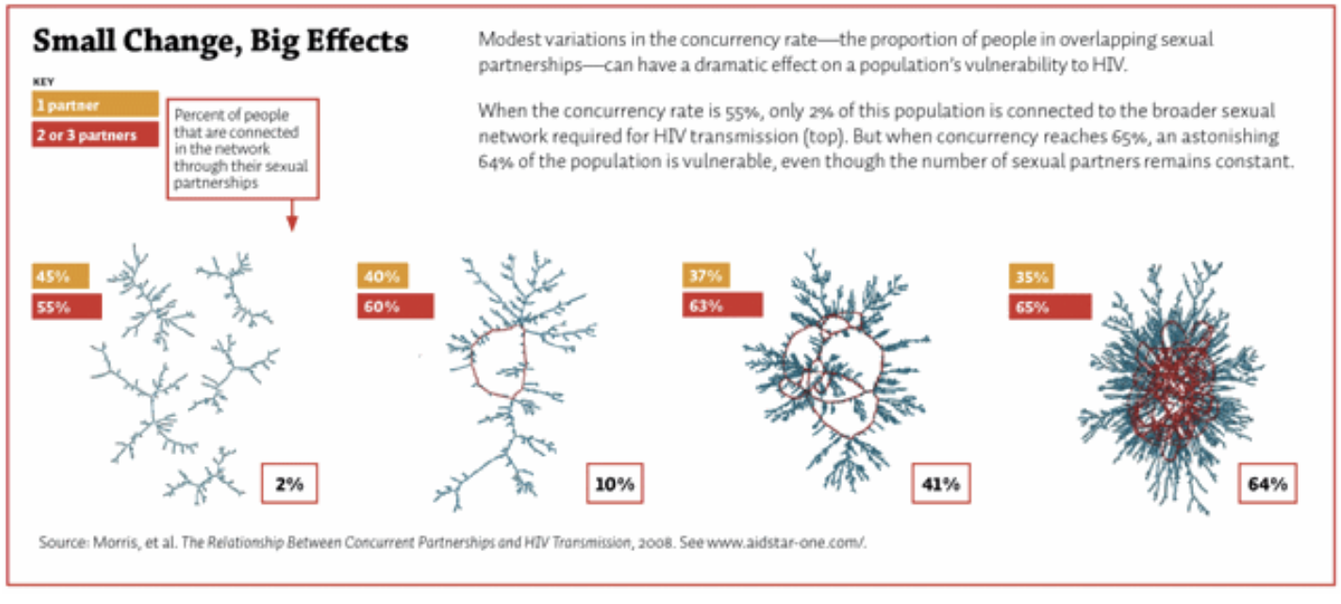




EPIDEMIOLOGY & POPULATION GENETICS







TAXONOMY FOR SIMULATION

Deterministic: fluid mechanics, quantum state evolution, ODEs and PDEs

- Often differentiable (at least in principle)

Stochastic: statistical physics (Ising model, etc.); particle scattering process, …

- Non-differentiable elements due to probabilistic control flow (eg. if/then/else conditions)

Measurement noise: may or may not be included

- eg. Use of ML for theoretical physics often treats system as if it can be exactly, directly observed

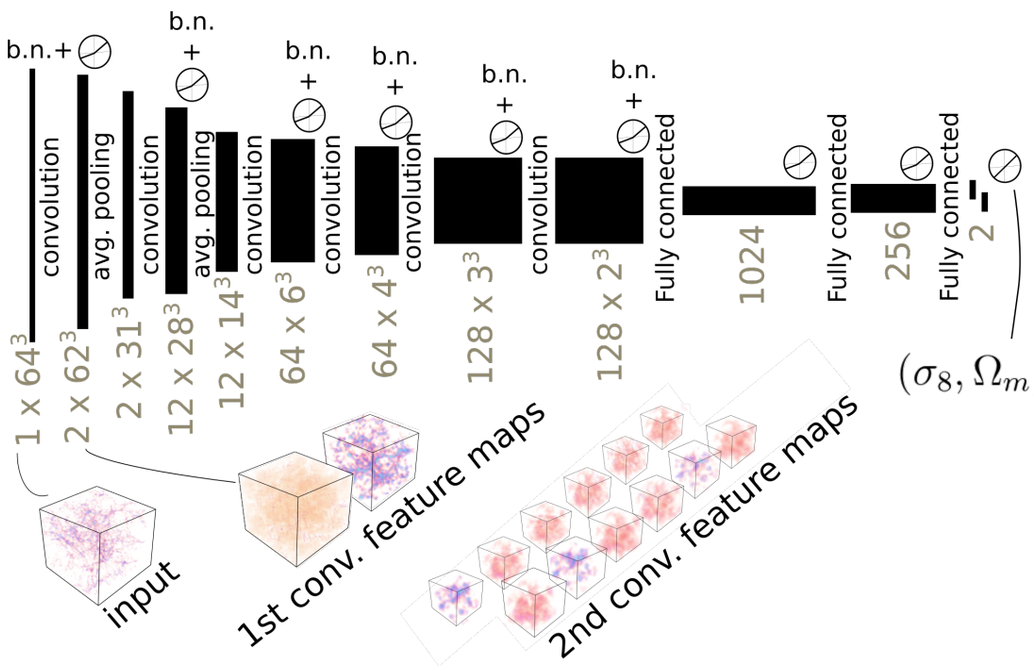
Simulators can produce labeled training data for supervised learning

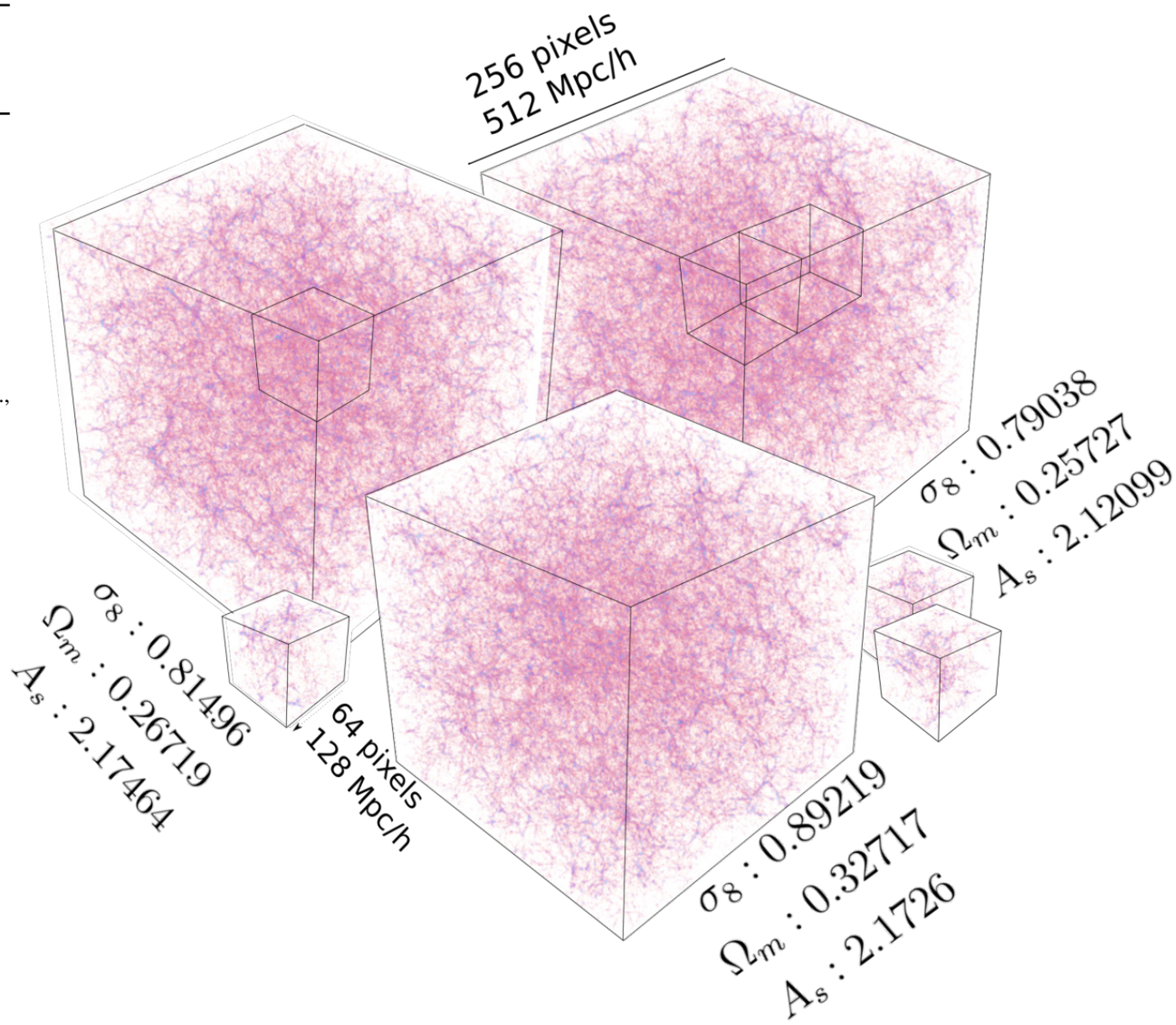
(note: some simulation are very computationally expensive)

Estimating Cosmological Parameters from the Dark Matter Distribution

Siamak Ravanbakhsh⋆ MRAVANBA@CS.CMU.EDU
Junier Oliva⋆ JOLIVA@CS.CMU.EDU
Sebastien Fromenteau† SFROMENT@ANDREW.CMU.EDU
Layne C. Price† LAYNEP@ANDREW.CMU.EDU
Shirley Ho† SHIRLEYH@ANDREW.CMU.EDU
Jeff Schneider⋆ JEFF.SCHNEIDER@CS.CMU.EDU
Barnabás Póczos⋆ BAPOCZOS@CS.CMU.EDU

⋆ School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
† McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Carnegie 5000 Forbes Ave., Pittsburgh, PA 15213, USA





Simulations are expensive, very few training examples

- each one is huge, crop into smaller boxes
- cropping looses larger scale structure

LATTICE QUANTUM CHROMO DYNAMICS

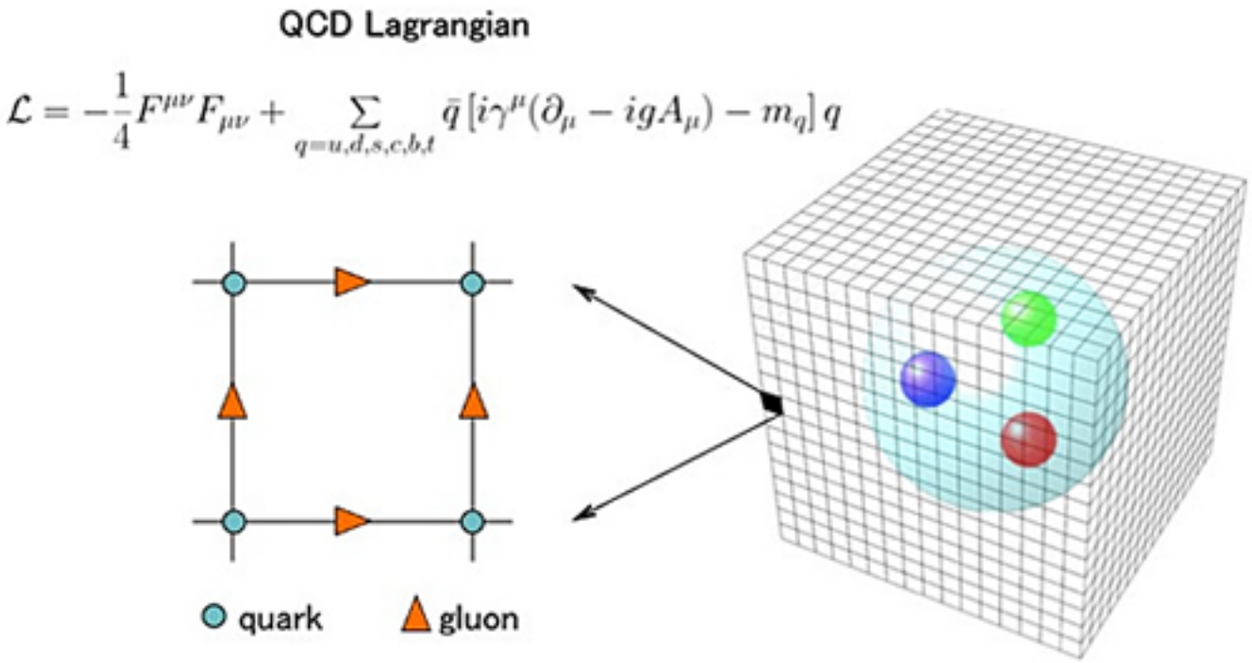
arXiv:1801.05784

Machine learning action parameters in lattice quantum chromodynamics

Phiala E. Shanahan,[1,2] Daniel Trewartha,[2] and William Detmold[3]

[1]*Department of Physics, College of William and Mary, Williamsburg, VA 23187-8795, USA*
[2]*Jefferson Laboratory, 12000 Jefferson Avenue, Newport News, VA 23606, USA*
[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Dated: January 18, 2018)

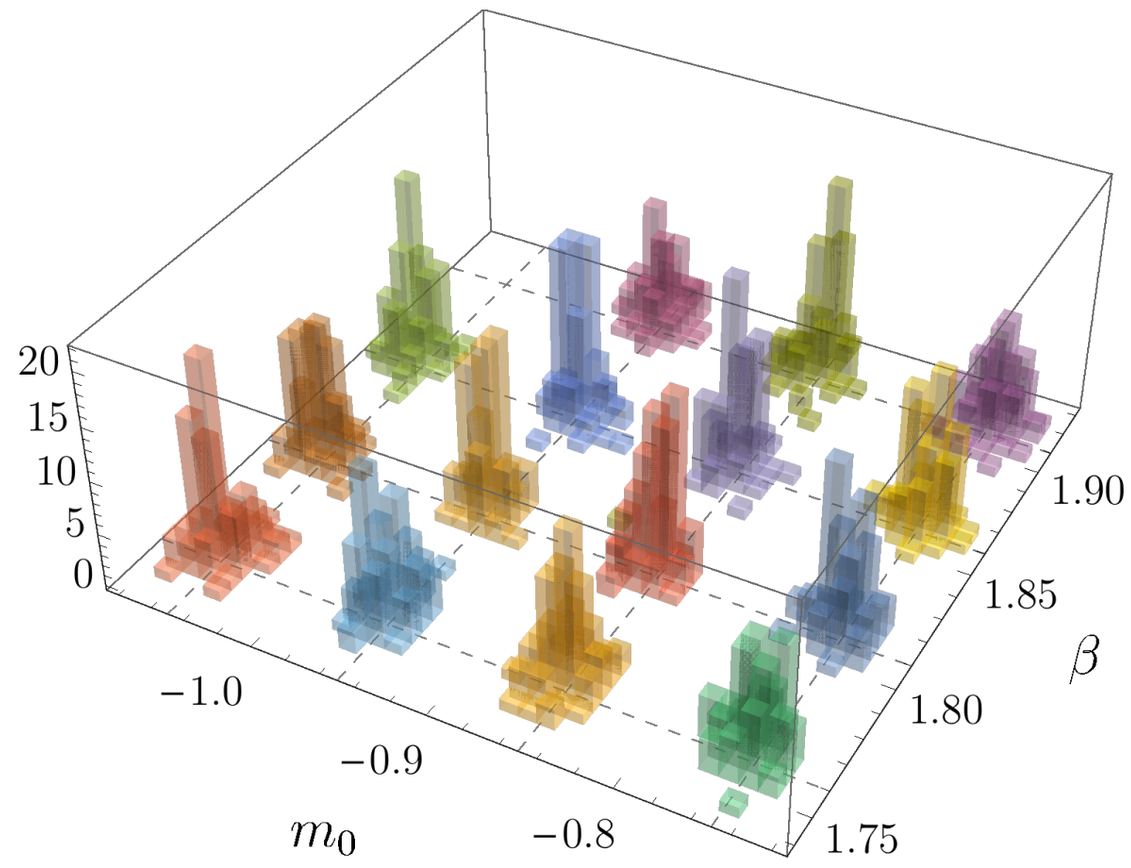


Gibbs distribution over lattice configurations

- 10^7 lattice sites
- $x_i \in \mathbb{R}^{32}$ at each site
- Local gauge symmetry SU(3)

Challenges:

- $O(100)$ of training examples
- each one is huge
- cropping is a problem because data is hierarchical



BEYOND POINT ESTIMATES

Much of the "low hanging fruit" is the use of supervised learning for classification and regression that is trained from simulated data (from a pre-existing physical model)

- Training data from simulator {x,y} ~ p(x,y)
- Leads to function $\hat{\mathbf{y}}(x) = NN(X)$
  - A predictive model or point estimate of $\hat{\mathbf{y}}$ given x
  - eg. A maximum likelihood estimate

But in science we would like to have point estimate and notion of uncertainty

- Likelihood p(x|y) or posterior p(y|x) — much harder!

NOTATION / TERMINOLOGY

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

p(x, z | θ, ν)

x
observed data
simulated data

z
latent variables

ν
nuisance parameters

INFERENCE

inverse problem
measurement
parameter estimation

GALTON BOARD

Say we want to infer θ, the probability to bounce right

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

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

GALTON BOARD

Say we want to infer θ, the probability to bounce right

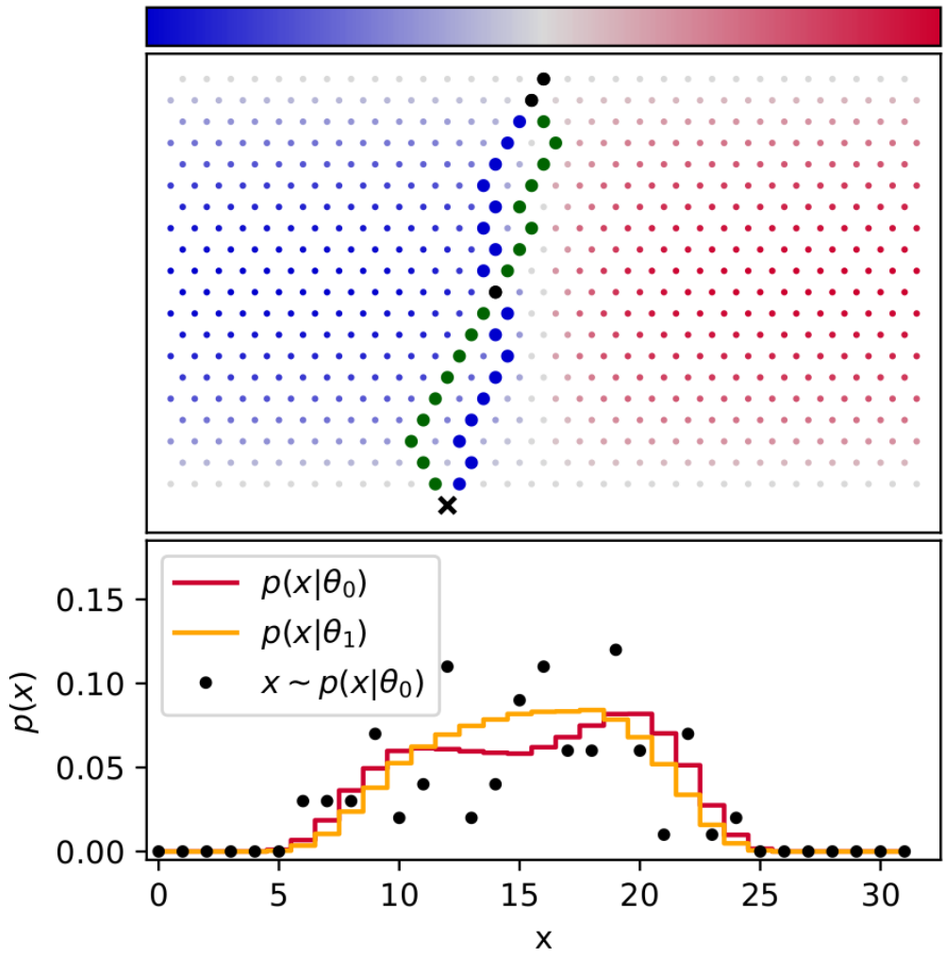
The probability of ending in bin x corresponds to the total probability of all the paths z from start to x.

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

WHAT IF WE SHIFT EACH PIN'S LOCATION?

The probability of ending in bin x still corresponds to the cumulative probability of all the paths from start to x:

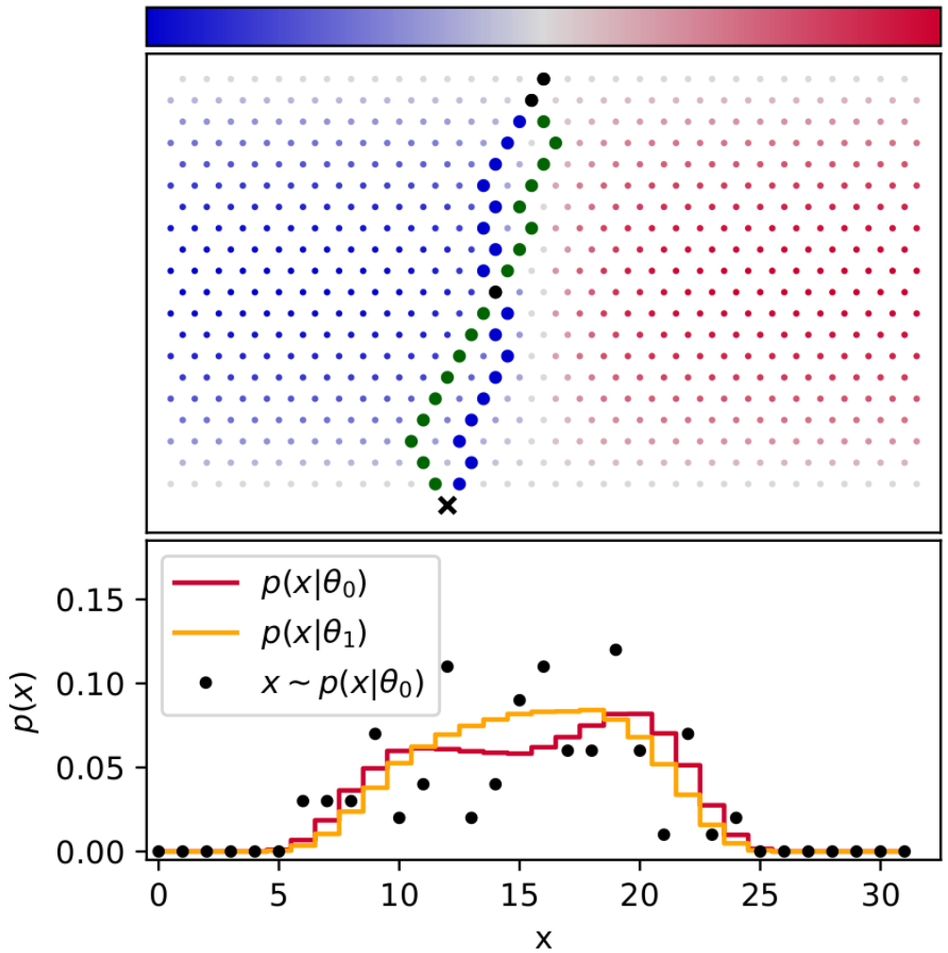
$$p(x|\theta) = \int p(x, z|\theta) dz$$



WHAT IF WE SHIFT EACH PIN'S LOCATION?

The probability of ending in bin x still corresponds to the cumulative probability of all the paths from start to x:

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



- But this integral can no longer be simplified analytically!
- As n grows larger, evaluating $p(x|\theta)$ becomes intractable since the number of paths grows combinatorially.
- Generating observations remains easy: drop the balls.

Since $p(x|\theta)$ cannot be evaluated, does this mean inference is no longer possible?

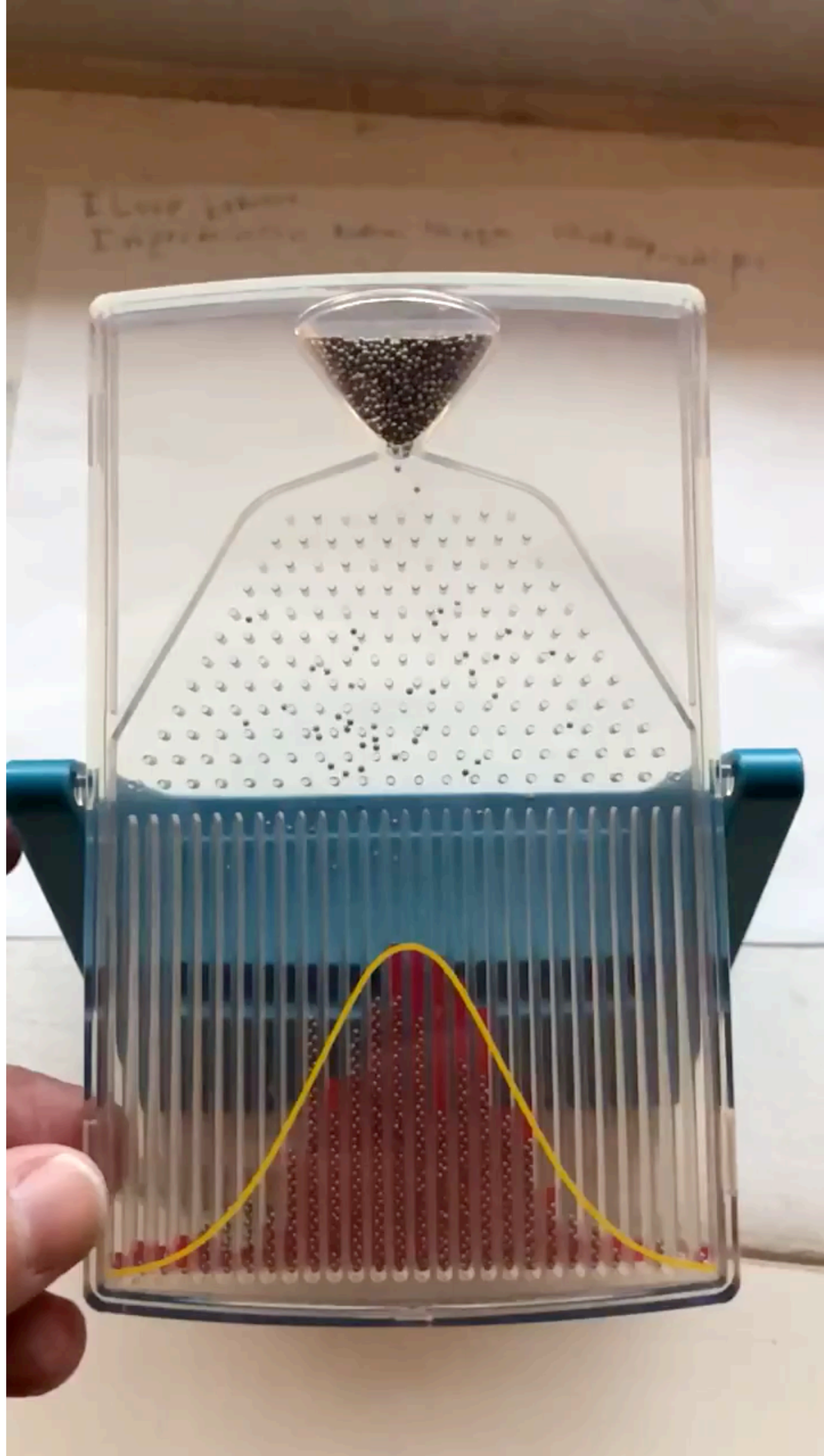
UH OH!

The actual situation is much more complicated.

It's not a Binomial distribution!

What is it?

I have no idea, but I could simulate it!



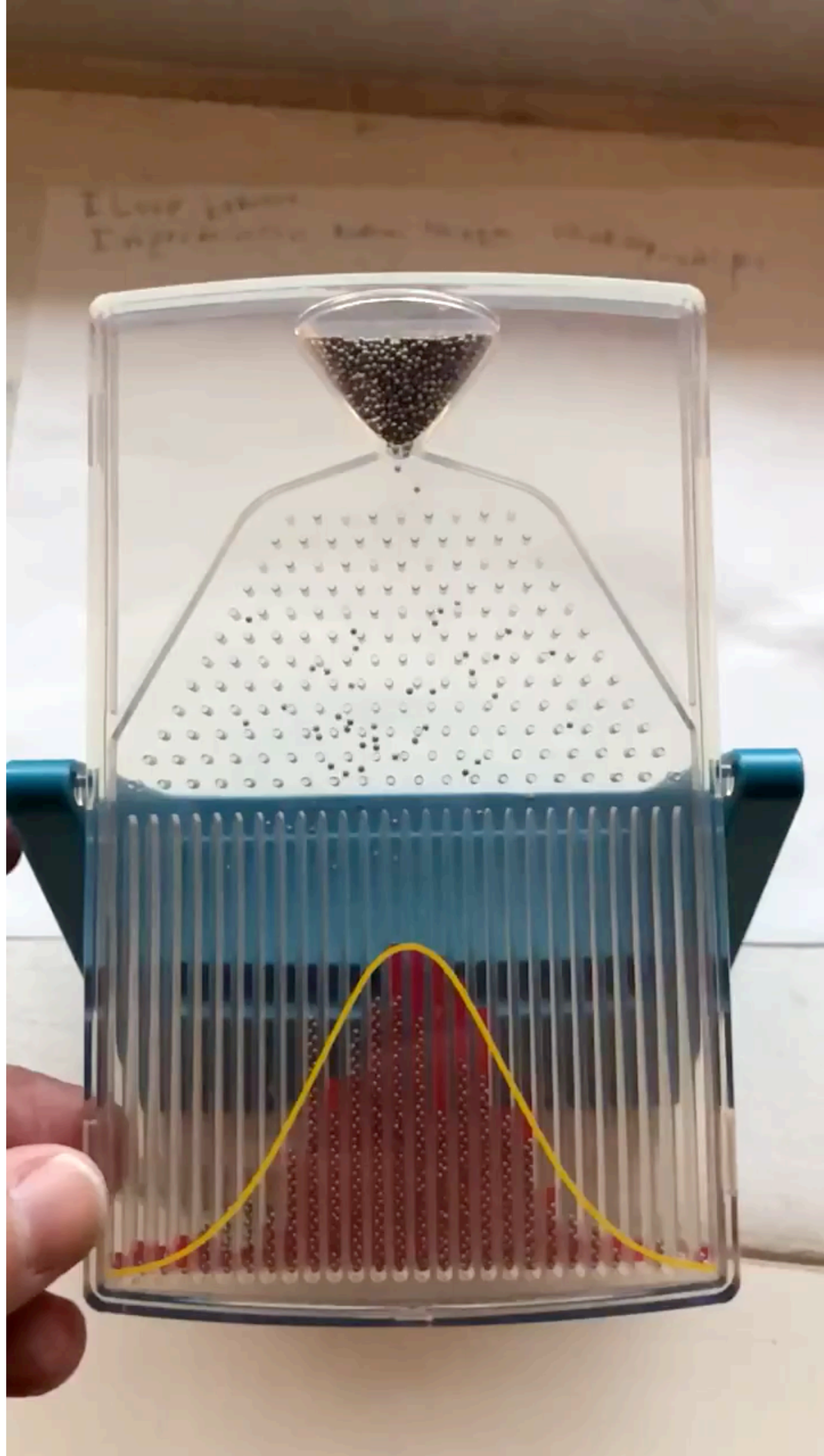
UH OH!

The actual situation is much more complicated.

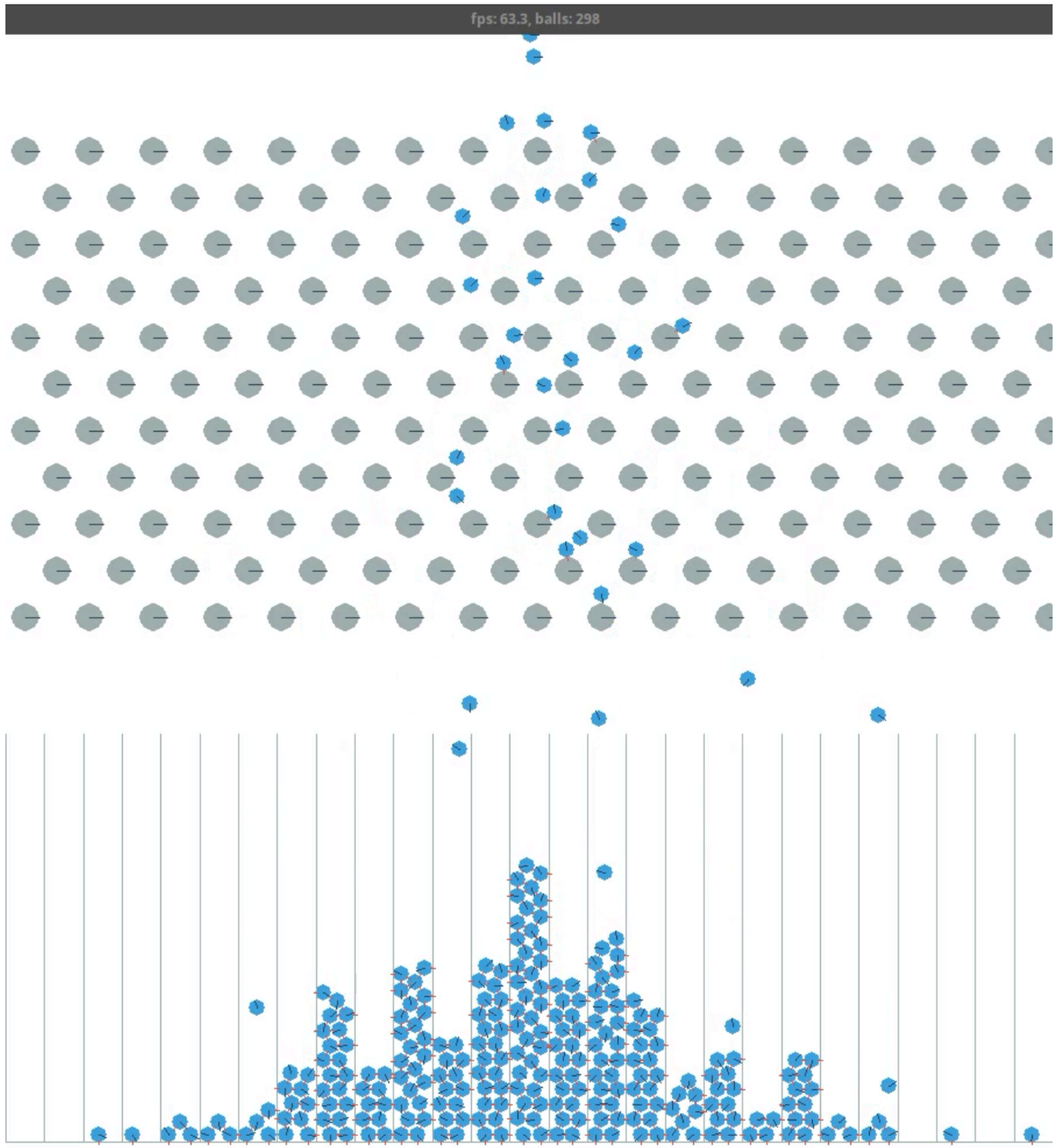
It's not a Binomial distribution!

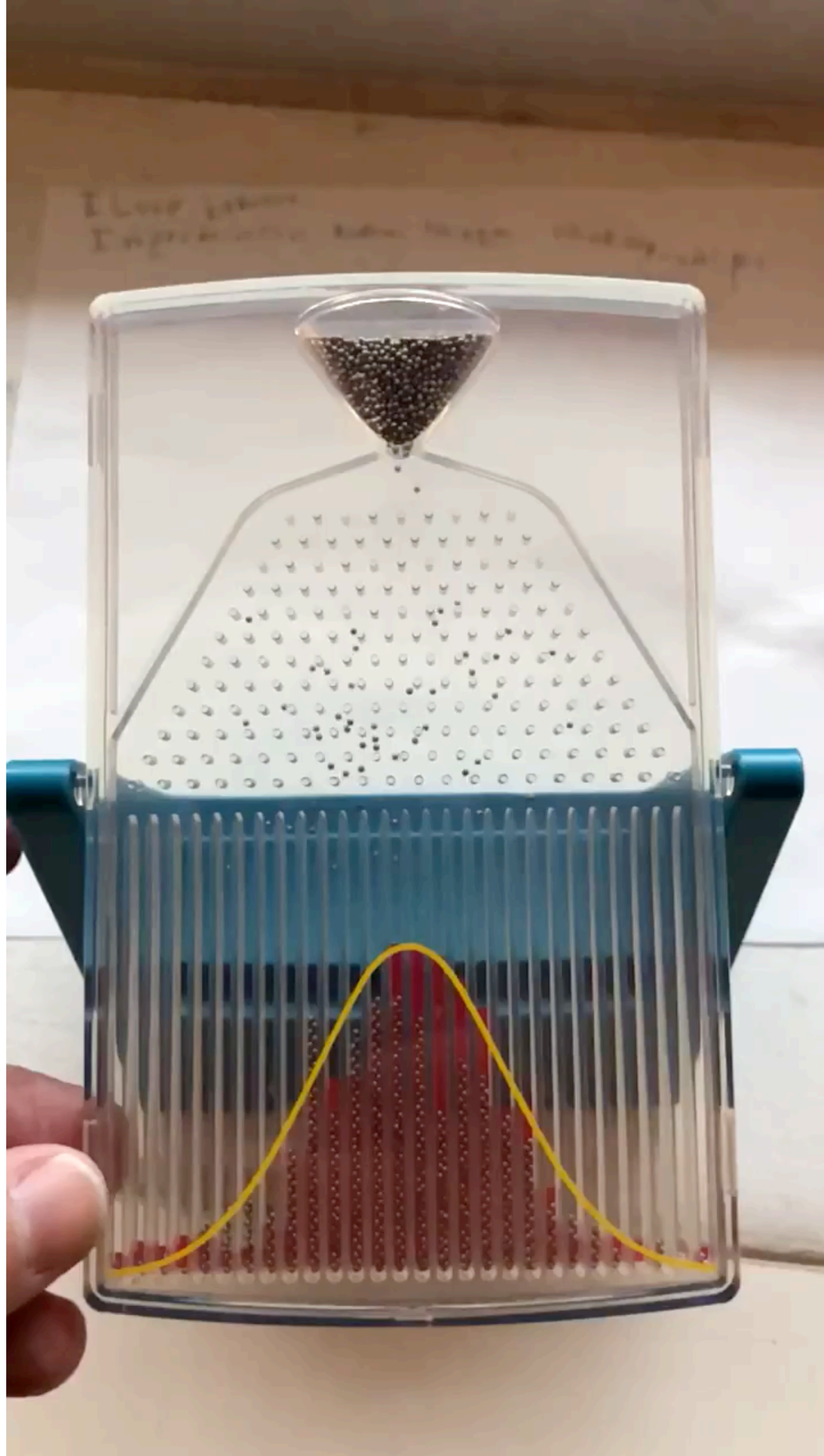
What is it?

I have no idea, but I could simulate it!



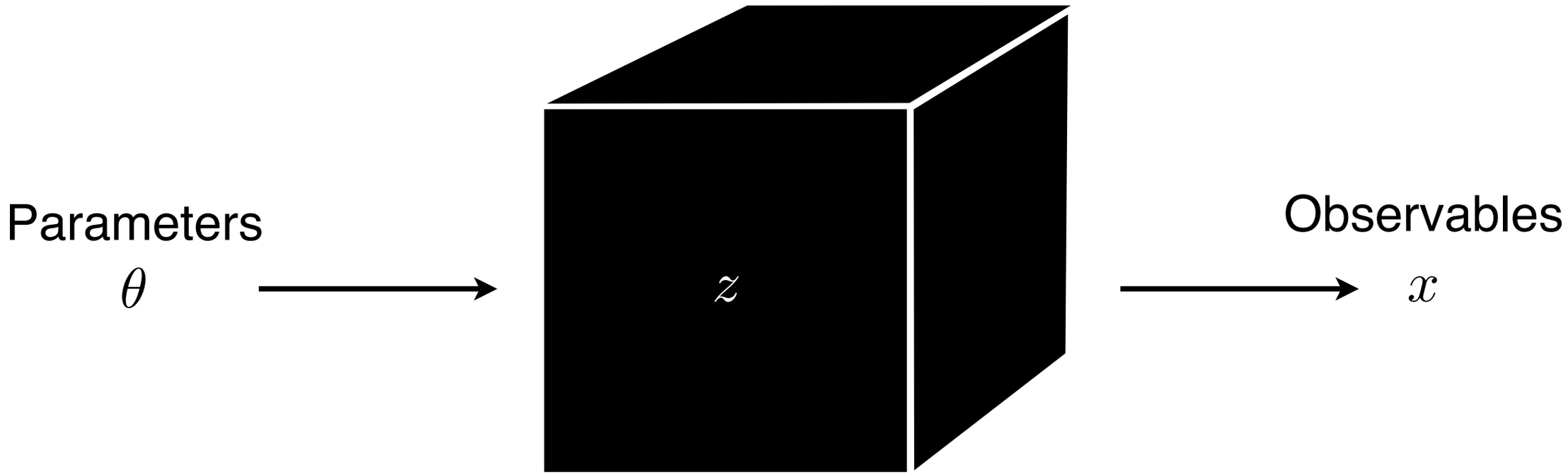
UH OH!





ANIMATION BY ATILIM GÜNEŞ BAYDIN

LIKELIHOOD-FREE INFERENCE



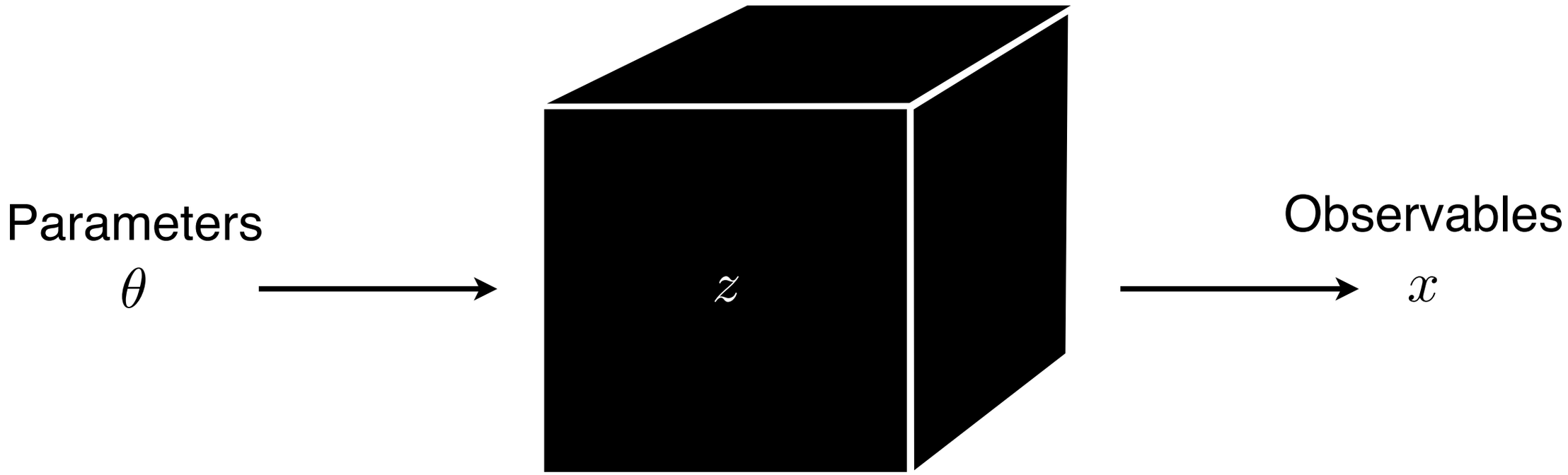
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Inference:

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

LIKELIHOOD-FREE INFERENCE



Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Inference:

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

A COMMON THEME, A COMMON LANGUAGE

ABC

resources on approximate Bayesian computational methods

Search

Home

Home

This website keeps track of developments in approximate Bayesian computation (ABC) (a.k.a. likelihood-free), a class of computational statistical methods for Bayesian inference under intractable likelihoods. The site is meant to be a resource both for biologists and statisticians who want to learn more about ABC and related methods. Recent publications are under Publications 2012. A comprehensive list of publications can be found under Literature. If you are unfamiliar with ABC methods see the Introduction. Navigate using the menu to learn more.

ABC in Montreal | ABC in Montreal (2014)

ABC in Montreal

Approximate Bayesian computation (ABC) or likelihood-free (LF) methods have developed mostly beyond the radar of the machine learning community, but are important tools for a large and diverse segment of the scientific community. This is particularly true for systems and population biology, computational neuroscience, computer vision, healthcare sciences, but also many others.

Interaction between the ABC and machine learning community has recently started and contributed to important advances. In general, however, there is still significant room for more intense interaction and collaboration. Our workshop aims at being a place for this to happen.

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:

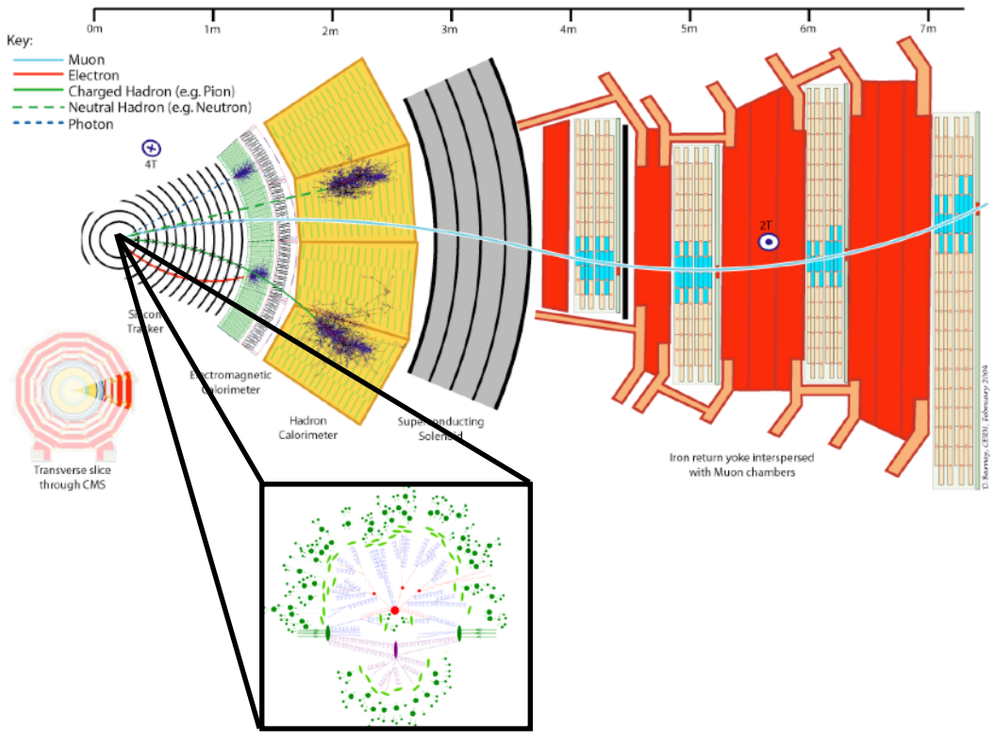
- **Generative adversarial networks** (a NIPS 2016 workshop) are implicit models with an adversarial training scheme.
- Recent advances in **variational inference** (a NIPS 2015 and 2016 workshop) have leveraged implicit models for more accurate approximations.
- **Approximate Bayesian computation** (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.

APPROACHES TO LIKELIHOOD-FREE INFERENCE

Use simulator

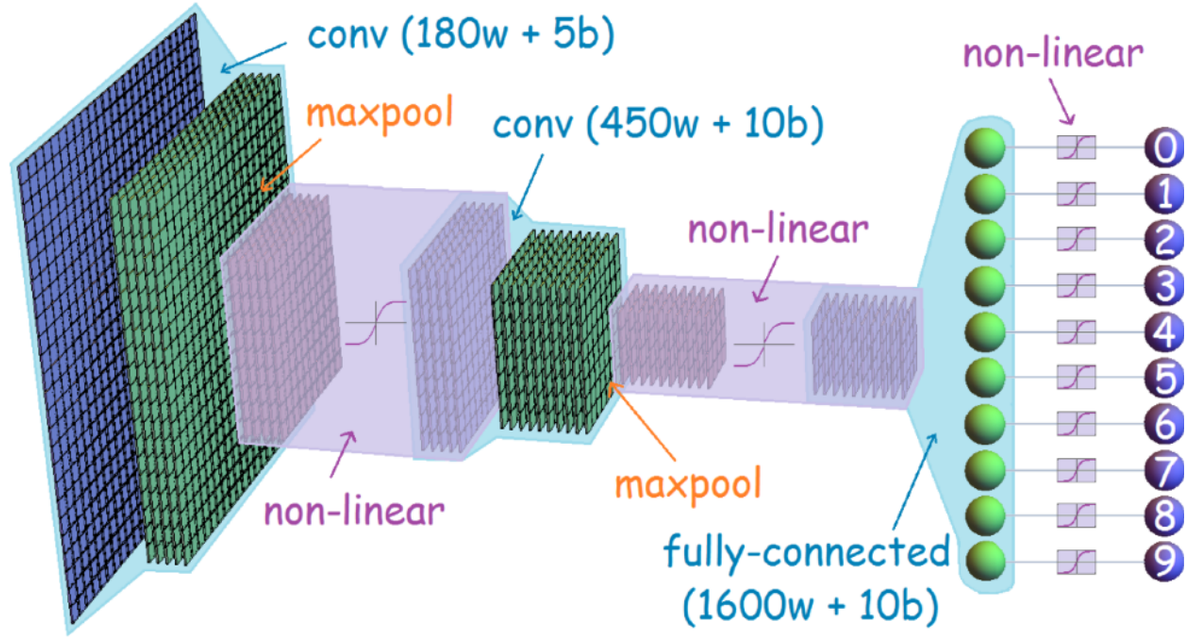
(much more efficiently)



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

Learn simulator

(with deep learning)



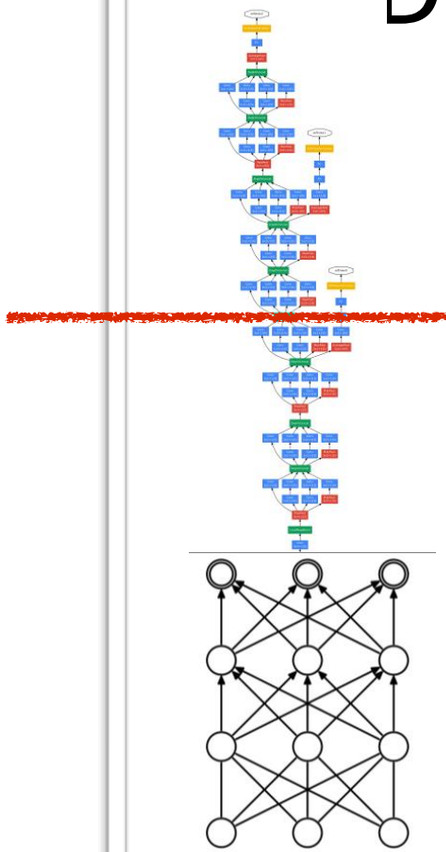
- Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAE)
- Likelihood ratio from classifiers (CARL)
- Autoregressive models, Normalizing Flows

[image credit: A.P. Goucher]

TWO PARADIGMS IN MACHINE LEARNING

Max
Welling

Discriminative or Generative?



-Deep Learning

-Kernel Methods

-Random Forests

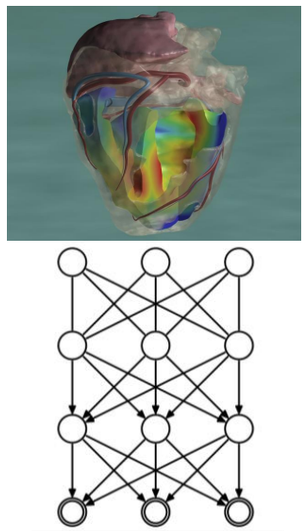
-Boosting

Variational Auto-Encoder

-Bayesian Networks

-Probabilistic Programs

-Simulator Models



- Advantages discriminative models:
  - Flexible map from input to target (low bias)
  - Efficient training algorithms available
  - Solve the problem you are evaluating on.
  - Very successful and accurate!

- Advantages generative models:
  - Inject expert knowledge
  - Model causal relations
  - Interpretable
  - Data efficient
  - More robust to domain shift
  - Facilitate un/semi-supervised learning

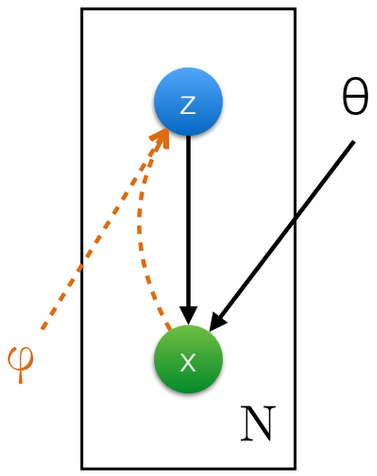
Deep Generative Models

VARIATIONAL AUTO-ENCODER

[Slides from D. Kingma NIPS 2015]

Auto-Encoding Variational Bayes

[Kingma and Welling, 2013/2014]
[Rezende et al, 2014]



- $q_\varphi(z|x) = N(\mu,\sigma^2)$
  $[\mu, \sigma^2] = f^{(z|x)}(x,\varphi) = \text{multilayer neural net}$
- Objective: lower bound of log p(x).
  - Jointly optimized w.r.t. φ and θ
  - This is approx. maximum likelihood
  - Simple SGD:
    - Sampling small minibatches of data
    - Sampling from approx. posterior
- This also minimizes an expected KL divergence
  $D_{KL}(q_\varphi(z|x)||p(z|x))$
  -> gives us cheap approx. inference for new datapoints

Kingma and Welling, Auto-encoding Variational Bayes, ICLR 2014
Rezende, Mohamed and Wierstra, Stochastic back-propagation and variational inference in deep latent Gaussian models, ICML 2014

Diederik (Durk) Kingma

Max Welling

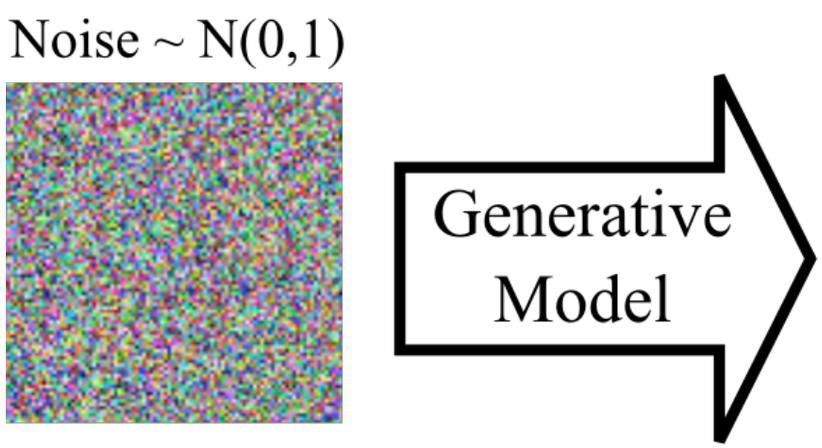
Danilo J. Rezende

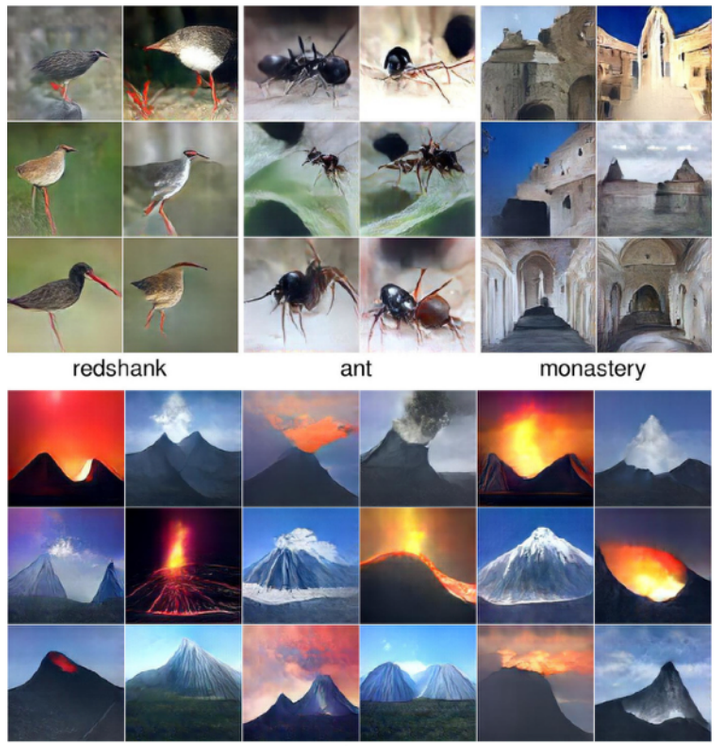
Conv. net as encoder/decoder, trained on faces

GENERATIVE ADVERSARIAL NETWORKS

Z

X

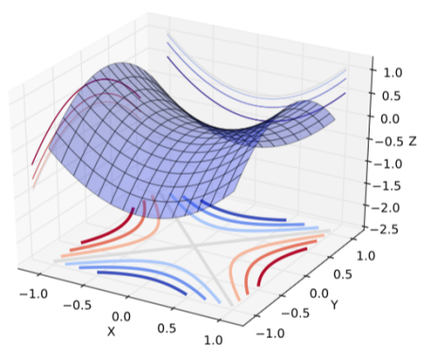




Leo is G Tom is D

- We want to
  - For fixed G, find D which maximizes $V(D, G)$,
  - For fixed D, find G which minimizes $V(D, G)$;
- In other words, we are looking for the *saddle point*

$$(D^*, G^*) = \max_D \min_G V(D, G).$$



TWO OBSERVATIONS

GANs and VAEs use deep neural network to transform latent Z to observed X

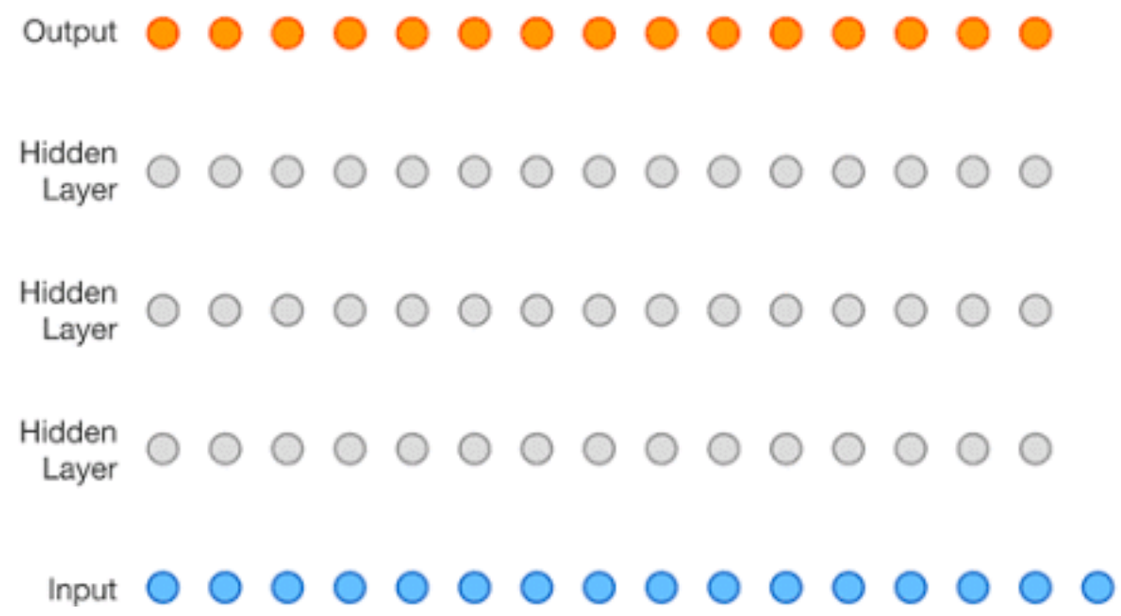
- But resulting density p(x) is intractable
  - Say the density is “implicit”
- And latent space z has no specific meaning or interpretation

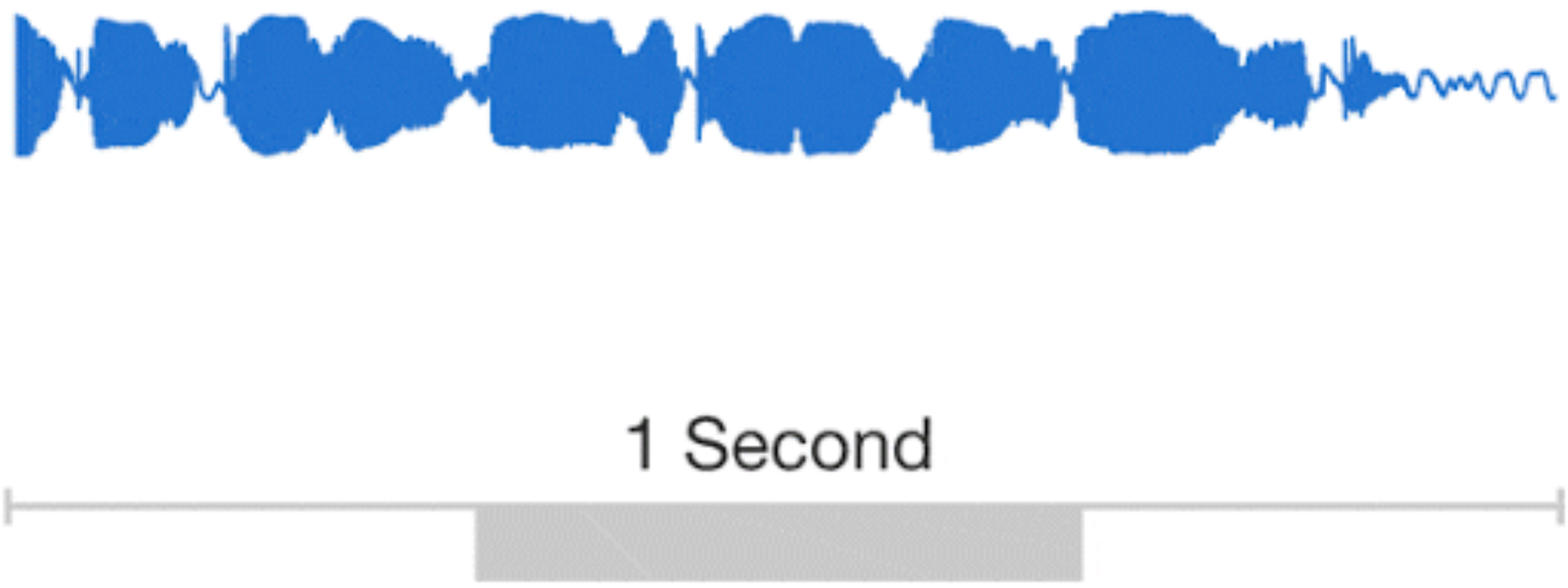
WAVENET: A GENERATIVE MODEL FOR RAW AUDIO

Autoregressive models defined by

$$p(x) = \prod_{t=1}^{T} p(x_t \mid x_{t-1}, \ldots, x_1).$$

have a tractable density. Train via maximum likelihood



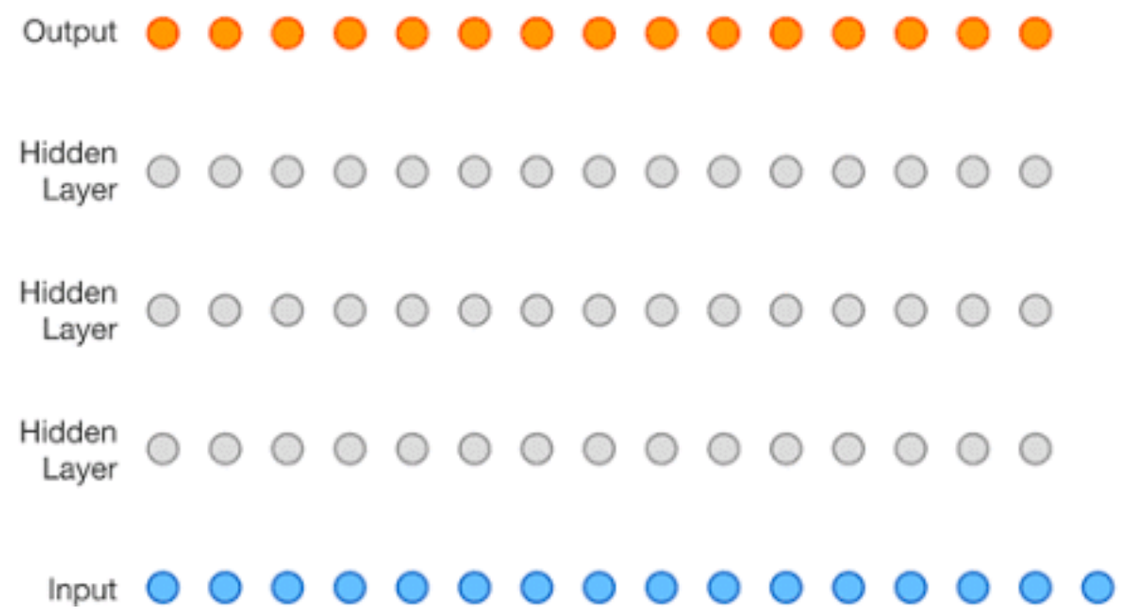


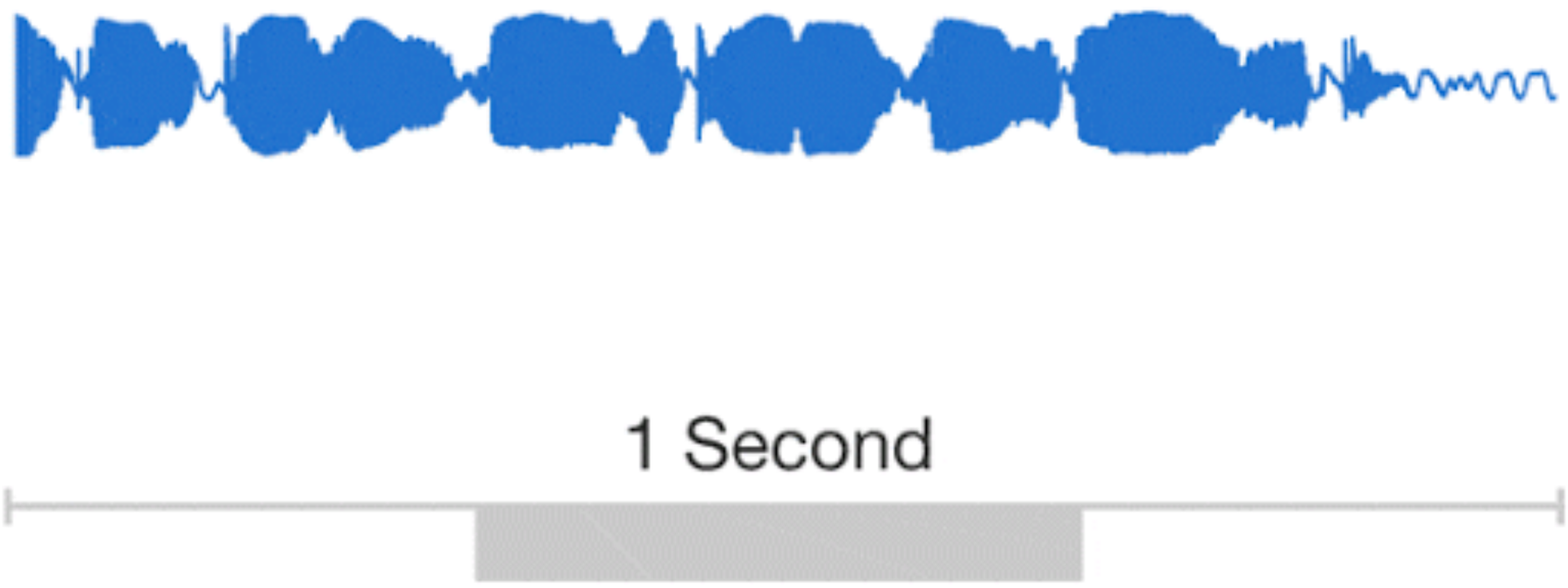
WAVENET: A GENERATIVE MODEL FOR RAW AUDIO

Autoregressive models defined by

$$p(x) = \prod_{t=1}^{T} p(x_t \mid x_{t-1}, \ldots, x_1).$$

have a tractable density. Train via maximum likelihood



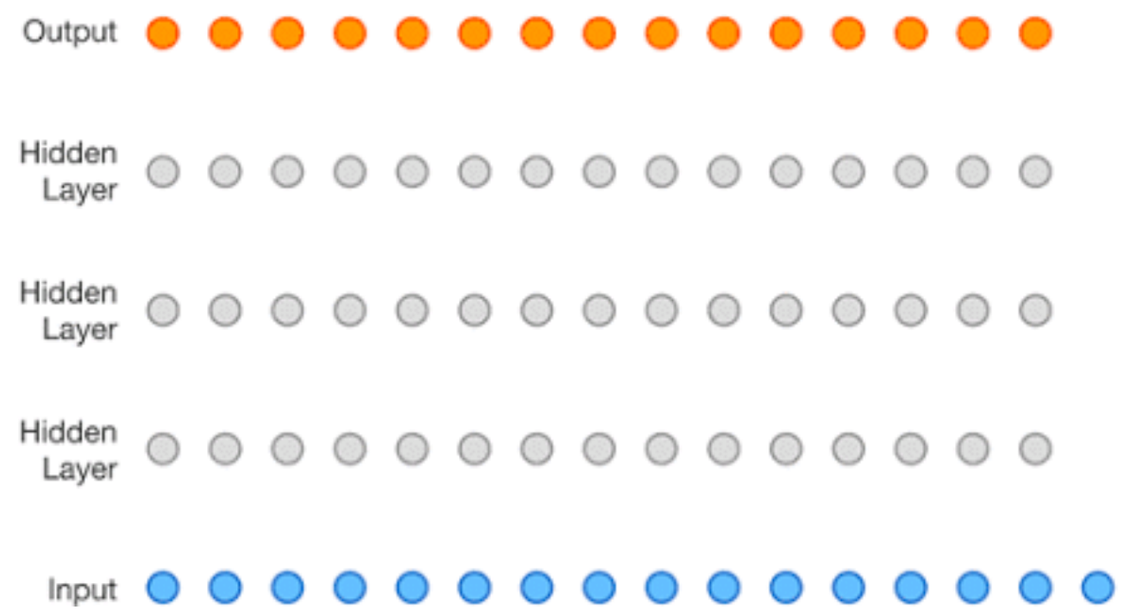


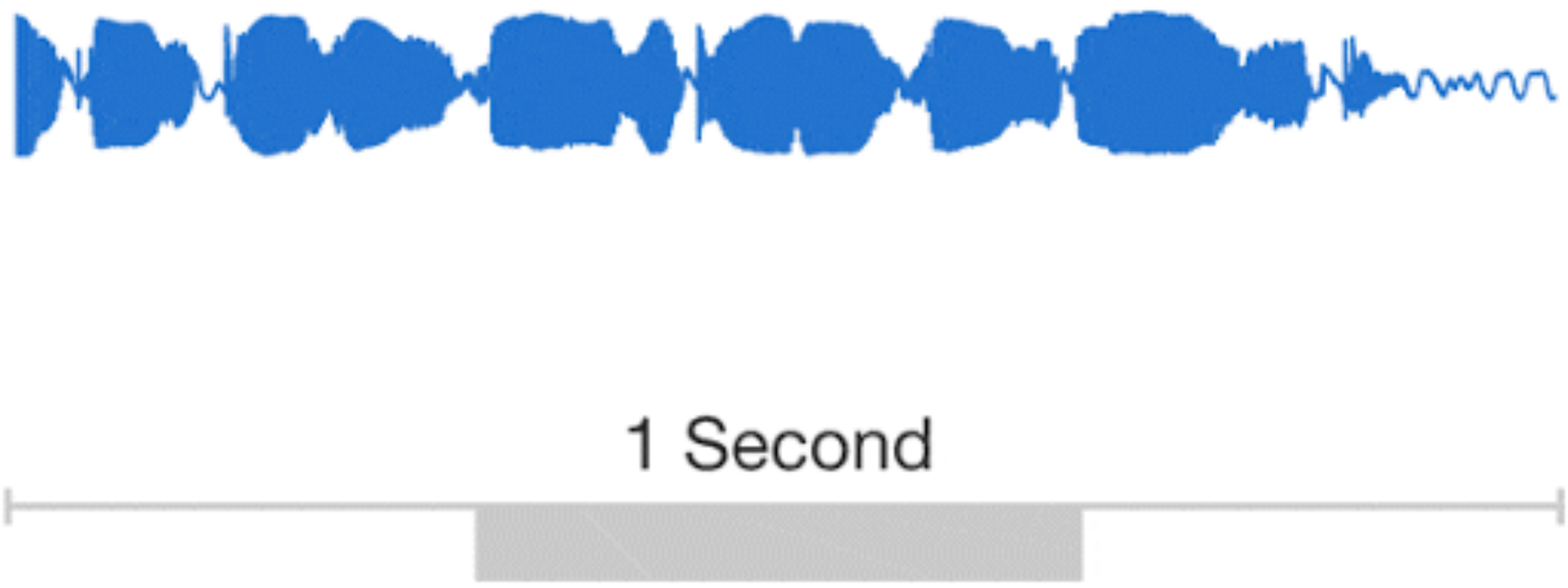
WAVENET: A GENERATIVE MODEL FOR RAW AUDIO

Autoregressive models defined by

$$p(x) = \prod_{t=1}^{T} p(x_t \mid x_{t-1}, \ldots, x_1).$$

have a tractable density. Train via maximum likelihood





Approximations using Change-of-variables

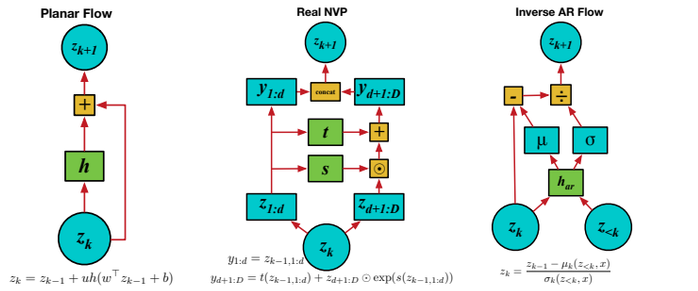
Exploit the rule for change of variables for random variables:

- Begin with an initial distribution $q_0(\mathbf{z}_0|\mathbf{x})$.
- Apply a sequence of K invertible functions f_k.

Choice of Transformation Function

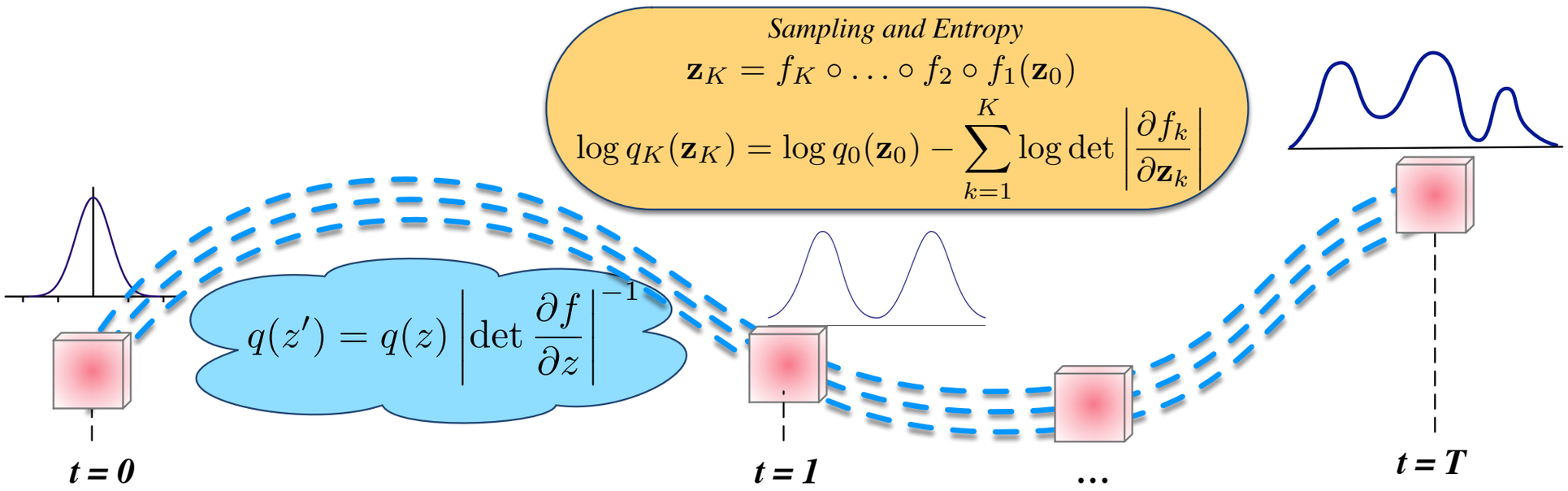
$$\mathcal{L} = \mathbb{E}_{q_0(\mathbf{z}_0)}[\log p(\mathbf{x}, \mathbf{z}_K)] - \mathbb{E}_{q_0(\mathbf{z}_0)}[\log q_0(\mathbf{z}_0)] - \mathbb{E}_{q_0(\mathbf{z}_0)}\left[\sum_{k=1}^{K} \log \det \left|\frac{\partial f_k}{\partial \mathbf{z}_k}\right|\right]$$

- Begin with a fully-factorised Gaussian and improve by change of variables.
- Triangular Jacobians allow for computational efficiency.



[Rezende and Mohamed, 2016; Dinh et al., 2016; Kingma et al., 2016]

Linear time computation of the determinant and its gradient.



Distribution flows through a sequence of invertible transforms

[Rezende and Mohamed, 2015]

FLOWS WITH CONTINUOUS TIME

FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models

Will Grathwohl[*†‡], **Ricky T. Q. Chen**[*†], **Jesse Bettencourt**[†], **Ilya Sutskever**[‡], **David Duvenaud**[†]

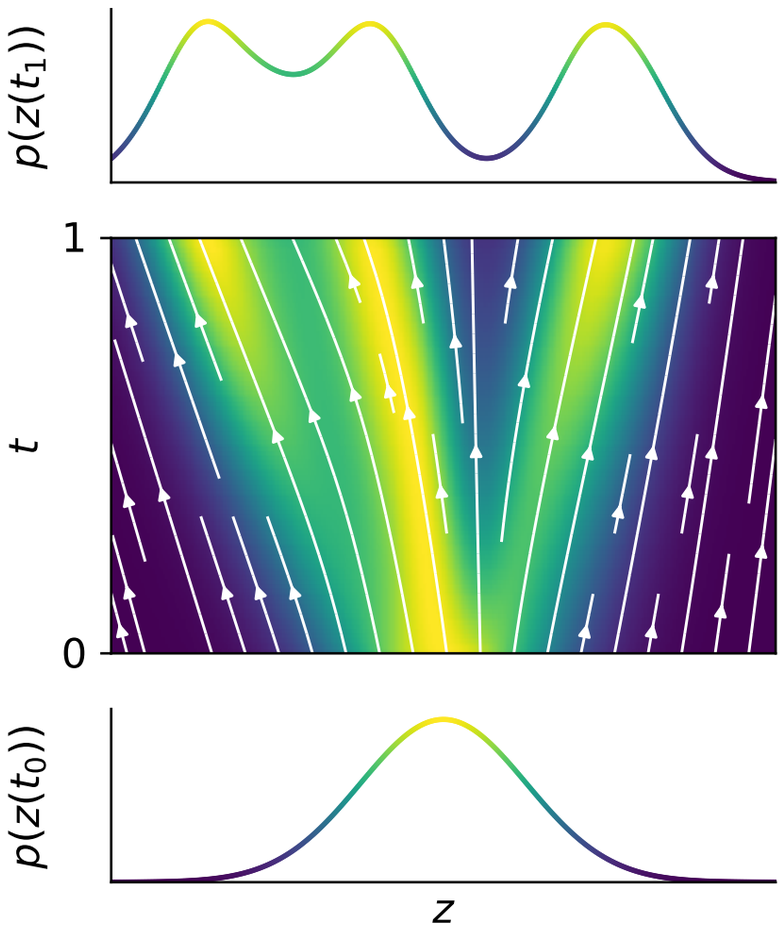
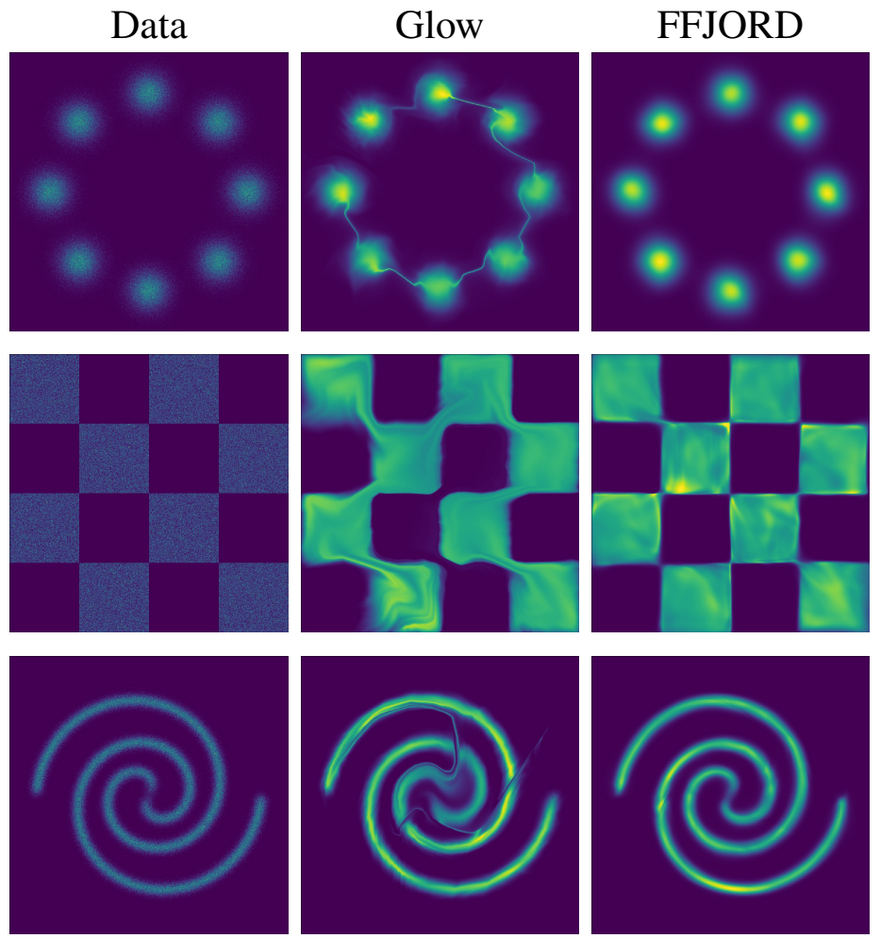


Figure 1: FFJORD transforms a simple base distribution at t_0 into the target distribution at t_1 by integrating over learned continuous dynamics.

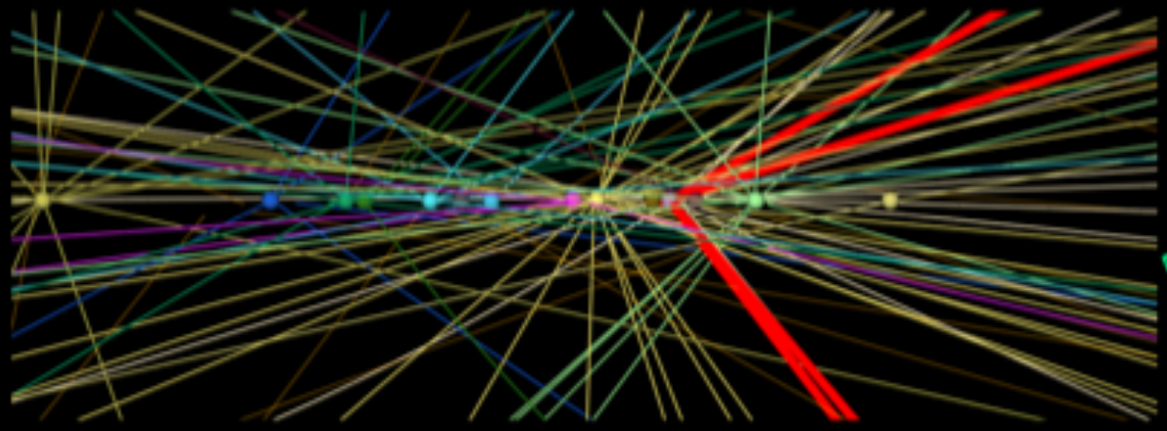


	Method	Train on data	One-pass Sampling	Exact log-likelihood	Free-form Jacobian
	Variational Autoencoders	✓	✓	✗	✓
	Generative Adversarial Nets	✓	✓	✗	✓
	Likelihood-based Autoregressive	✓	✗	✓	✗
Change of Variables	Normalizing Flows	✗	✓	✓	✗
	Reverse-NF, MAF, TAN	✓	✗	✓	✗
	NICE, Real NVP, Glow, Planar CNF	✓	✓	✓	✗
	FFJORD	✓	✓	✓	✓

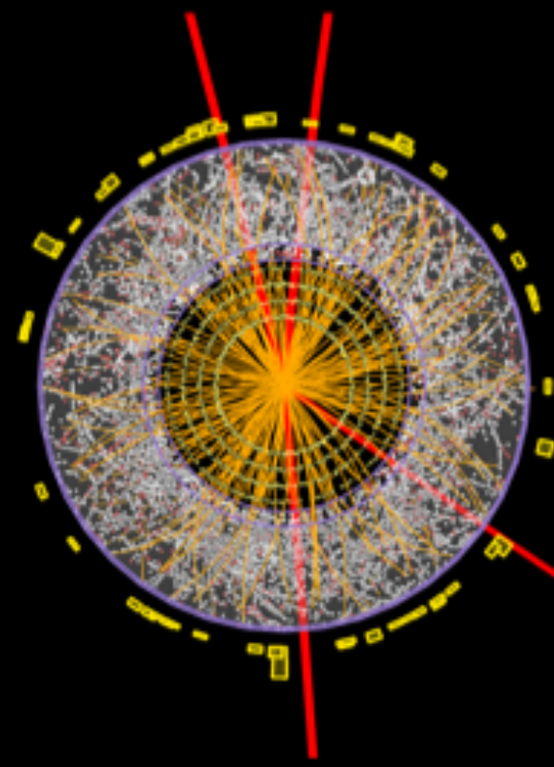
Table 1: A comparison of the abilities of generative modeling approaches.

Examples of Simulation-Based Inference
(aka Likelihood-free inference)

THE HIGGS BOSON



ATLAS EXPERIMENT
http://atlas.ch



Run: 204769
Event: 71902630
Date: 2012-06-10
Time: 13:24:31 CEST

THE CAUSAL, GENERATIVE MODEL

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

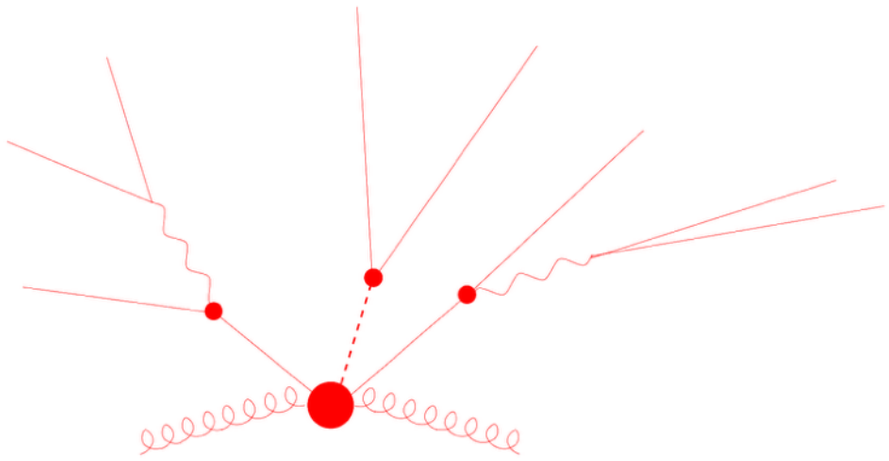
THE CAUSAL, GENERATIVE MODEL

$$\mathcal{L}_{SM} = \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}}_{\text{kinetic energies and self-interactions of the gauge bosons}}$$

$$+ \underbrace{\bar{L}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})L + \bar{R}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g'YB_{\mu})R}_{\text{kinetic energies and electroweak interactions of fermions}}$$

$$+ \underbrace{\frac{1}{2}\left|(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})\,\phi\right|^{2} - V(\phi)}_{W^{\pm},Z,\gamma,\text{and Higgs masses and couplings}}$$

$$+ \underbrace{g''(\bar{q}\gamma^{\mu}T_{a}q)\,G^{a}_{\mu}}_{\text{interactions between quarks and gluons}} + \underbrace{(G_{1}\bar{L}\phi R + G_{2}\bar{L}\phi_{c}R + h.c.)}_{\text{fermion masses and couplings to Higgs}}$$



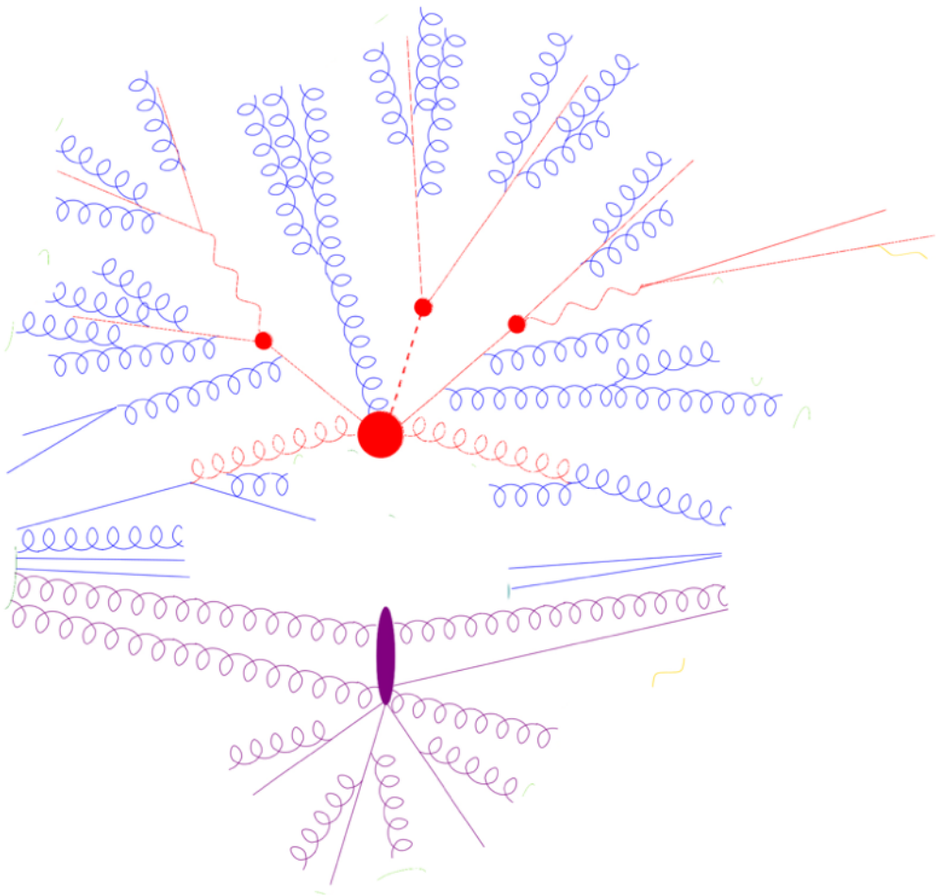
THE CAUSAL, GENERATIVE MODEL

$$\mathcal{L}_{SM} = \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}}_{\text{kinetic energies and self-interactions of the gauge bosons}}$$

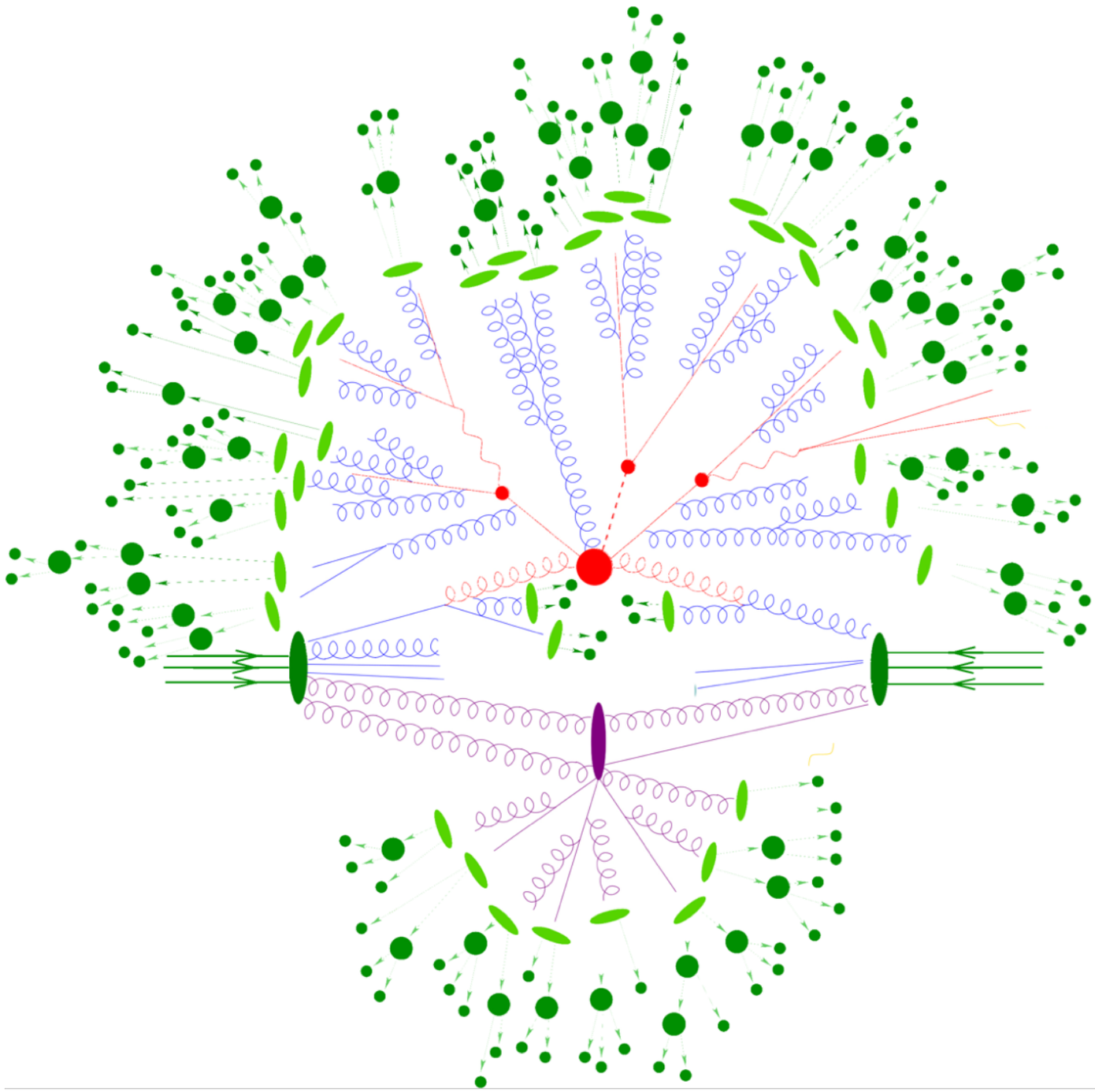
$$+ \underbrace{\bar{L}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})L + \bar{R}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g'YB_{\mu})R}_{\text{kinetic energies and electroweak interactions of fermions}}$$

$$+ \underbrace{\frac{1}{2}\left|(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})\,\phi\right|^{2} - V(\phi)}_{W^{\pm},Z,\gamma,\text{and Higgs masses and couplings}}$$

$$+ \underbrace{g''(\bar{q}\gamma^{\mu}T_{a}q)\,G^{a}_{\mu}}_{\text{interactions between quarks and gluons}} + \underbrace{(G_{1}\bar{L}\phi R + G_{2}\bar{L}\phi_{c}R + h.c.)}_{\text{fermion masses and couplings to Higgs}}$$



THE CAUSAL, GENERATIVE MODEL



$$\mathcal{L}_{SM} = \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}}_{\text{kinetic energies and self-interactions of the gauge bosons}}$$

$$+ \underbrace{\bar{L}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})L + \bar{R}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g'YB_{\mu})R}_{\text{kinetic energies and electroweak interactions of fermions}}$$

$$+ \underbrace{\frac{1}{2}\left|(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})\,\phi\right|^{2} - V(\phi)}_{W^{\pm},Z,\gamma,\text{and Higgs masses and couplings}}$$

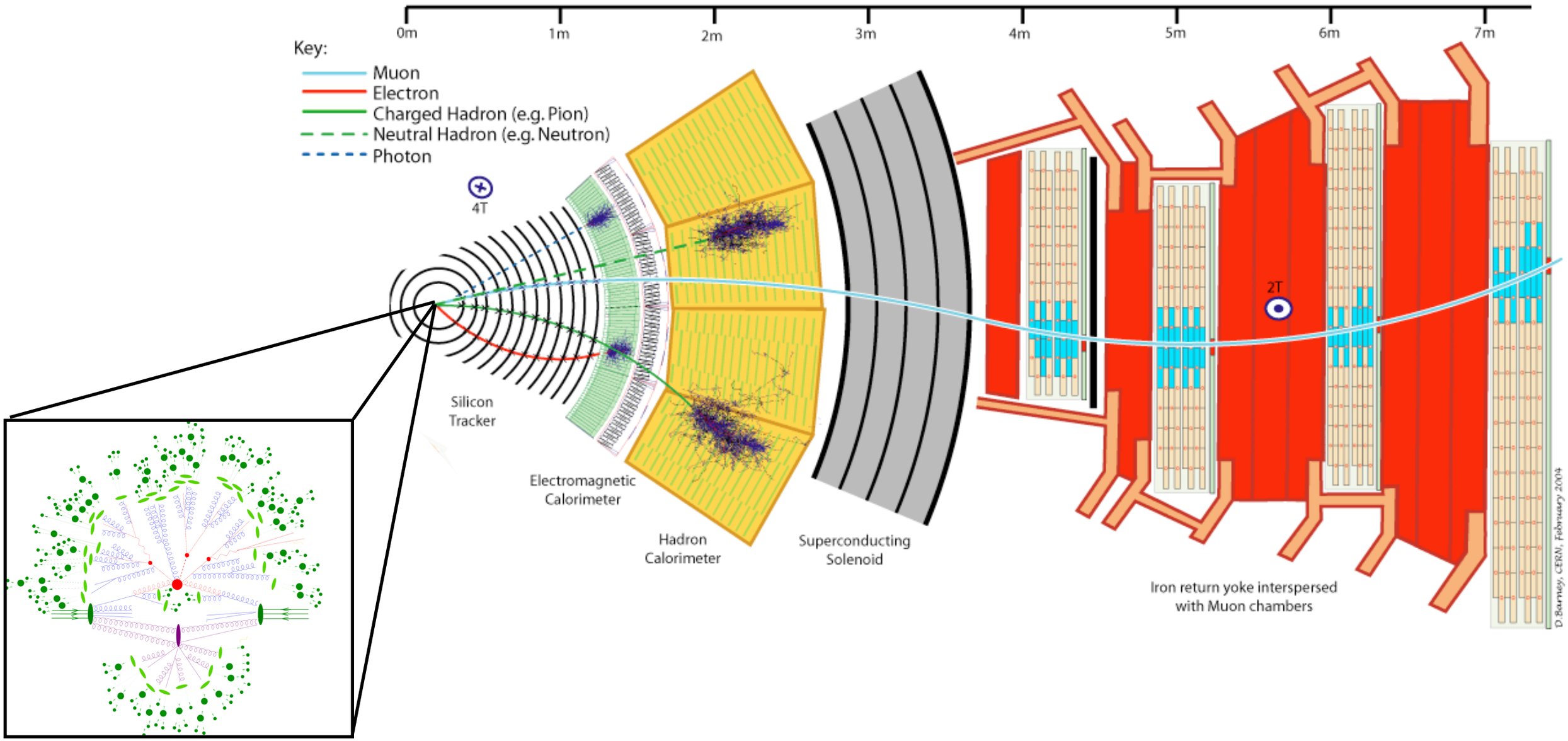
$$+ \underbrace{g''(\bar{q}\gamma^{\mu}T_{a}q)\,G^{a}_{\mu}}_{\text{interactions between quarks and gluons}} + \underbrace{(G_{1}\bar{L}\phi R + G_{2}\bar{L}\phi_{c}R + h.c.)}_{\text{fermion masses and couplings to Higgs}}$$

THE CAUSAL, GENERATIVE MODEL

Conceptually: Prob(detector response | particles)

Implementation: Monte Carlo integration over micro-physics

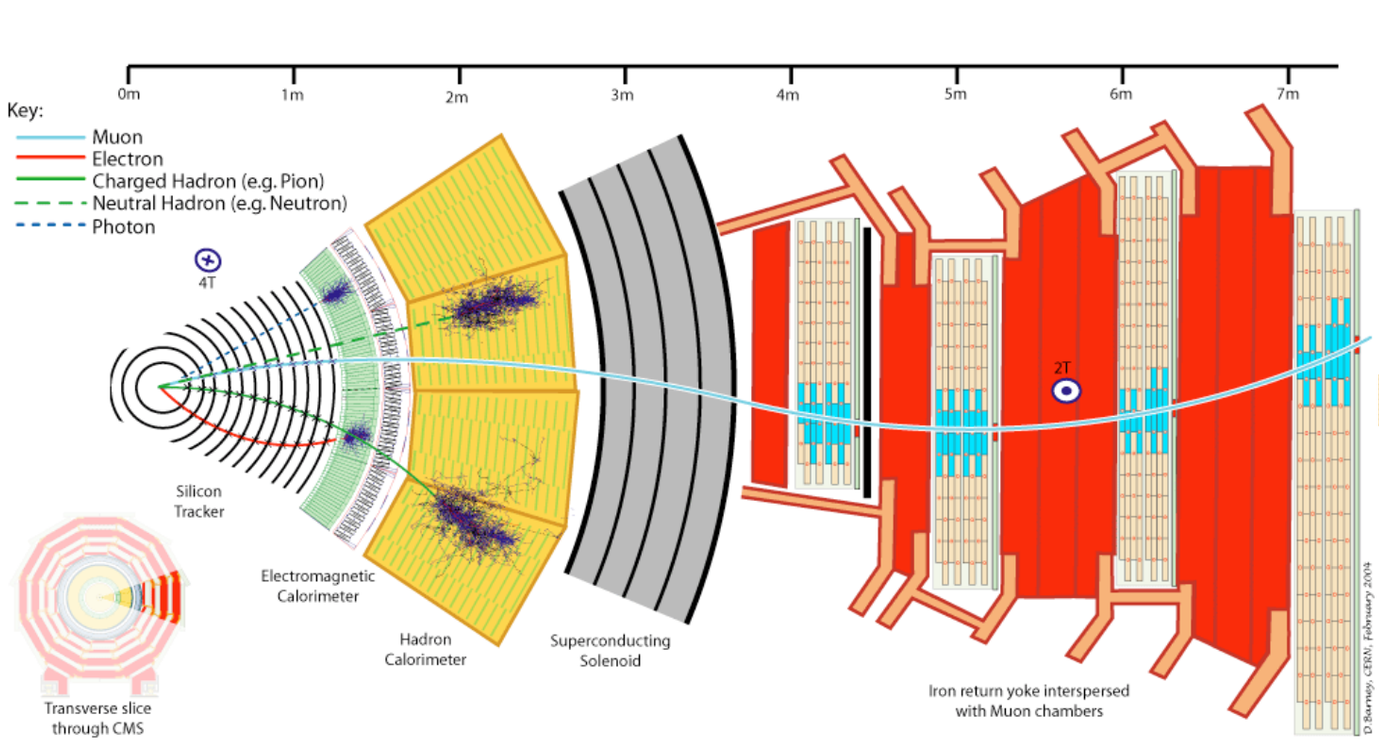
Consequence: evaluation of the likelihood is intractable

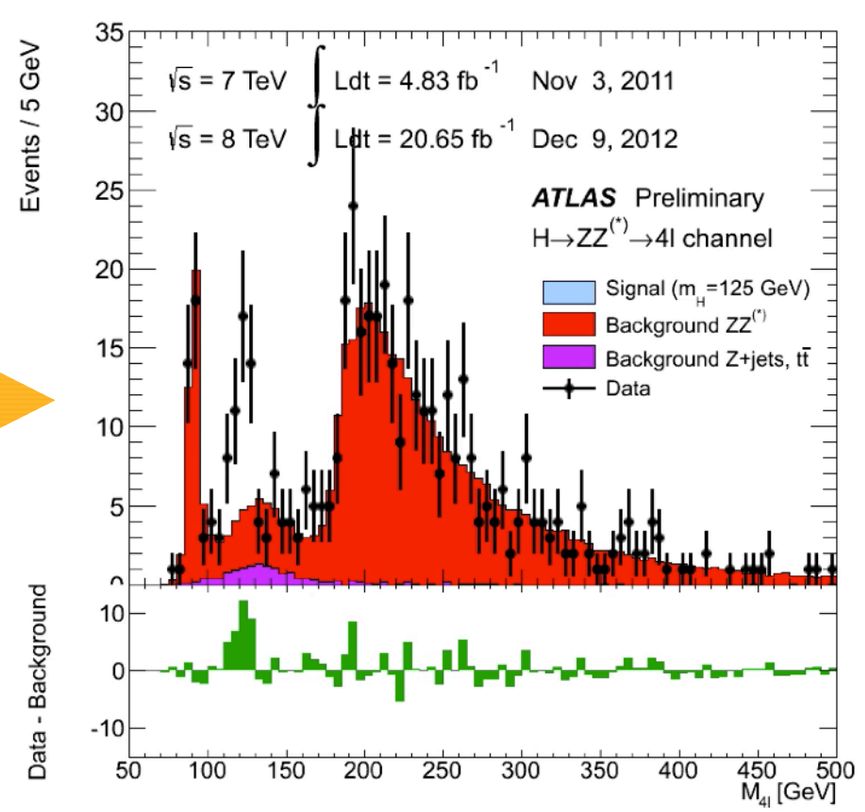


10^8 SENSORS → 1 REAL-VALUED QUANTITY

Most measurements and searches for new particles at the LHC are based on the distribution of a single summary statistic

- choosing a good summary statistic (feature engineering) is a task for a skilled physicist and tailored to the goal of measurement or new particle search
- likelihood $p(x|\theta)$ **approximated** using histograms (univariate density estimation)

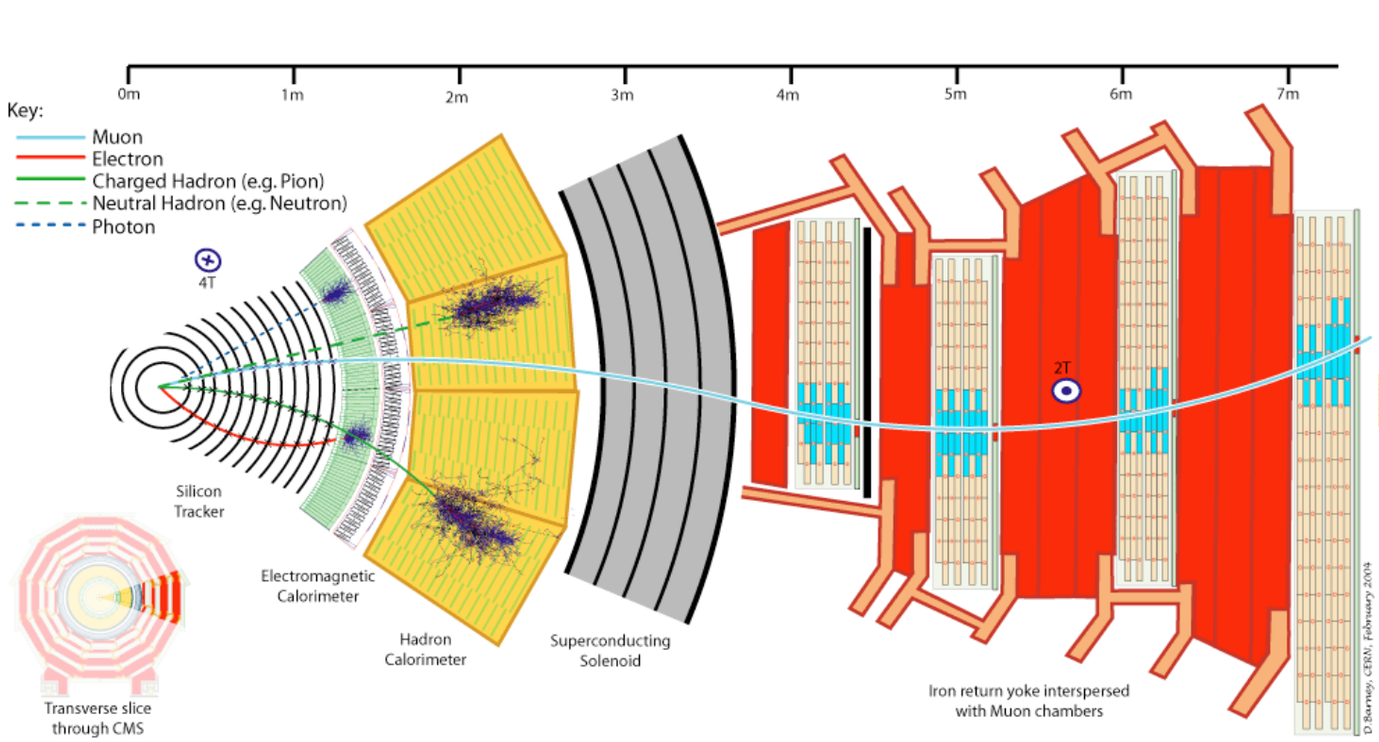


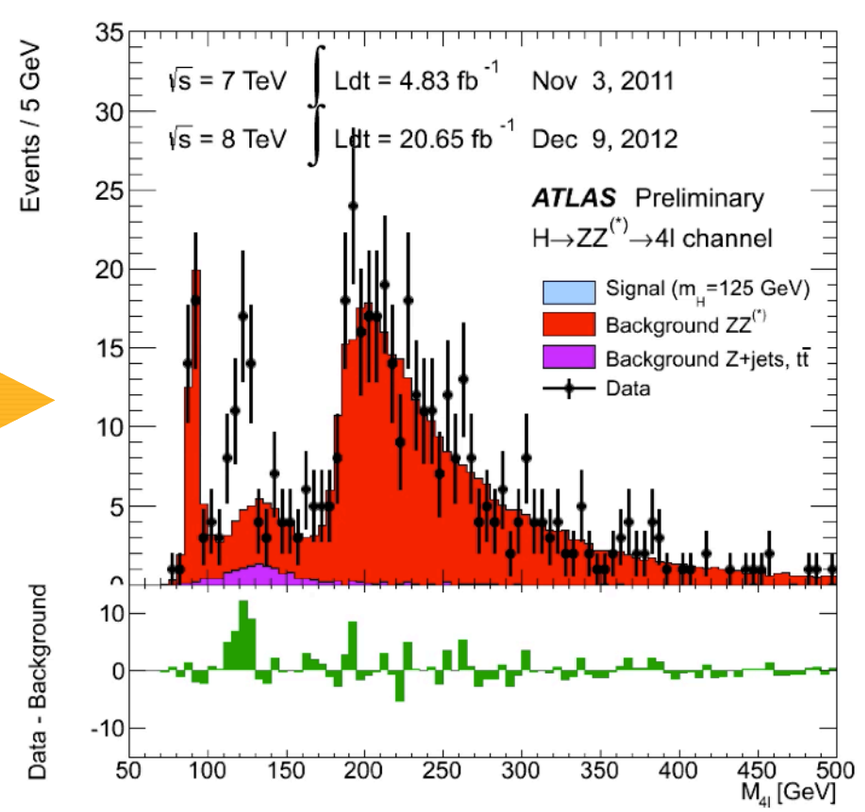


10^8 SENSORS → 1 REAL-VALUED QUANTITY

Most measurements and searches for new particles at the LHC are based on the distribution of a single summary statistic

- choosing a good summary statistic (feature engineering) is a task for a skilled physicist and tailored to the goal of measurement or new particle search
- likelihood $p(x|\theta)$ **approximated** using histograms (univariate density estimation)





THE CRUX, AN INTRACTABLE INTEGRAL

MC Sampling

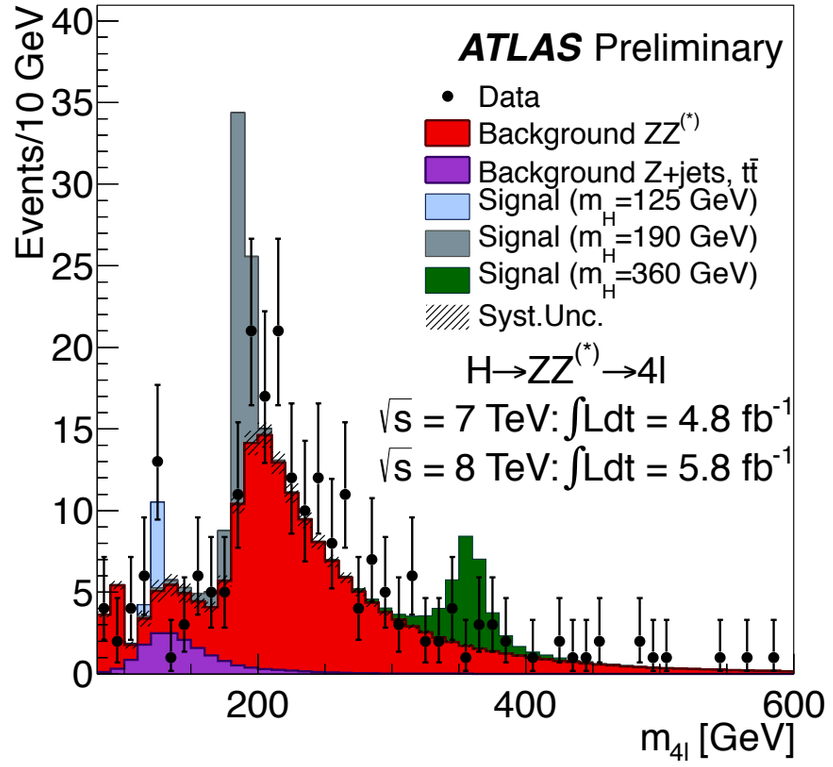
observed

simulation

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

$$\hat{p}(x|\theta)$$

histogram approximation



THE CRUX, AN INTRACTABLE INTEGRAL

observed

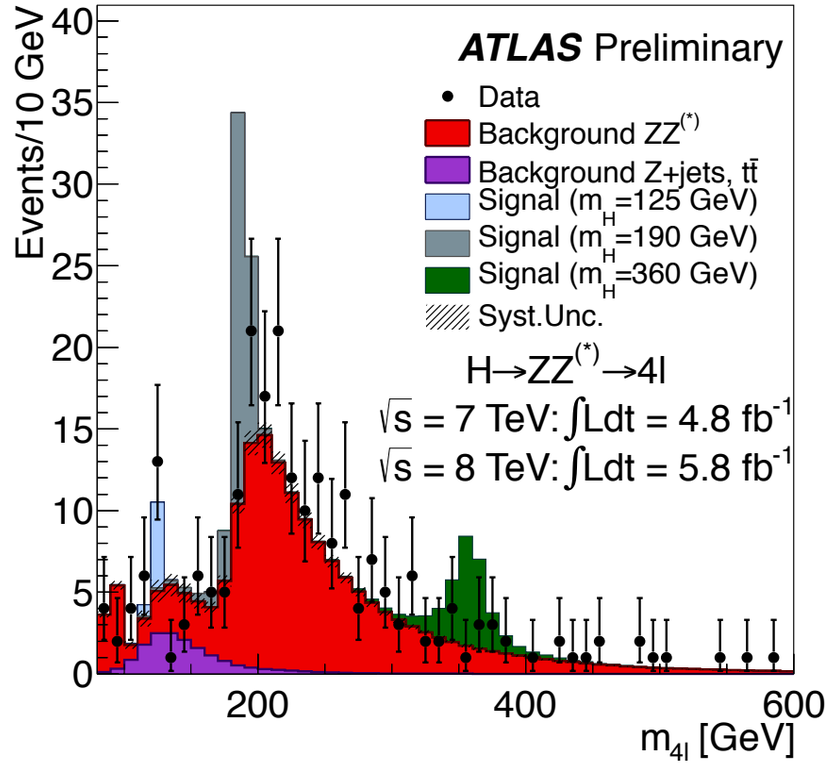
MC Sampling

simulation

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

$$\hat{p}(x|\theta)$$

histogram approximation

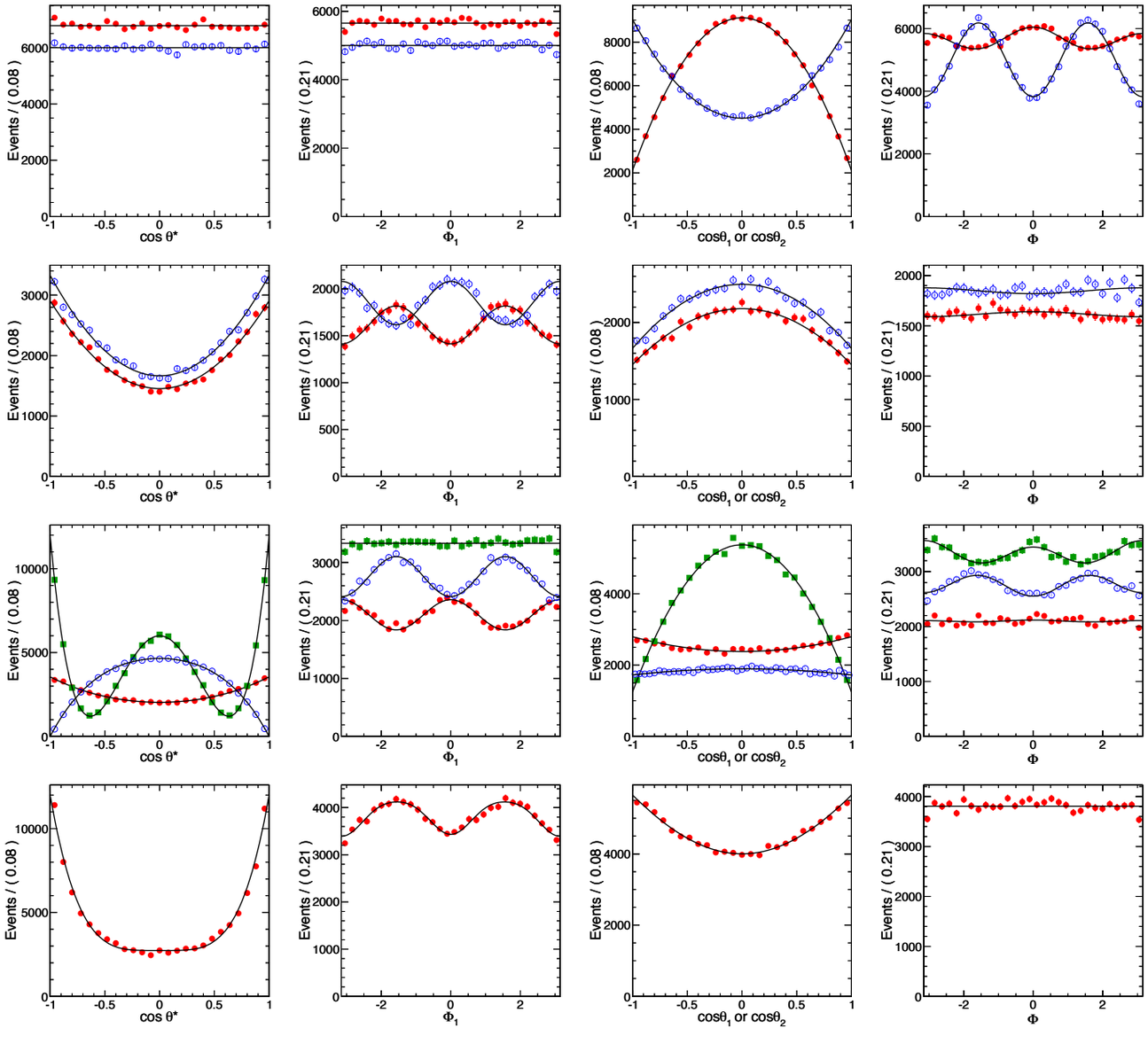


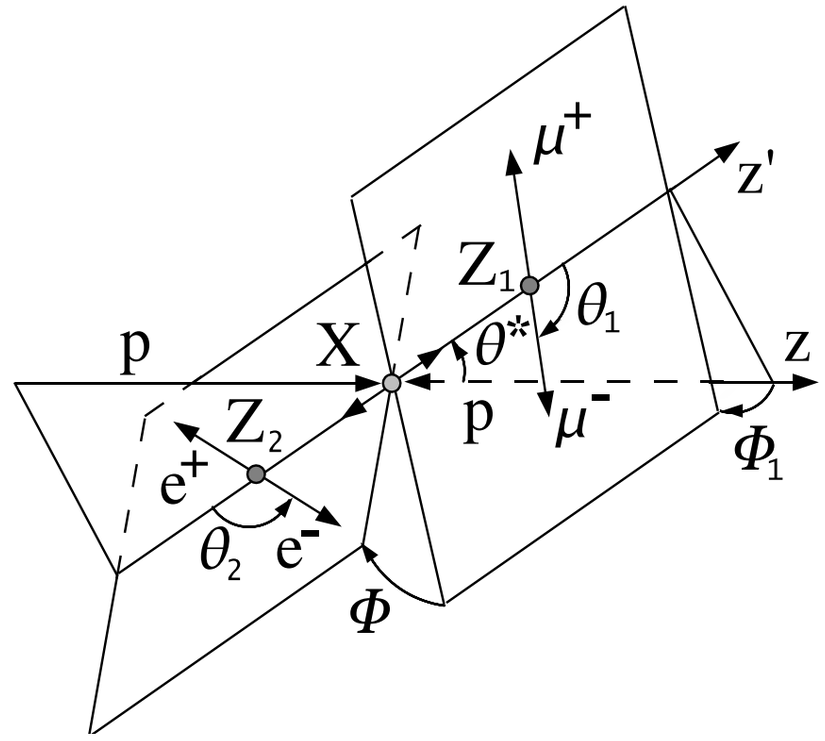
This doesn't scale if x is high dimensional!

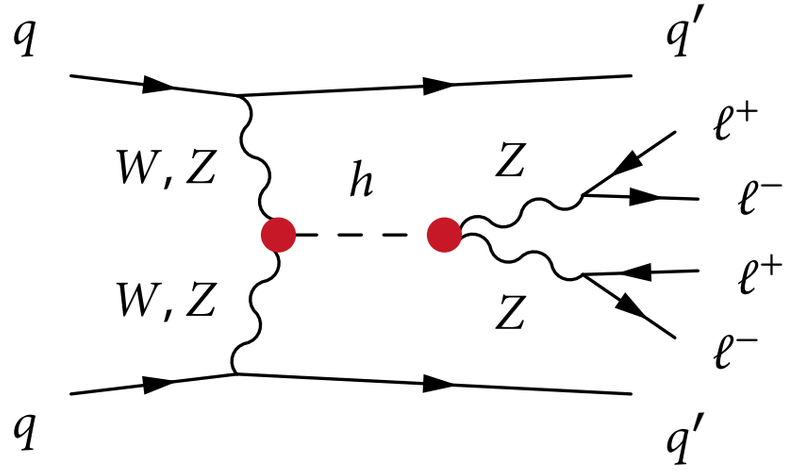
HIGH DIMENSIONAL EXAMPLE

When looking for deviations from the standard model Higgs, we would like leverage subtle kinematic correlations

- thus each observation **x** is high-dimensional







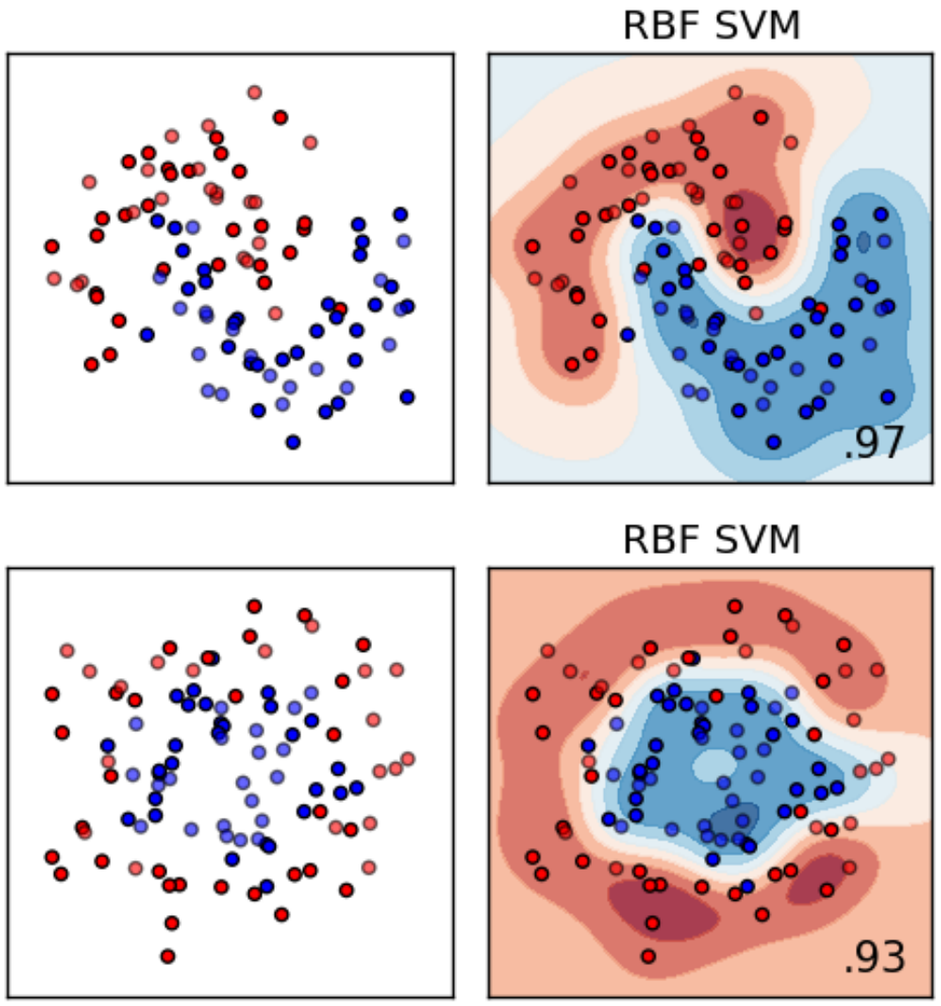
The Problem:

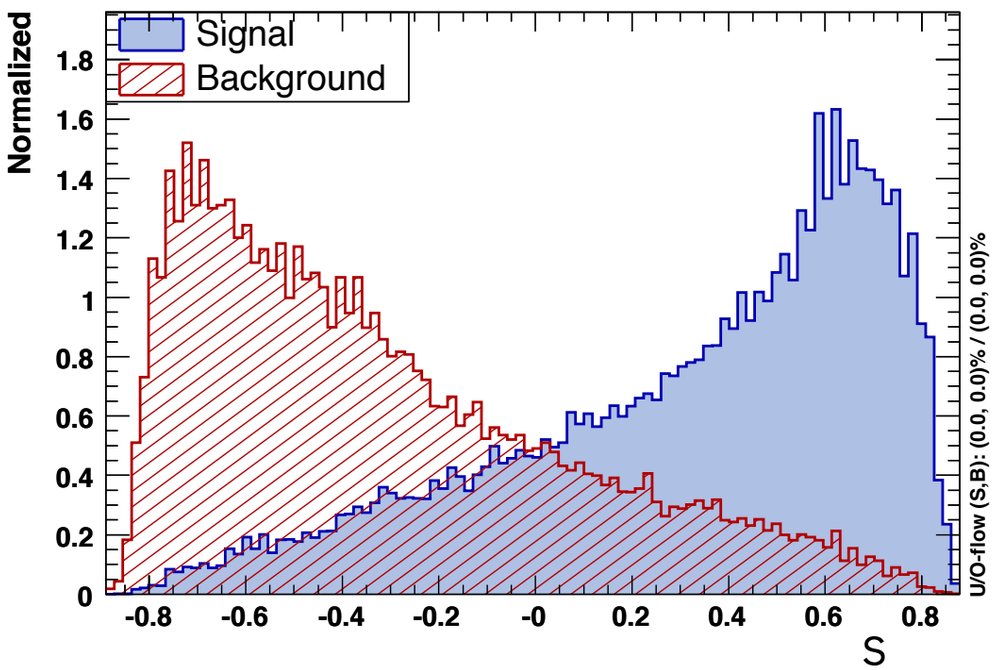
This doesn't scale if **x** is high dimensional!

How much are we loosing in feature engineering?

What if we don't know how to design a good feature?

LIKELIHOOD RATIO TRICK





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

$$L[s] = \int p(x|H_0)\,(0 - s(x))^2\,dx + \int p(x|H_1)\,(1 - s(x))^2 dx$$

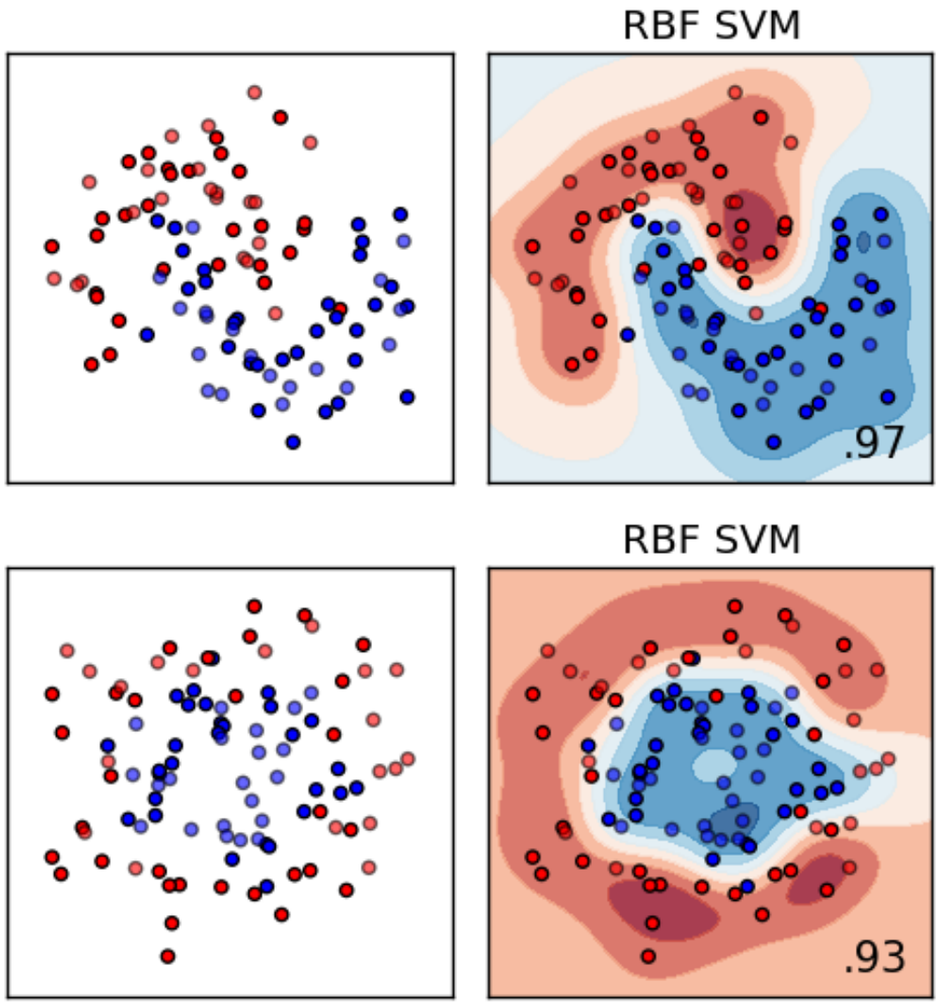
- i.e. approximate the Bayes optimal classifier

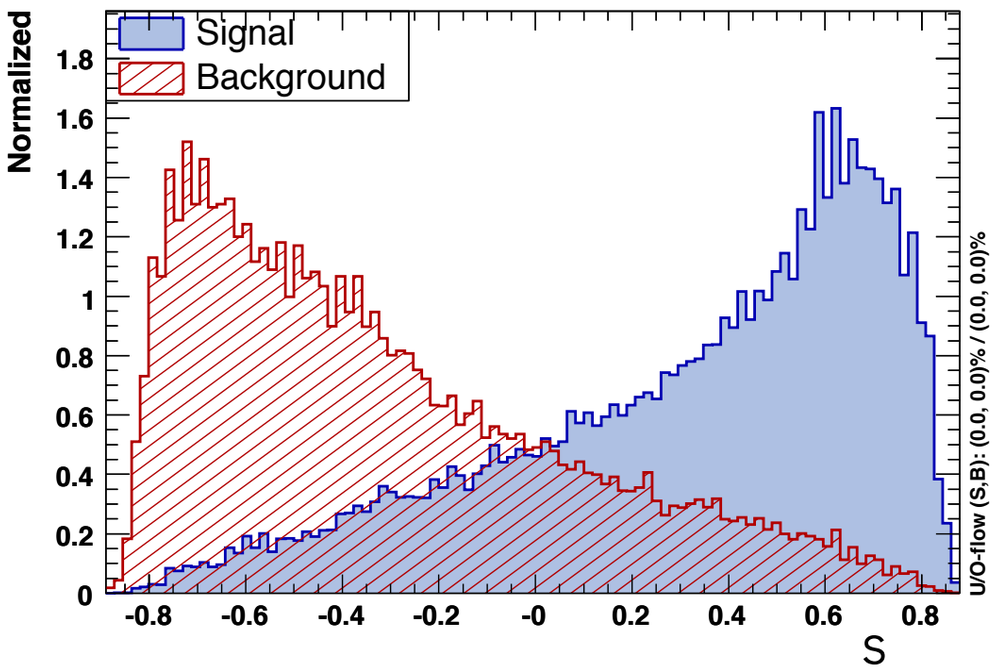
$$s(x) = \frac{p(x|H_1)}{p(x|H_0) + p(x|H_1)}$$

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

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

LIKELIHOOD RATIO TRICK





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

$$L[s] = \int p(x|H_0)\,(0 - s(x))^2\,dx + \int p(x|H_1)\,(1 - s(x))^2 dx$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}(y_i - s(x_i))^2$$

- i.e. approximate the Bayes 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

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

GANs AND THE LIKELIHOOD RATIO TRICK

The discriminator of a GAN approximates

$$s(x) = \frac{p(x|G)}{p(x|D) + p(x|G)}$$

Which is one-to-one with the likelihood ratio

$$\frac{p(x|D)}{p(x|G)} = 1 - \frac{1}{s(x)}$$

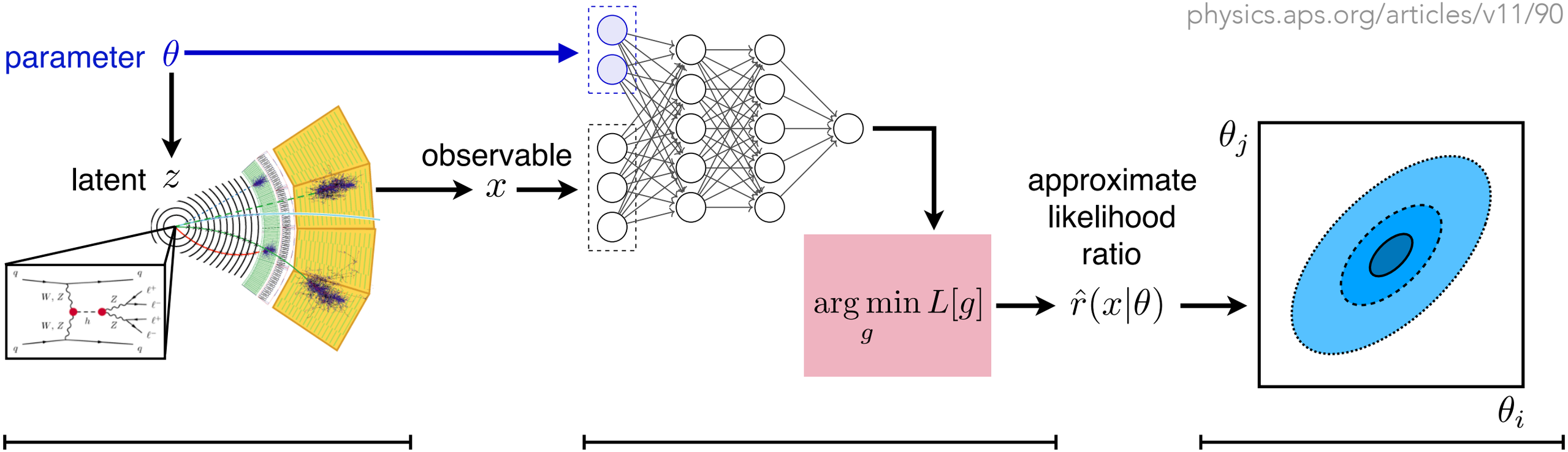
Can do the same thing for any two points θ_0 & θ_1 in parameter space Θ. I call this a **parametrized classifier**

$$s(x;\theta_0,\theta_1) = \frac{p(x|\theta_1)}{p(x|\theta_0) + p(x|\theta_1)}$$

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

LIKELIHOOD-FREE INFERENCE

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



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

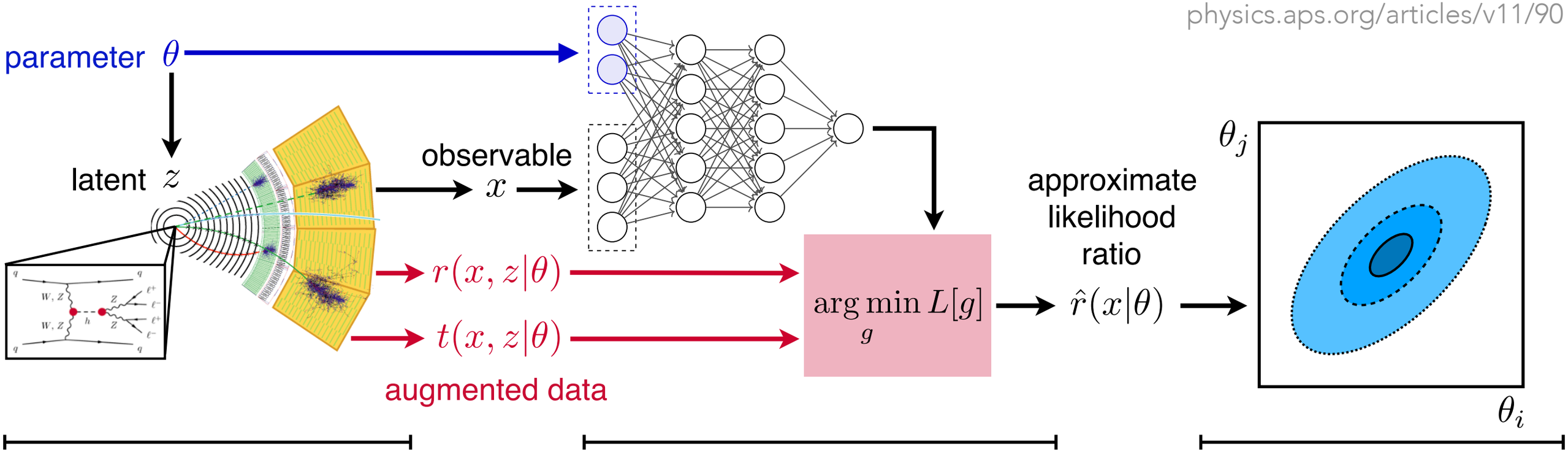
Currently a 2-stage process:

1. learning surrogate
2. Inference on parameters of simulator

Wanted: theory with **joint treatment** of the two stages

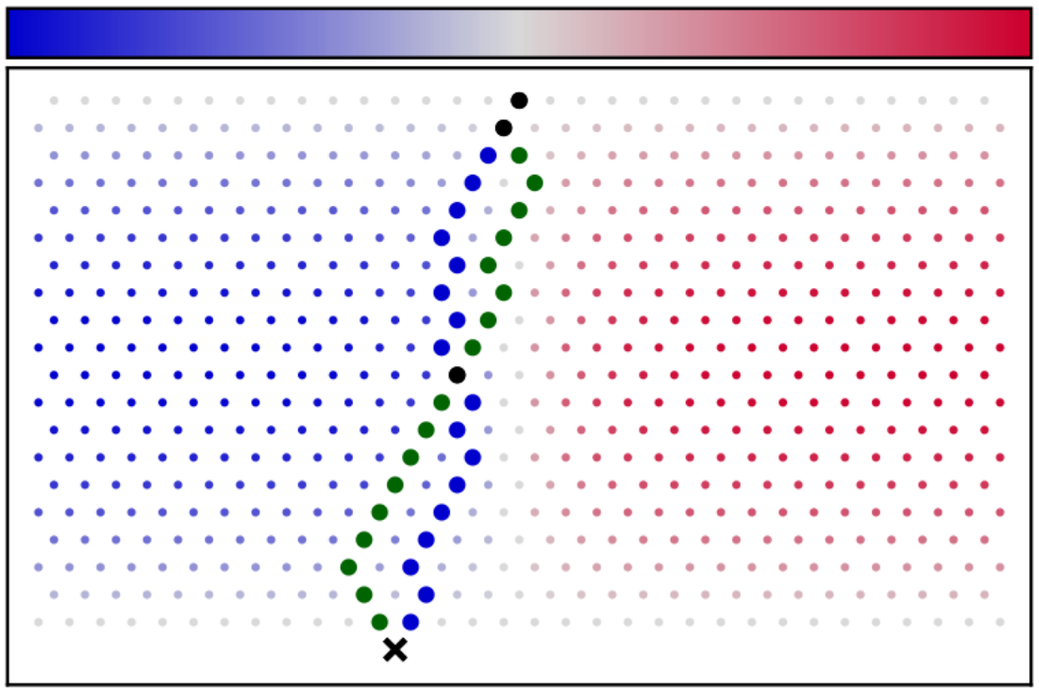
LEARNING THE LIKELIHOOD RATIO

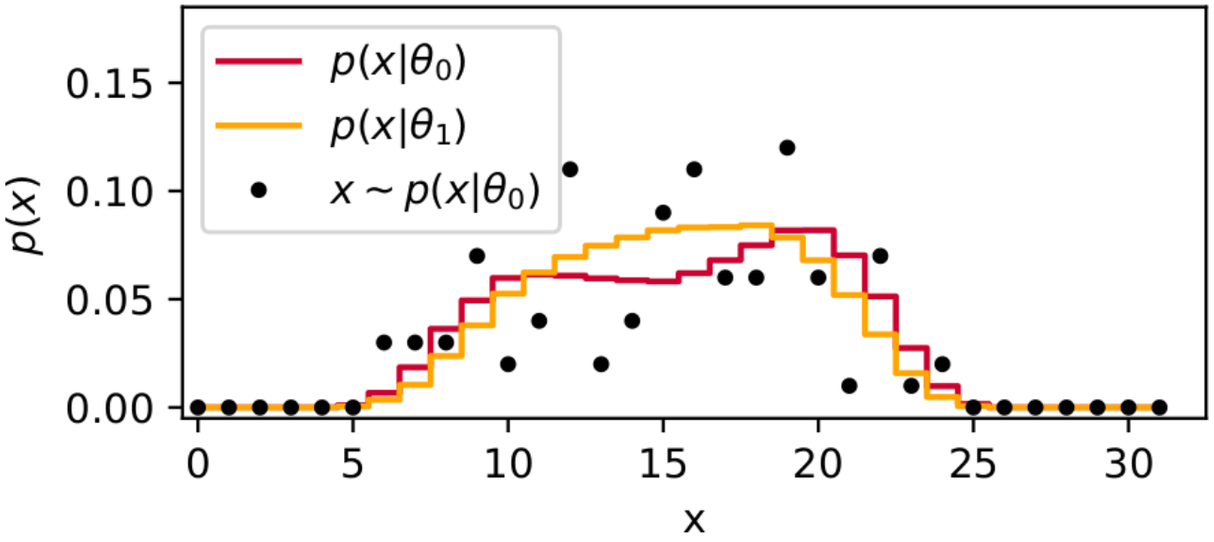
arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90



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

MINING GOLD





While implicit density is intractable

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

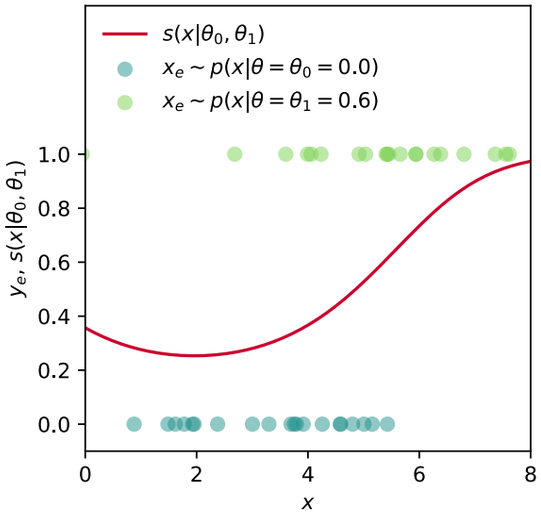
Some quantities conditioned on latent z are tractable:

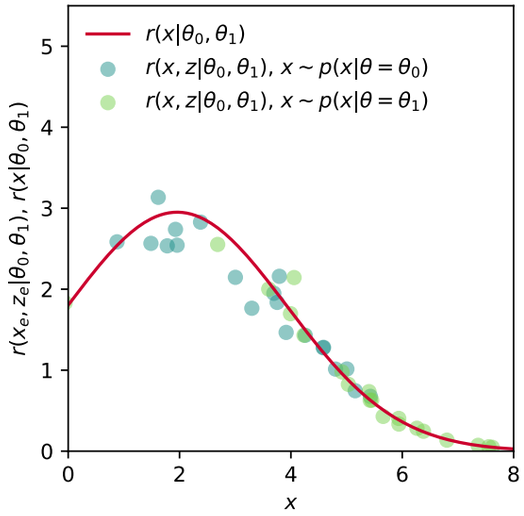
$$r(x, z|\theta_0, \theta_1) = \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)}$$

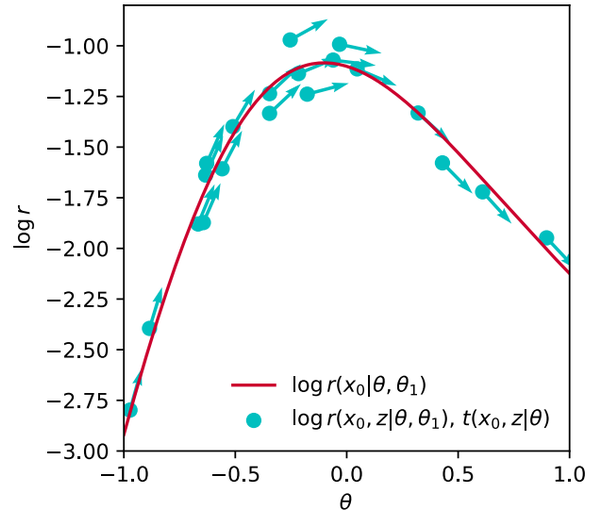
and similar to REINFORCE policy gradient

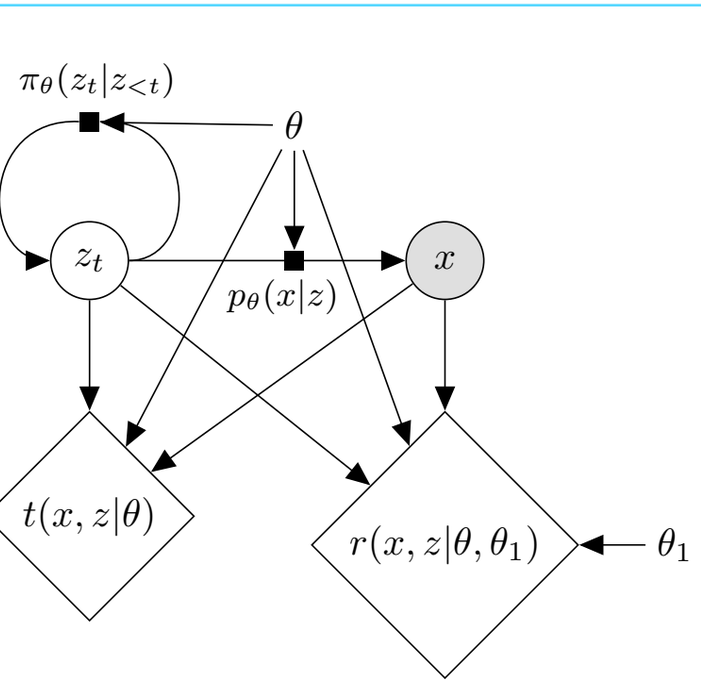
$$t(x, z|\theta_0) = \frac{\nabla_\theta p(x, z|\theta)|_{\theta_0}}{p(x, z|\theta_0)} = \nabla_\theta \log p(x, z|\theta)|_{\theta_0}$$

PUTTING IT ALL TOGETHER

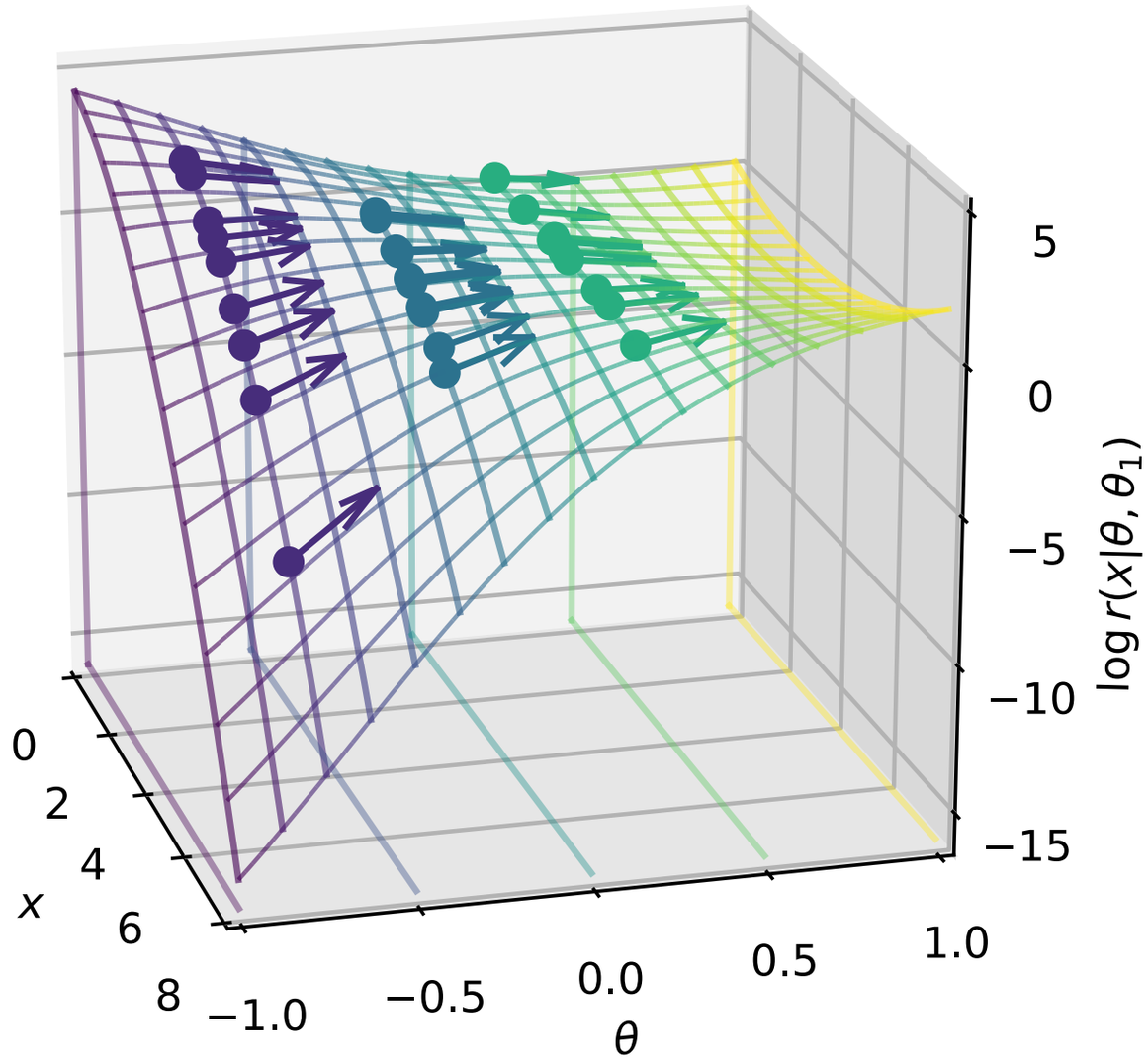








can think of simulator as policy πθ in language of reinforcement learning



HOW DO YOU USE THE GOLD?

We have joint likelihood ratio

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

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

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

One can show it is minimized by

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

We want likelihood ratio

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

LEARNING THE SCORE

Similar to the joint likelihood ratio, we can calculate the joint score

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

We want score

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

LEARNING THE SCORE

Similar to the joint likelihood ratio, we can calculate the joint score

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

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

$$t(x|\theta_0) = \arg\min_{\hat{t}(x|\theta_0)} L_t[\hat{t}(x|\theta_0)] \; .$$

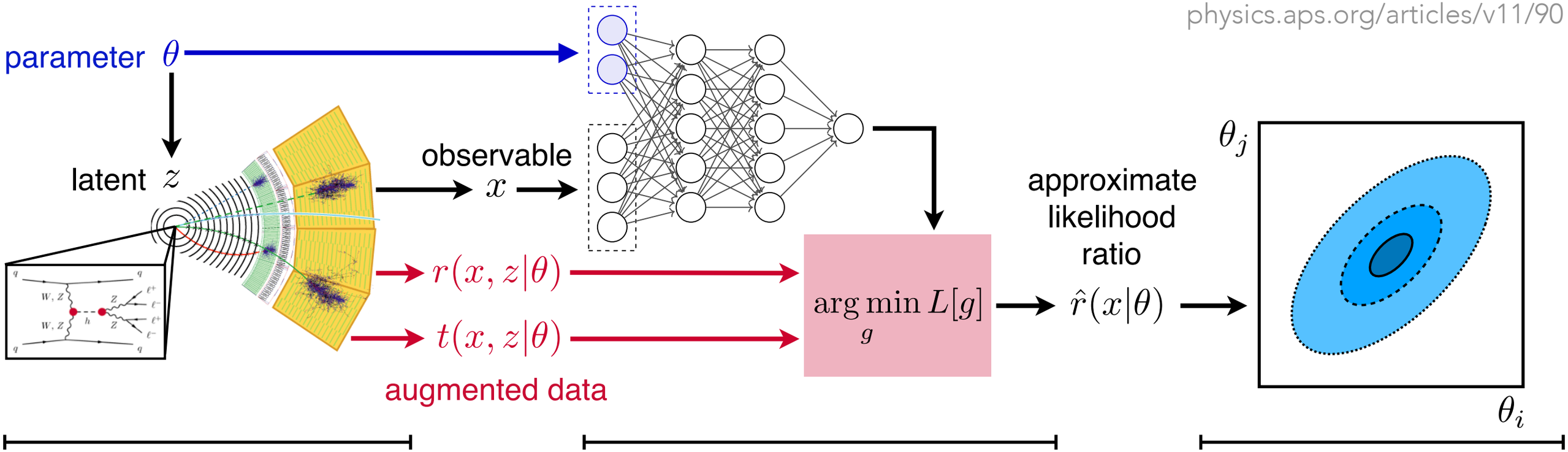
We want score

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

Again, we implement this minimization through machine learning

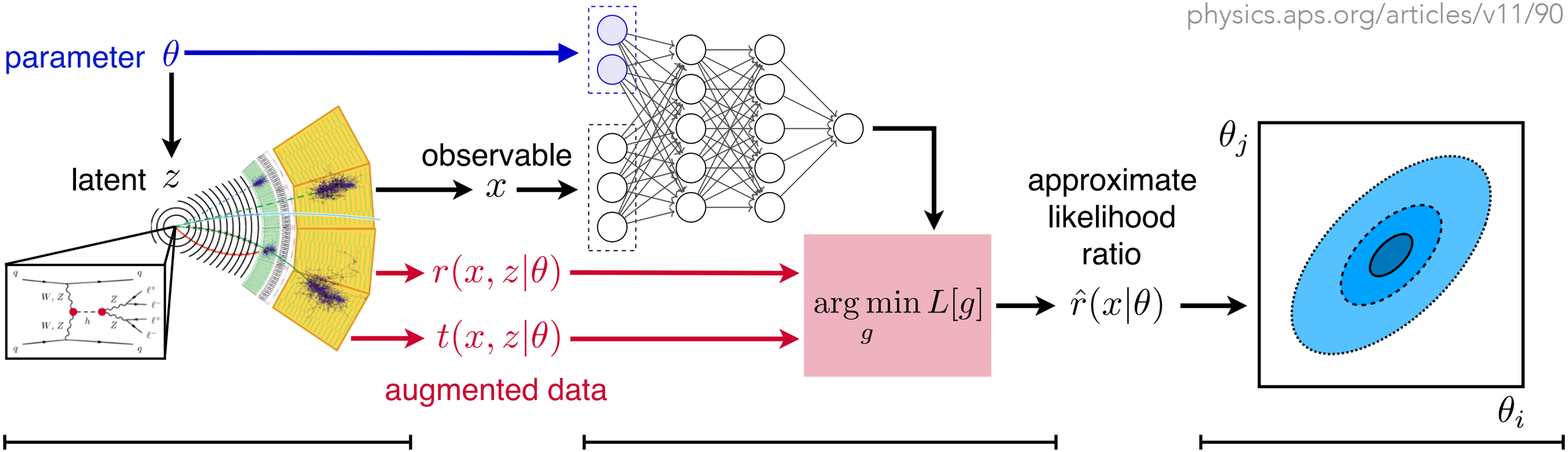
LEARNING THE LIKELIHOOD RATIO

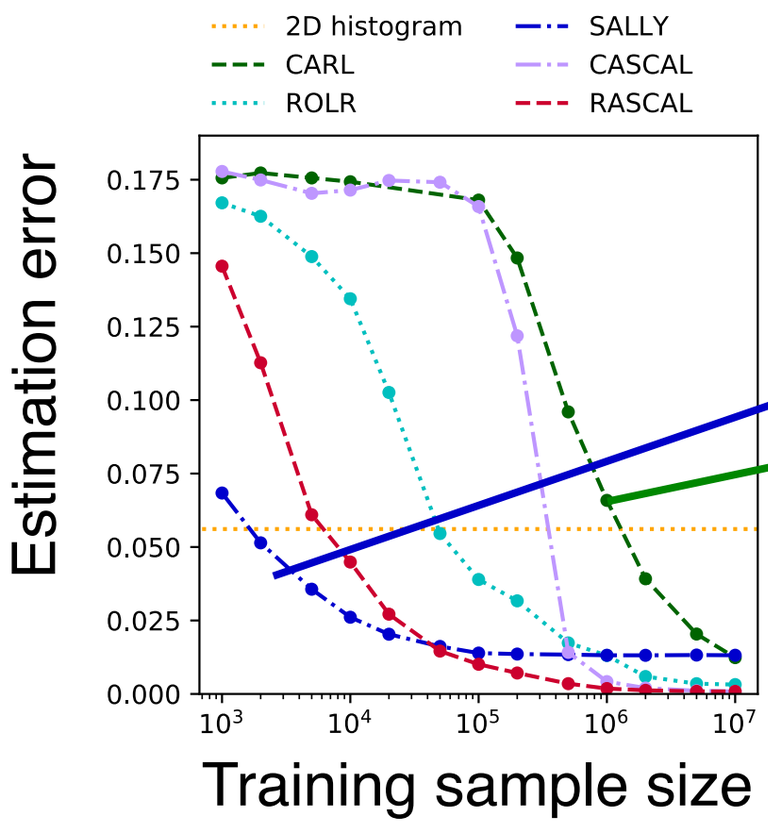
arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90



LEARNING THE LIKELIHOOD RATIO

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



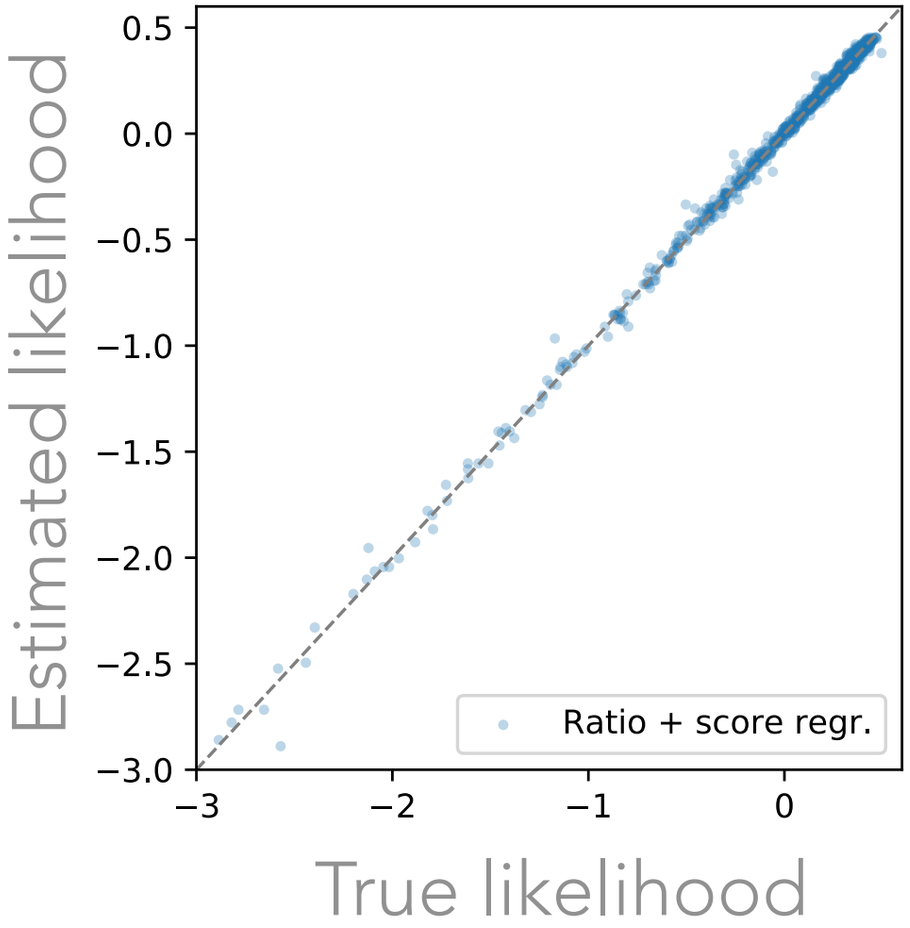


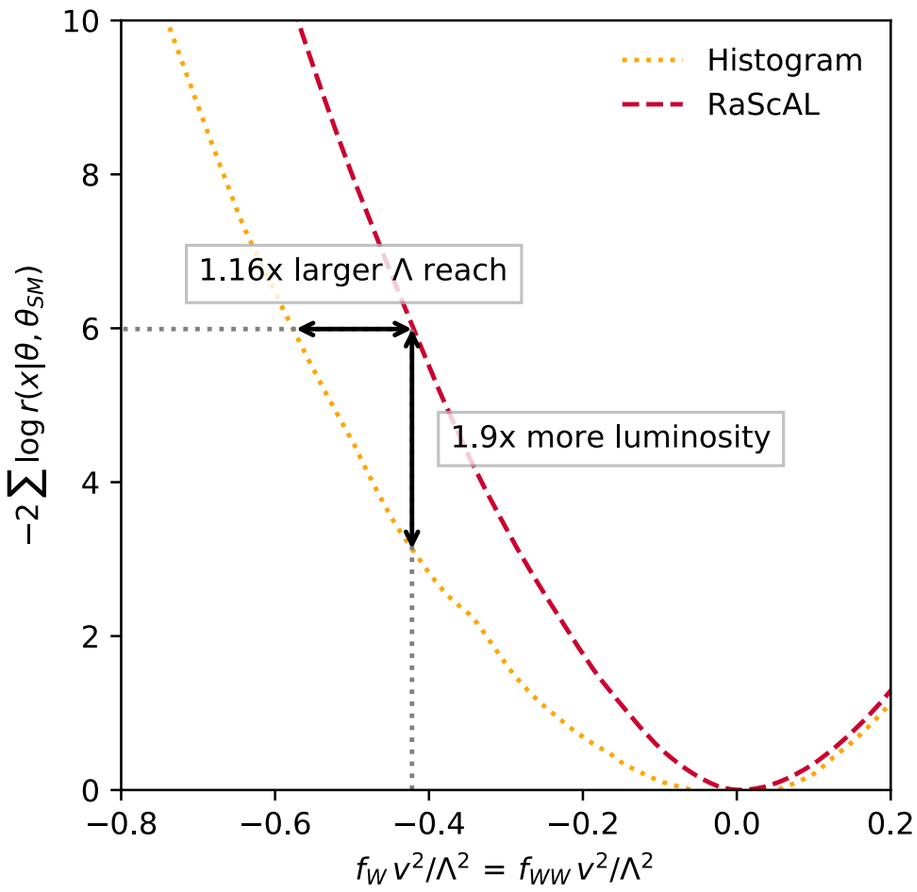
New techniques
require less data than
without augmented data

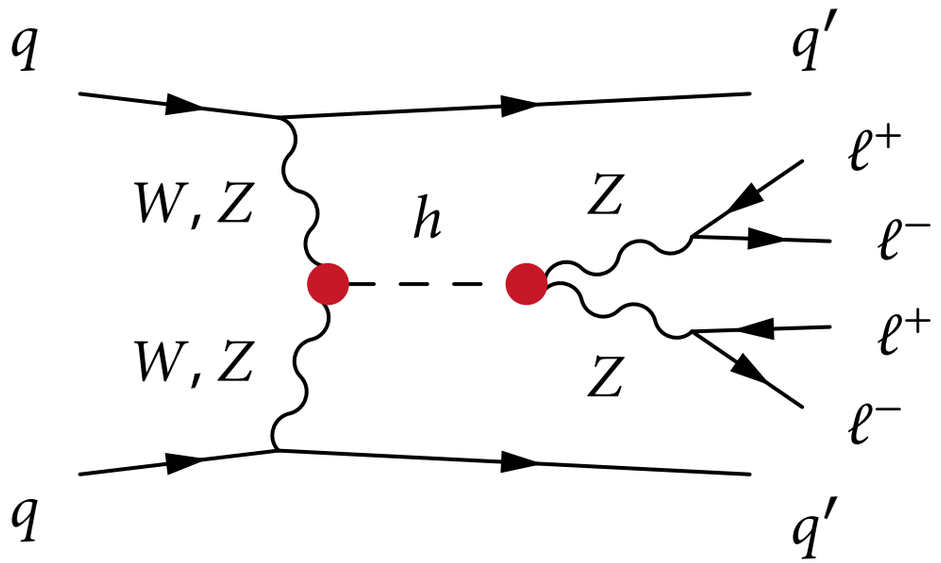
Traditional Approach no NN

IMPACT ON STUDIES OF THE HIGGS BOSON

(based on a 42-Dim observation ***x***)





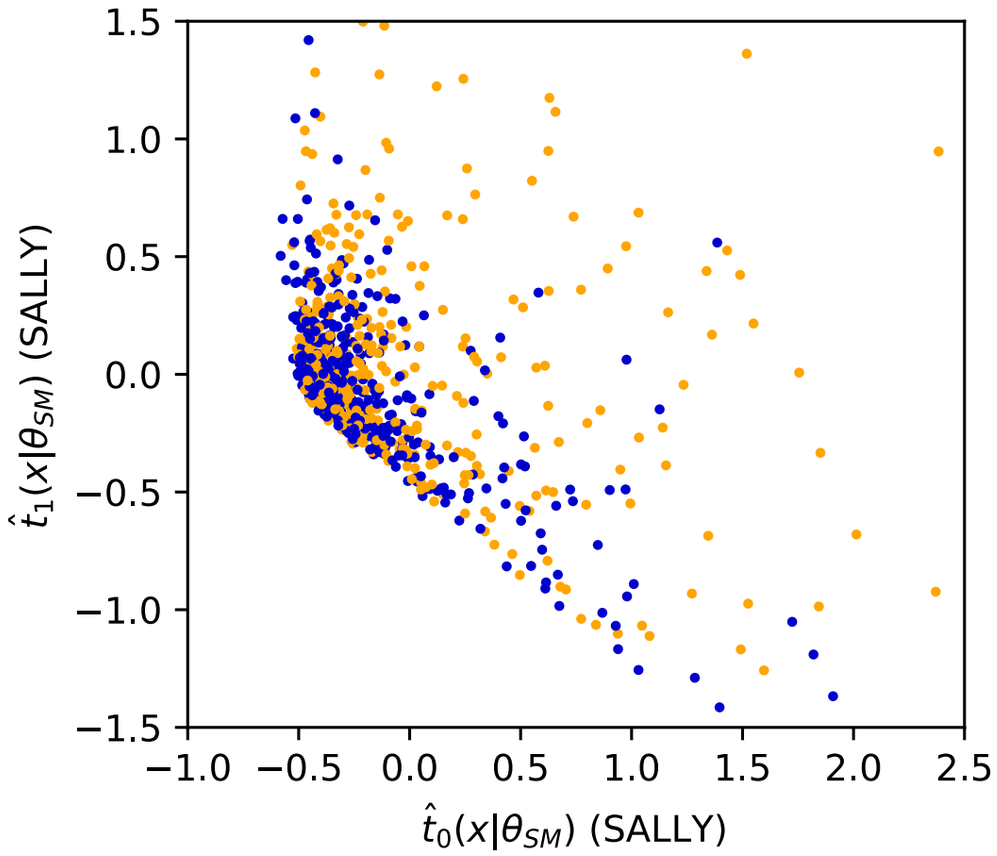


J Brehmer, J Pavez, G Louppe, K.C. PRL & PRD 2018 [arXiv:1805.00013 & arXiv:1805.00020]
"Better Higgs Measurements Through Information Geometry" [arXiv:1612.05261] & CARL [arxiv:1506.02169]

LOCALLY SUFFICIENT STATISTICS

One of the initial motivations for using ML to approximate the likelihood is that most engineered features loose information.

However, the score provides "locally sufficient statistics" that capture all the information in the region of neighborhood of θ_0 (aka the standard model)



One summary statistic per parameter

$$p(x|\theta) \sim e^{t(x|\theta_{SM})\cdot(\theta-\theta_{SM})}$$

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

SIMILAR

MNRAS **000**, 1–14 (2017) Preprint 28 March 2018 Compiled using MNRAS LATEX style file v3.0

Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology

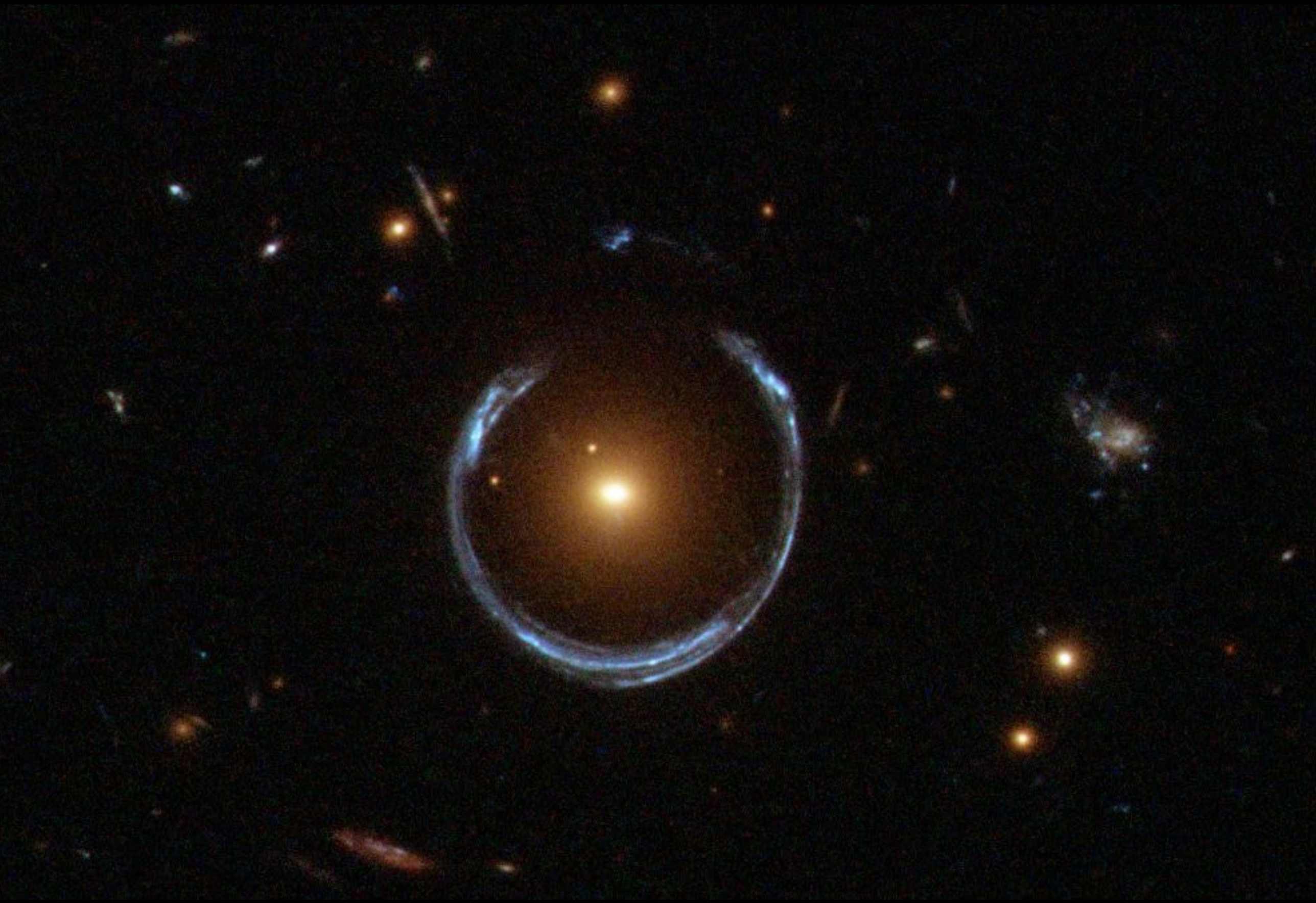
Justin Alsing,[1,2]★ Benjamin Wandelt[1,3,4,5] and Stephen Feeney[1]
[1] *Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York City, NY 10010, USA*
[2] *Imperial Centre for Inference and Cosmology, Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[3] *Institut d'Astrophysique de Paris (IAP), UMR 7095, CNRS UPMC Universite Paris 6, Sorbonne Universitse, 98bis boulevard Arago, F-75014 Paris, France*
[4] *Institut Lagrange de Paris (ILP), Sorbonne Universitse, 98bis boulevard Arago, F-75014 Paris, France*
[5] *Department of Physics and Astronomy, University of Illinois at Urbana-Champaign, 1002 W Green St, Urbana, IL 61801, USA*

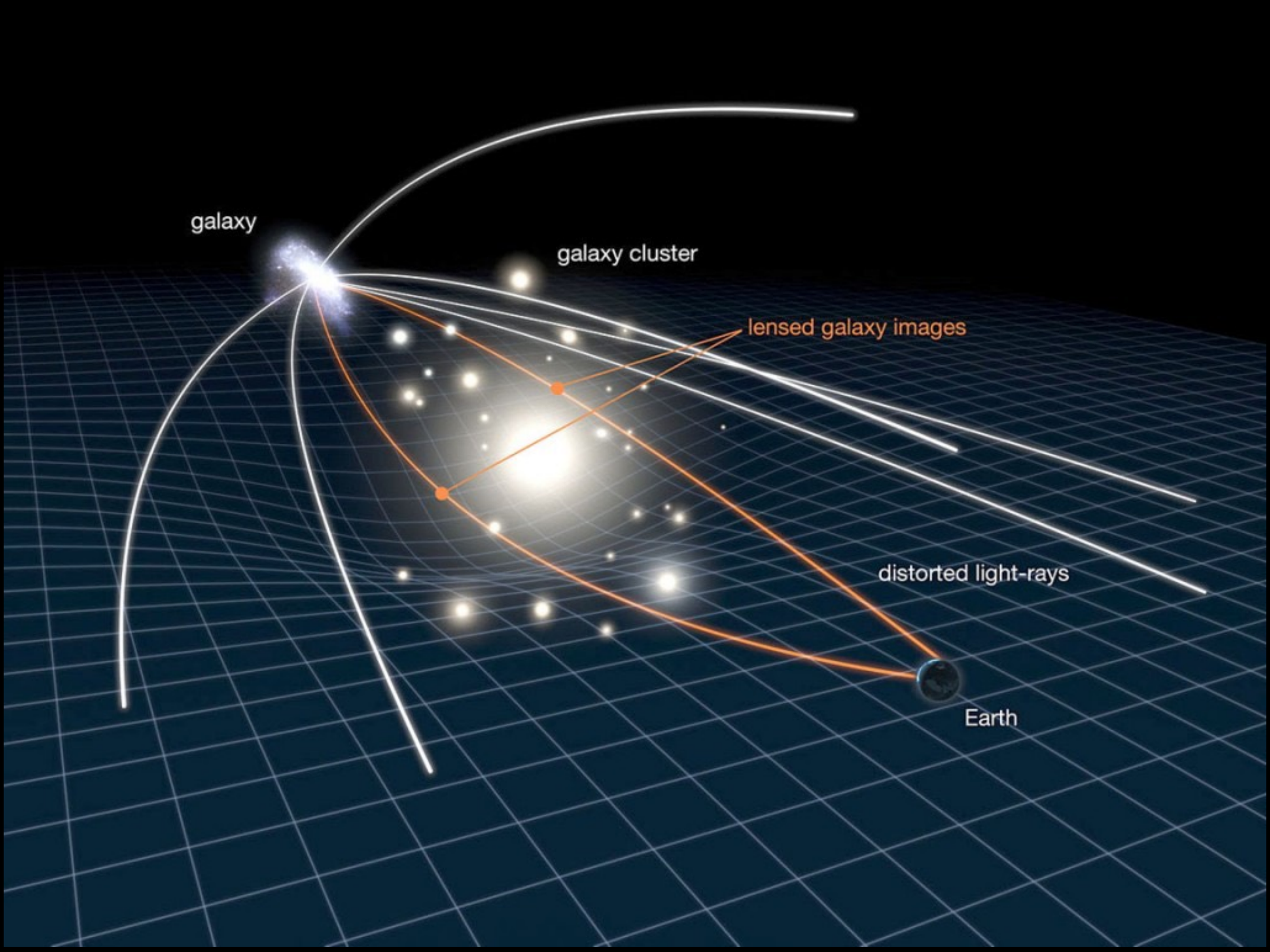
Automatic physical inference with information maximising neural networks

Tom Charnock*
Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France

Guilhem Lavaux†
Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France and Sorbonne Universités, Institut Lagrange de Paris, 98 bis boulevard Arago, 75014 Paris, France

Benjamin D. Wandelt‡
Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA
Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France
Sorbonne Universités, Institut Lagrange de Paris, 98 bis boulevard Arago, 75014 Paris, France and Department of Astrophysical Sciences, 4 Ivy Lane, Princeton University, Princeton, NJ 08544, USA



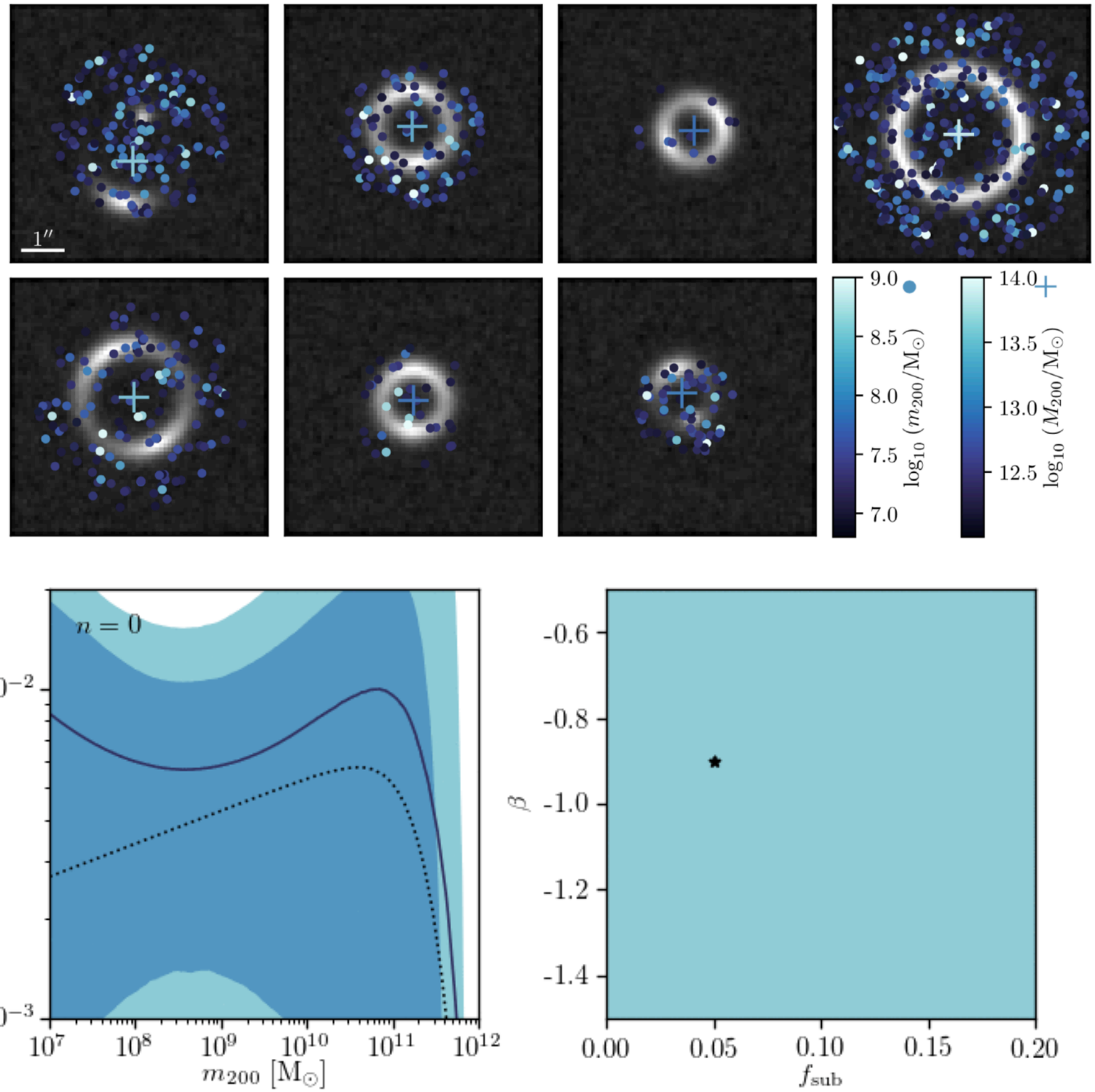

galaxy
galaxy cluster
lensed galaxy images
distorted light-rays
Earth

arXiv:1909.02005

Latent space Z:

Number of dark matter sub halos and their mass and location lead to complex latent space for each image.

Goal is inference at the population-level

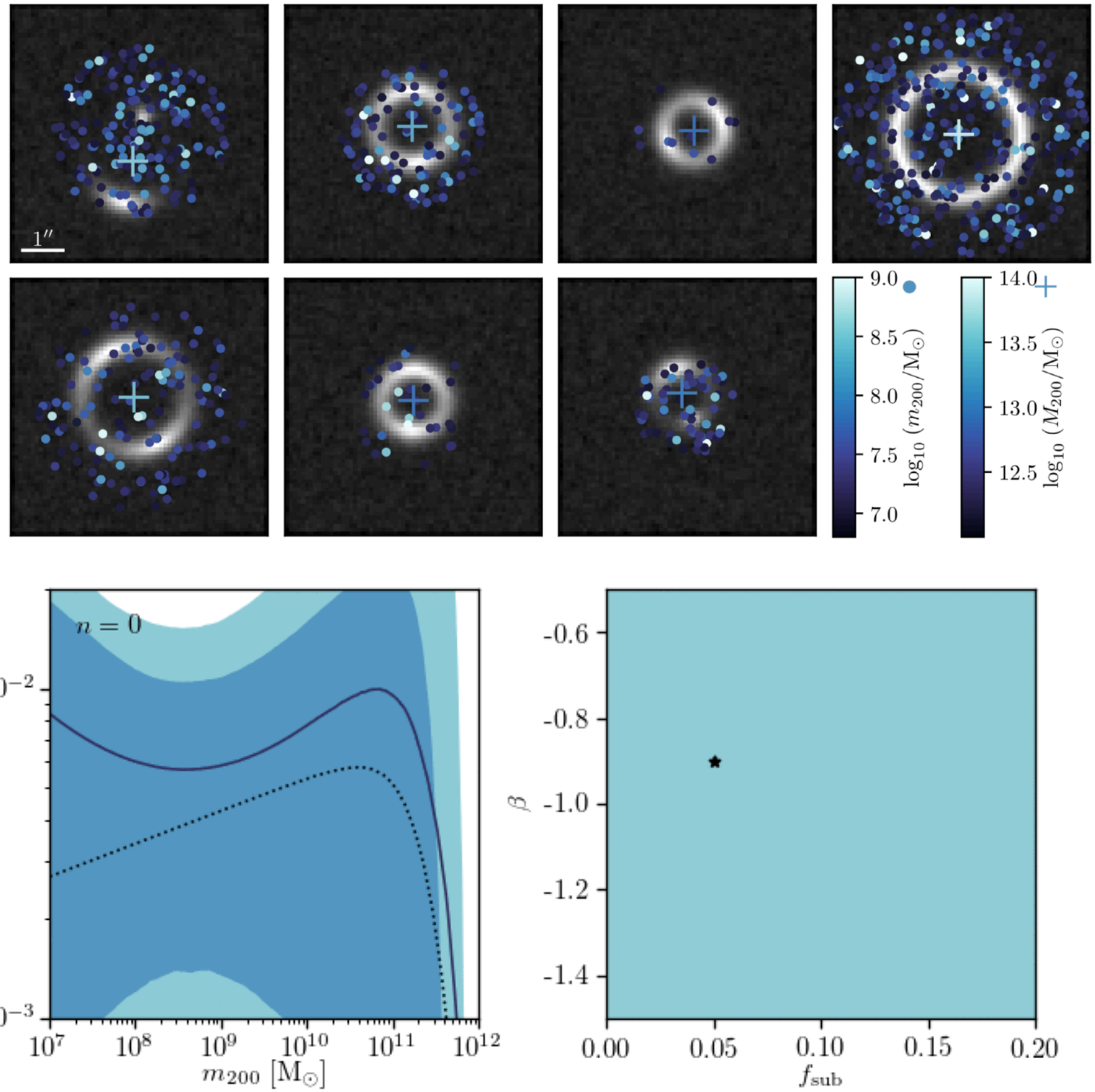


arXiv:1909.02005

Latent space Z:

Number of dark matter sub halos and their mass and location lead to complex latent space for each image.

Goal is inference at the population-level

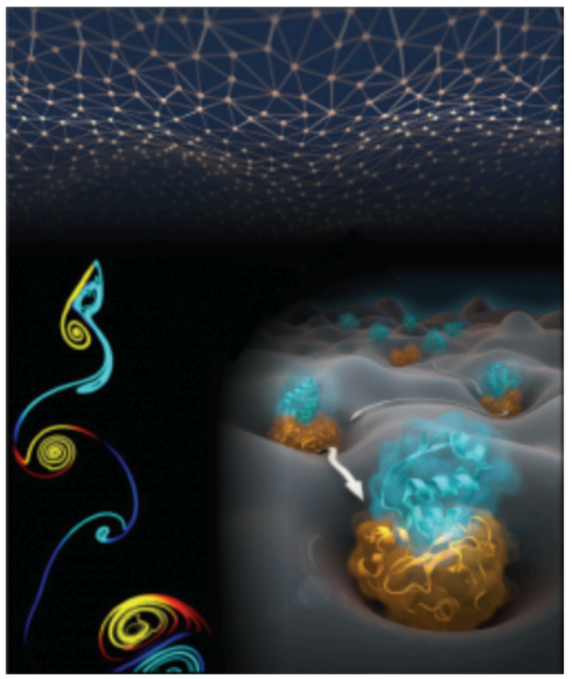


Machine Learning for Physics and the Physics of Learning

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.

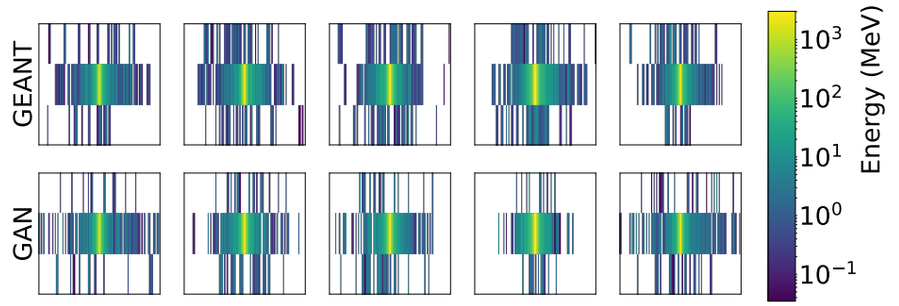
The exchange between fields can go in both directions. Since its beginning, machine learning has been inspired by methods from statistical physics. Many modern machine learning tools, such as variational inference and maximum entropy, are refinements of techniques invented by physicists. Physics, information 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.

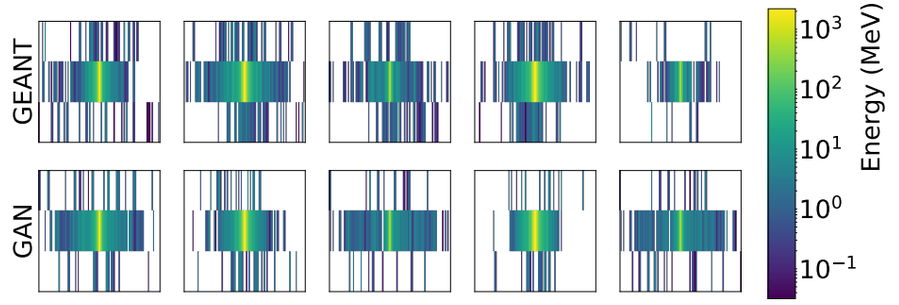
Generative Models

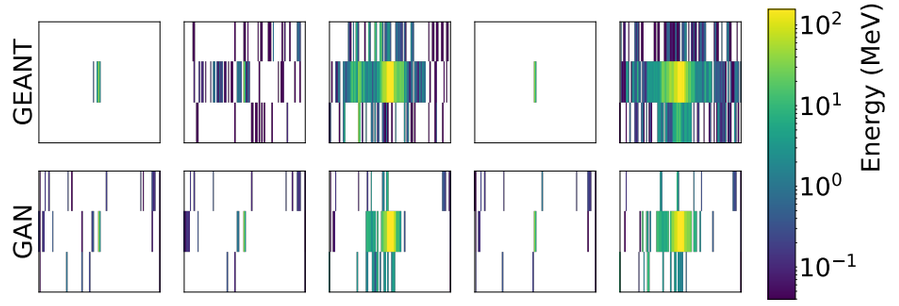
"What I cannot create, I do not understand."

—RICHARD FEYNMAN

GENERATIVE MODEL FOR IMAGES







Note, same NN can model birds, ants, volcanos, and calorimeters!

Is that good or bad? Did they learn underlying model?

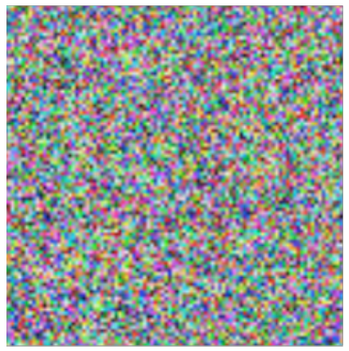
Correlation ≠ Causation

DEEP GENERATIVE MODEL VS. SIMULATION

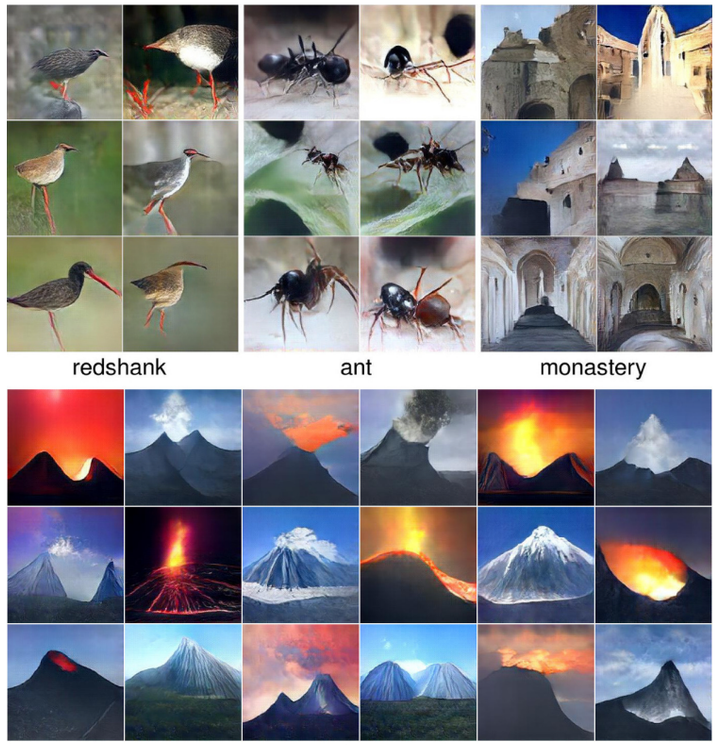
Z

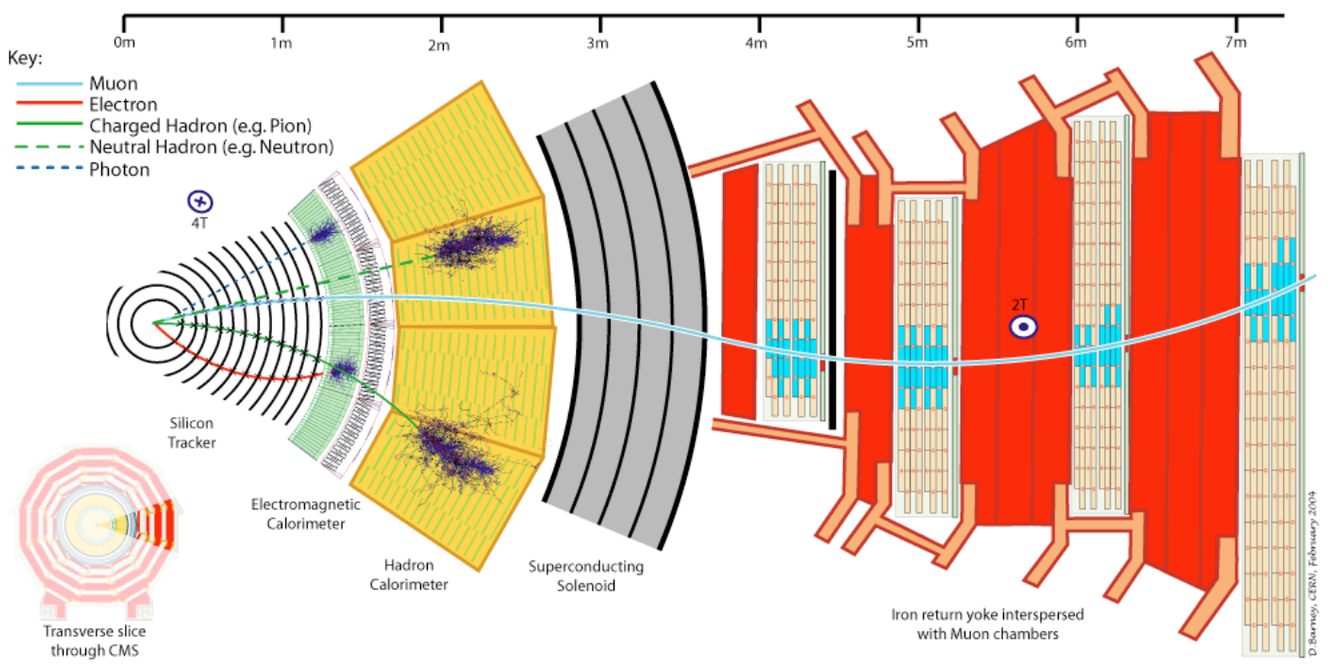
X

Noise ~ N(0,1)



Generative Model





http://torch.ch/blog/2015/11/13/gan.html

Paganini, et. al. Phys.Rev.Lett. 120 (2018) no.4, 042003 ; & ATL-SOFT-PUB-2018-001

Judea Pearl

Figure 1. The causal hierarchy. Questions at level 1 can be answered only if information from level i or higher is available.

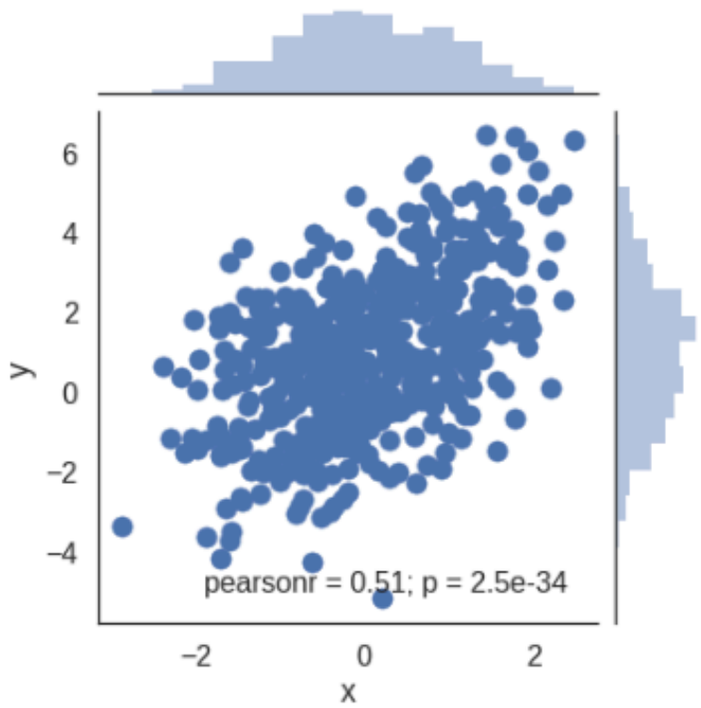
Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief inY?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing, Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

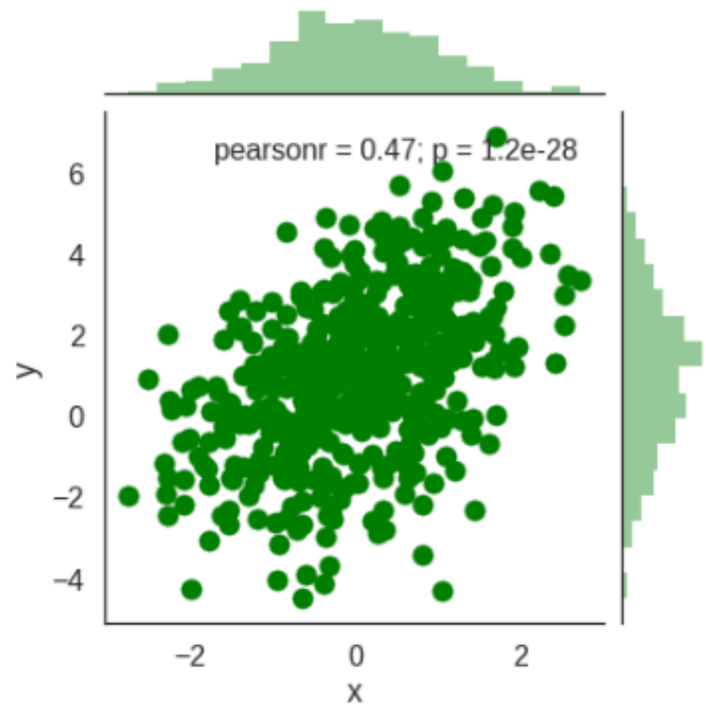
SAME JOINT DISTRIBUTION, DIFFERENT CAUSAL MODEL

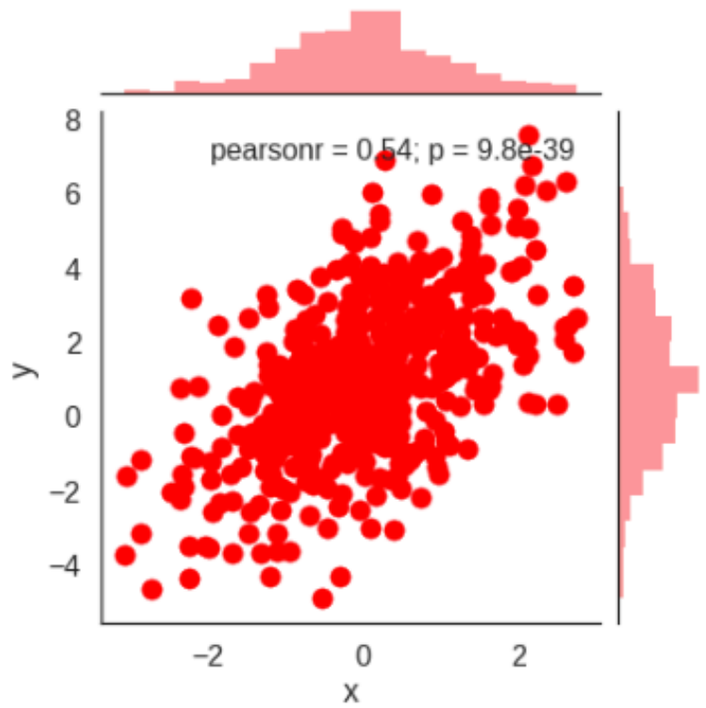
```
x = randn()
y = x + 1 + sqrt(3)*randn()
```

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

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



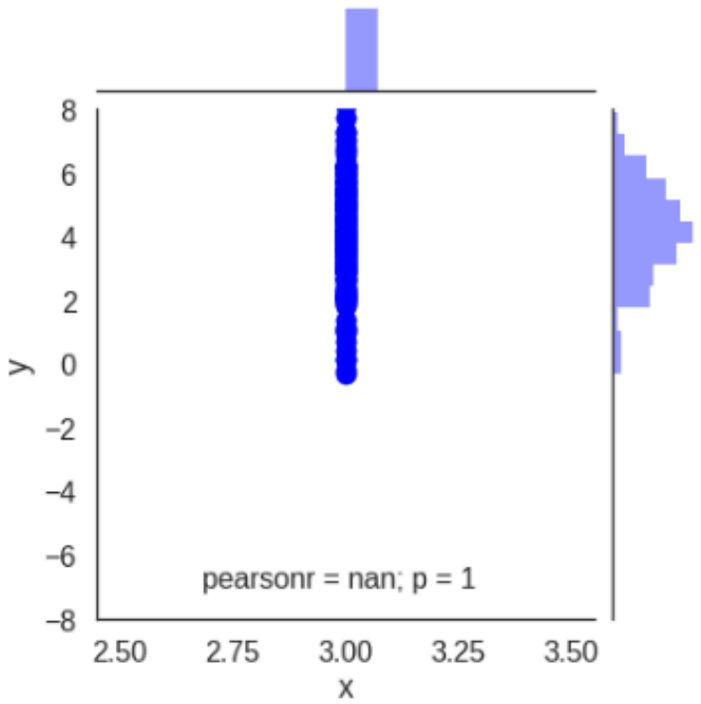




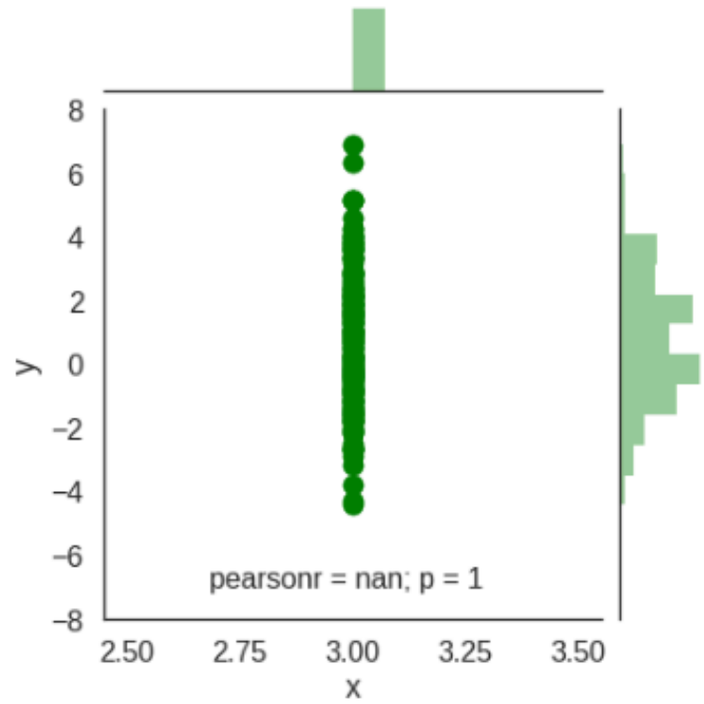
https://colab.research.google.com/drive/1rjjjA7teiZVHJCMTVD8KlZNu3EjS7Dmu#scrollTo=TlzzvcGOZdvW
https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/

CAUSATION > CORRELATION

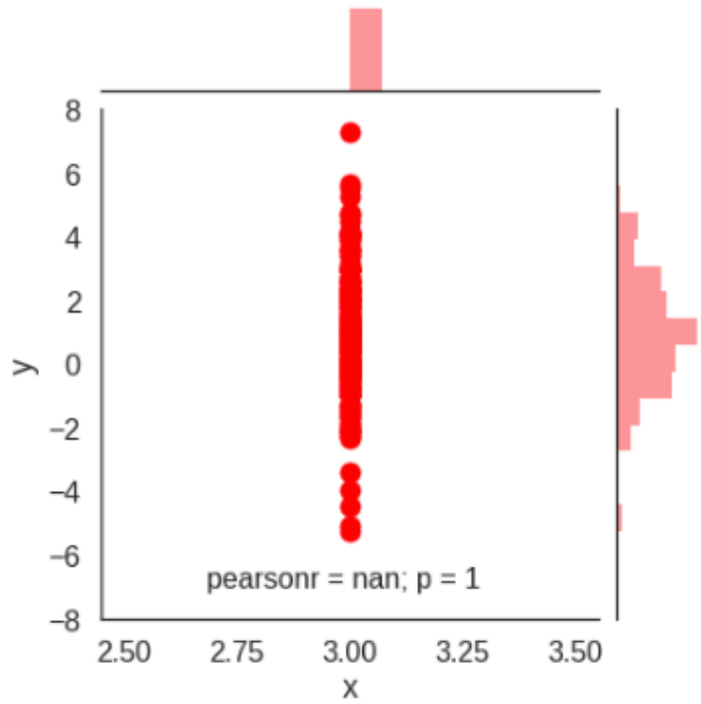
x = randn()
x = 3
y = x + 1 + sqrt(3)*randn()
x = 3



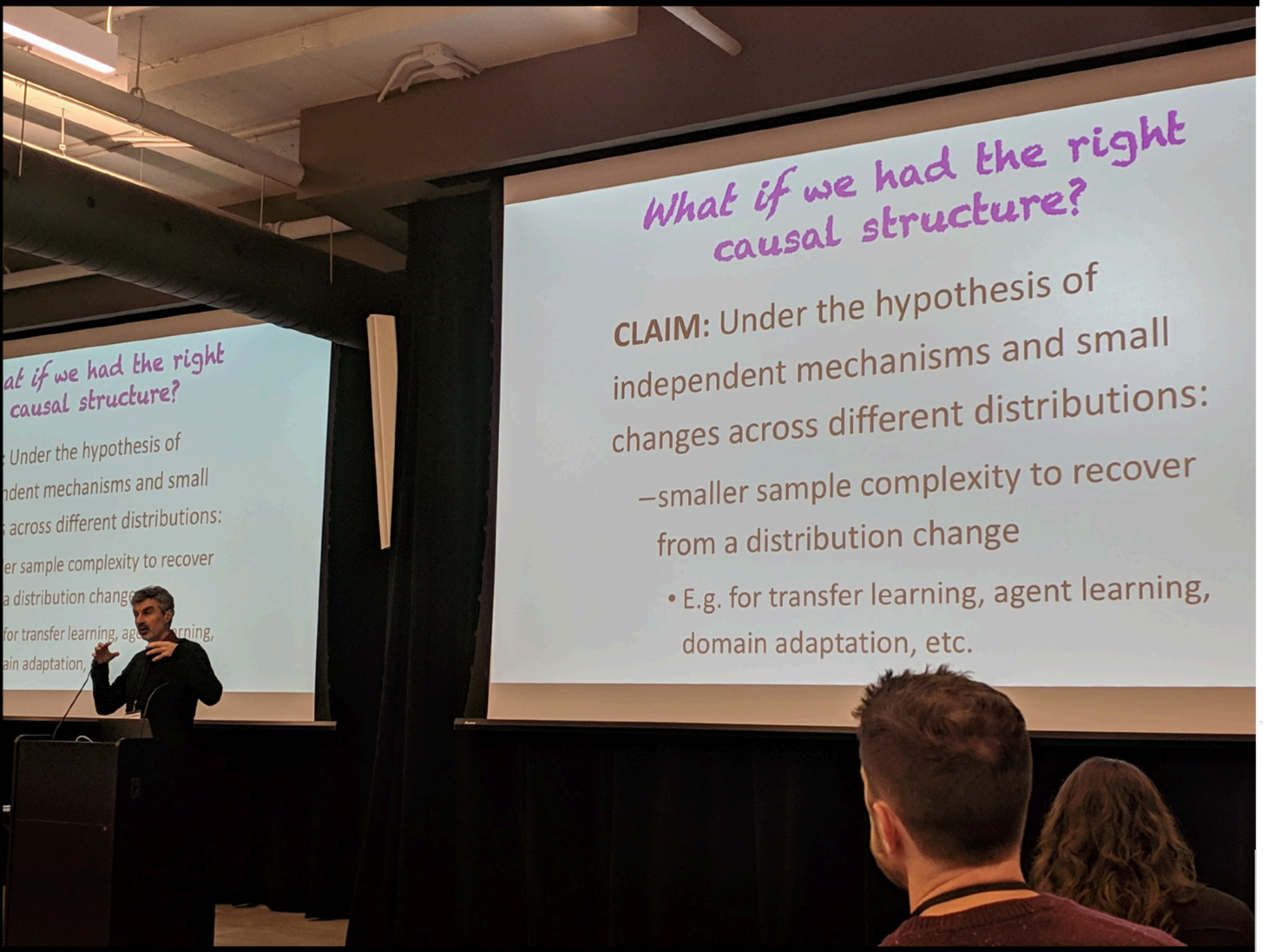
y = 1 + 2*randn()
x = 3
x = (y-1)/4 + sqrt(3)*randn()/2
x = 3



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



https://colab.research.google.com/drive/1rjjjA7teiZVHJCMTVD8KlZNu3EjS7Dmu#scrollTo=TlzzvcGOZdvW
https://www.inference.vc/causal-inference-2-illustrating-interventions-in-a-toy-example/



Yoshua Bengio on [arXiv:1901.10912]
and public FB discussion

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).

Said like this, it feels pretty trivial. Yoshua proposes to use this to infer the right causal model from a plurality of observed distributions.

Dan Roy Max Welling yes. He's been arguing for generative models with causal structure for years as the way to extract information for rich environments. So not this

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 😉

The message from human cognition:

Richly structured models of objects and their relations are a powerful tool for reasoning about, and interacting with, the world.

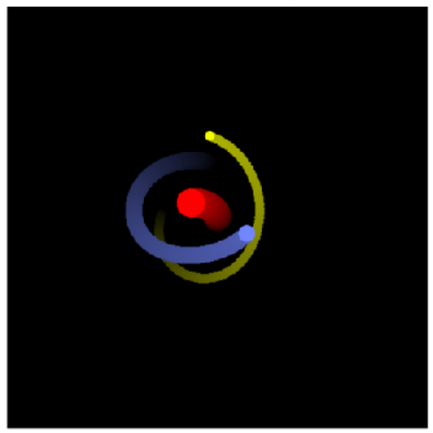
- Objects and relations reflect *decisions* made by evolution, experience, and task demands about how to represent the world in an *efficient and useful way*
- Intelligence is about *model-building*, beyond just recognizing patterns (Tenenbaum)
- *Combinatorial generalization* via abstraction and compositionality ("infinite use of finite means")

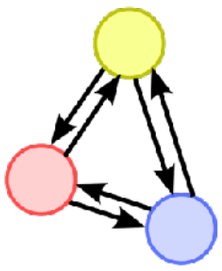
Peter Battaglia

Insight of data generating process informs inductive bias on architecture

Physical systems as graphs

n-body

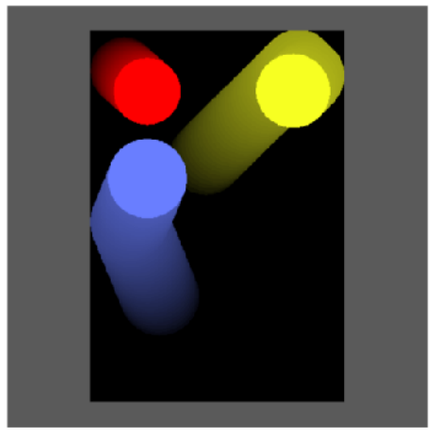


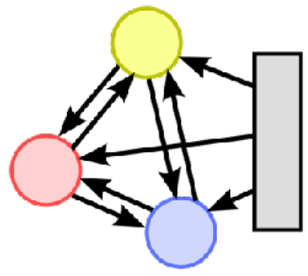


Nodes: bodies

Edges: gravitational forces

Balls

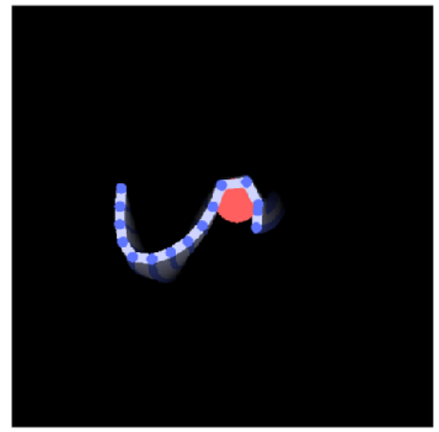


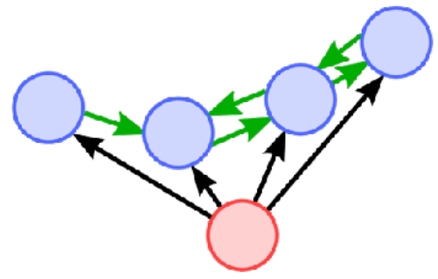


Nodes: balls

Edges: rigid collisions between balls, and walls

String



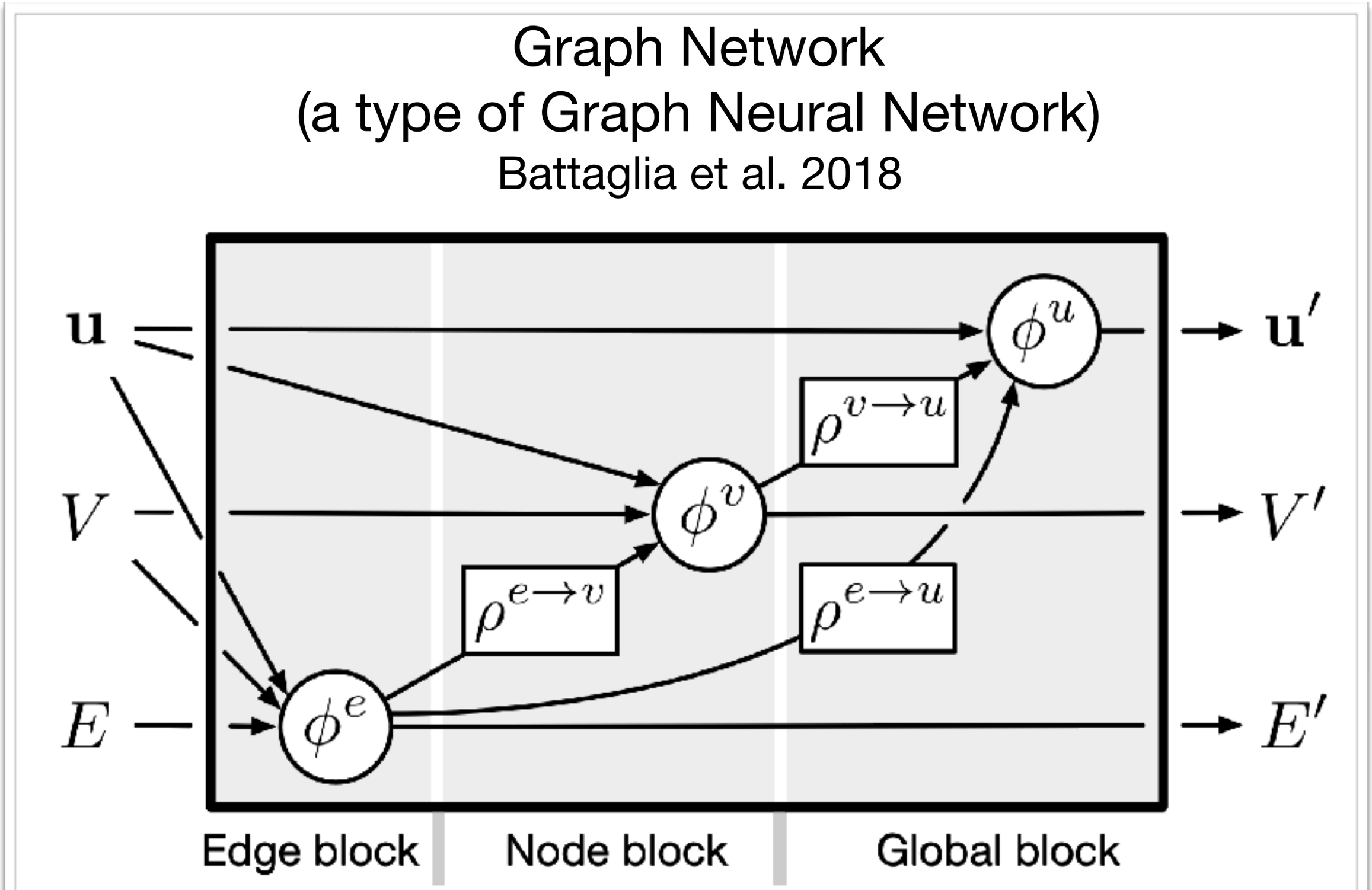


Nodes: masses

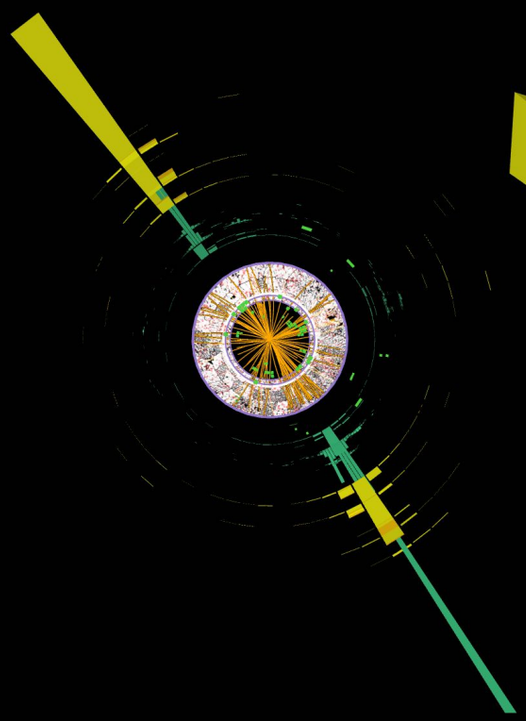
Edges: springs and rigid collisions

Battaglia et al., 2016, NeurIPS

Insight of data generating process informs inductive bias on architecture



JETS

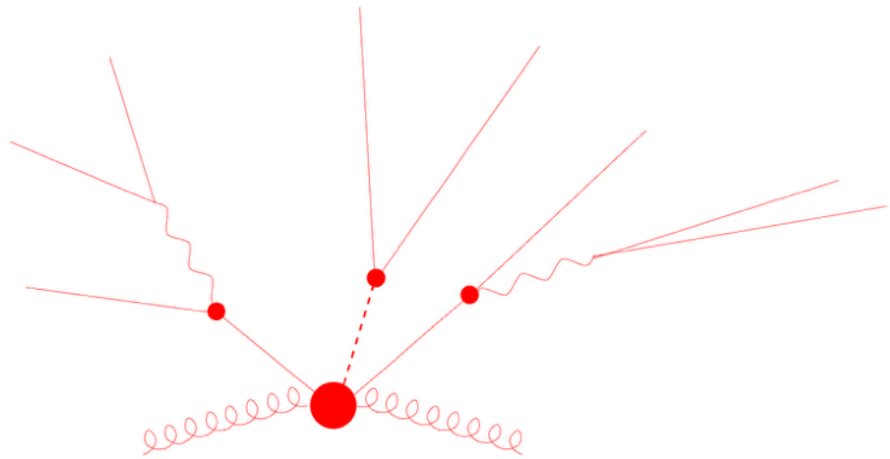


Run: 329716
Event: 857582452
2017-07-14 10:48:51 CEST

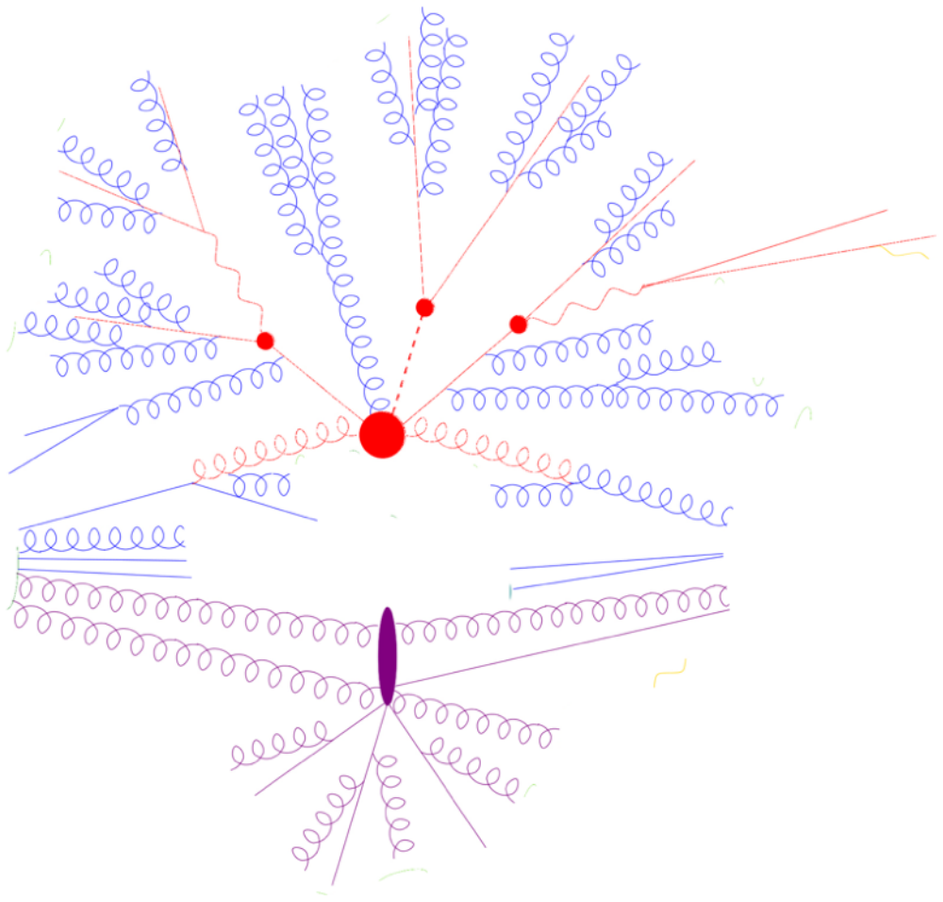
ATLAS
EXPERIMENT

CAUSAL, GENERATIVE MODEL FOR JETS

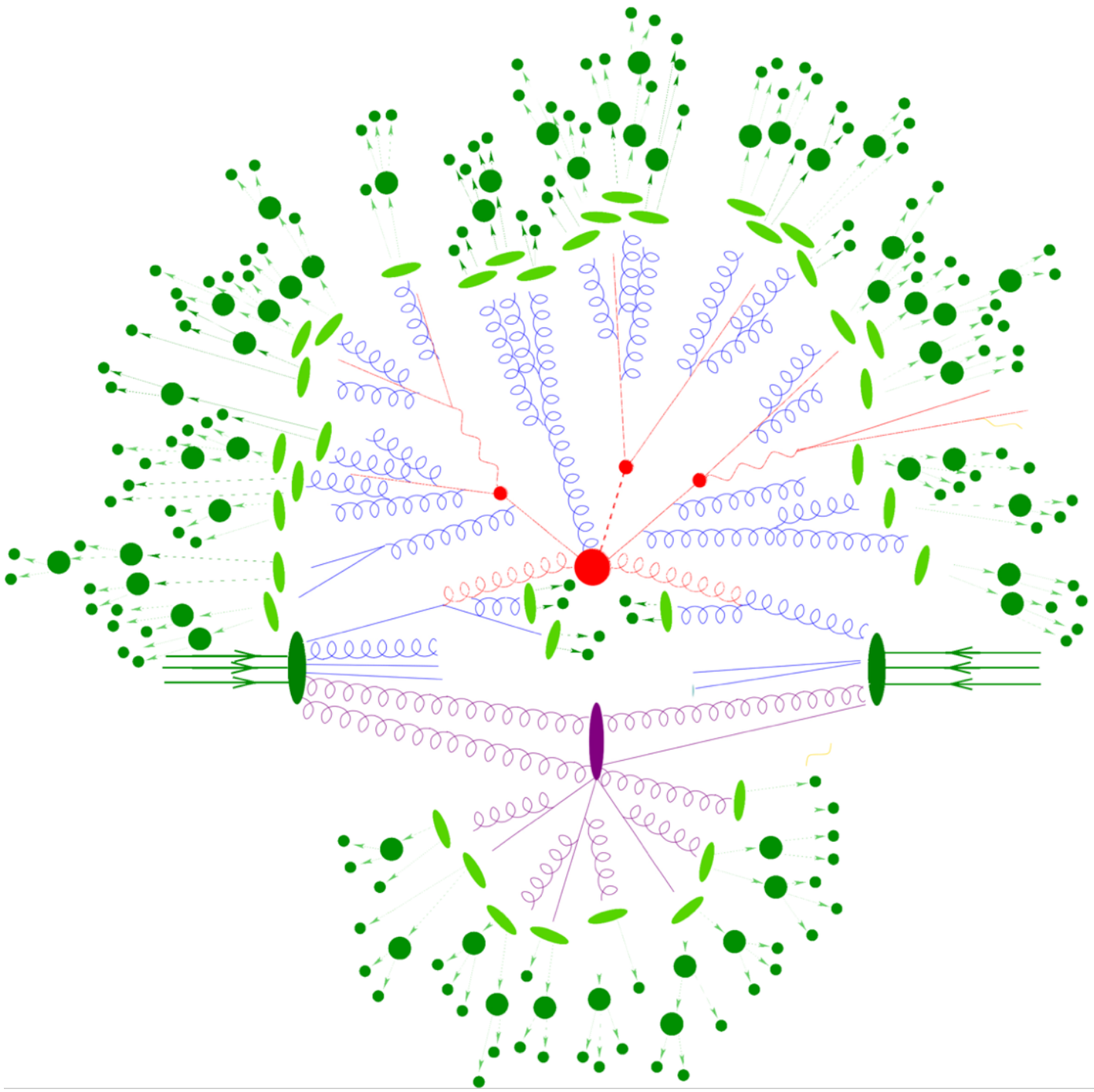
CAUSAL, GENERATIVE MODEL FOR JETS



CAUSAL, GENERATIVE MODEL FOR JETS

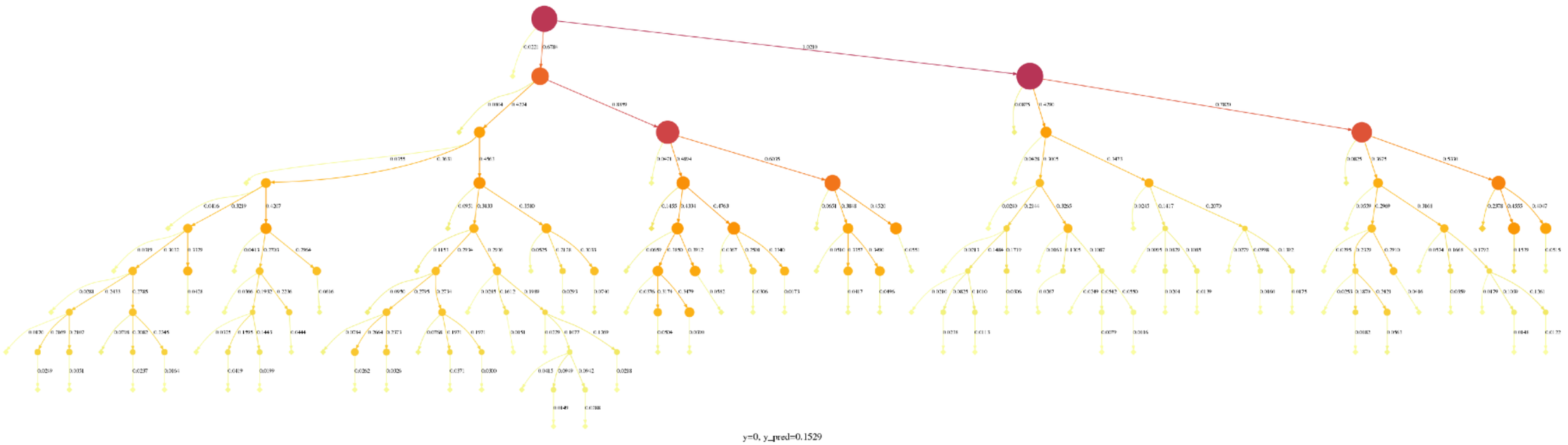


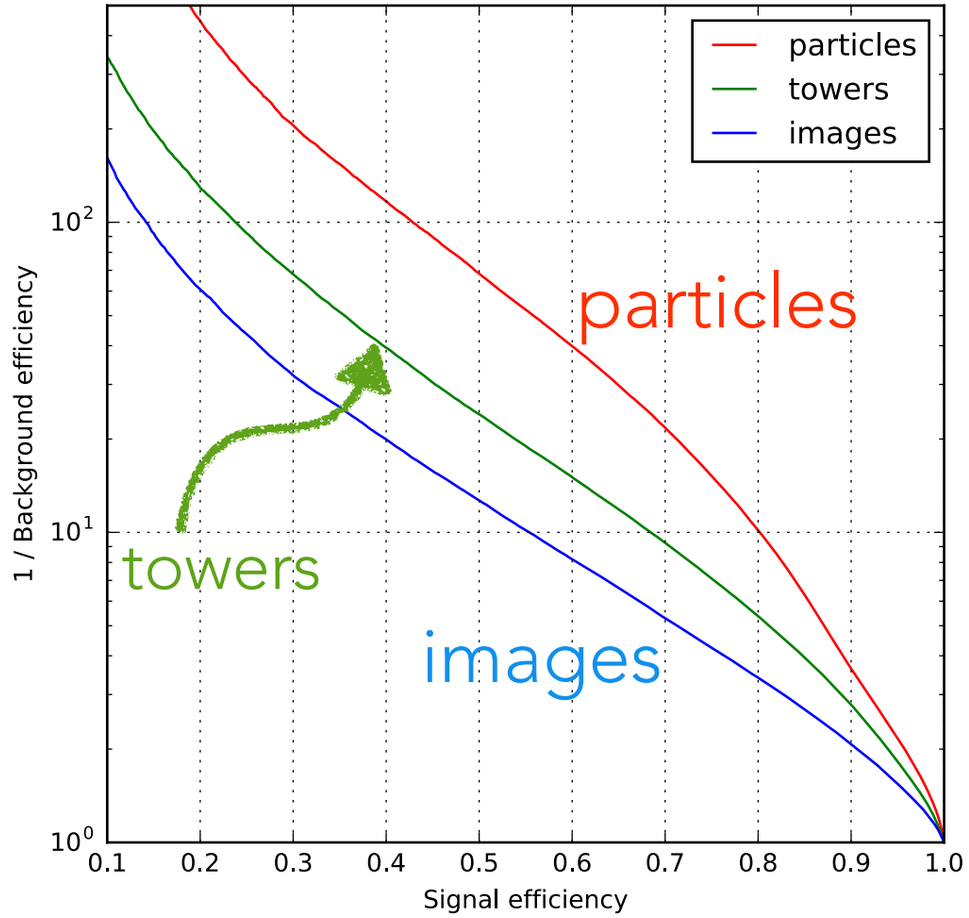
CAUSAL, GENERATIVE MODEL FOR JETS



QCD-INSPIRED RECURSIVE NEURAL NETWORKS

Insight of data generating process informs inductive bias on architecture





- Generative process is a tree-like, ~stationary Markov Process
- Physics algorithms exist to estimate the tree
- Tree-RNN needs much less data to train!

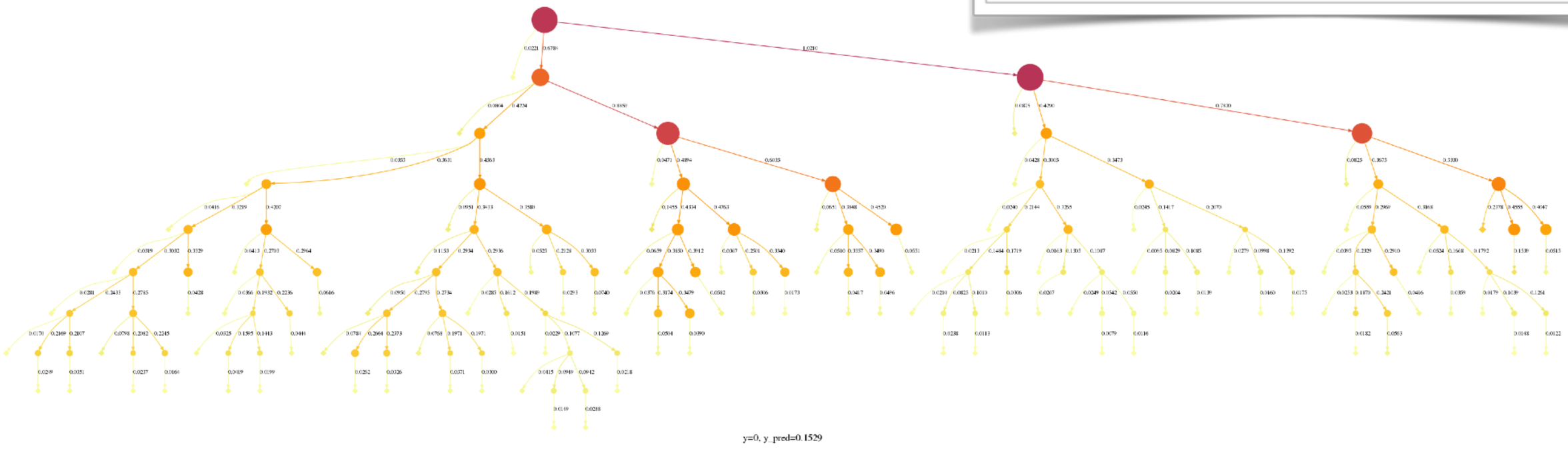
Louppe, Cho, Becot, KC [arXiv:1702.00748]

INDUCTIVE BIAS

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]

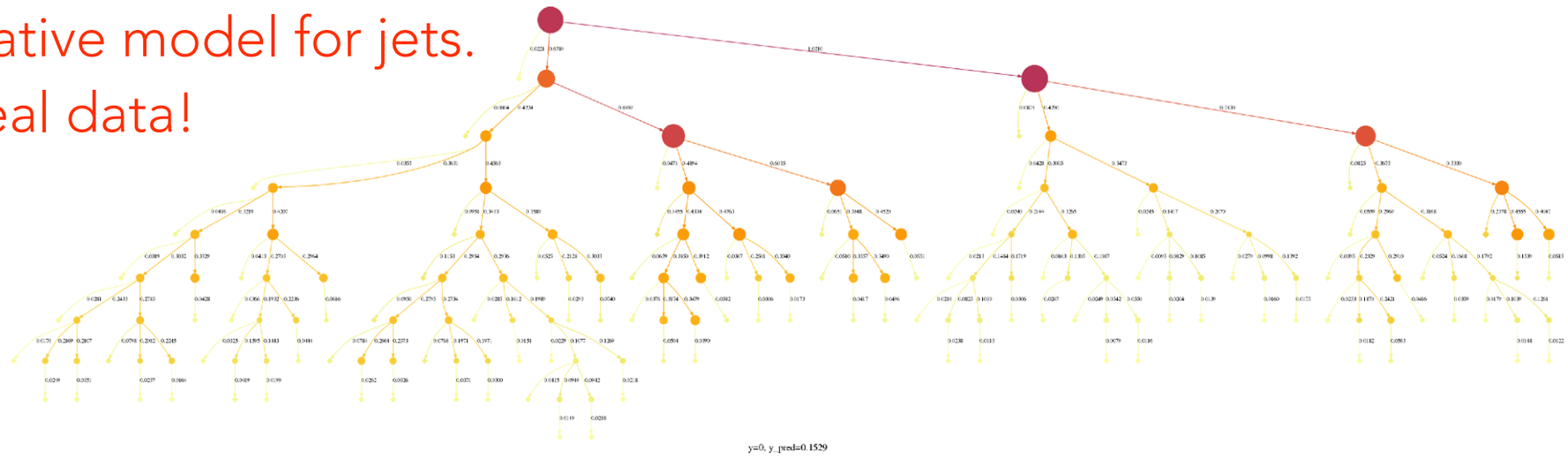


	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	966±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1246±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	378± 5	391 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	802±12	783±13	57k
Multi-body 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

CAUSAL, GENERATIVE MODELS FOR JETS

JUNIPR is a causal, generative model for jets.

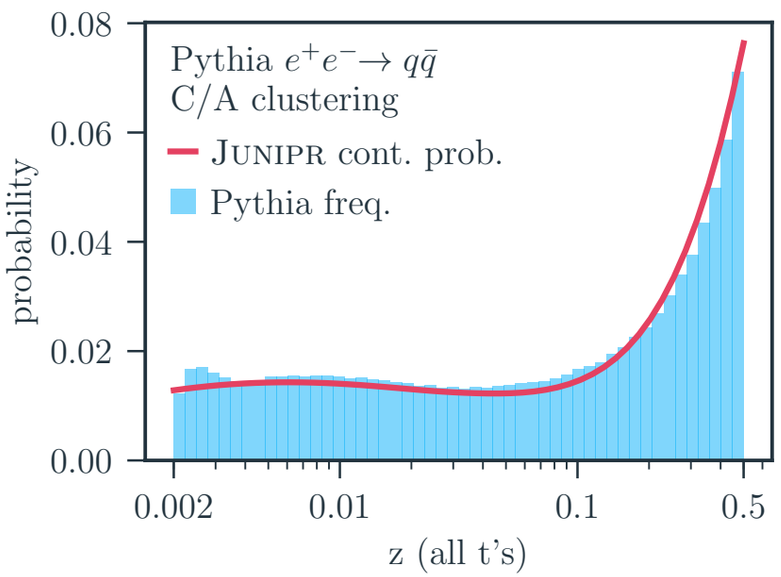
Can train on real data!

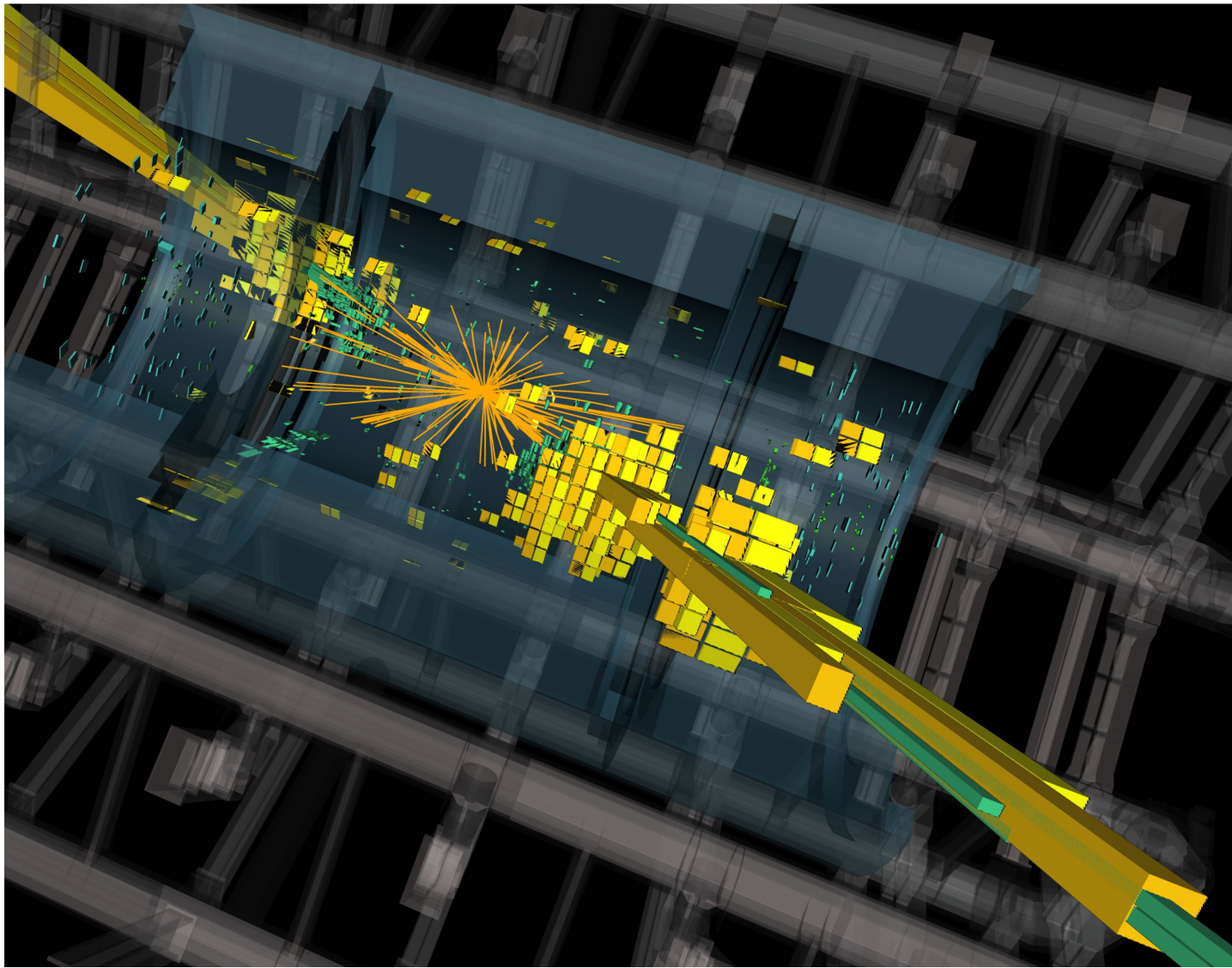


tractable likelihood

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

… and it is interpretable





Andreassen, Feige, Frye, Schwartz arXiv:1804.09720

BABY STEPS

Before we are able to discover new models on experimental data, should be able to recover model from simulation

- should be able to recover ground truth with increasingly fewer hints (in less restricted model space)

- Simulators have causal structure, can perform interventions and test different approaches to causal discovery

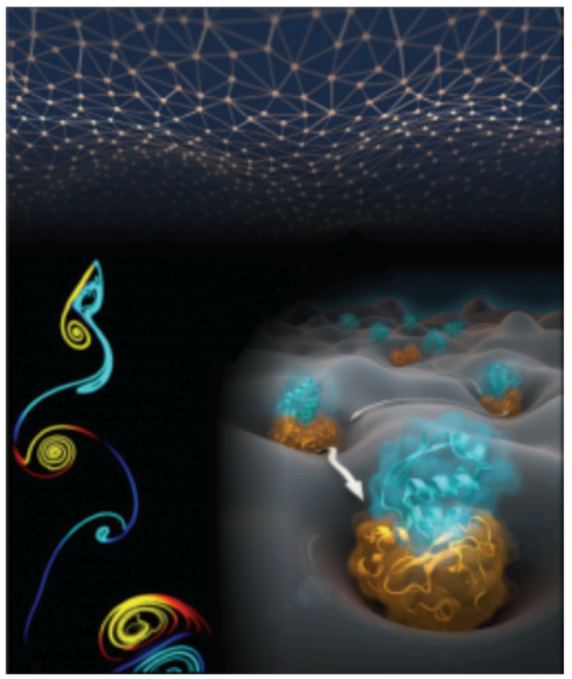
The ability to systematically improve on an existing simulator model with real data may be easier than discovering new model from scratch, and may be even more valuable in practice

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.



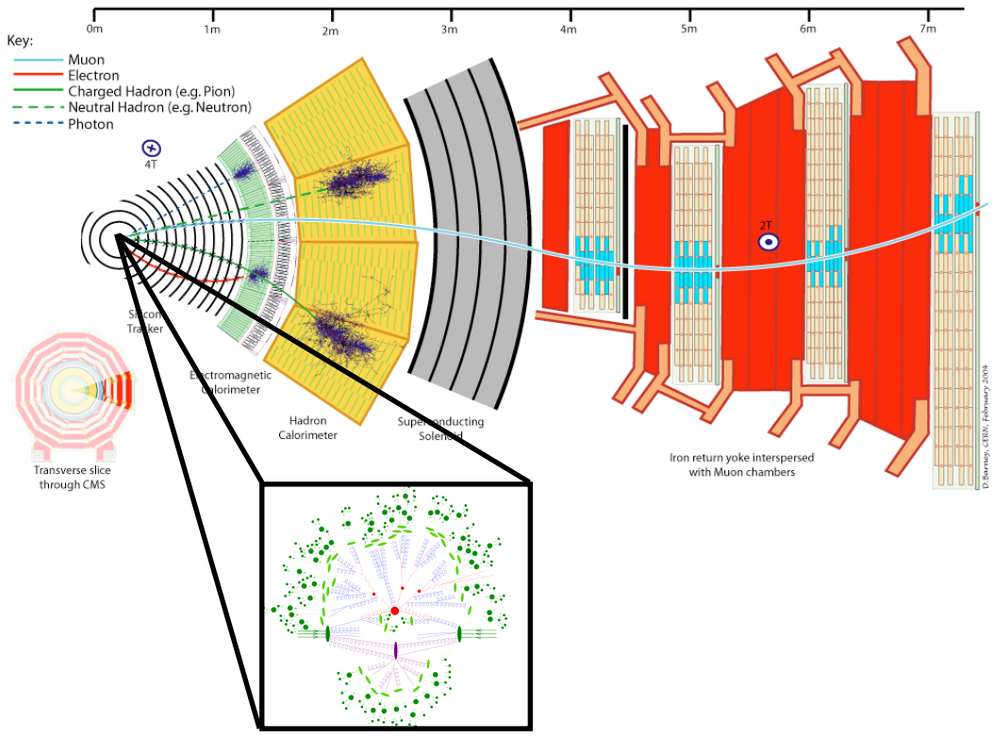
As yet, most applications of machine learning to physical sciences have been limited to the "low-hanging fruits," as they have mostly been focused on fitting pre-existing physical models to data and on discovering strong signals. We believe that machine learning also provides an exciting opportunity 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.

TWO APPROACHES TO LIKELIHOOD FREE INFERENCE

Use simulator

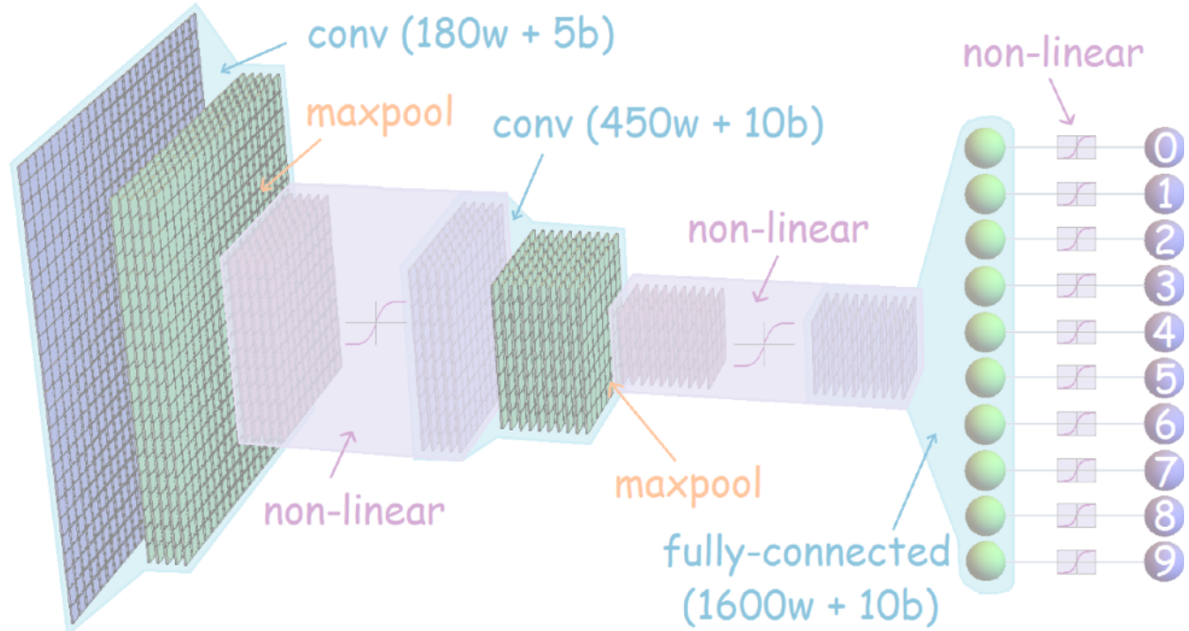
(much more efficiently)



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

Learn simulator

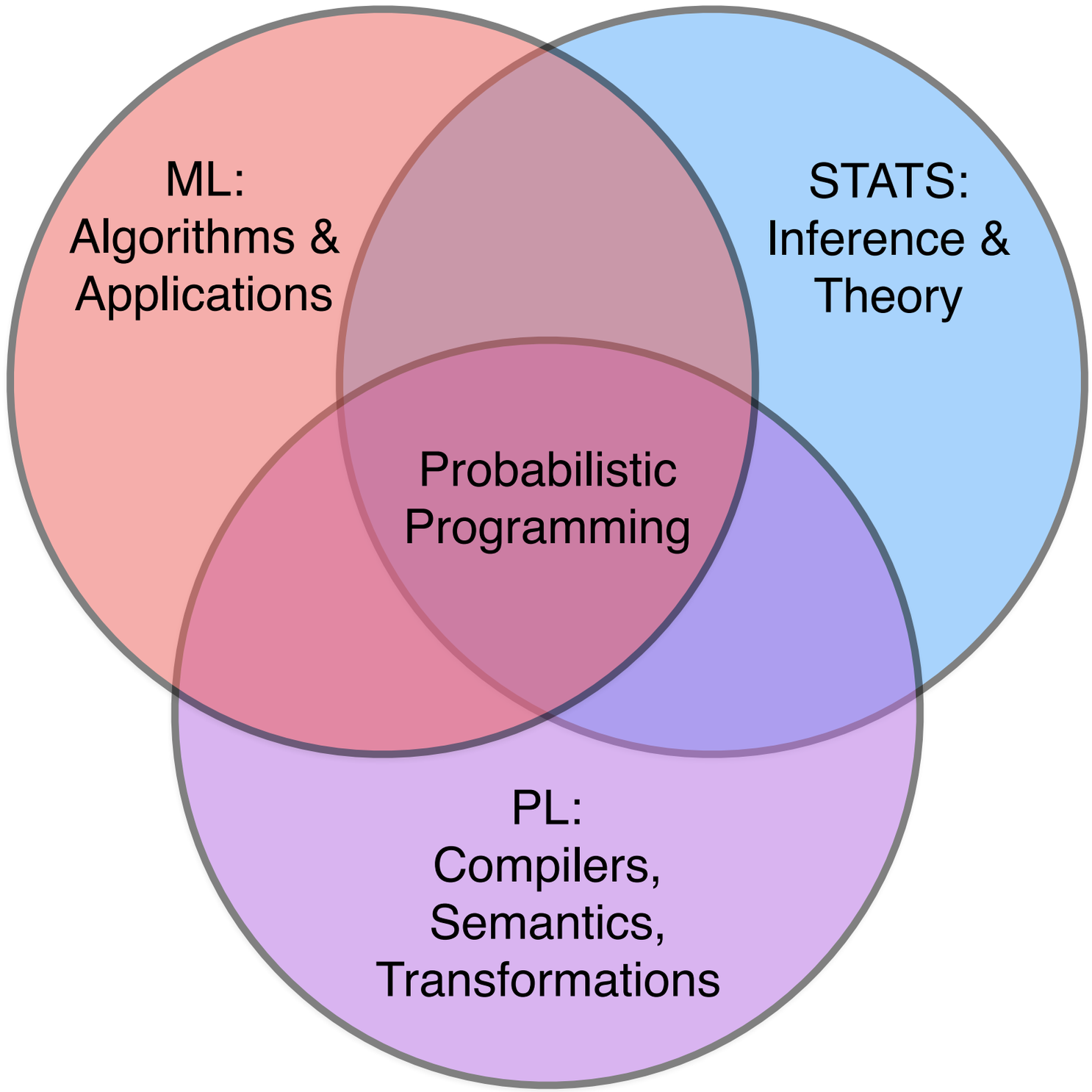
(with deep learning)



- Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAE)
- Likelihood ratio from classifiers (CARL)
- Autogregressive models, Normalizing Flows

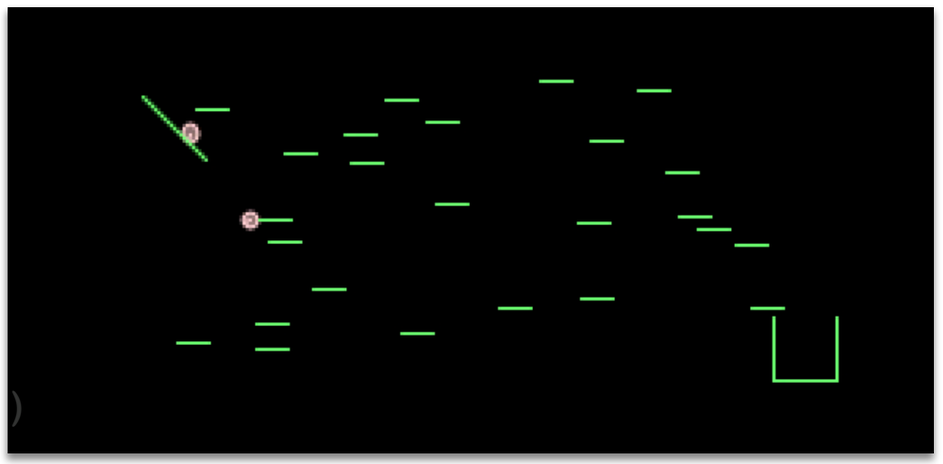
[image credit: A.P. Goucher]

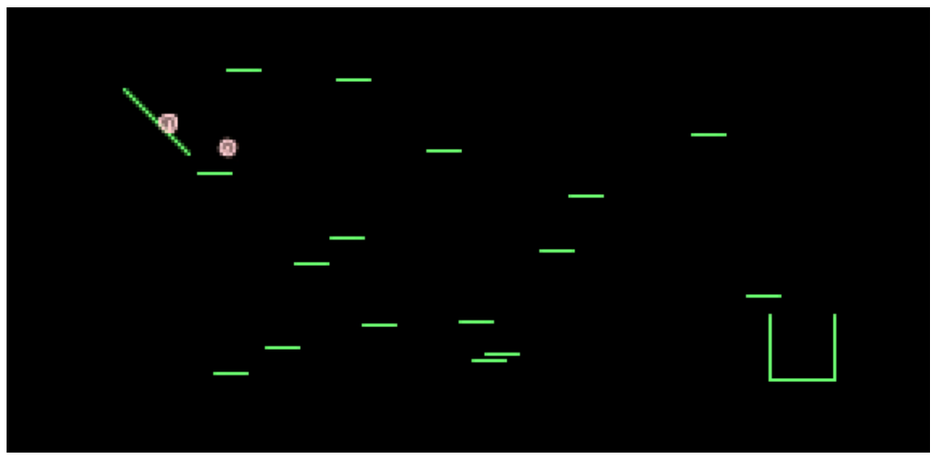
Probabilistic Programming

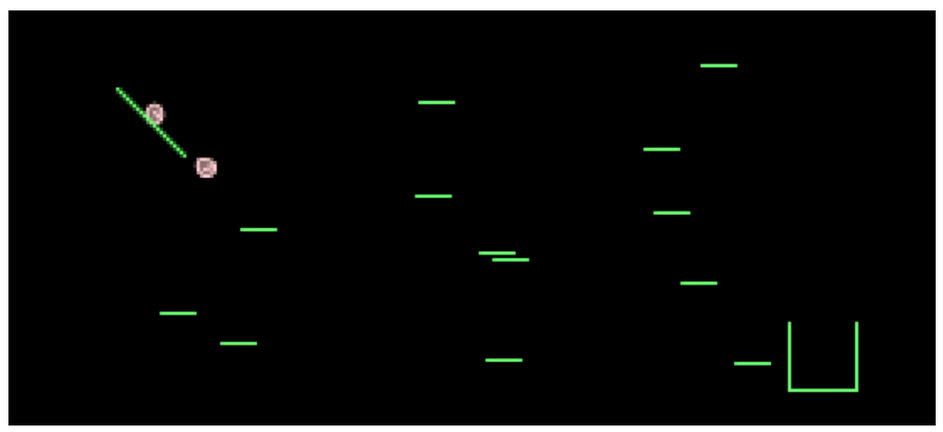


PROBABILISTIC PROGRAMMING EXAMPLE

```
(defquery arrange-bumpers []
    (let [number-of-bumpers (sample (poisson 20))
          bumpydist (uniform-continuous 0 10)
          bumpxdist (uniform-continuous -5 14)
          bumper-positions (repeatedly
                            number-of-bumpers
                            #(vector (sample bumpxdist)
                                     (sample bumpydist)))

          ;; code to simulate the world
          world (create-world bumper-positions)
          end-world (simulate-world world)
          balls (:balls end-world)

          ;; how many balls entered the box?
          num-balls-in-box (balls-in-box end-world)]

      {:balls balls
       :num-balls-in-box num-balls-in-box
       :bumper-positions bumper-positions}))
```





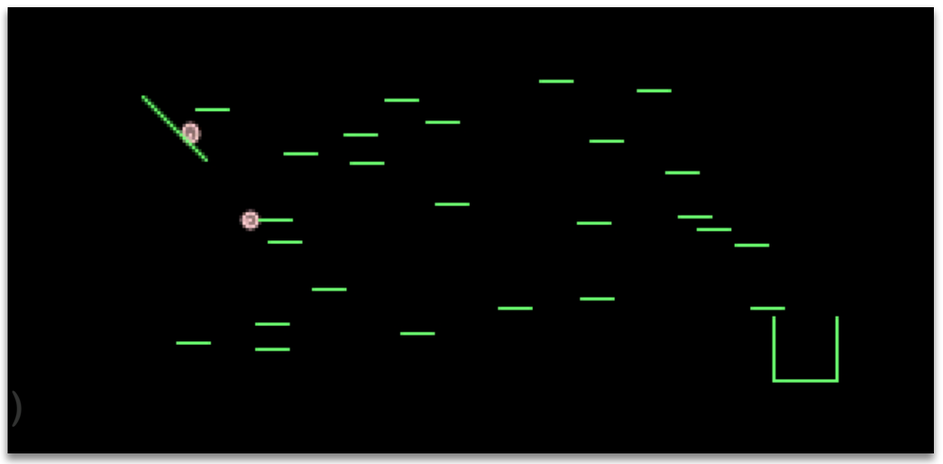


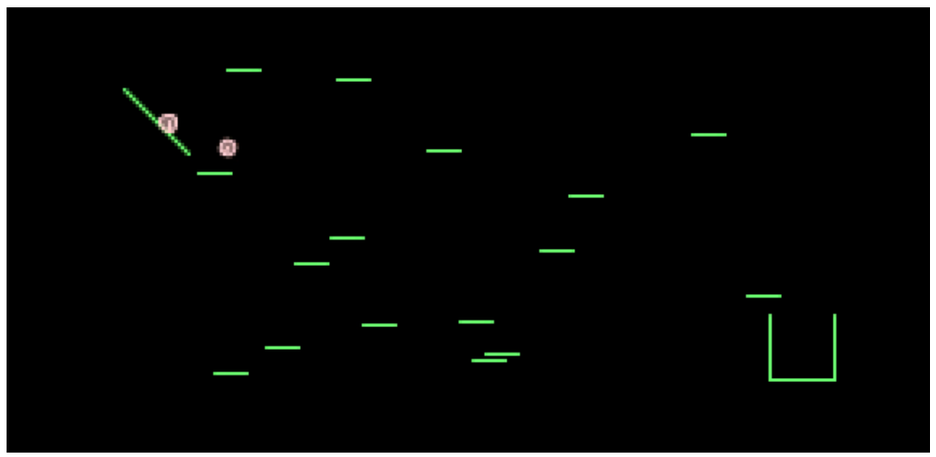
3 examples generated from simulator

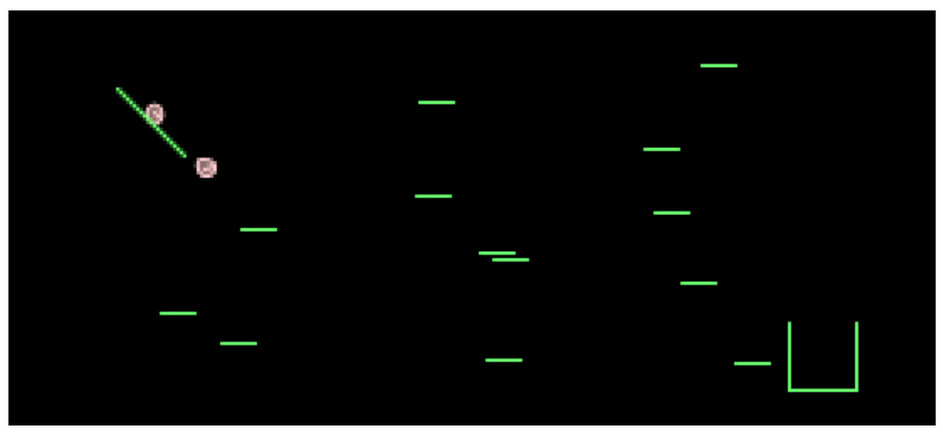
[slides, Frank Wood]

PROBABILISTIC PROGRAMMING EXAMPLE

```
(defquery arrange-bumpers []
    (let [number-of-bumpers (sample (poisson 20))
          bumpydist (uniform-continuous 0 10)
          bumpxdist (uniform-continuous -5 14)
          bumper-positions (repeatedly
                            number-of-bumpers
                            #(vector (sample bumpxdist)
                                     (sample bumpydist)))

          ;; code to simulate the world
          world (create-world bumper-positions)
          end-world (simulate-world world)
          balls (:balls end-world)

          ;; how many balls entered the box?
          num-balls-in-box (balls-in-box end-world)]

      {:balls balls
       :num-balls-in-box num-balls-in-box
       :bumper-positions bumper-positions}))
```





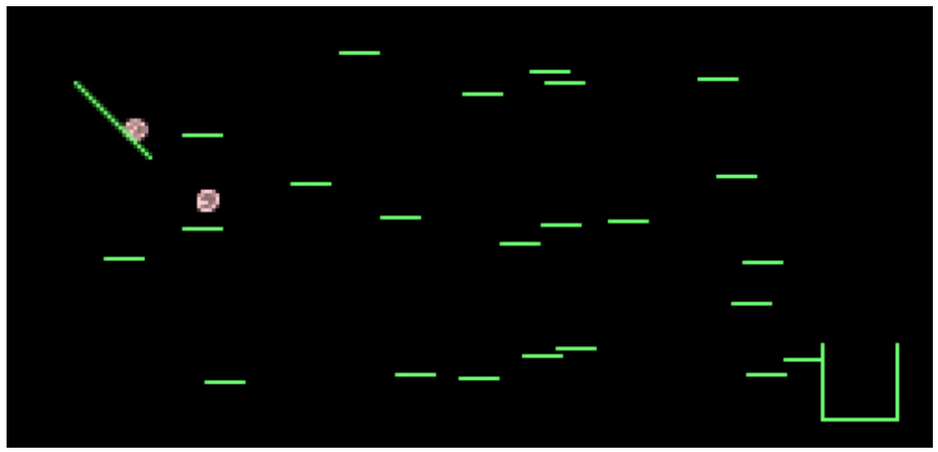


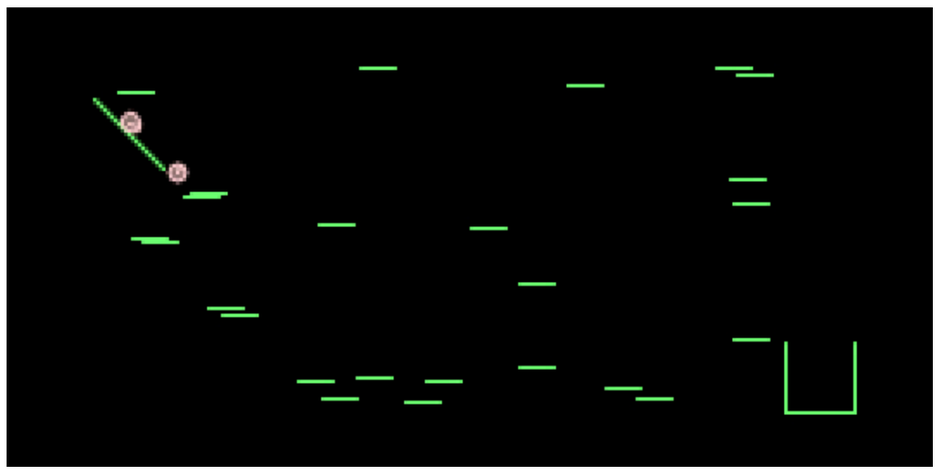
3 examples generated from simulator

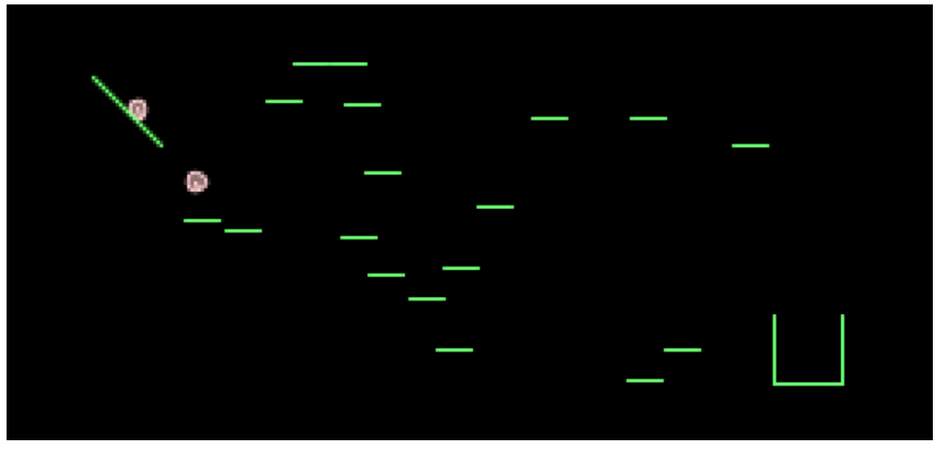
[slides, Frank Wood]

PROBABILISTIC PROGRAMMING EXAMPLE

```
(defquery arrange-bumpers []
    (let [number-of-bumpers (sample (poisson 20))
          bumpydist (uniform-continuous 0 10)
          bumpxdist (uniform-continuous -5 14)
          bumper-positions (repeatedly
                            number-of-bumpers
                            #(vector (sample bumpxdist)
                                     (sample bumpydist)))

          ;; code to simulate the world
          world (create-world bumper-positions)
          end-world (simulate-world world)
          balls (:balls end-world)

          ;; how many balls entered the box?
          num-balls-in-box (balls-in-box end-world)

          obs-dist (normal 4 0.1)]

    (observe obs-dist num-balls-in-box)
```





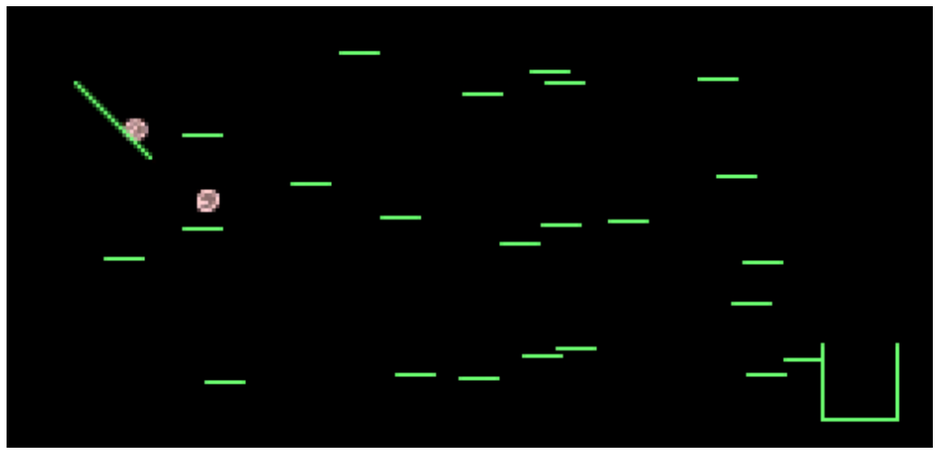


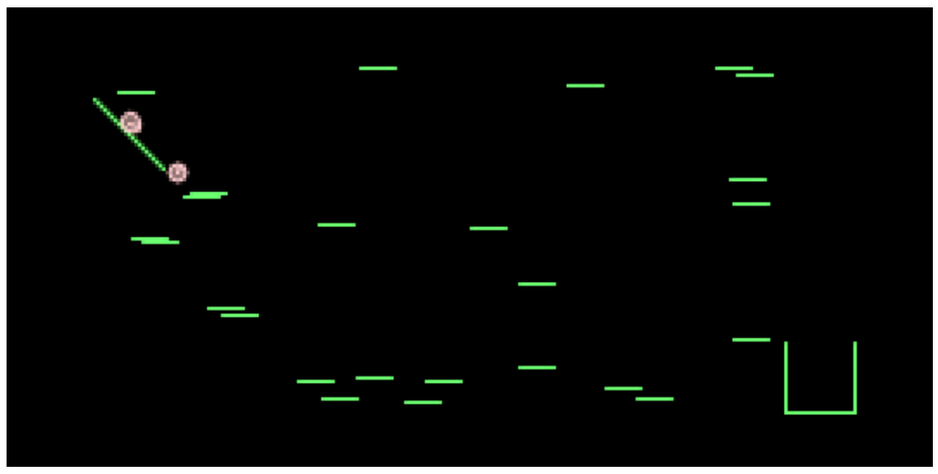
3 examples generated from simulator
conditioned on ~20% of balls land in box

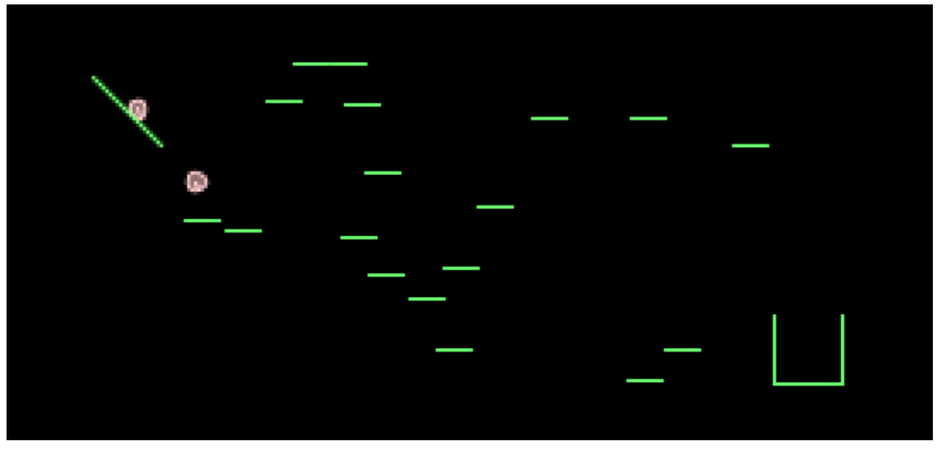
[slides, Frank Wood]

PROBABILISTIC PROGRAMMING EXAMPLE

```
(defquery arrange-bumpers []
    (let [number-of-bumpers (sample (poisson 20))
          bumpydist (uniform-continuous 0 10)
          bumpxdist (uniform-continuous -5 14)
          bumper-positions (repeatedly
                            number-of-bumpers
                            #(vector (sample bumpxdist)
                                     (sample bumpydist)))

          ;; code to simulate the world
          world (create-world bumper-positions)
          end-world (simulate-world world)
          balls (:balls end-world)

          ;; how many balls entered the box?
          num-balls-in-box (balls-in-box end-world)

          obs-dist (normal 4 0.1)]

    (observe obs-dist num-balls-in-box)
```







3 examples generated from simulator
conditioned on ~20% of balls land in box

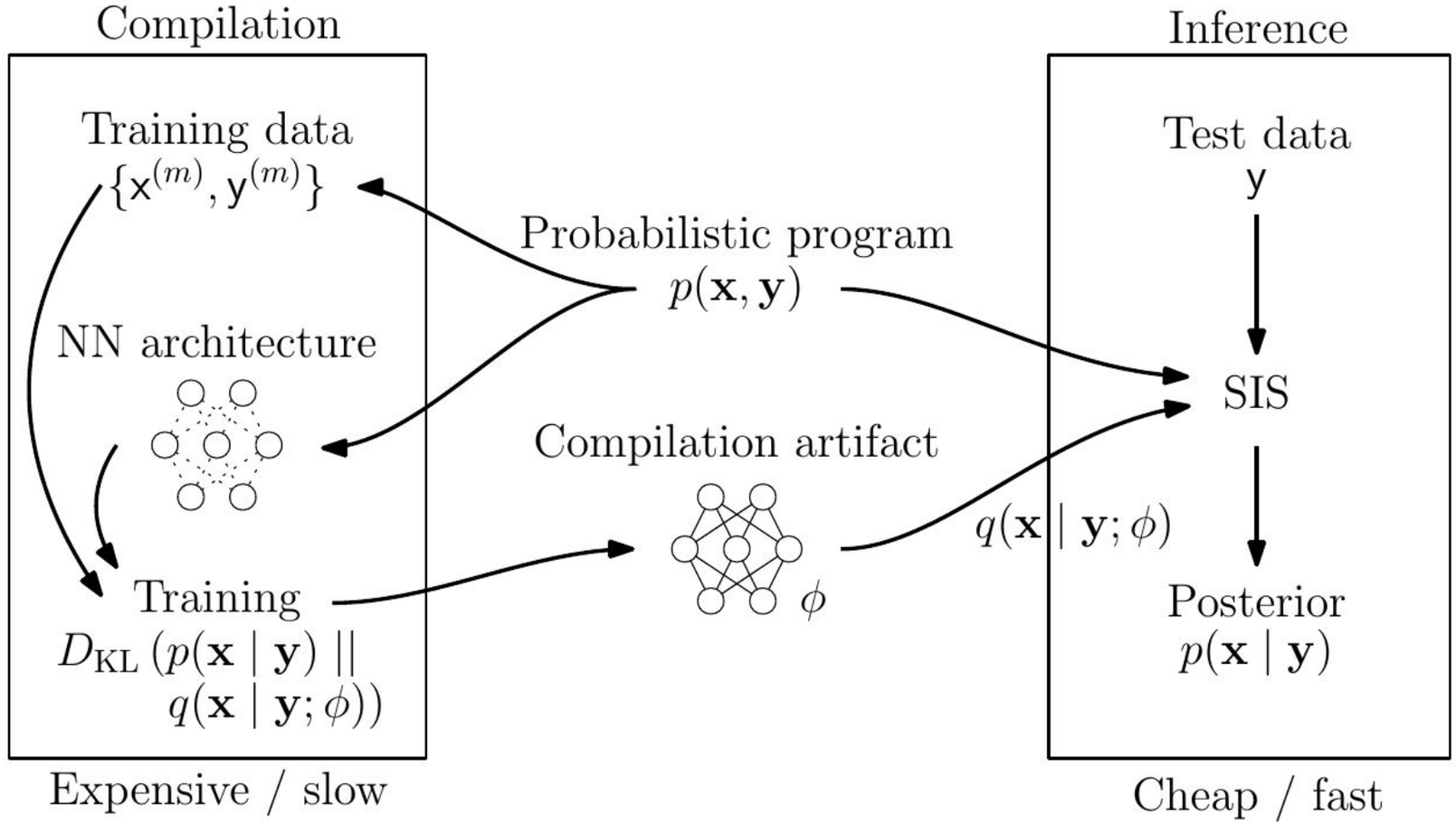
[slides, Frank Wood]

PROB PROG: HOW DOES IT WORK?

In short: hijack the random number generators and use NN's to perform a *very* smart type of importance sampling

Input: an inference problem denoted in a universal PPL (Anglican, CPProb)

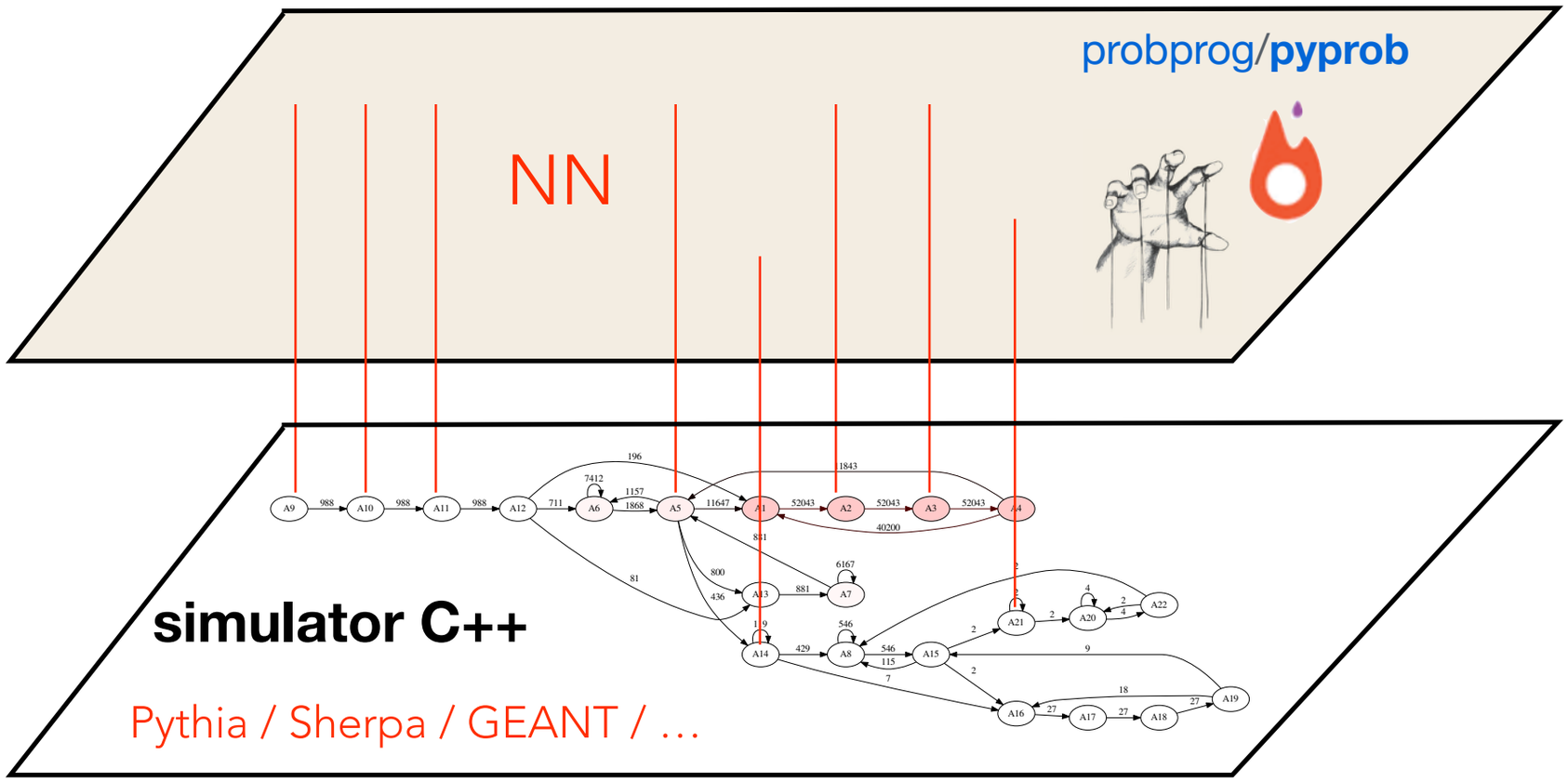
Output: a trained inference network, or "compilation artifact" (Torch, PyTorch)



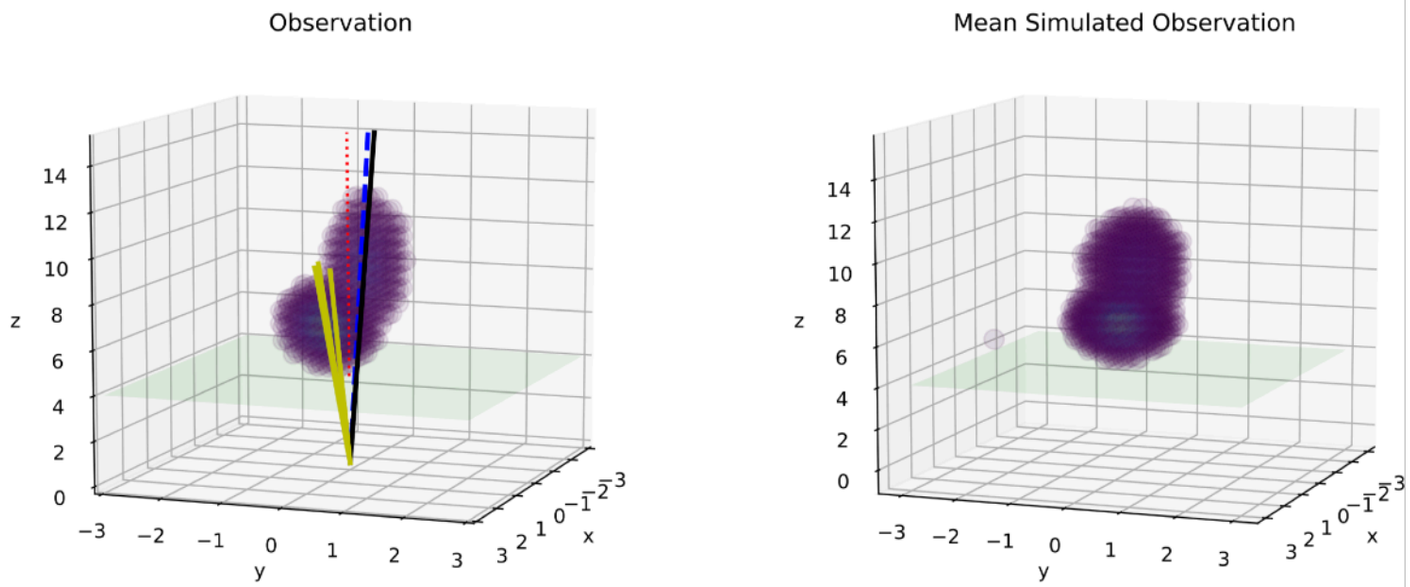
Le, Baydin and Wood. Inference Compilation and Universal Probabilistic Programming. AISTATS 2017. *arXiv:1610.09900*

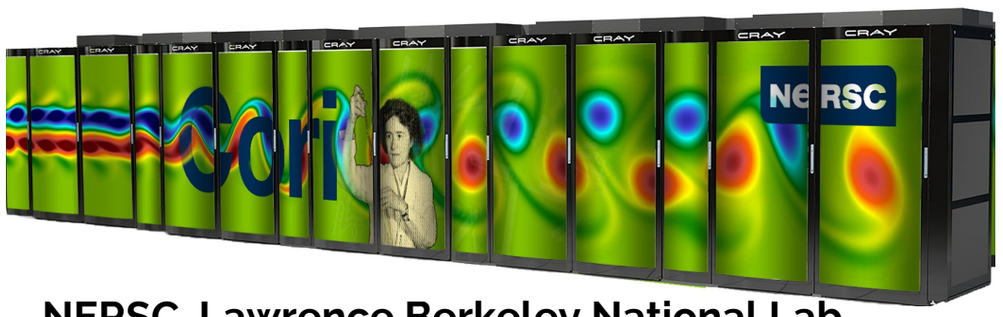
PROBABILISTIC PROGRAMING

Idea: hijack the random number generators and use Neural Network to perform a *very* fancy type of importance sampling



- Neural Network powered inference engine (python)
- real-world scientific simulator (C++)





NERSC, Lawrence Berkeley National Lab

arXiv:1807.07706 & 1907.03382

CODE

pyprob

https://github.com/probprog/pyprob

A PyTorch-based PPL

Inference engines:

- Markov chain Monte Carlo
  - Lightweight Metropolis Hastings (LMH)
  - Random-walk Metropolis Hastings (RMH)
- Importance Sampling
  - Regular (proposals from prior)
  - **Inference compilation (IC)**

Le, Baydin and Wood. Inference Compilation and Universal Probabilistic Programming. AISTATS 2017 *arXiv:1610.09900*.

PPX

https://github.com/probprog/ppx

Probabilistic **P**rogramming e**X**ecution protocol

- Cross-platform, via flatbuffers: http://google.github.io/flatbuffers/
- Supported languages: C++, C#, Go, Java, JavaScript, PHP, Python, TypeScript, Rust, Lua
- Similar to Open Neural Network Exchange (ONNX) for deep learning

Enables inference engines and simulators to be

- implemented in different programming languages
- executed in separate processes, separate machines across networks

[slide: Atılım Güneş Baydin]

etalumis simulate

Atılım Güneş Baydin
Bradley Gram-Hansen

Lukas Heinrich

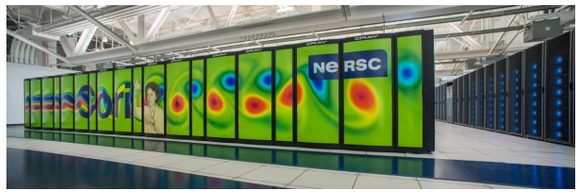
Kyle Cranmer

Frank Wood
Andreas Munk
Saeid Naderiparizi

Wahid Bhimji
Jialin Liu
Prabhat

Gilles Louppe

Lei Shao
Larry Meadows
Victor Lee



Cori supercomputer, Lawrence Berkeley Lab
2,388 Haswell nodes (32 cores per node)
9,688 KNL nodes (68 cores per node)

25

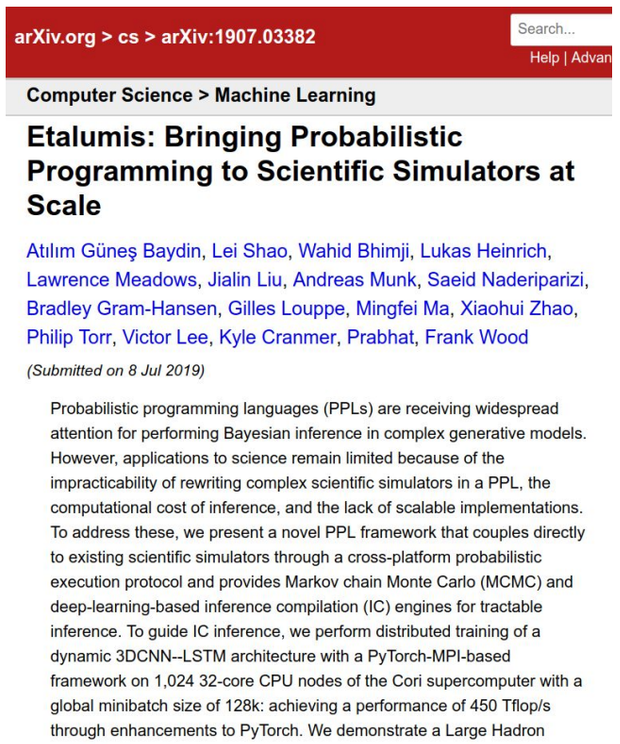
Message:
This was an heroic effort… and simple compared to uncovering the underlying theory for the data, which is a ***much*** harder problem.

Papers: computing (best paper finalist, SC19)

International Conference for High Performance Computing, Networking, Storage, and Analysis (SC19), Denver, CO, November 17–22, 2019

https://arxiv.org/abs/1907.03382

- Novel PPL framework; execution of existing stochastic simulators; HPC features including handling multi-TB data and distributed training and inference
- The largest scale posterior inference in a Turing-complete PPL; approximately 25,000 latents expressed by Sherpa code base of ~1M lines of code in C++, TBs of distributions
- Synchronous data parallel training of a dynamic 3DCNN–LSTM NN architecture using PyTorch MPI, at the scale of 1,024 nodes (32,768 CPU cores) with a global minibatch size of 128k. Largest scale use of PyTorch MPI, largest minibatch size for this form of NN

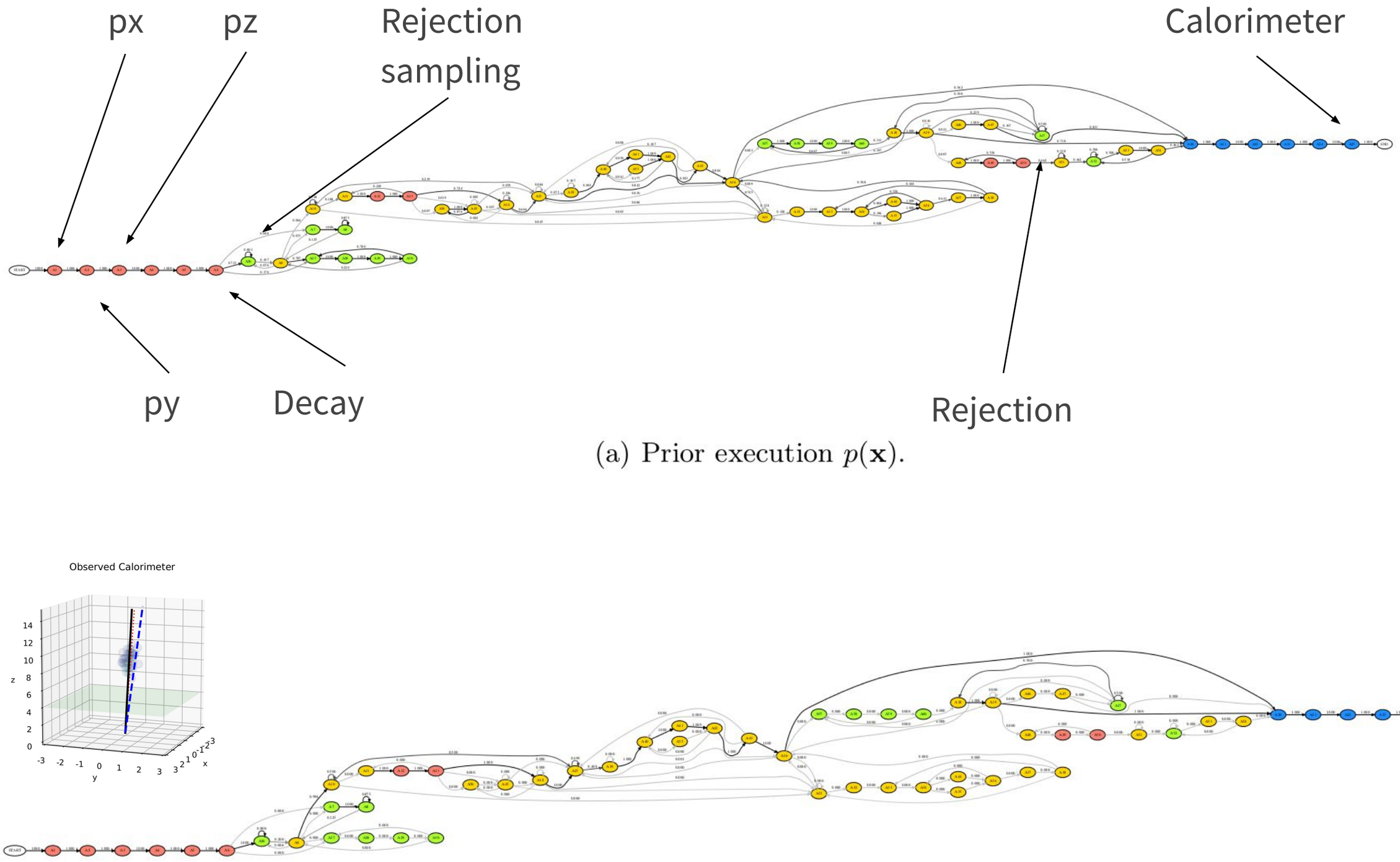


50

DEEP INTERPRETABILITY

[slide: Atılım Güneş Baydin]

Latent probabilistic structure of 250 most frequent trace types



(a) Prior execution $p(\mathbf{x})$.

(b) Posterior execution $p(\mathbf{x}|\mathbf{y})$ conditioned on a given calorimeter observation $\mathbf{y}$.

WHY DO WE CARE ABOUT INTERPRETABILITY?

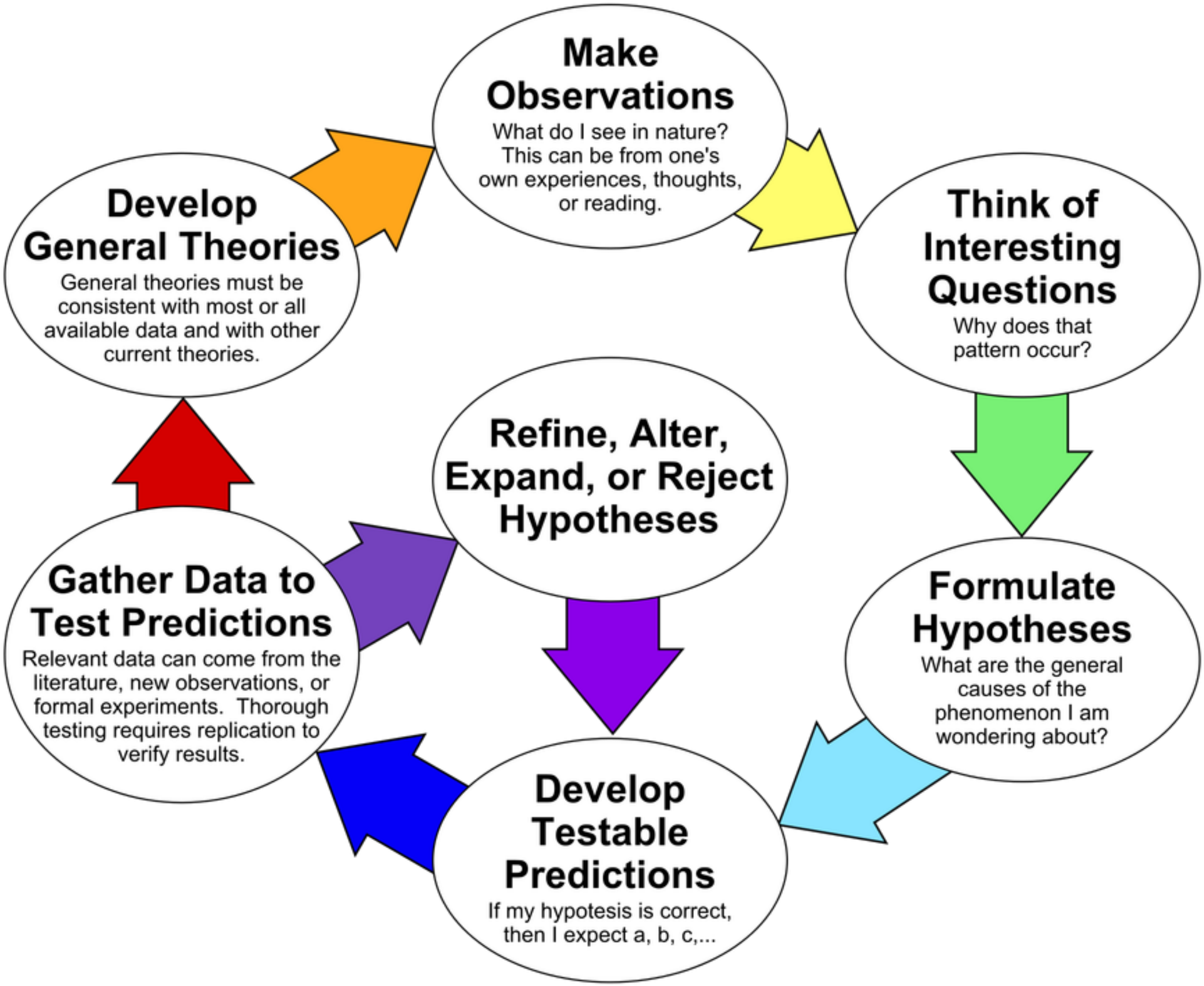
For a fixed task, one might not care about interpretability as long as the performance on the task is good

- Depending on context, "good" may mean that it generalizes well, is robust to domain shift, performance can be characterized and validated to be within some tolerance, etc…

But for progress in science, we don't just want to solve today's task well.

- For science to progress we need to be able to generate new hypotheses, design experiments, etc.

The Scientific Method as an Ongoing Process



Why physics problems are interesting for those trying to understand theory of deep learning

Growing appreciation that learning algorithm & model architecture (parametrization) interact

* Lenka Zdeborova

EFFECT OF DATA DISTRIBUTION

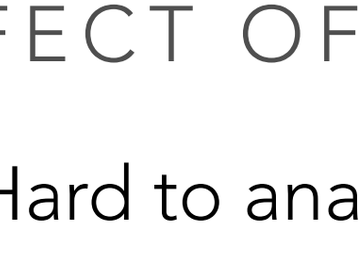
Hard to analyze effect of data structure for real-world data sources.

Toy models are useful!

SIMPLER QUESTION: WHEN CAN A NEURAL NETWORK LEARN A TEACHER-NEURAL NETWORK?

Teacher-network

- Generates data X, n samples of p dimensional data, e.g. random input vectors.
- Generates weights w*, e.g. iid random.
- Generates labels y.



Student-network

- Observes X, y, the architecture of the network.
- How does the best achievable generalisation error depend on the number of samples n?

input on random manifolds,
output from a teacher network

Teacher → Causal, Generative Model (Simulator)

Richer set of problems can be investigated.

Insight of data generating process informs inductive bias on architecture

PHYSICS PROBLEMS HAVE A LOT OF STRUCTURE

- Causal structure (we take it for granted)
- Hierarchical / compositional structure
- Rich symmetries (in data & internal to generative process)
- We can compare vast array of experiments in context of theoretical model (“transfer learning on steroids”)
- Well understood correlations
- Non-trivial “noise models” (aka detector response)

CONCLUSION

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

- leverage expert knowledge of the generative process
- learn surrogates that extract relevant features for inference task

Learning underlying physical models for data implies going beyond statistical correlations

- Need to learn a causal, generative model
- Need to make interventions, design experiments, generate hypotheses. This is one reason why interpretability is important
- Attempting to learn simulators with known ground truth is a good starting point

Some Quotes

Three principles — the conformability of nature to herself, the applicability of the criterion of simplicity, and the "unreasonable effectiveness" of certain parts of mathematics in describing physical reality — are thus consequences of the underlying law of the elementary particles and their interactions.

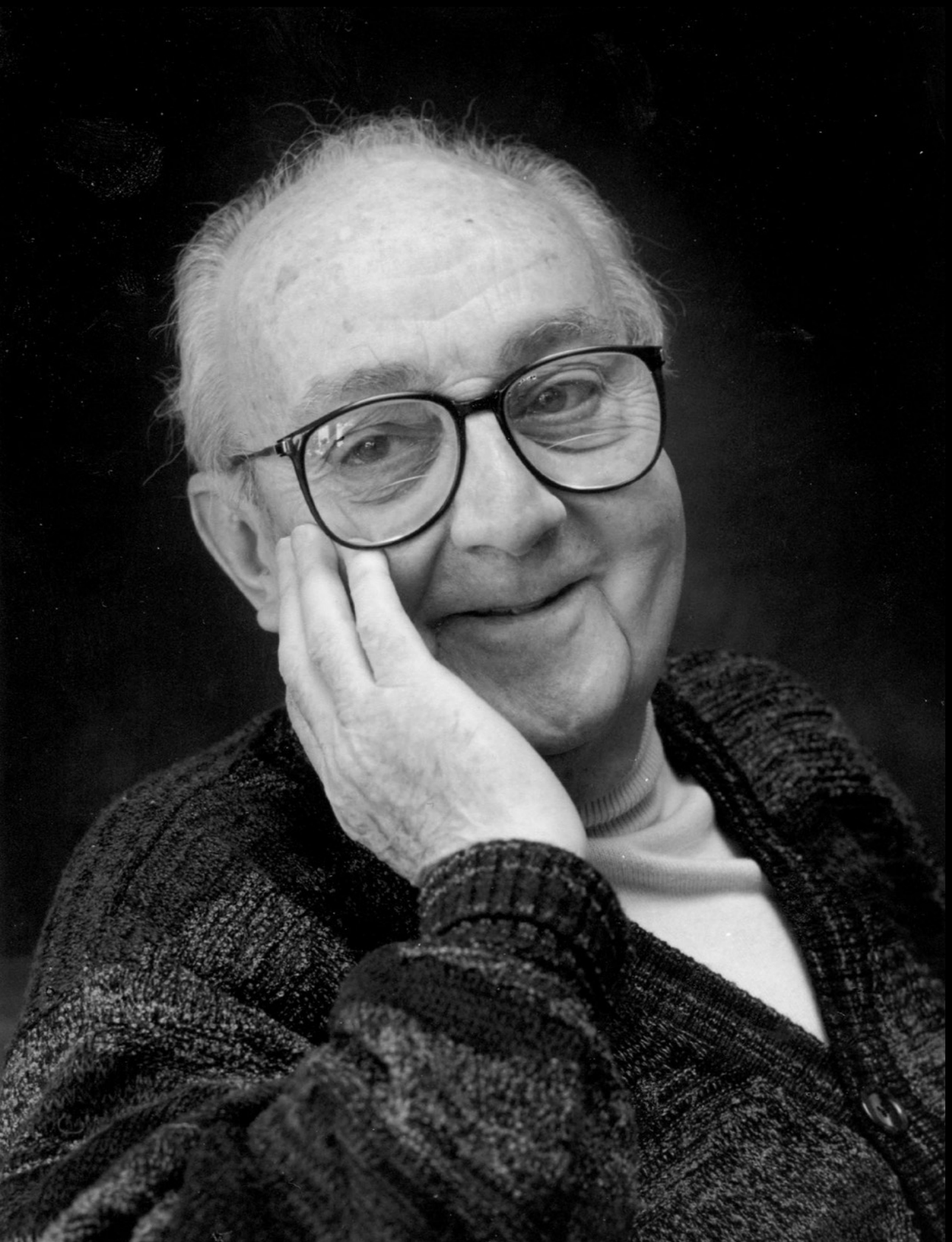
–MURRAY GELL-MANN

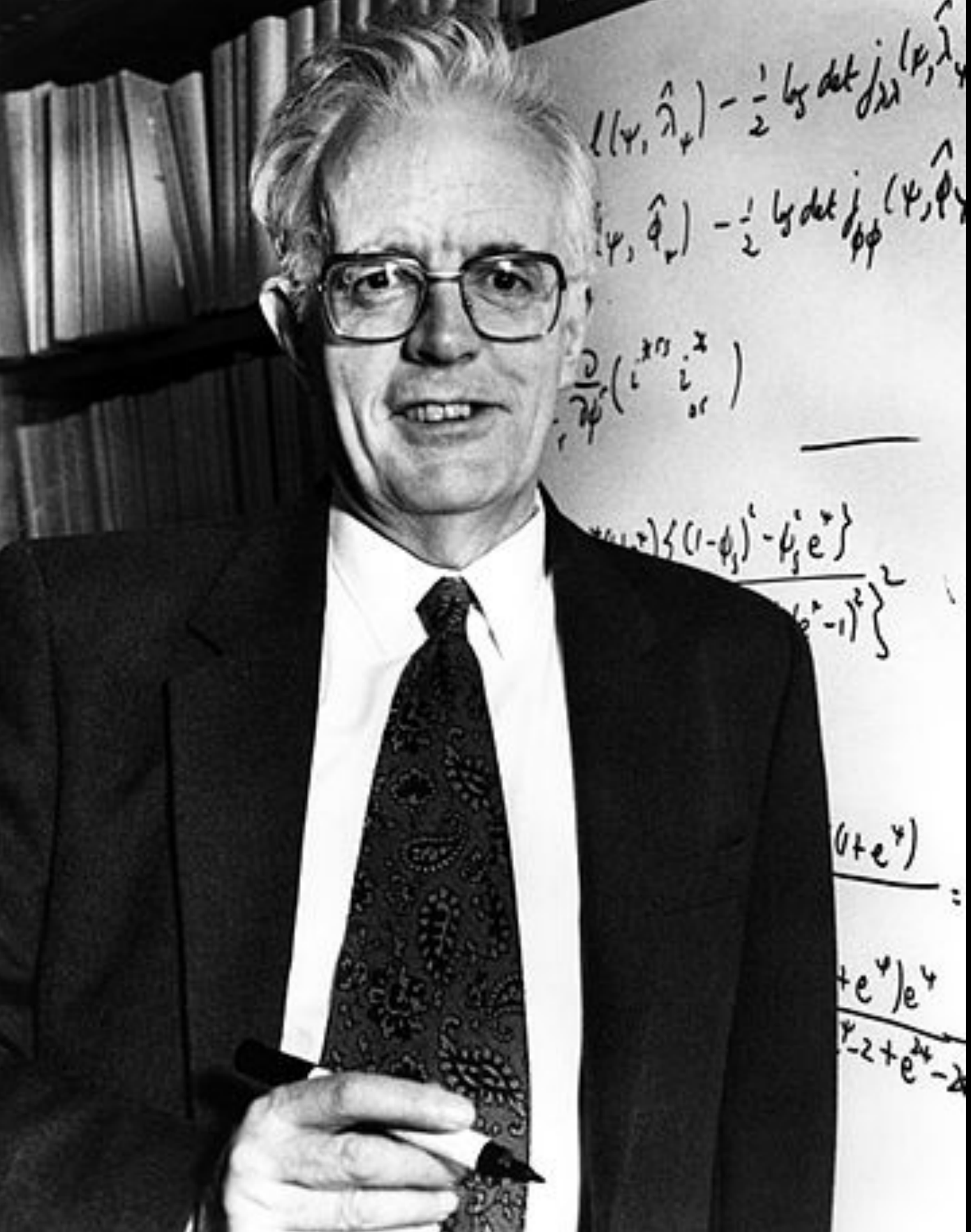
It doesn't matter how beautiful your guess is,
if it disagrees with experiment, it's wrong.

– RICHARD P. FEYNMAN

All models are wrong, but some are are useful.

–GEORGE BOX



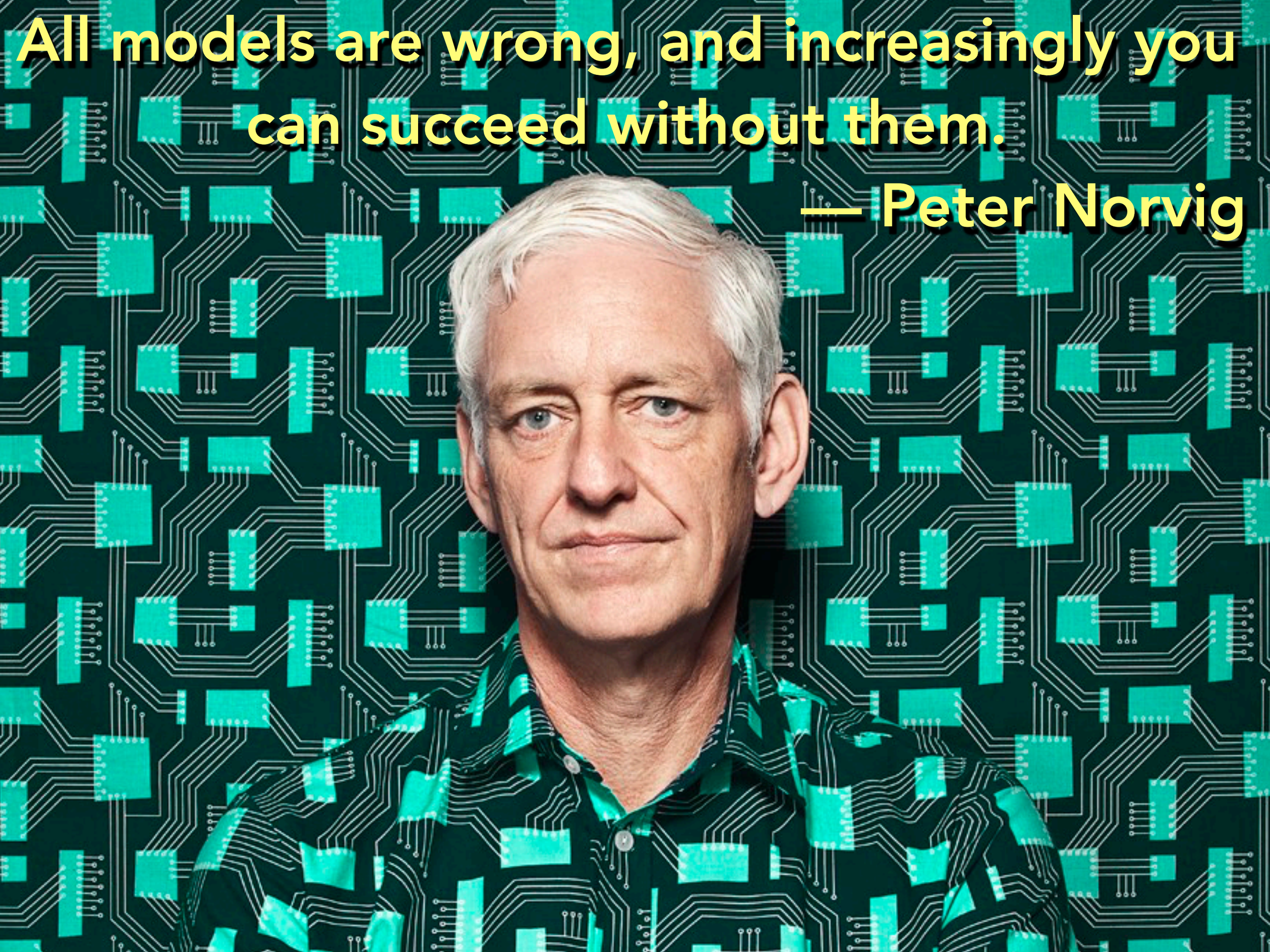


The very word model implies simplification and idealization. **The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd.** The construction of idealized representations that capture important stable aspects of such systems is, however, a vital part of general scientific analysis and statistical models, especially substantive ones, do not seem essentially different from other kinds of model.

–SIR DAVID COX

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

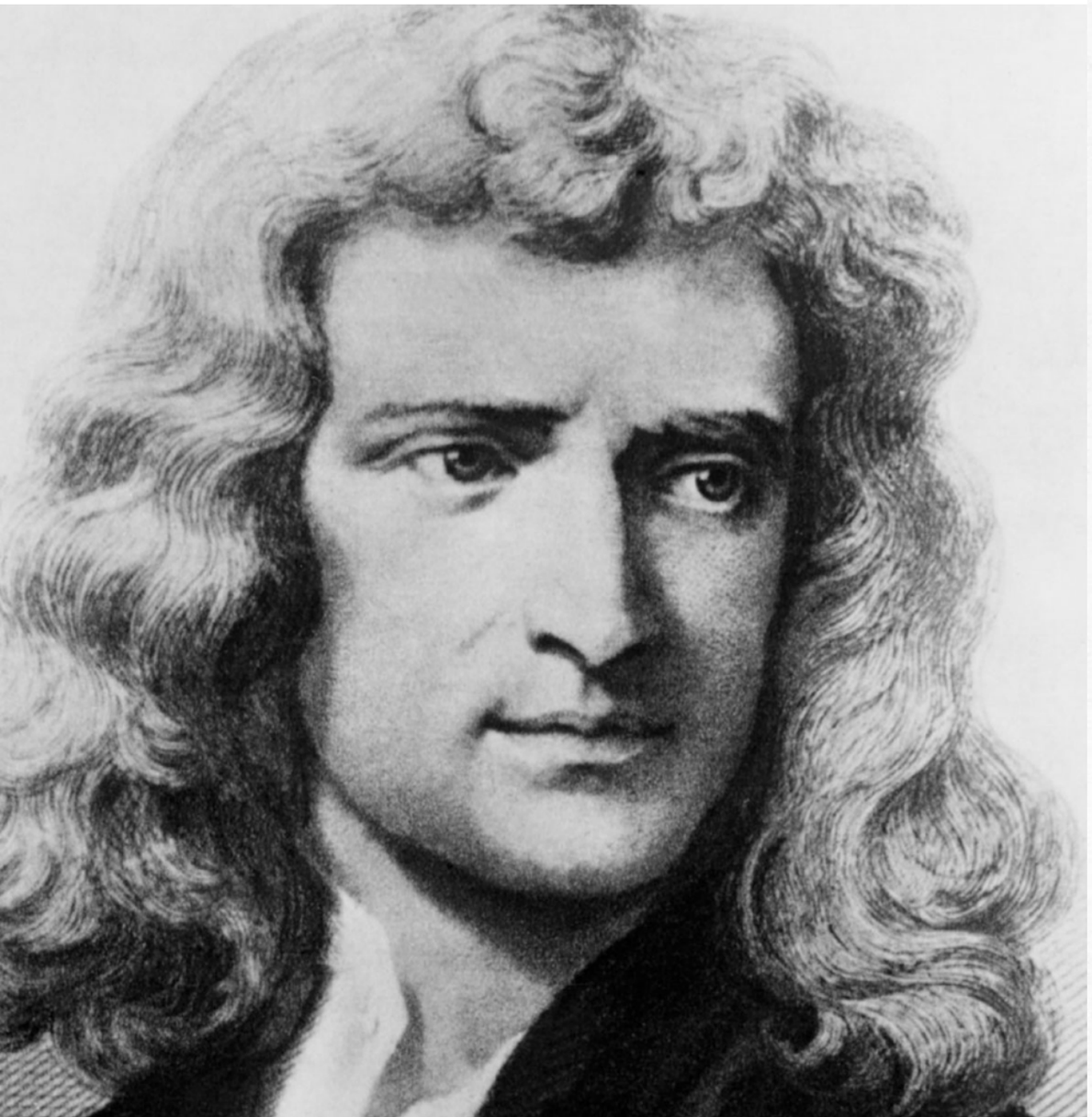
— Peter Norvig



CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

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

Empiricism (Newton) vs. Rationalism (Gottfried Leibniz)



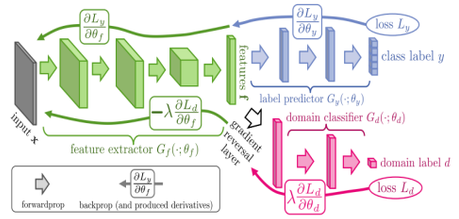
Learning Representations using Causal Invariance

LÉON BOTTOU

FACEBOOK AI RESEARCH & NEW YORK UNIVERSITY

3- Adversarial Domain Adaptation

- The goal is to learn a classifier that does not depend on the environment.
- An adversarial term makes it hard to recover the environment label e from the representation $\phi(x)$.



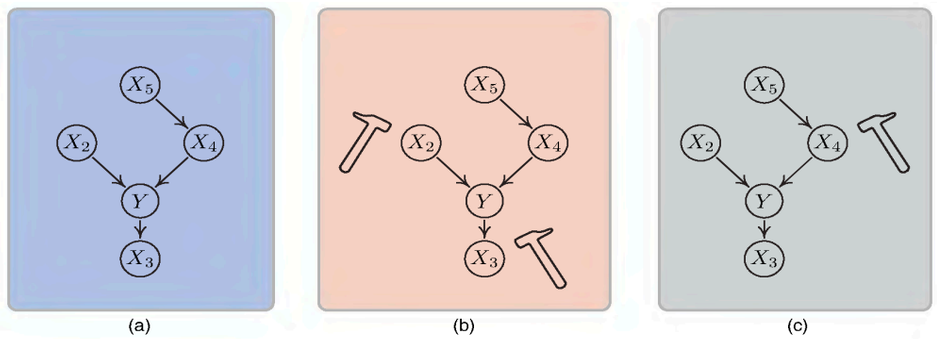
- This implies that $\mathbb{P}\big(\phi(X_e)\big)$ does not depend on e.
  Therefore $\mathbb{P}\{f(X_e)\}$ does not depend on e either. But $\mathbb{P}\{Y_e\}$ might..
- Conditional ADA stratifies on Y to achieve $\mathbb{P}(\phi(X_e)|Y_e) \perp\!\!\!\perp e$.
  Hence $\mathbb{E}(\phi(X_e)|Y_e) \perp\!\!\!\perp e$ instead of $\mathbb{E}\big(Y_e|\phi(X_e)\big) \perp\!\!\!\perp e$.

(Ganin et al., 2015; Edwards & Storkey, 2015; Louppe et al., 2016; Li et al,

2- Invariant causal prediction

Inspiration

- Environments result from interventions on a causal graph.



- The set of variables in the graph is assumed known.
- Representations ϕ merely select a subset of the variables.

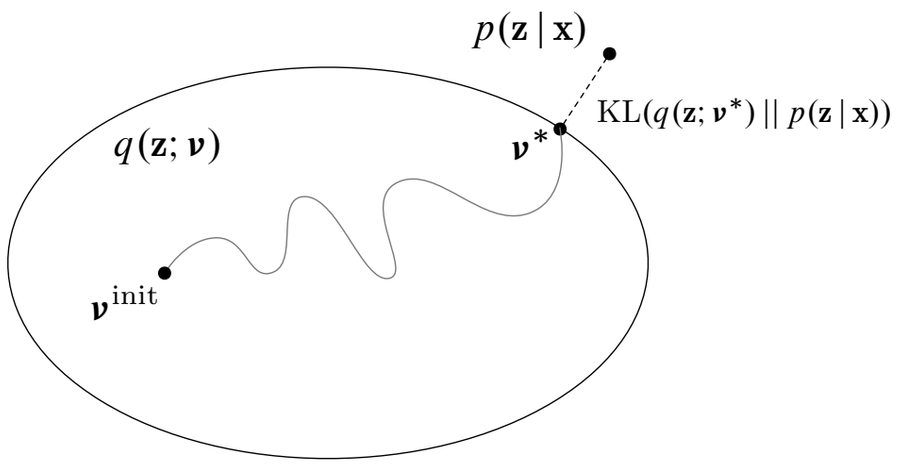
If we find an invariant representation,
we have recovered the direct causes of Y.

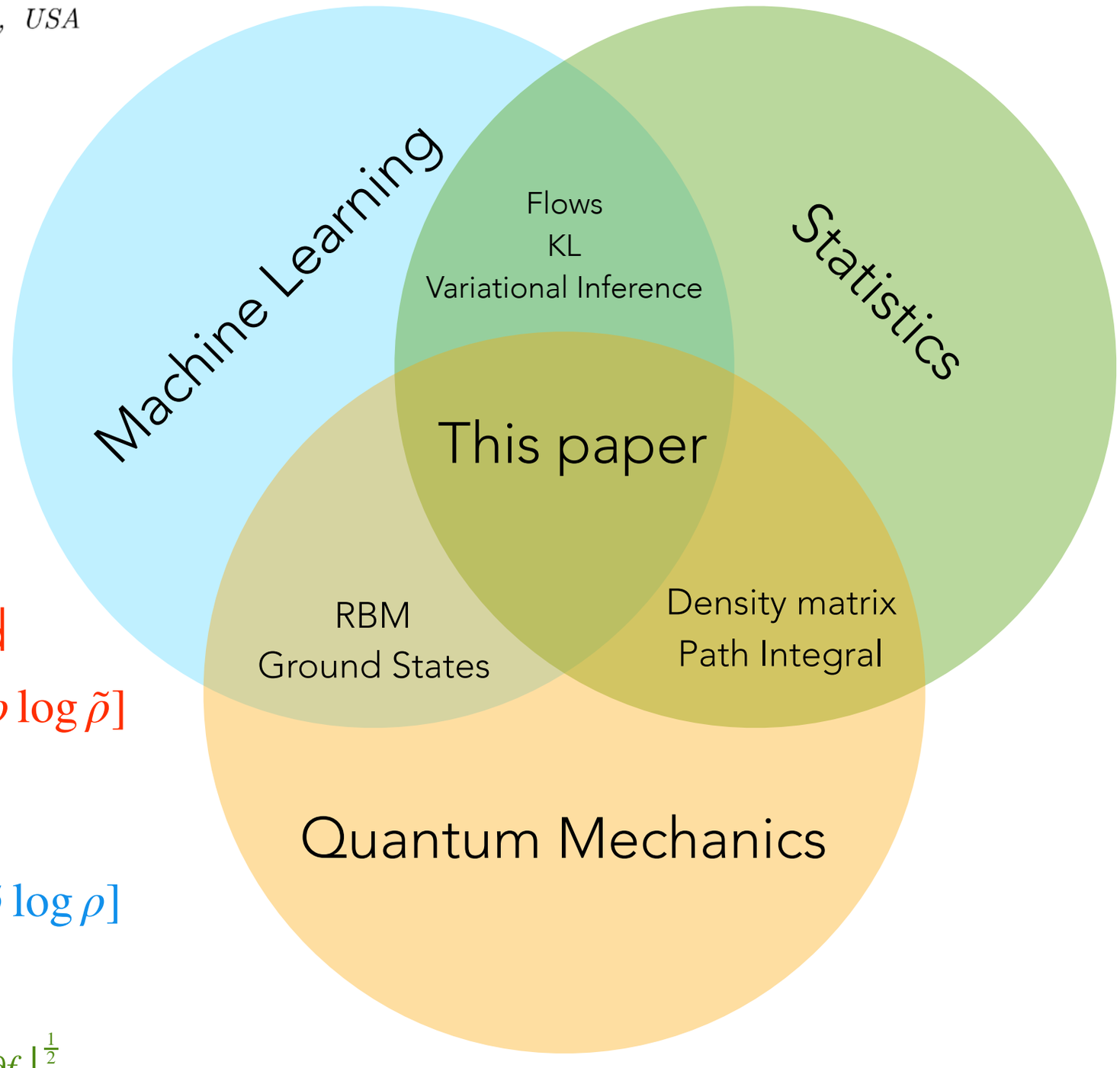
(Peters, Bühlman, Meinshausen, 2016)

Inferring the quantum density matrix with machine learning

Kyle Cranmer* and Siavash Golkar†
Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA.

Duccio Pappadopulo‡
Bloomberg LP, New York, NY 10022, USA





Quantum Maximum Likelihood

$$\mathrm{KL}[p\|\,q] \rightarrow S[\rho\|\,\tilde{\rho}] = \mathrm{Tr}[\rho\log\rho] - \mathrm{Tr}[\rho\log\tilde{\rho}]$$

Quantum Variational Inference

$$\mathrm{KL}[q\|\,p] \rightarrow S[\tilde{\rho}\|\,\rho] = \mathrm{Tr}[\tilde{\rho}\log\tilde{\rho}] - \mathrm{Tr}[\tilde{\rho}\log\rho]$$

Quantum Normalizing Flows

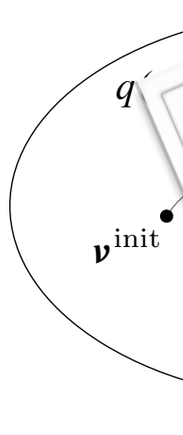
$$p(x) = p(f(x))\left|\det\frac{\partial f}{\partial x}\right| \rightarrow \psi_i(x) = \phi_i(f(x))\left|\det\frac{\partial f}{\partial x}\right|^{\frac{1}{2}}$$

KC, Golkar, Pappadopulo, doi:10.6084/m9.figshare.6197069; arXiv:1904.05903
Pfau, Petersen, Agarwal, Barrett, Stachenfeld, arXiv: 1806.02215

Inferring the quantum density matrix with machine learning

Kyle Cranmer* and Siavash Golkar†
Center for Cosmology and Particle Physics, Department of Ph
New York University, New York, NY 10003, US

Duccio Pappadopul
Bloomberg LP, New York

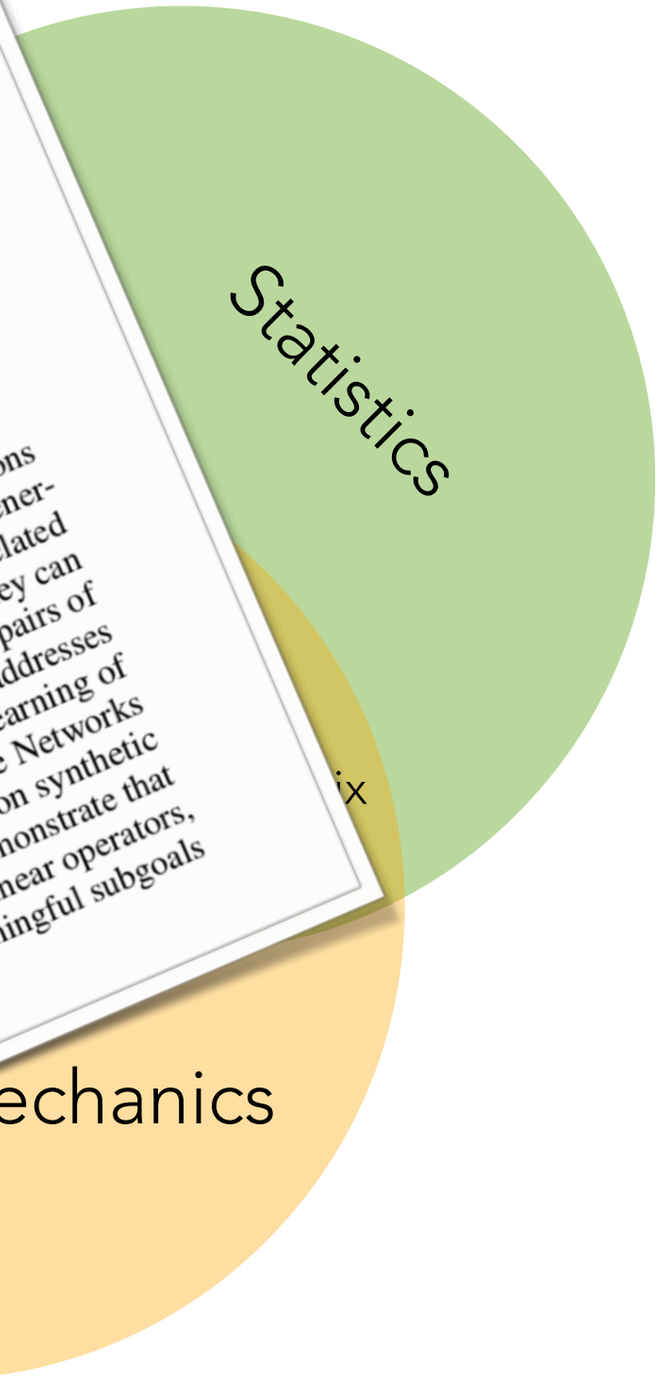


Spectral Inference Networks: Unifying Spectral Methods With Deep Learning

David Pfau[1], Stig Petersen[1], Ashish Agarwal[2], David Barrett[1] and Kim Stachenfeld[1]
[1]DeepMind London, UK
[2]Google Brain Mountain View, CA, USA
{pfau, svp, agarwal, barrettdavid, stachenfeld}@google.com

Abstract

We present Spectral Inference Networks, a framework for learning eigenfunctions of linear operators by stochastic optimization. Spectral Inference Networks generalize Slow Feature Analysis to generic symmetric operators, and are closely related to Variational Monte Carlo methods from computational physics. As such, they can be a powerful tool for unsupervised representation learning from video or pairs of data. We derive a training algorithm for Spectral Inference Networks that addresses the bias in the gradients due to finite batch size and allows for online learning of multiple eigenfunctions. We show results of training Spectral Inference Networks on problems in quantum mechanics and feature learning for videos on synthetic datasets as well as the Arcade Learning Environment. Our results demonstrate that Spectral Inference Networks accurately recover eigenfunctions of linear operators, can discover interpretable representations from video and find meaningful subgoals in reinforcement learning environments.



Quantum Maxi… KL[p‖ q] → S[ρ‖ $\tilde{\rho}$] …

Quantum Variation… KL[q‖ p] → S[$\tilde{\rho}$‖ ρ] = Tr[…

Quantum Normalizing …

$$p(x) = p(f(x)) \left|\det \frac{\partial f}{\partial x}\right| \quad \rightarrow \quad \psi_i(x) = q \ldots \left|\det \frac{\partial f}{\partial x}\right|^2$$

KC, Golkar, Pappadopulo, doi:10.6084/m9.figshare.6197069; arXiv:1904.05903
Pfau, Petersen, Agarwal, Barrett, Stachenfeld, arXiv: 1806.02215