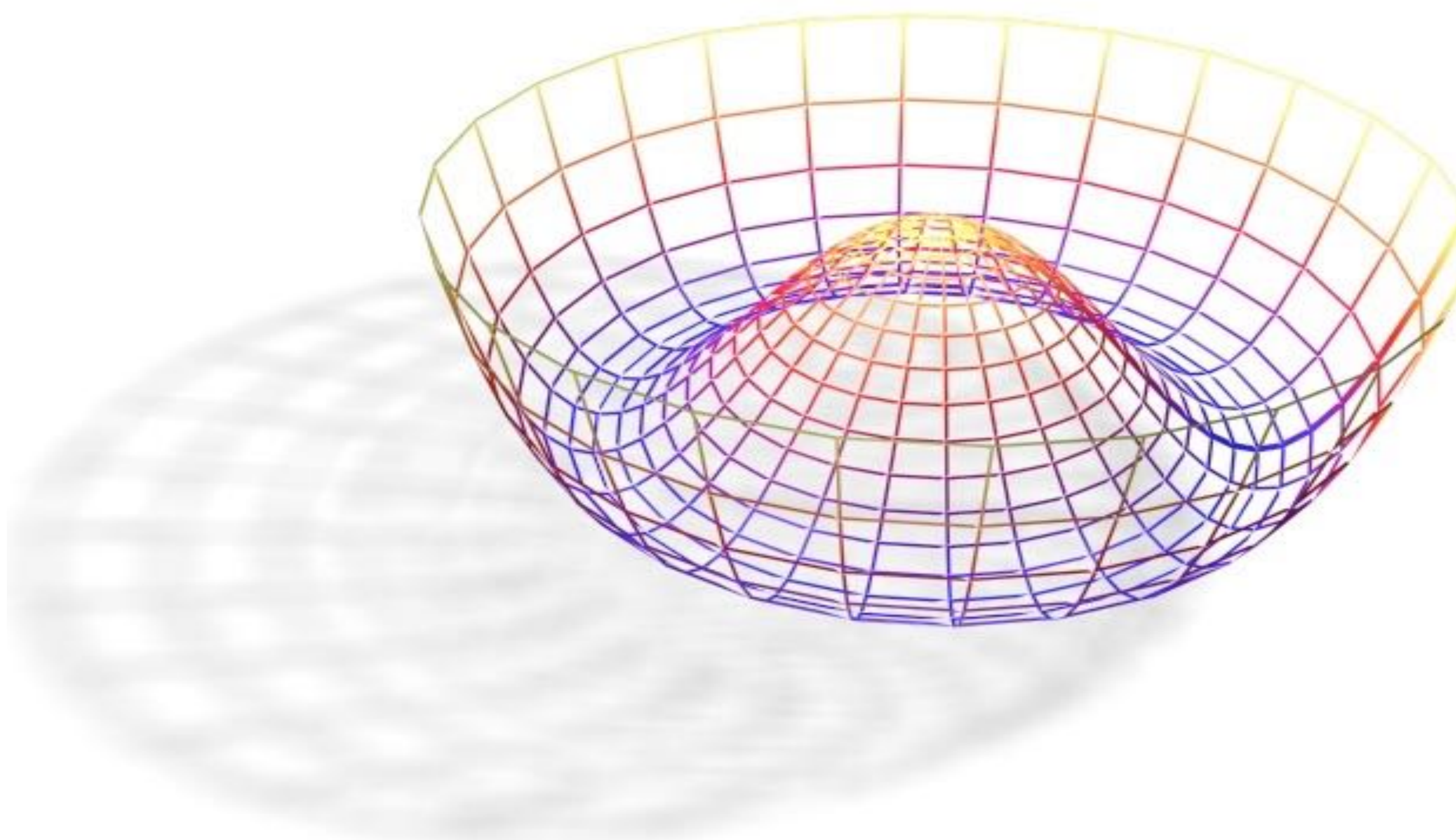
NYU CENTER FOR DATA SCIENCE

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

LITTLE DECISIONS WITH BIG DATA

BIG DECISIONS WITH LITTLE DATA



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

THE CHARGE

"We would delighted if you would come and talk about the Big Data and Particle Physics at a level which would excite our graduate students. "

Buzz words

- Big Data
- Data Science
- Machine Learning & Artificial Intelligence

Big Data

video: https://cds.cern.ch/record/1541893 video: https://cds.cern.ch/record/1541893

BIG DATA

Yes, of course the LHC has "Big Data", but not much in common with "Big Data" in industry:

- our data is very clean and organized; industry data is messy (missing values, heterogenous formats, etc.)
- we are usually testing specific theories; in industry, often trying to learn some predictive model from the data
- our data is collected under stable conditions (i.i.d.); in industry data collection biases the data in unknown ways, causation vs. correlation is difficult
- industry largely leap-frogged HEP Grid computing with the technologies like map-reduce, cloud computing, containers, etc.

For these reasons, HEP has not been particularly relevant to the discussions around "Big Data"

FROM BIG TO LITTLE DATA

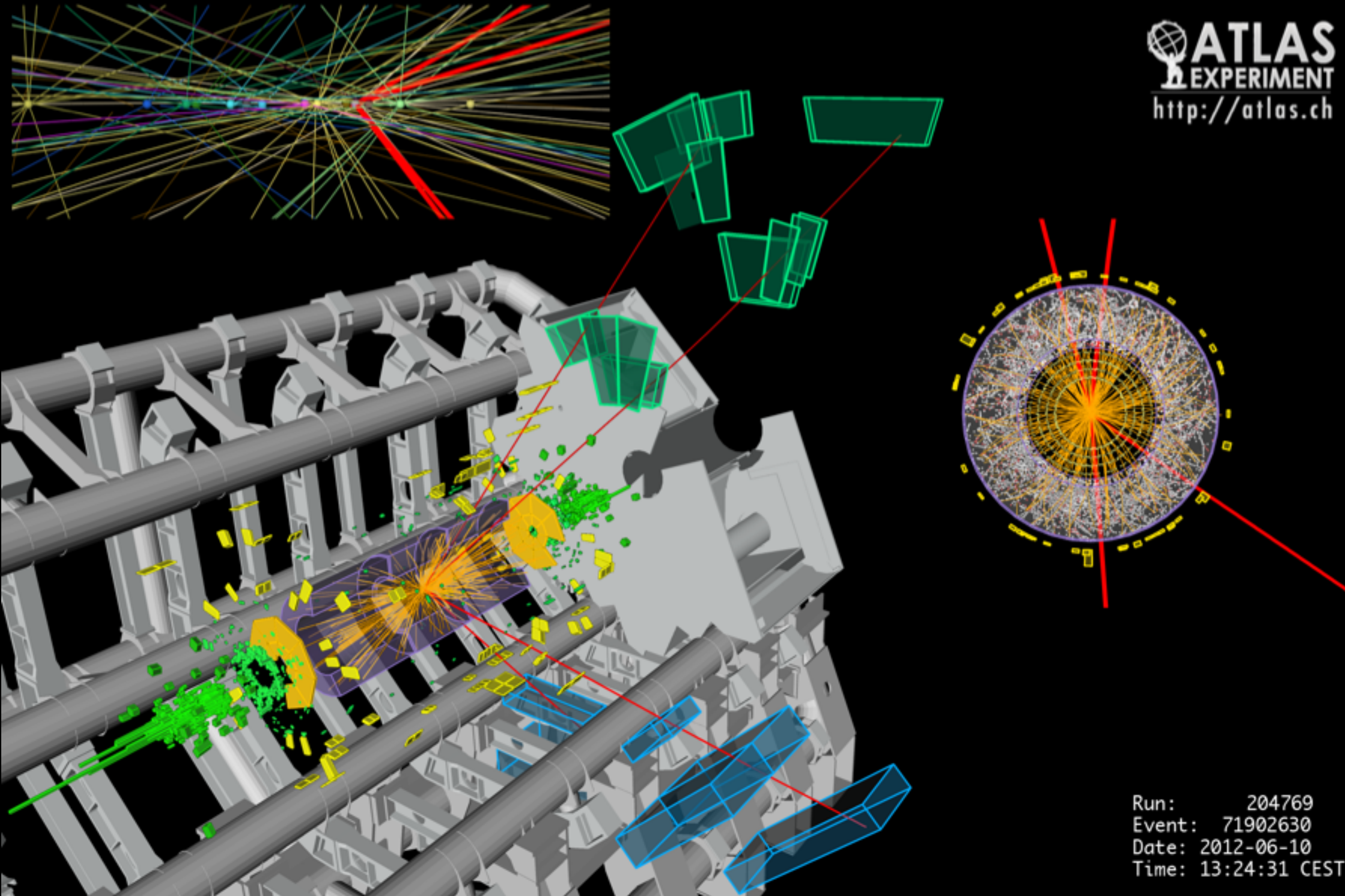
Our "big data" processing involves "little decisions"

- **Trigger**: quickly decide to keep or not an event based on a fairly simple criterion
- **Particle Identification**: quickly decide which particle type is most consistent with an energy deposit in the detector
- **Reconstruction**: quickly estimate the energy and momentum of that particle

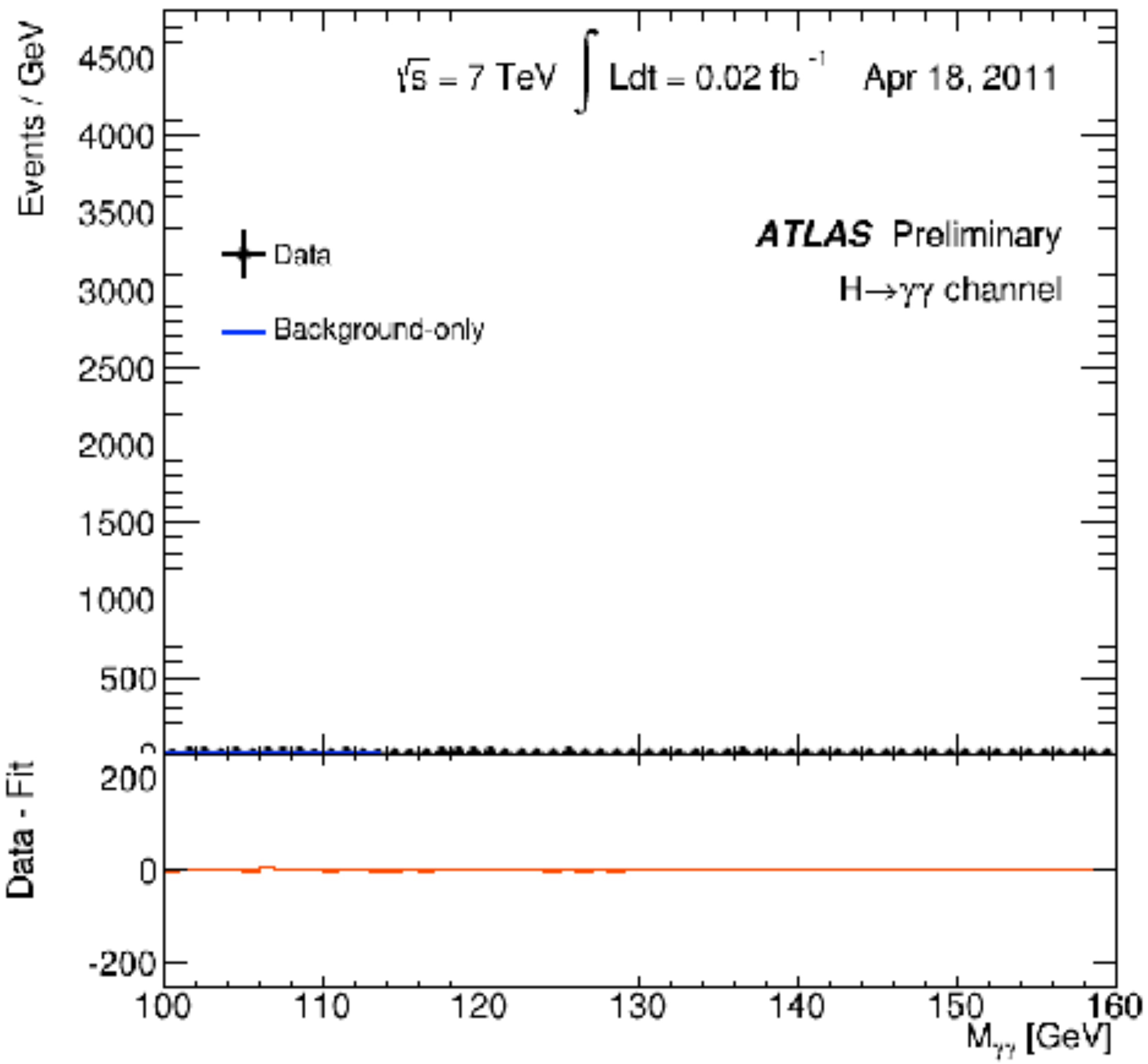
Generally, data volume reduces as decisions become more significant

- **skim**: quick selection of events of interest for an analysis based on simple properties like the number of electrons
- **pre-selection:** gradually increasing level of sophistication in requirements
- **final event selection**: highly-optimized cuts on most sophisticated variables, may involve machine learning techniques

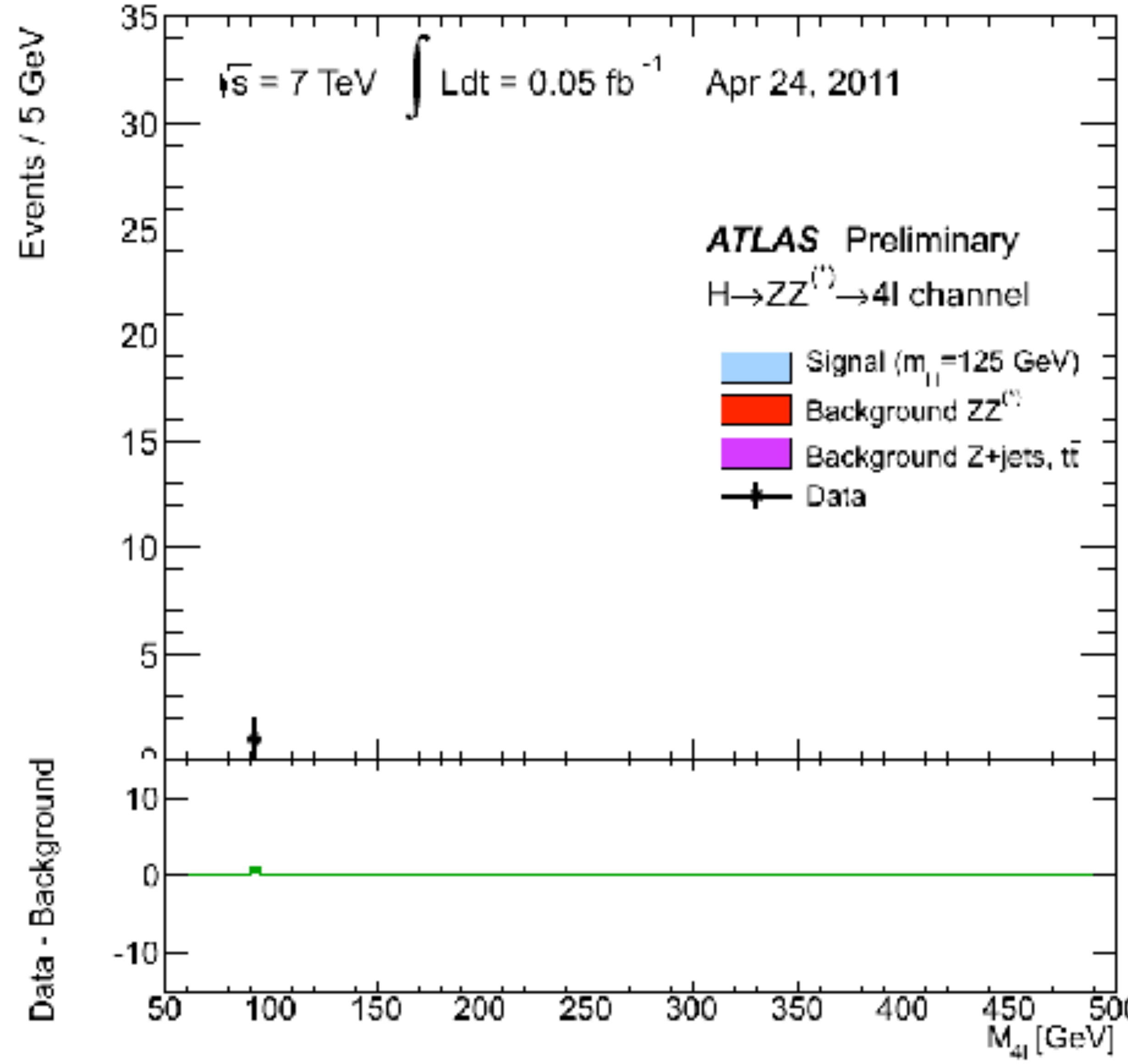
$$H \to ZZ \to 4l$$


ATLAS
EXPERIMENT
http://atlas.ch
Run: 204769
Event: 71902630
Date: 2012-06-10
Time: 13:24:31 CEST

Discovery!

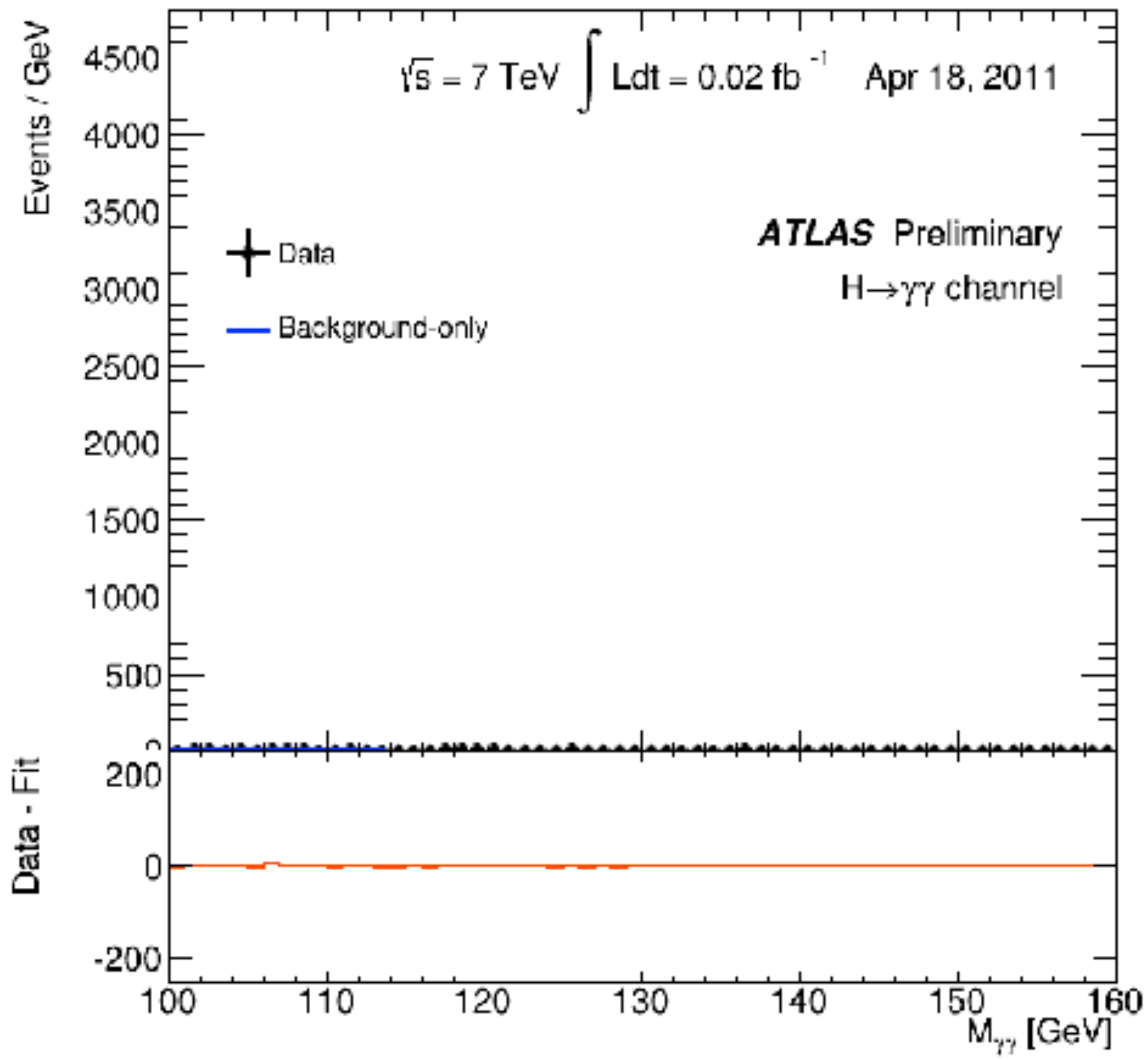


√s = 7 TeV ∫ Ldt = 0.02 fb⁻¹ Apr 18, 2011
ATLAS Preliminary
H→γγ channel
Data
Background-only
Events / GeV
Data - Fit
M γγ [GeV]

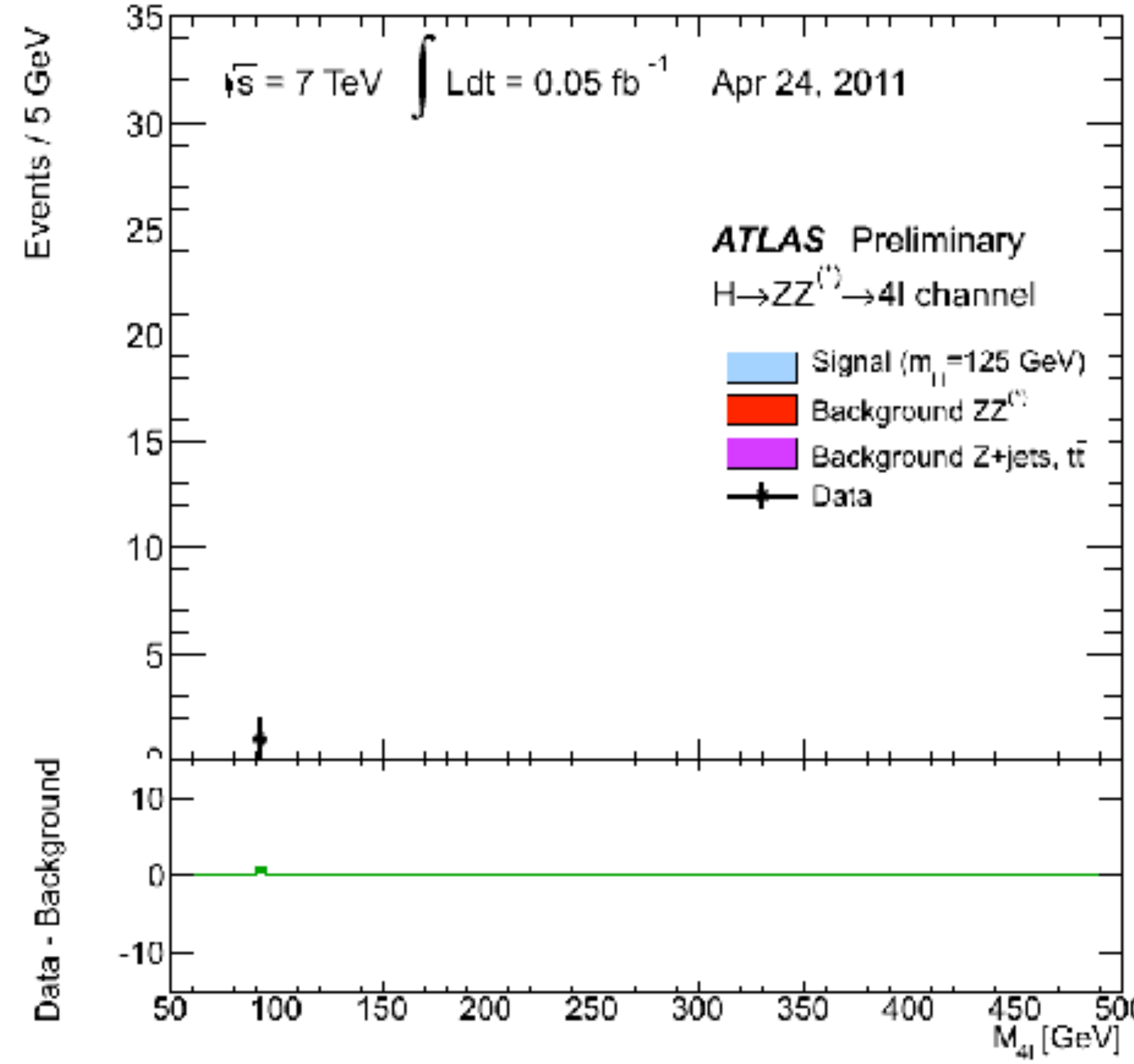


√s = 7 TeV ∫ Ldt = 0.05 fb⁻¹ Apr 24, 2011
ATLAS Preliminary
H→ZZ(*)→4l channel
Signal (mH=125 GeV)
Background ZZ(*)
Background Z+jets, tt̄
Data
Events / 5 GeV
Data - Background
M 4l [GeV]

Discovery!



√s = 7 TeV ∫ Ldt = 0.02 fb⁻¹ Apr 18, 2011
ATLAS Preliminary
H→γγ channel
Data
Background-only
Events / GeV
Data - Fit
M_γγ [GeV]



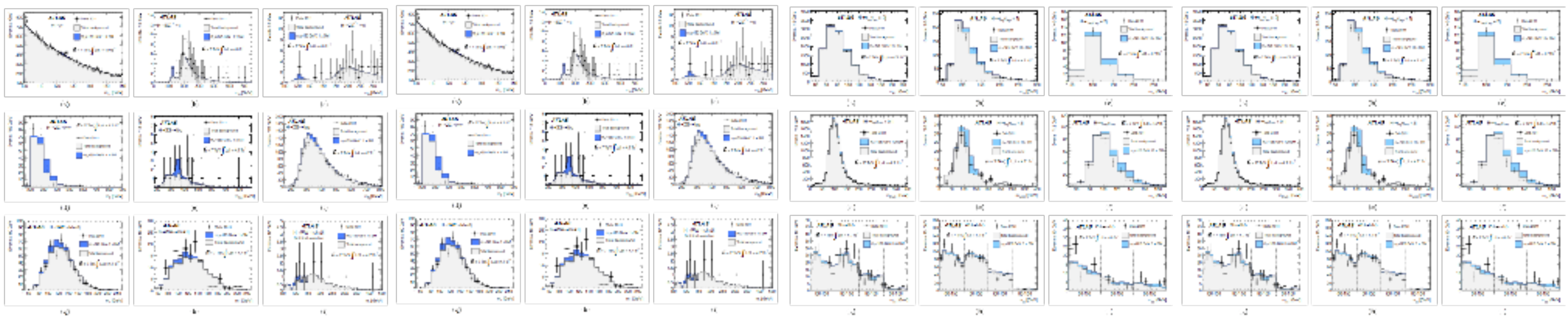
√s = 7 TeV ∫ Ldt = 0.05 fb⁻¹ Apr 24, 2011
ATLAS Preliminary
H→ZZ(*)→4l channel
Signal (m_H=125 GeV)
Background ZZ(*)
Background Z+jets, tt̄
Data
Events / 5 GeV
Data - Background
M_4l [GeV]

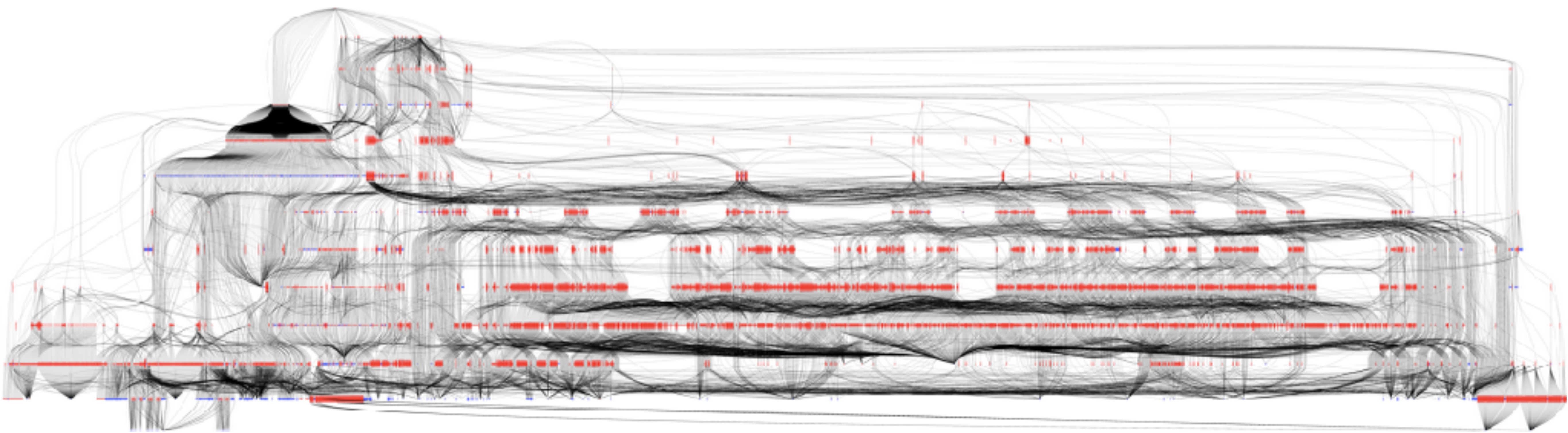
The Big Decisions are based on statistics

- **Exclusion Limits:** likelihood ratio tests
- **Discovery**: hypothesis tests based on likelihood ratio
- **Measurements**: maximum likelihood estimators

In each case, we need a statistical model for the data.

Collaborative Statistical Modeling





$$\mathbf{f}_{\mathrm{tot}}(\mathcal{D}_{\mathrm{sim}}, \mathcal{G}|\boldsymbol{\alpha}) = \prod_{c \in \mathrm{channels}} \left[\mathrm{Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha}) \right] \cdot \prod_{p \in \mathbb{S}} f_p(a_p|\alpha_p)$$

The Nobel Prize in Physics 2013

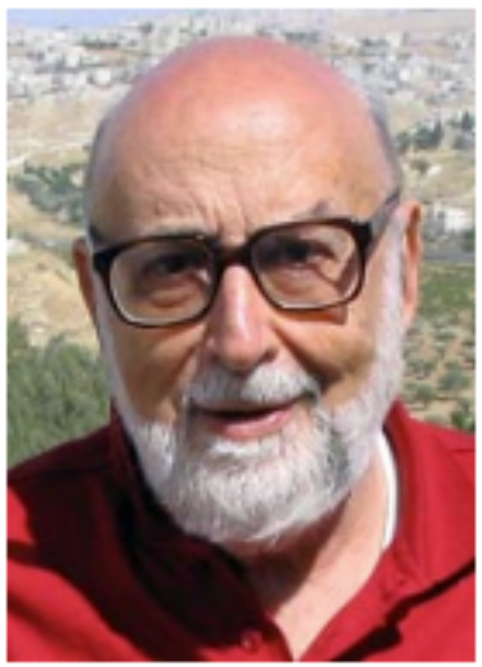


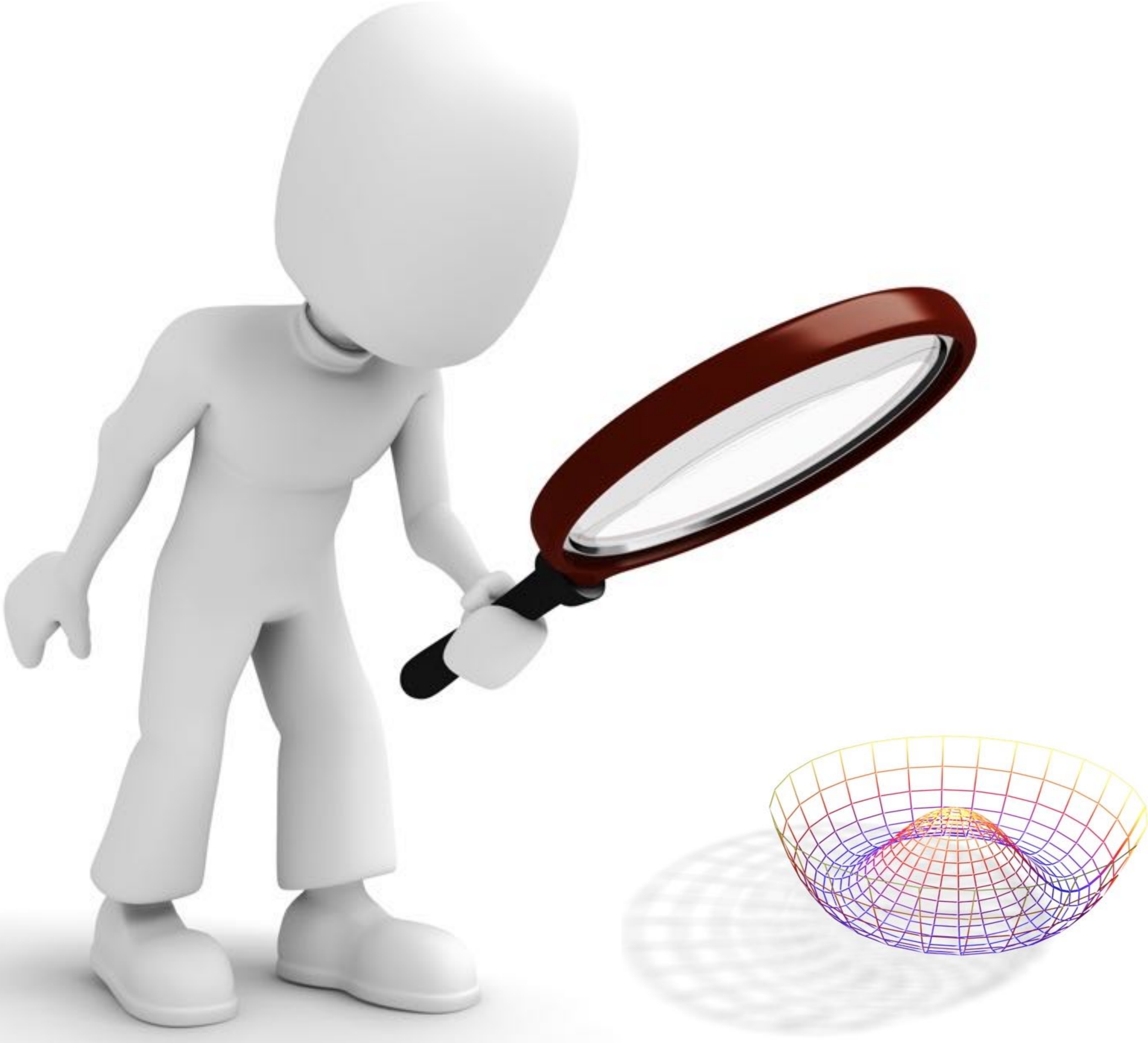
Photo: Pnicolet via Wikimedia Commons

François Englert

Photo: G-M Greuel via Wikimedia Commons

Peter W. Higgs

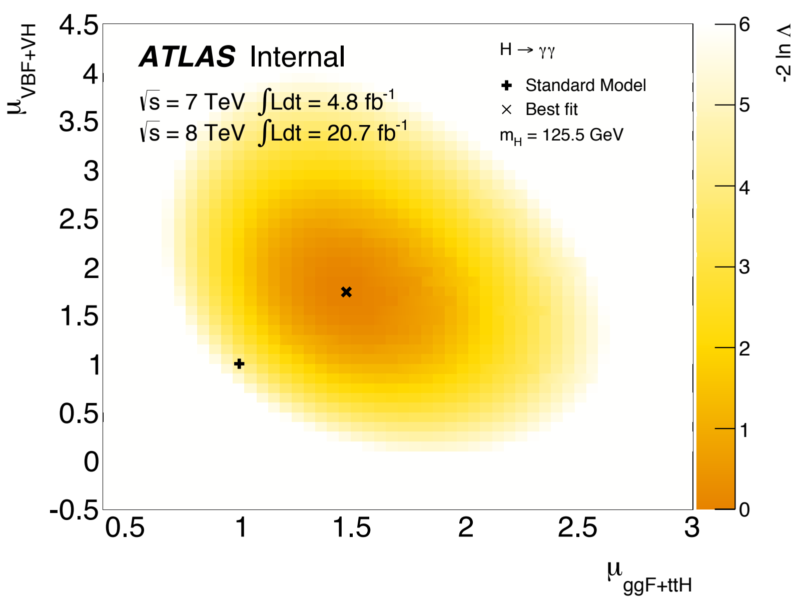
The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"*



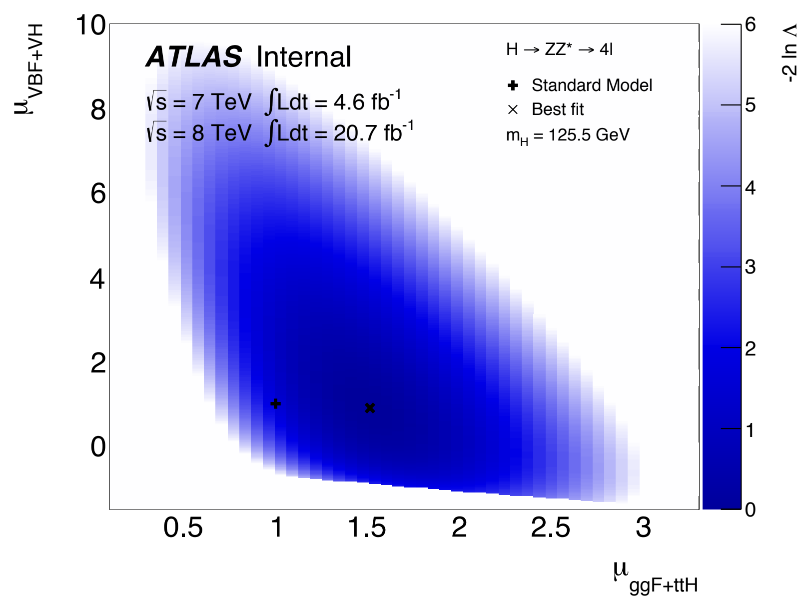
WHAT INFO AND HOW TO RETRIEVE IT

Through collaboration with theoretical community, we were able to identify a targeted form of data sharing that balanced generality &

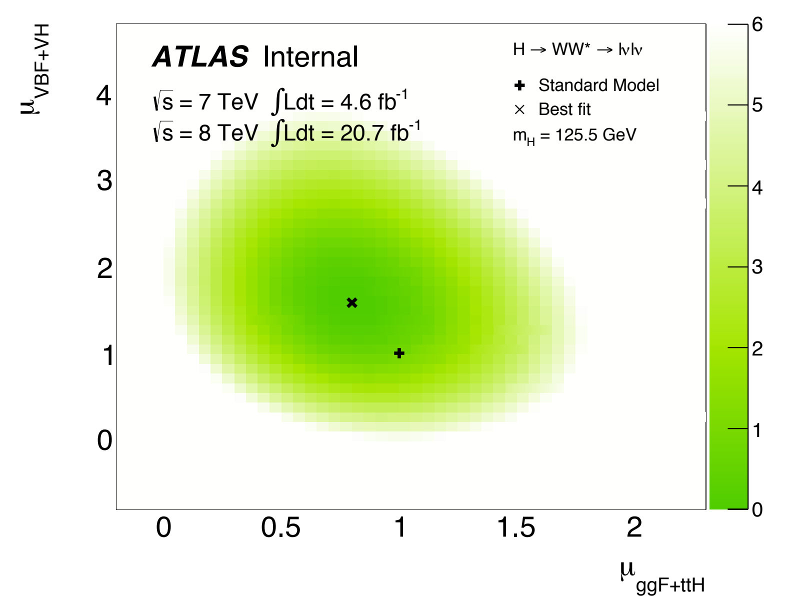
http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44



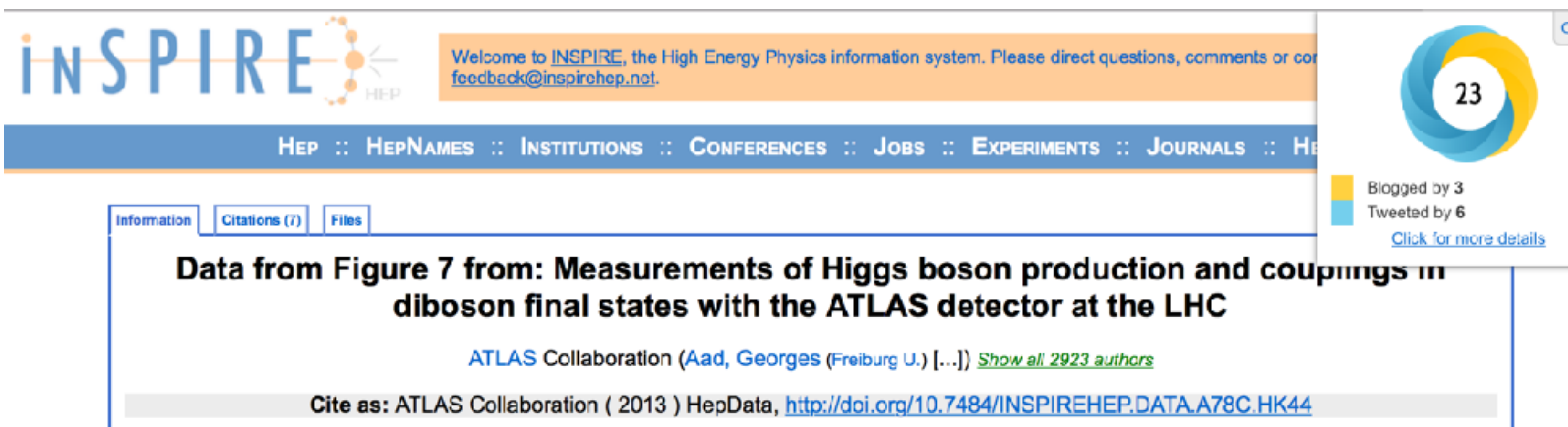
http://doi.org/10.7484/INSPIREHEP.DATA.RF5P.6M3K



http://doi.org/10.7484/INSPIREHEP.DATA.26B4.TY5F

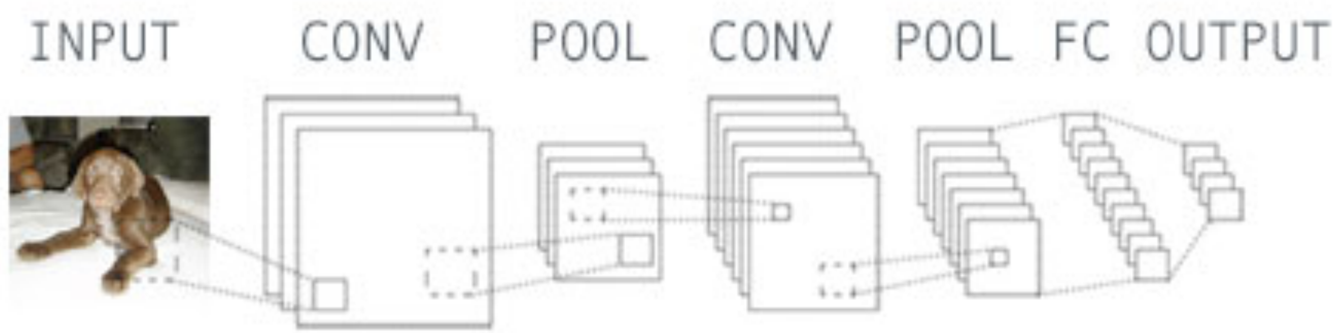


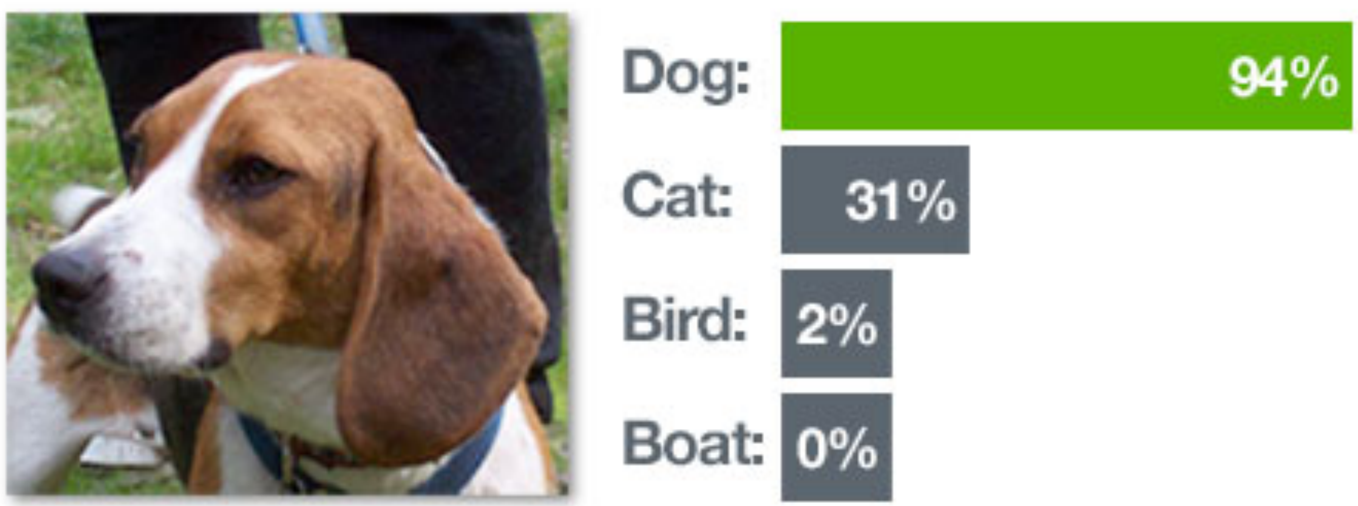
These data are directly linked to the paper in INSPIRE and have been cited:

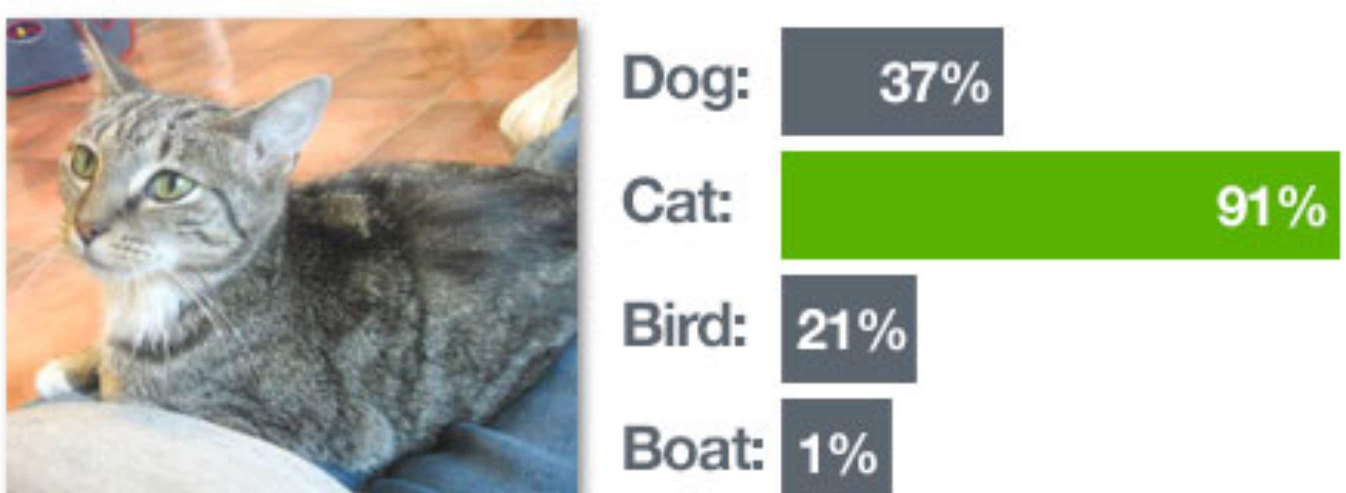


Rethinking HEP Data Analysis
Post Higgs Discovery

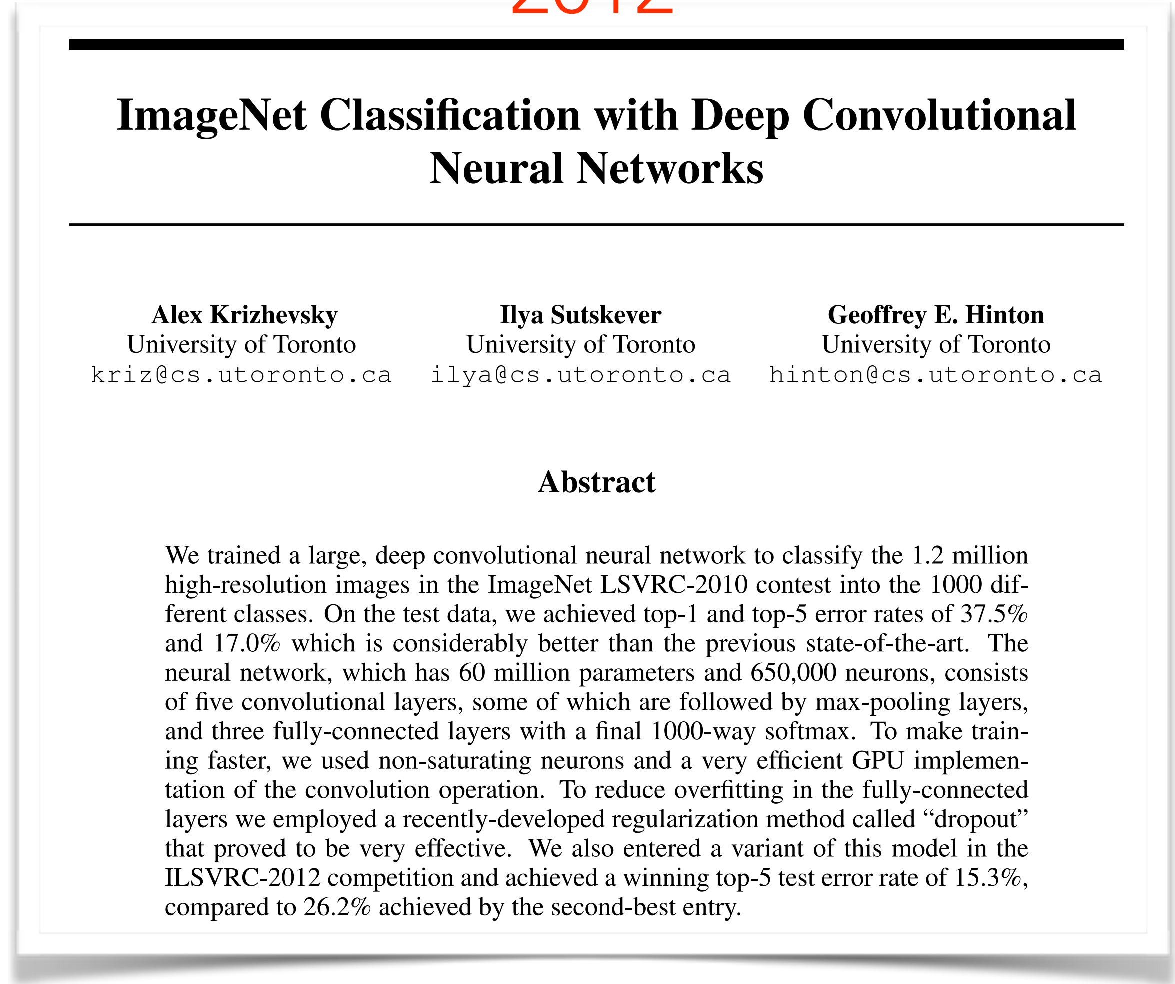
DEEP LEARNING REVOLUTION







2012



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

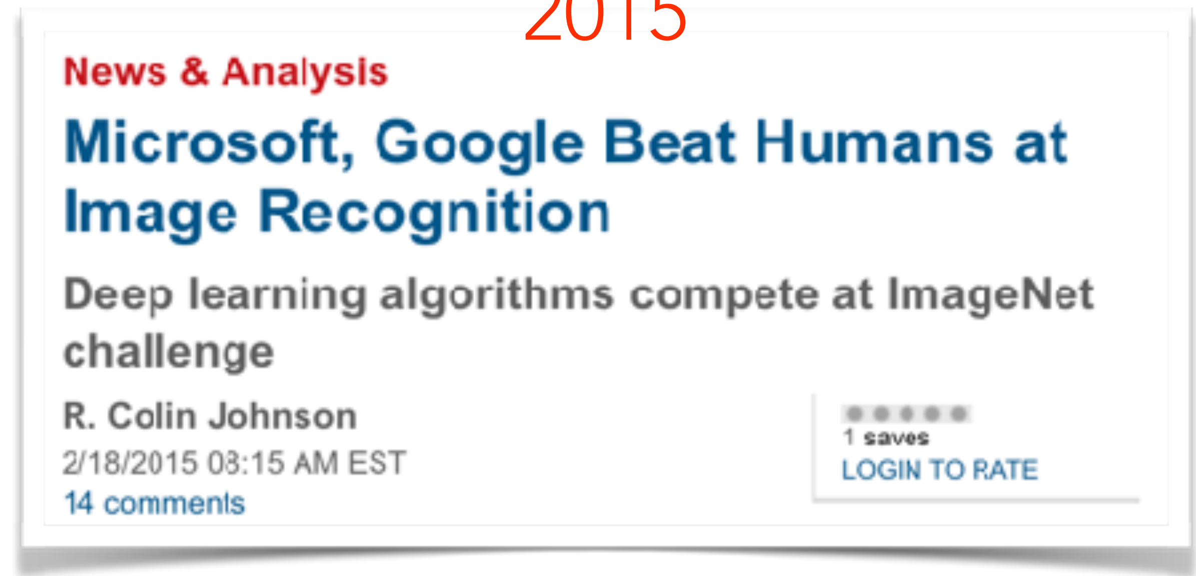
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called “dropout” that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

2015



News & Analysis

Microsoft, Google Beat Humans at Image Recognition

Deep learning algorithms compete at ImageNet challenge

R. Colin Johnson
2/18/2015 03:15 AM EST
14 comments

1 saves
LOGIN TO RATE

THE PLAYERS

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

p(x, z | θ, ν)

x
observed data
simulated data

z
latent variables
Monte Carlo truth

ν
nuisance parameters

INFERENCE

inverse problem
measurement
parameter estimation

THE FORWARD MODEL

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

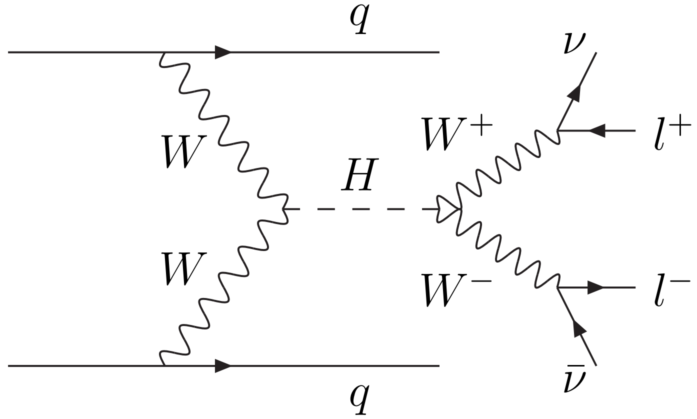
1) We begin with Quantum Field Theory

The Forward Model

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

1) We begin with Quantum Field Theory

2) Theory gives detailed prediction for high-energy collisions



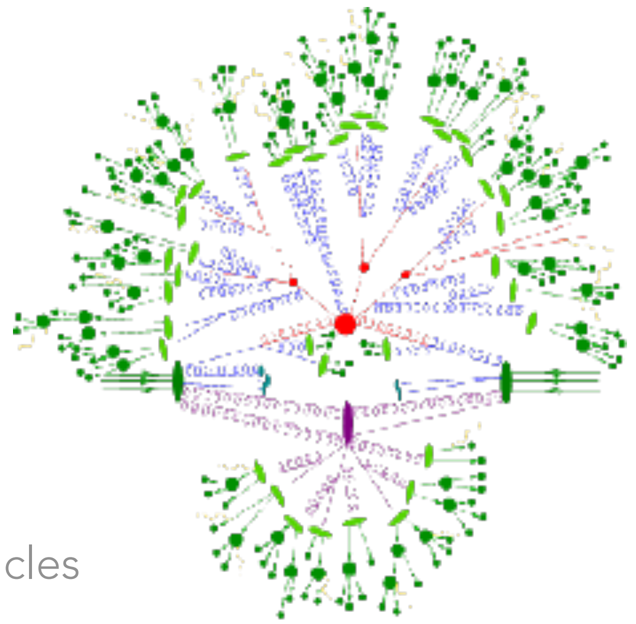
hierarchical: 2 → O(10) → O(100) particles

The Forward Model

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

1) We begin with Quantum Field Theory

2) Theory gives detailed prediction for high-energy collisions

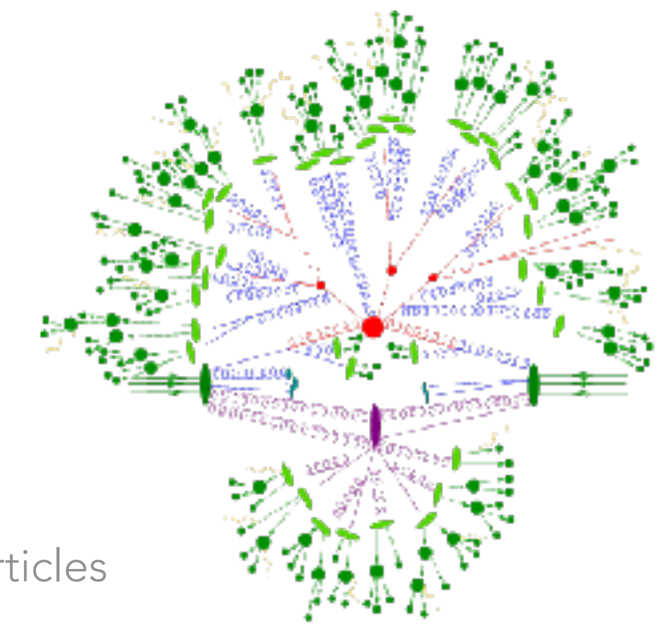


hierarchical: 2 → O(10) → O(100) particles

THE FORWARD MODEL

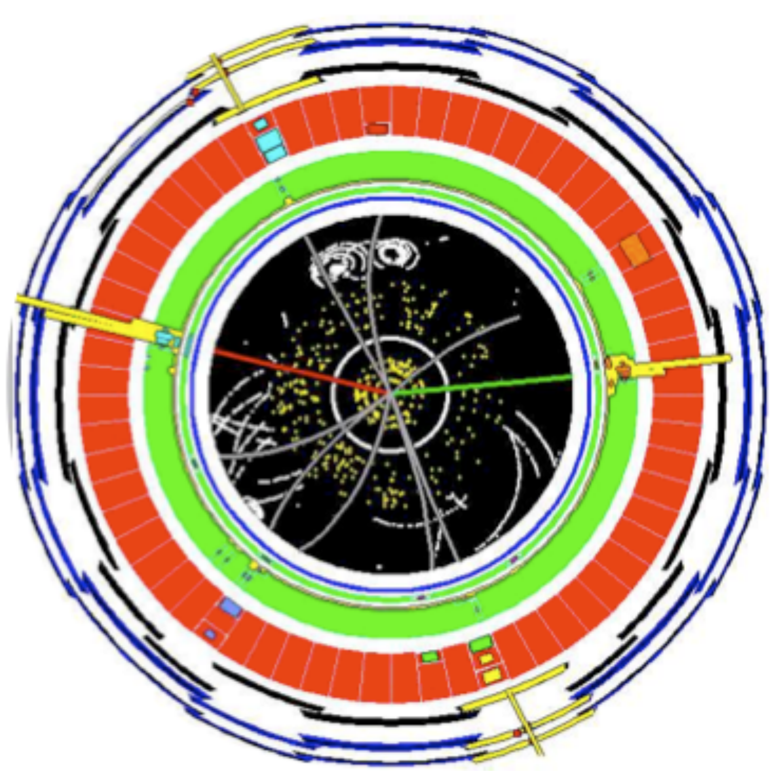
$$
\begin{aligned}
\mathcal{L}_{SM} = & \underbrace{-\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a}_{\text{kinetic energies and self-interactions of the gauge bosons}} \\
& + \underbrace{\bar{L}\gamma^\mu(i\partial_\mu -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

1) We begin with Quantum Field Theory



2) Theory gives detailed prediction for high-energy collisions

hierarchical: 2 → O(10) → O(100) particles



3) The interaction of outgoing particles with the detector is simulated.

>100 million sensors

THE FORWARD MODEL

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

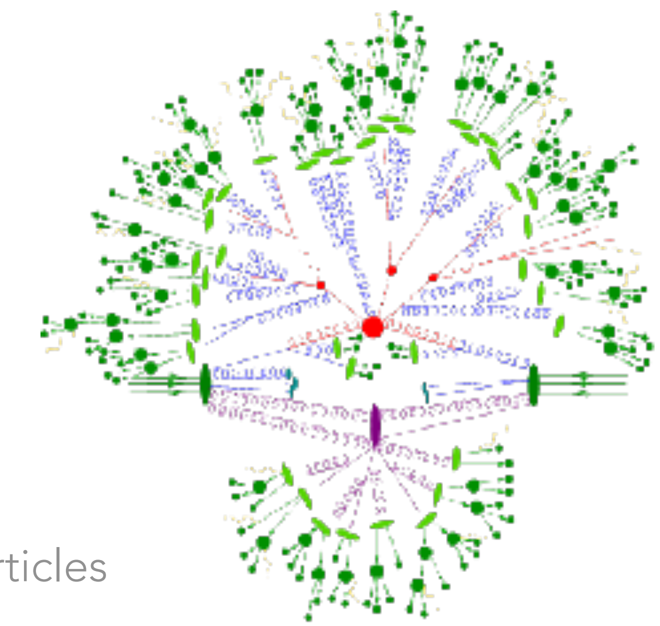
$$+ \underbrace{\bar{L}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g\tau\cdot\mathbf{W}_{\mu} - \frac{1}{2}g'YB_{\mu})L + \bar{R}\gamma^{\mu}(i\partial_{\mu} - \frac{1}{2}g'YB_{\mu})R}_{\text{kinetic energies and electroweak interactions of fermions}}$$

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

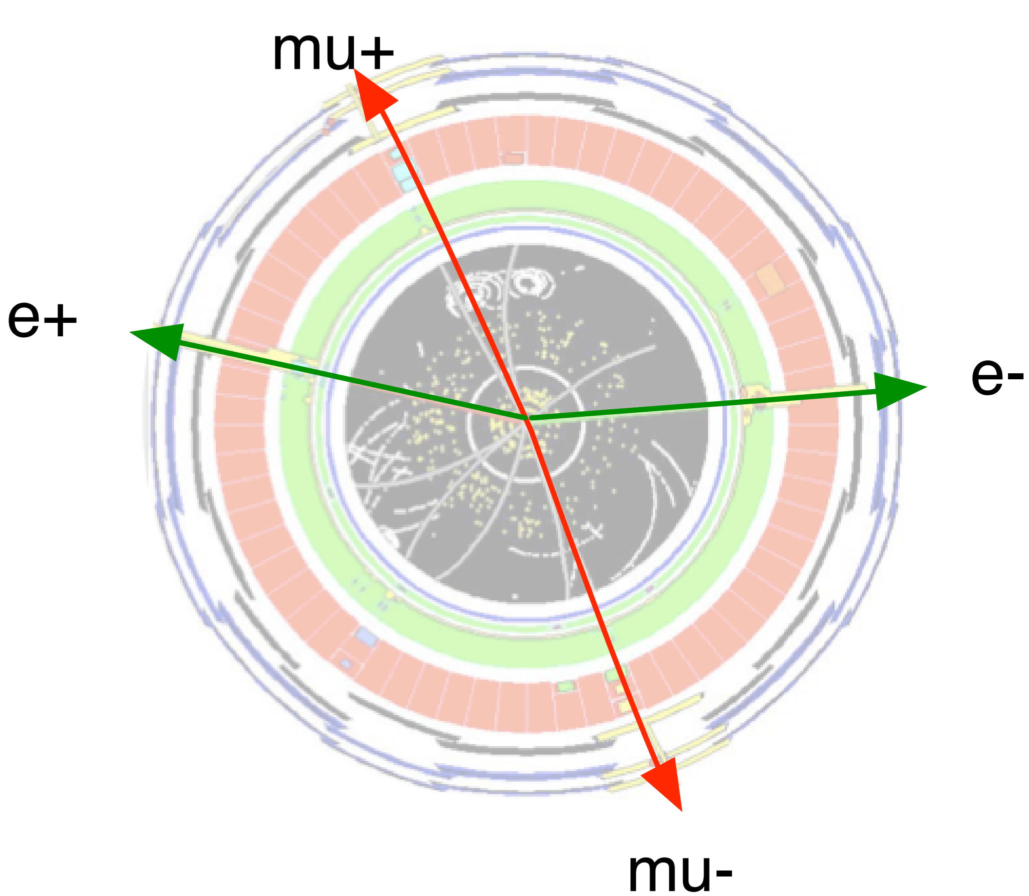
$$+ \underbrace{g''(\bar{q}\gamma^{\mu}T_{a}q)G^{a}_{\mu}}_{\text{interactions between quarks and gluons}} + \underbrace{(G_{1}\bar{L}\phi R + G_{2}\bar{L}\phi_{c}R + h.c.)}_{\text{fermion masses and couplings to Higgs}}$$

1) We begin with Quantum Field Theory

2) Theory gives detailed prediction for high-energy collisions



hierarchical: 2 → O(10) → O(100) particles



3) The interaction of outgoing particles with the detector is simulated.

>100 million sensors

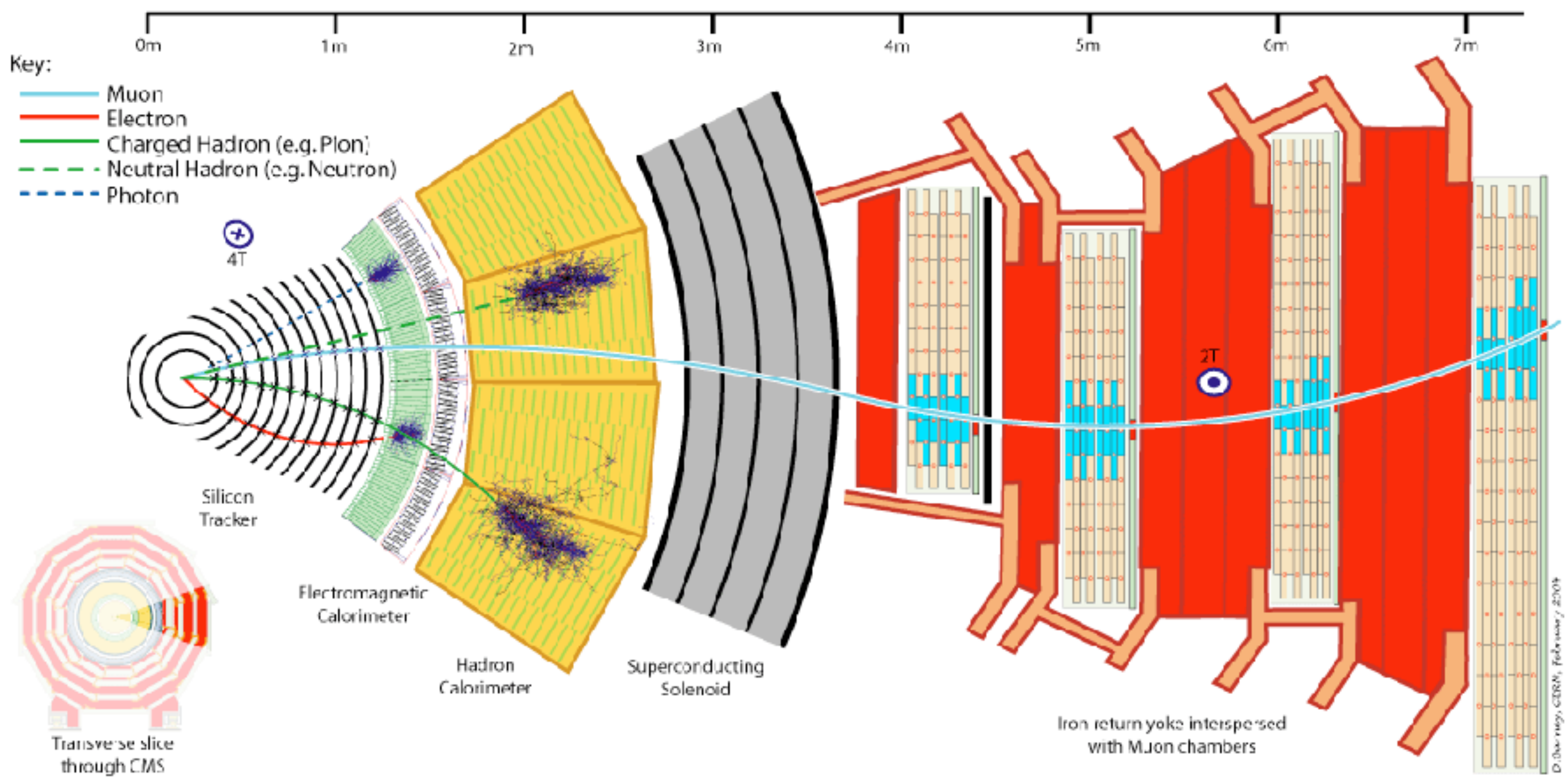
4) Finally, we run particle identification and feature extraction algorithms on the simulated data as if they were from real collisions.

~10-30 features describe interesting part

Conceptually: Prob(detector response | particles)

Implementation: Monte Carlo integration over micro-physics

Consequence: evaluation of the likelihood is intractable



Conceptually: Prob(detector response | particles)

Implementation: Monte Carlo integration over micro-physics

Consequence: evaluation of the likelihood is intractable

This motivates a new class of algorithms for what is called **likelihood-free inference**, which only require ability to generate samples from the simulation in the "forward mode"

A COMMON THEME

ABC

resources on approximate Bayesian computational methods

Search

Home

Home

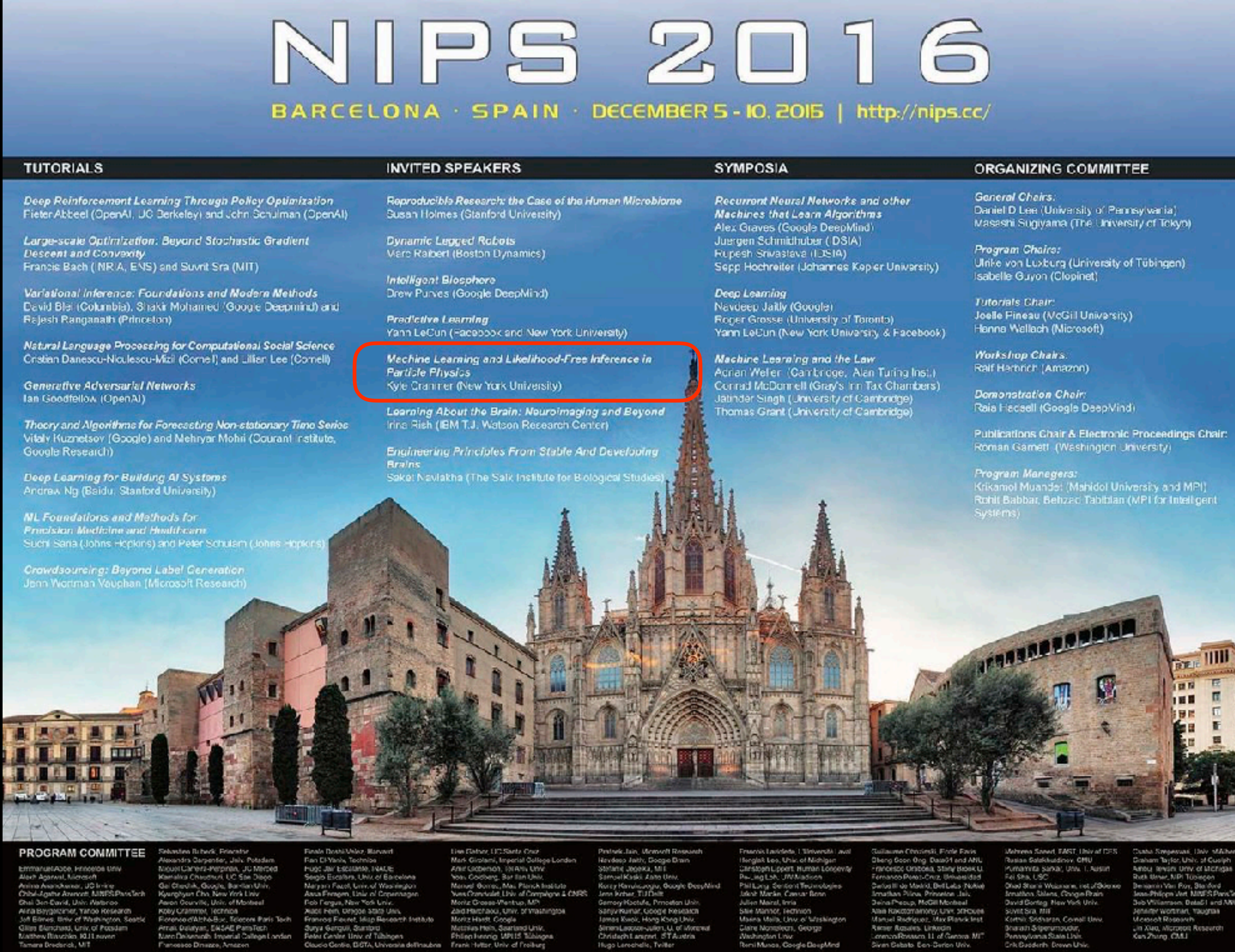
This website keeps track of developments in approximate Bayesian computation (ABC) (a.k.a. likelihood-free), a class of computational statistical methods for Bayesian inference under intractable likelihoods. The site is meant to be a resource both for biologists and statisticians who want to learn more 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.


NIPS 2016
BARCELONA · SPAIN · DECEMBER 5 - 10, 2016 | http://nips.cc/
TUTORIALS
Deep Reinforcement Learning Through Policy Optimization
Pieter Abbeel (OpenAI, UC Berkeley) and John Schulman (OpenAI)
Large-scale Optimization: Beyond Stochastic Gradient Descent and Convexity
Francis Bach (INRIA, ENS) and Suvrit Sra (MIT)
Variational Inference: Foundations and Modern Methods
David Blei (Columbia), Shakir Mohamed (Google Deepmind) and Rajesh Ranganath (Princeton)
Natural Language Processing for Computational Social Science
Cristian Danescu-Niculescu-Mizil (Cornell) and Lillian Lee (Cornell)
Generative Adversarial Networks
Ian Goodfellow (OpenAI)
Theory and Algorithms for Forecasting Non-stationary Time Series
Vitaly Kuznetsov (Google) and Mehryar Mohri (Courant Institute, Google Research)
Deep Learning for Building AI Systems
Andrew Ng (Baidu, Stanford University)
ML Foundations and Methods for Precision Medicine and Healthcare
Suchi Saria (Johns Hopkins) and Peter Schulam (Johns Hopkins)
Crowdsourcing: Beyond Label Generation
Jenn Wortman Vaughan (Microsoft Research)
INVITED SPEAKERS
Reproducible Research: the Case of the Human Microbiome
Susan Holmes (Stanford University)
Dynamic Legged Robots
Marc Raibert (Boston Dynamics)
Intelligent Biosphere
Drew Purves (Google DeepMind)
Predictive Learning
Yann LeCun (Facebook and New York University)
Machine Learning and Likelihood-Free Inference in Particle Physics
Kyle Cranmer (New York University)
Learning About the Brain: Neuroimaging and Beyond
Irina Rish (IBM T.J. Watson Research Center)
Engineering Principles From Stable And Developing Brains
Saket Navlakha (The Salk Institute for Biological Studies)
SYMPOSIA
Recurrent Neural Networks and other Machines that Learn Algorithms
Alex Graves (Google DeepMind)
Juergen Schmidhuber (IDSIA)
Rupesh Srivastava (IDSIA)
Sepp Hochreiter (Johannes Kepler University)
Deep Learning
Navdeep Jaitly (Google)
Roger Grosse (University of Toronto)
Yann LeCun (New York University & Facebook)
Machine Learning and the Law
Adrian Weller (Cambridge, Alan Turing Inst.)
Conrad McDonnell (Gray's Inn Tax Chambers)
Jatinder Singh (University of Cambridge)
Thomas Grant (University of Cambridge)
ORGANIZING COMMITTEE
General Chairs:
Daniel D Lee (University of Pennsylvania)
Masashi Sugiyama (The University of Tokyo)
Program Chairs:
Ulrike von Luxburg (University of Tübingen)
Isabelle Guyon (Clopinet)
Tutorials Chair:
Joelle Pineau (McGill University)
Hanna Wallach (Microsoft)
Workshop Chairs:
Ralf Herbrich (Amazon)
Demonstration Chair:
Raia Hadsell (Google DeepMind)
Publications Chair & Electronic Proceedings Chair:
Roman Garnett (Washington University)
Program Managers:
Krikamol Muandet (Mahidol University and MPI)
Rohit Babbar, Behzad Tabibian (MPI for Intelligent Systems)
PROGRAM COMMITTEE
Emmanuel Abbe, Princeton Univ
Alekh Agarwal, Microsoft
Anima Anandkumar, UC Irvine
Chloé-Agathe Azencott, MINES ParisTech
Shai Ben-David, Univ. Waterloo
Jeff Bilmes, Univ. of Washington, Seattle
Gilles Blanchard, Univ. of Potsdam
Matthew Blaschko, KU Leuven
Tamara Broderick, MIT
Alexandra Carpentier, Univ. Potsdam
Miguel Carreira-Perpiñán, UC Merced
Kamalika Chaudhuri, UC San Diego
Gal Chechik, Google, Bar-Ilan Univ.
Kyunghyun Cho, New York Univ
Aaron Courville, Univ. of Montreal
Koby Crammer, Technion
Florence d'Alché-Buc, Telecom Paris Tech
Arnak Dalalyan, ENSAE ParisTech
Marc Deisenroth, Imperial College London
Francesco Dinuzzo, Amazon
Ran El-Yaniv, Technion
Sergio Escalera, Univ. of Barcelona
Maryam Fazel, Univ. of Washington
Aasa Feragen, Univ. of Copenhagen
Rob Fergus, New York Univ.
Surya Ganguli, Stanford
Peter Gehler, Univ. of Tübingen
Mark Girolami, Imperial College London
Yoav Goldberg, Bar Ilan Univ.
Moritz Grosse-Wentrup, MPI
Zaid Harchaoui, Univ. of Washington
Moritz Hardt, Google
Matthias Hein, Saarland Univ.
Philipp Hennig, MPI IS Tübingen
Frank Hutter, Univ. of Freiburg
Prateek Jain, Microsoft Research
Navdeep Jaitly, Google Brain
Stefanie Jegelka, MIT
Samuel Kaski, Aalto Univ.
Koray Kavukcuoglu, Google DeepMind
Samory Kpotufe, Princeton Univ.
Sanjiv Kumar, Google Research
James Kwok, Hong Kong Univ.
Simon Lacoste-Julien, U. of Montreal
Christoph Lampert, IST Austria
Hugo Larochelle, Twitter
Honglak Lee, Univ. of Michigan
Christoph Lippert, Human Longevity
Po-Ling Loh, UW Madison
Phil Long, Sentient Technologies
Julien Mairal, Inria
Shie Mannor, Technion
Marina Meila, Univ. of Washington
Claire Monteleoni, George Washington Univ.
Remi Munos, Google DeepMind
Cheng Soon Ong, Data61 and ANU
Francesco Orabona, Stony Brook U.
Fernando Perez-Cruz, Universidad Carlos III de Madrid, Bell Labs (Nokia)
Jonathan Pillow, Princeton Univ.
Doina Precup, McGill Montreal
Manuel Rodriguez, Max Planck Inst.
Romer Rosales, LinkedIn
Lorenzo Rosasco, U. of Genova, MIT
Sivan Sabato, Ben-Gurion Univ.
Ruslan Salakhutdinov, CMU
Purnamrita Sarkar, Univ. T. Austin
Fei Sha, USC
Jonathon Shlens, Google Brain
David Sontag, New York Univ.
Suvrit Sra, MIT
Karthik Sridharan, Cornell Univ.
Bharath Sriperumbudur, Pennsylvania State Univ.
Erik Sudderth, Brown Univ.
Graham Taylor, Univ. of Guelph
Ambuj Tewari, Univ. of Michigan
Ruth Urner, MPI Tübingen
Benjamin Van Roy, Stanford
Jean-Philippe Vert, MINES ParisTech
Bob Williamson, Data61 and ANU
Jennifer Wortman Vaughan, Microsoft Research
Lin Xiao, Microsoft Research
Kun Zhang, CMU

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.

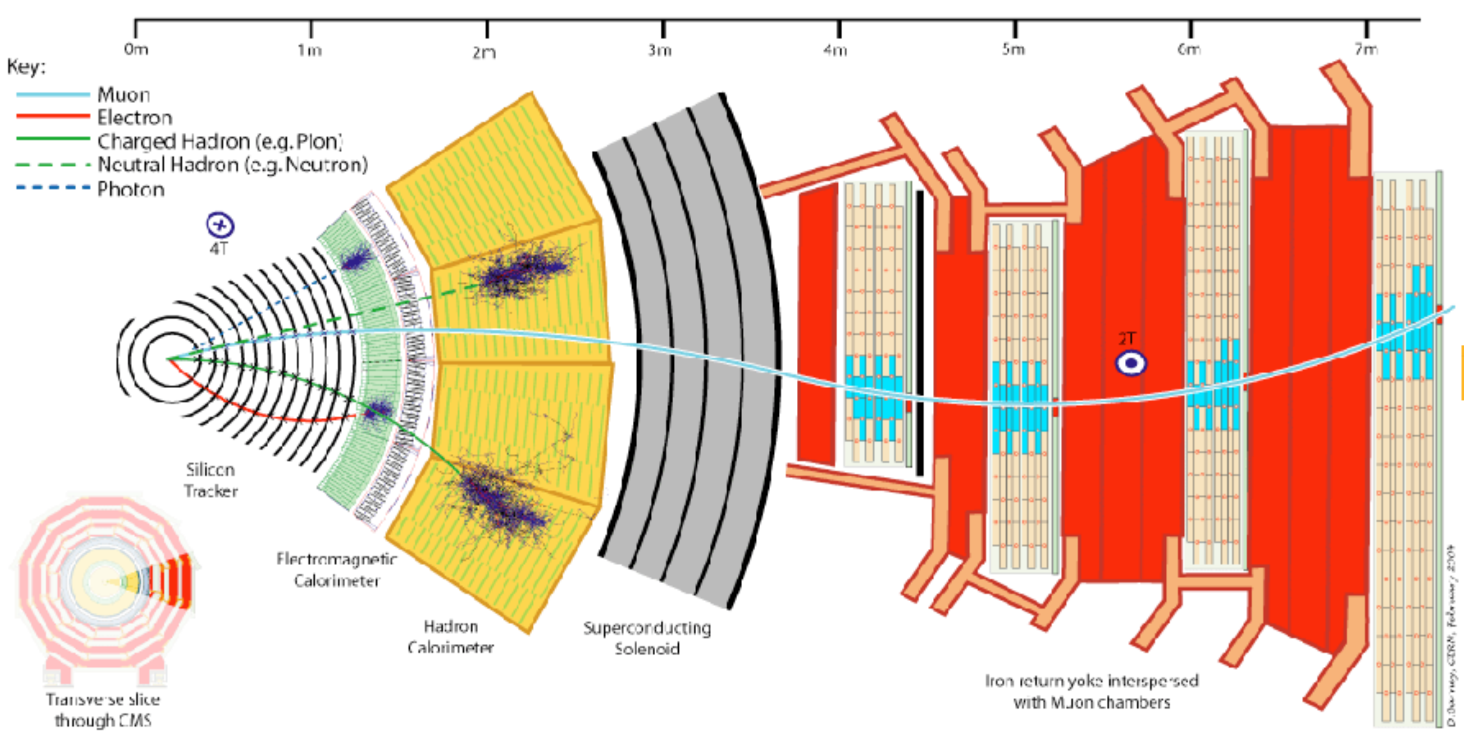
We hope to bring together these different views on implicit models, identifying their core challenges and combining their innovations.

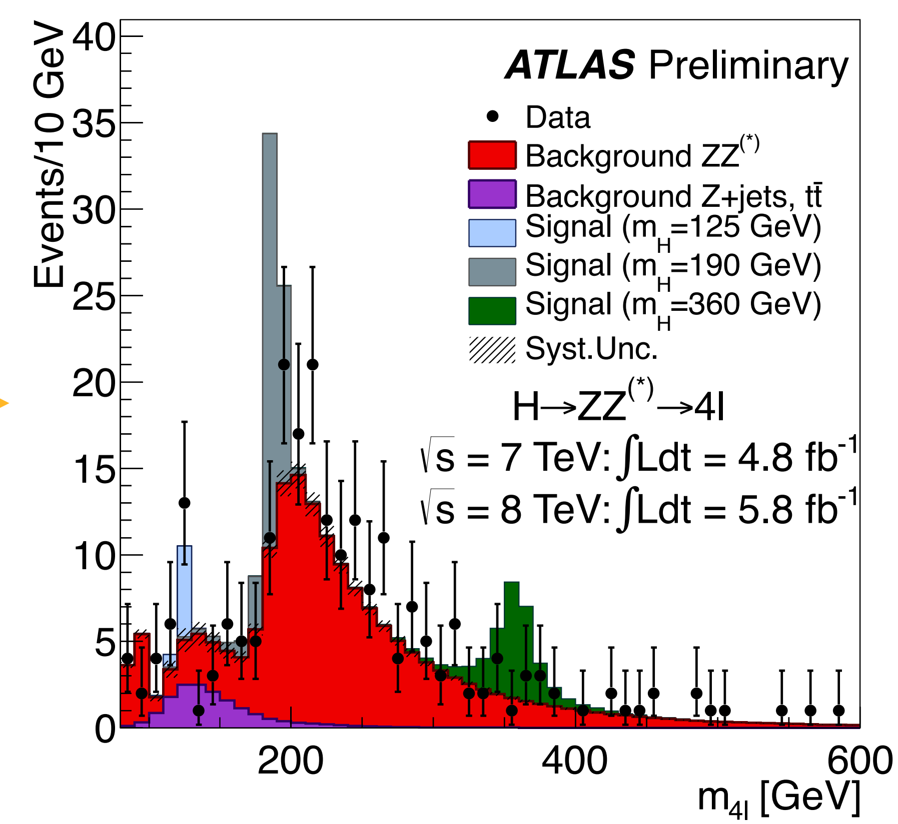
But wait… this hasn't stopped us so far.

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 variable or feature

- choosing a good variable (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)

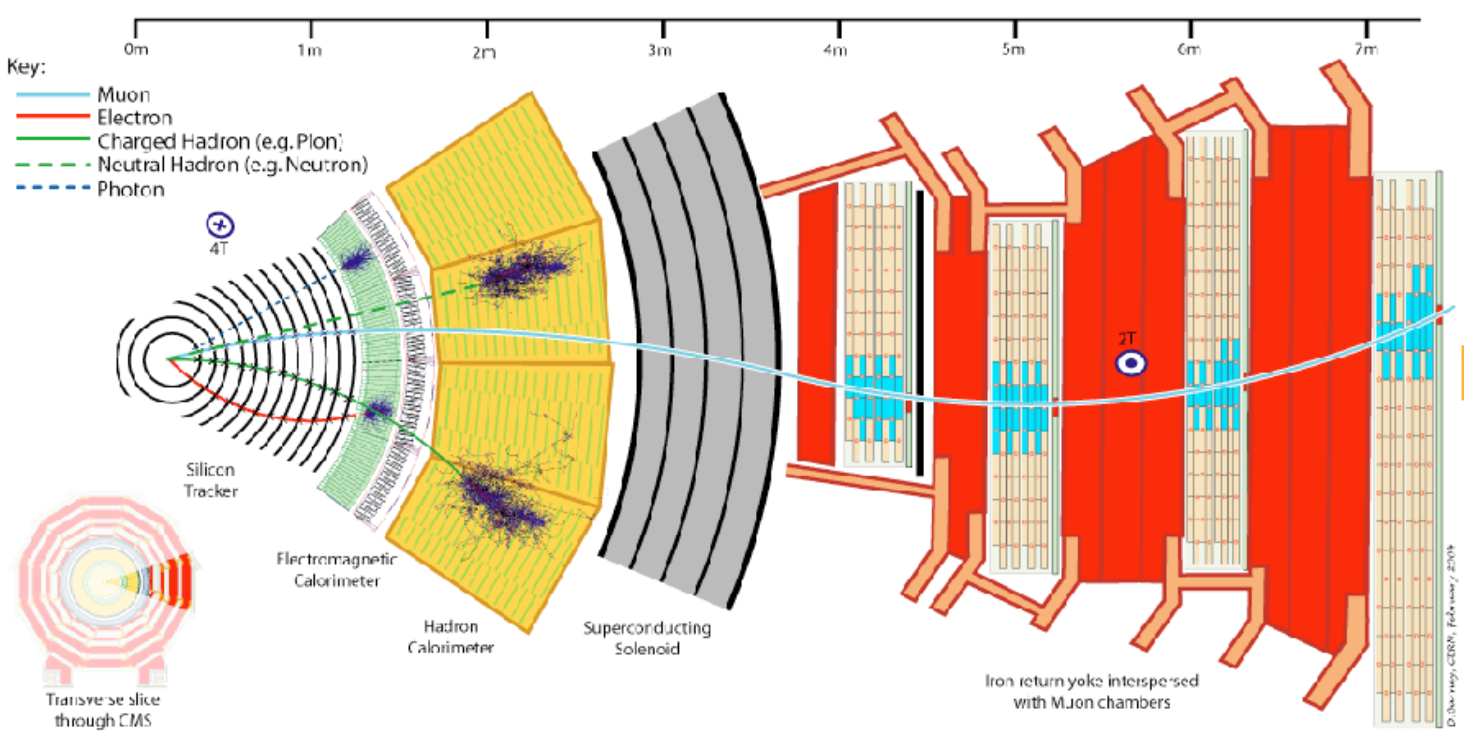


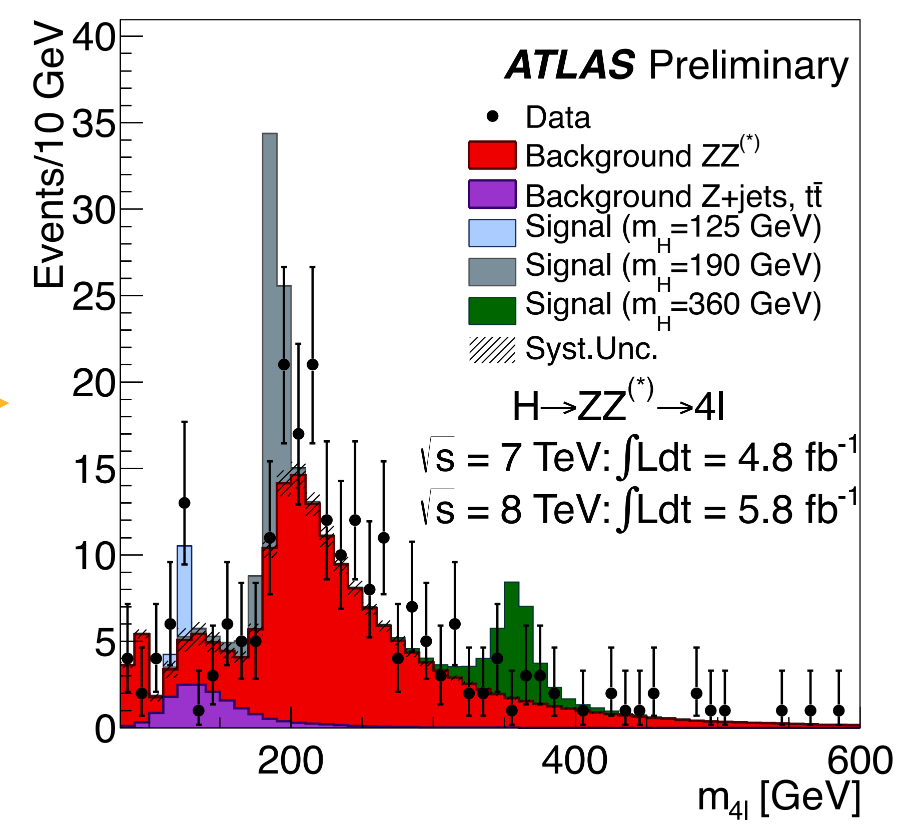


10^8 SENSORS → 1 REAL-VALUED QUANTITY

Most measurements and searches for new particles at the LHC are based on the distribution of a single variable or feature

- choosing a good variable (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)





This doesn't scale if x is high dimensional!

THE CRUX, AN INTRACTABLE INTEGRAL

Monte Carlo Sampling

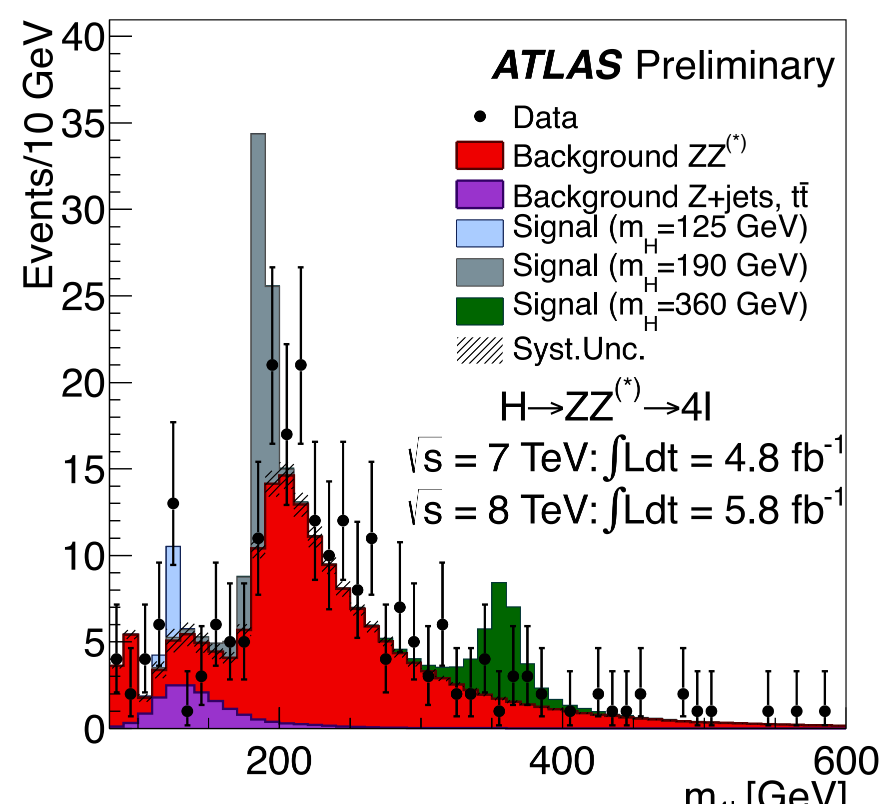
observed

latent state in simulation

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

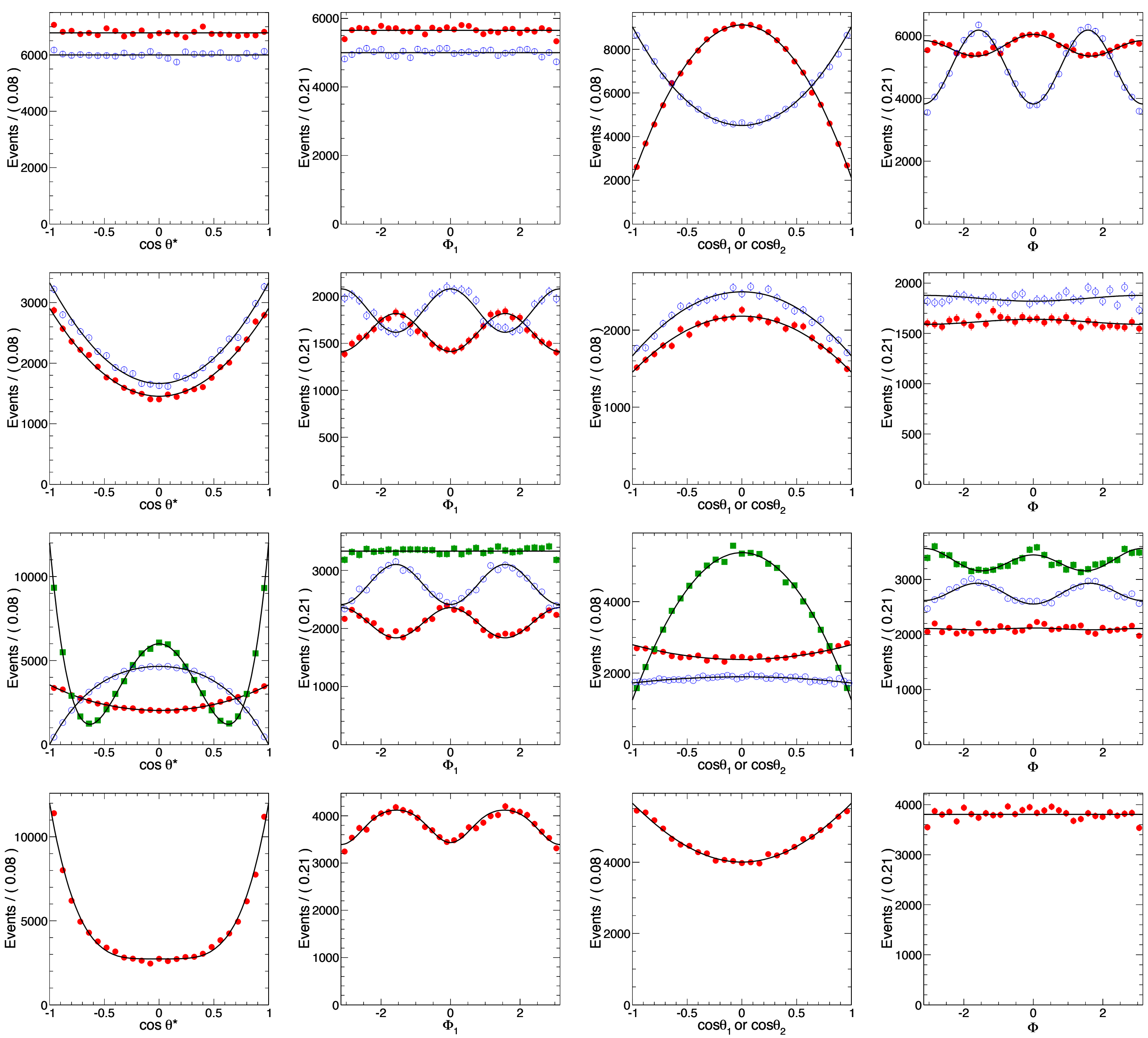
$$\hat{p}(x|\theta) \longrightarrow$$

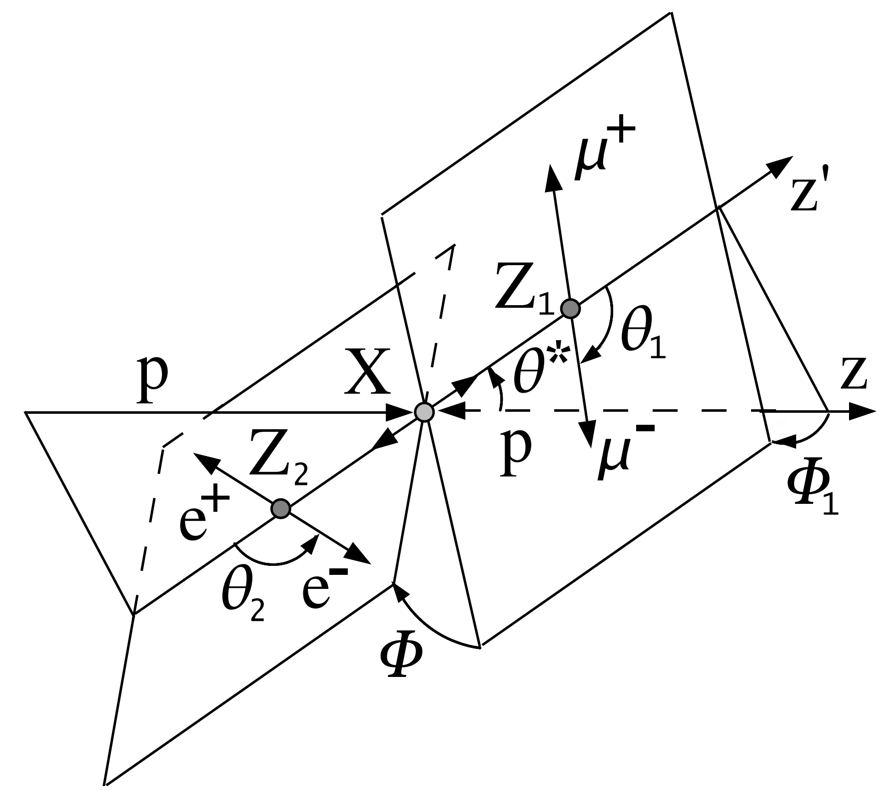
histogram approximation

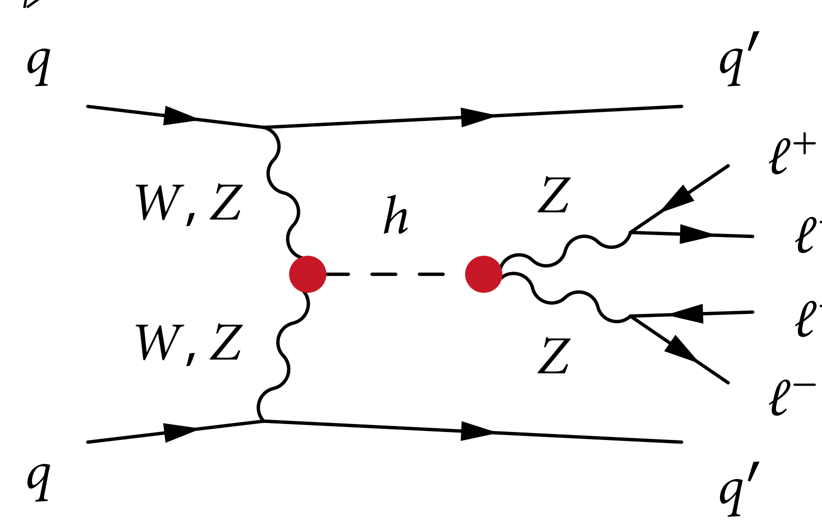


When looking for deviations from the standard model Higgs, we would like to look at all sorts of kinematic correlations

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







Information geometry provides a very powerful tool for phenomenology of EFT

- formal bounds on how well parameters can be measured
- exploit fully differential cross-section
- Global fit (eg. 13 parameters) & can profile/marginalize parameters you aren't interested in (eg. CP violating vs. CP conserving)

For an effective Higgs-gauge Lagrangian truncated at mass dimension six,

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \frac{f_i}{\Lambda^2}\mathcal{O}_i \tag{9}$$

our CP-even reference scenario consists of the renormalizable Standard Model Lagrangian combined with the five CP-even dimension-six operators in the HISZ basis [6, 7, 35],

$$\begin{aligned}
\mathcal{O}_B &= i\frac{g}{2}\,(D^\mu\phi^\dagger)(D^\nu\phi)\,B_{\mu\nu} & \mathcal{O}_W &= i\frac{g}{2}\,(D^\mu\phi)^\dagger\sigma^k(D^\nu\phi)\,W^k_{\mu\nu}\\
\mathcal{O}_{BB} &= -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,B_{\mu\nu}\,B^{\mu\nu} & \mathcal{O}_{WW} &= -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\,W^{\mu\nu\,k}\\
\mathcal{O}_{\phi,2} &= \frac{1}{2}\,\partial^\mu(\phi^\dagger\phi)\,\partial_\mu(\phi^\dagger\phi)\,. & &
\end{aligned} \tag{10}$$

At the same mass dimension, CP-odd couplings are described by operators

$$\begin{aligned}
\mathcal{O}_{B\tilde{B}} &= -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,\widetilde{B}_{\mu\nu}\,B^{\mu\nu} \equiv -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,\epsilon_{\mu\nu\rho\sigma}B^{\rho\sigma}\,B^{\mu\nu}\\
\mathcal{O}_{W\widetilde{W}} &= -\frac{g^2}{4}\,(\phi^\dagger\phi)\,\widetilde{W}^k_{\mu\nu}\,W^{\mu\nu\,k} \equiv -\frac{g^2}{4}\,(\phi^\dagger\phi)\,\epsilon_{\mu\nu\rho\sigma}W^{\rho\sigma\,k}\,W^{\mu\nu\,k}\,.
\end{aligned} \tag{11}$$

With the Levi-Civita tensor, these operators break down as C-conserving and P-violating.

$$I_{ij} = \begin{array}{cc}
\begin{array}{ccccccccccccc} f_{\phi,2} & f_W & f_B & f_{WW} & f_{BB} & f_{W\widetilde{W}} & f_{B\tilde{B}} & \text{Im}\,f_W & \text{Im}\,f_B & \text{Im}\,f_{WW} & \text{Im}\,f_{BB} & \text{Im}\,f_{W\widetilde{W}} & \text{Im}\,f_{B\tilde{B}} \end{array} & \\
\begin{pmatrix}
4942 & -968 & -50 & 54 & 2 & -7 & 0 & -1 & 0 & 2 & 0 & 36 & 0\\
-968 & 715 & 35 & -191 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & -55 & -1\\
-50 & 35 & 6 & -9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0\\
54 & -191 & -9 & 321 & 3 & -1 & 0 & 0 & 0 & 1 & 0 & 72 & 1\\
2 & -3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\
-7 & 1 & 0 & -1 & 0 & 359 & 4 & 41 & 1 & -81 & -1 & -1 & 0\\
0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\
-1 & 0 & 0 & 0 & 0 & 41 & 0 & 6 & 0 & -12 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
2 & 0 & 0 & 1 & 0 & -81 & -1 & -12 & 0 & 23 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
36 & -55 & -2 & 72 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 21 & 0\\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} &
\begin{array}{c} f_{\phi,2}\\ f_W\\ f_B\\ f_{WW}\\ f_{BB}\\ f_{W\widetilde{W}}\\ f_{B\tilde{B}}\\ \text{Im}\,f_W\\ \text{Im}\,f_B\\ \text{Im}\,f_{WW}\\ \text{Im}\,f_{BB}\\ \text{Im}\,f_{W\widetilde{W}}\\ \text{Im}\,f_{B\tilde{B}} \end{array}
\end{array}, \tag{30}$$

HIGGS EFT

"Better Higgs Measurements Through Information Geometry" [arXiv:1612.05261]

- Theory language: dimension-6 operators of SM EFT, $\mathcal{L} \supset \sum_i \frac{f_i}{\Lambda^2}\mathcal{O}_i$

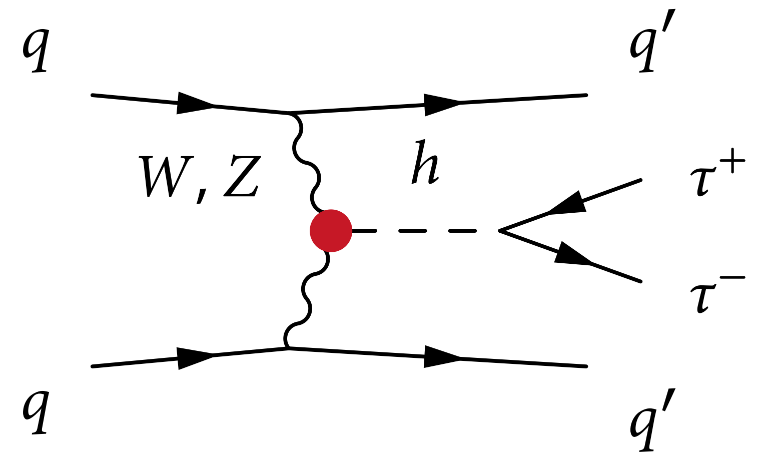
  [W. Buchmuller, D. Wyler 85; K. Hagiwara, S. Ishihara, S. R. Szalapski, D. Zeppenfeld 93; B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

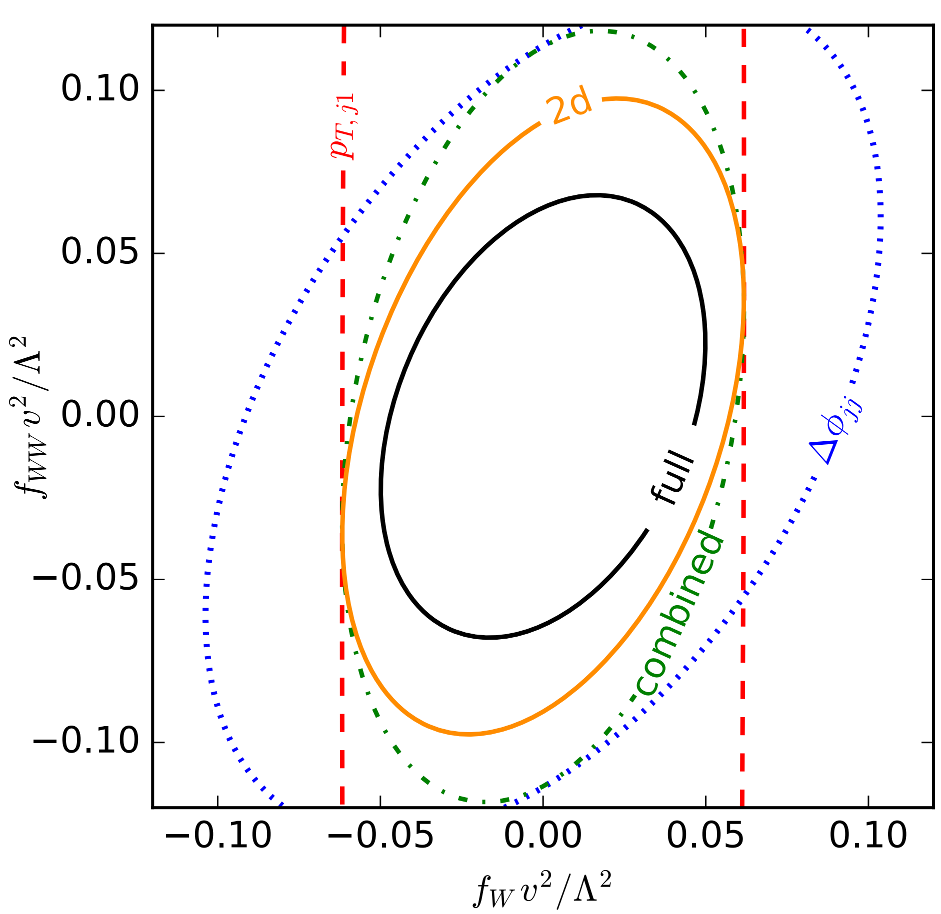
  - Total rate: $\mathcal{O}_{\phi,2} = \frac{1}{2}\,\partial^\mu(\phi^\dagger\phi)\,\partial_\mu(\phi^\dagger\phi)$
  - New kinematic structures:

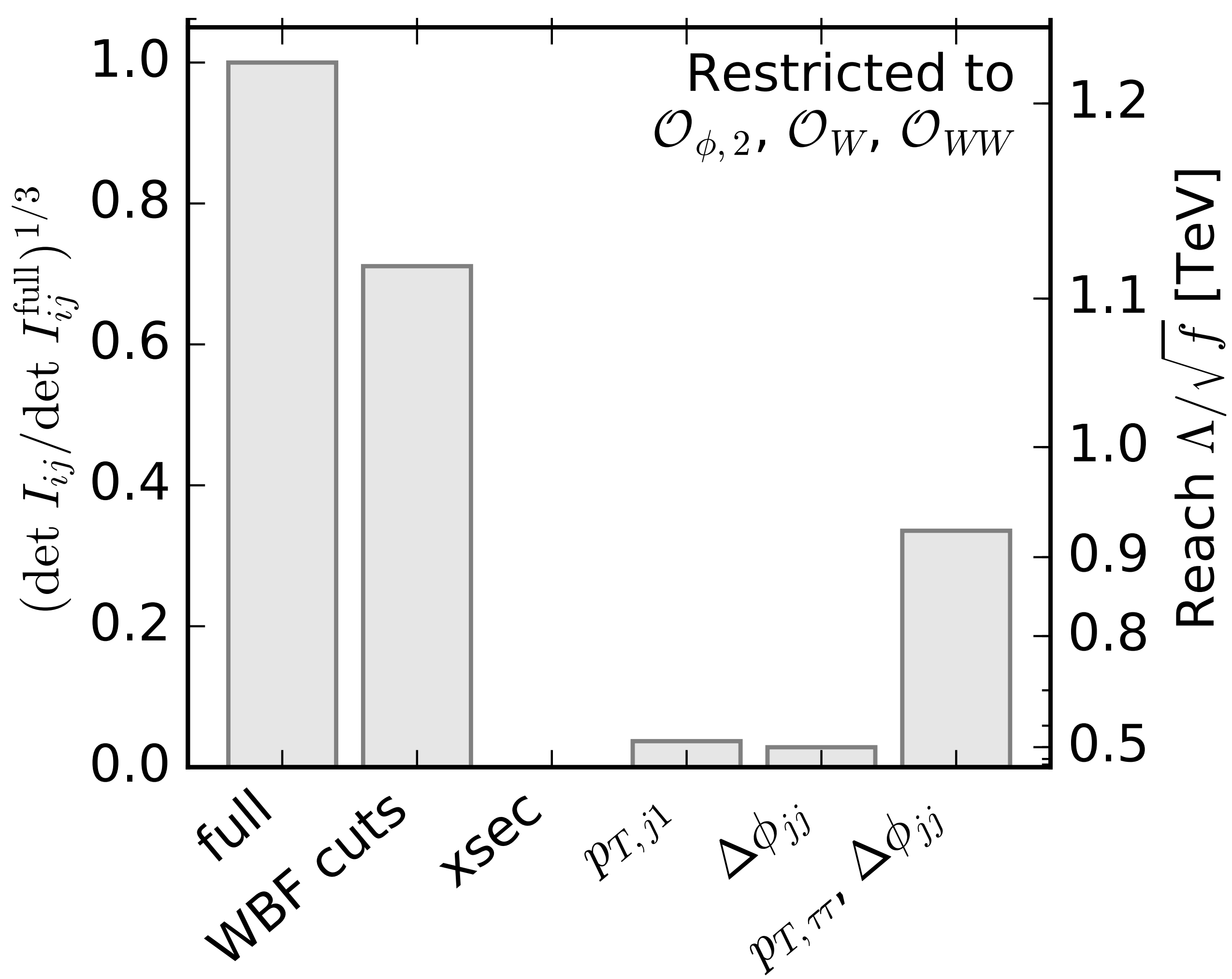
$$\mathcal{O}_B = \mathrm{i}\frac{g}{2}\,(D^\mu\phi^\dagger)(D^\nu\phi)\,B_{\mu\nu} \qquad \mathcal{O}_W = \mathrm{i}\frac{g}{2}\,(D^\mu\phi)^\dagger\sigma^k(D^\nu\phi)\,W^k_{\mu\nu}$$

$$\mathcal{O}_{BB} = -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,B_{\mu\nu}B^{\mu\nu} \qquad \mathcal{O}_{WW} = -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}W^{\mu\nu\,k}$$

  - CP violation: $\mathcal{O}_{W\widetilde{W}} = -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\widetilde{W}^{\mu\nu\,k}$
  - Others strongly constrained by EWPD or redundant



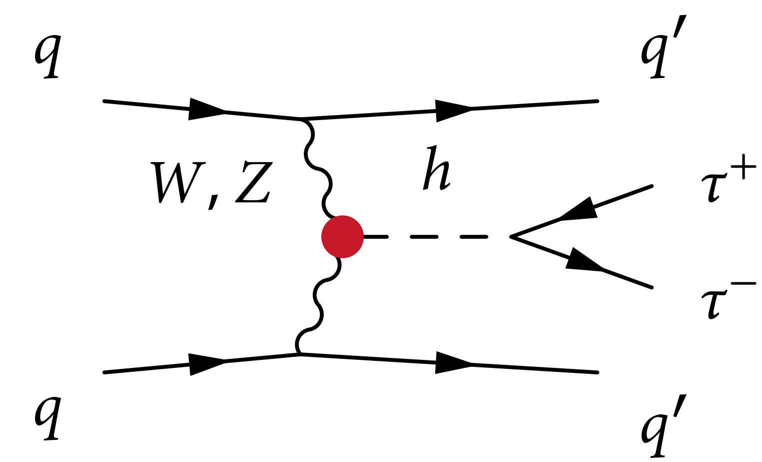




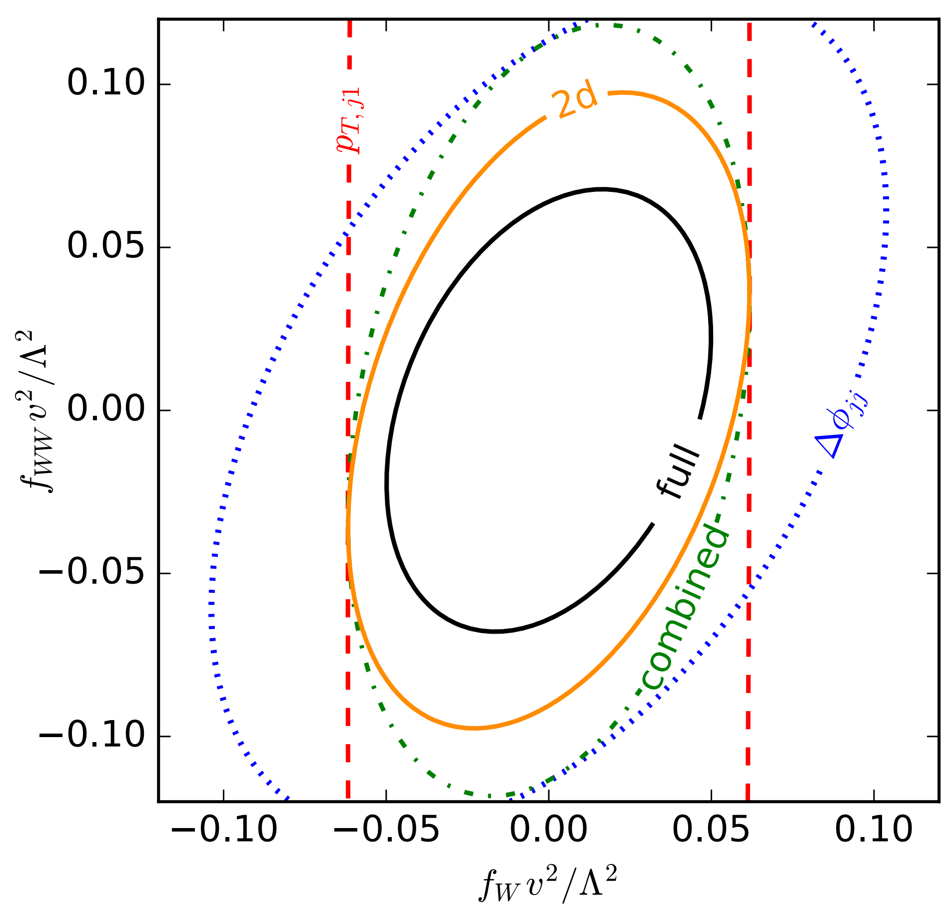
HIGGS EFT

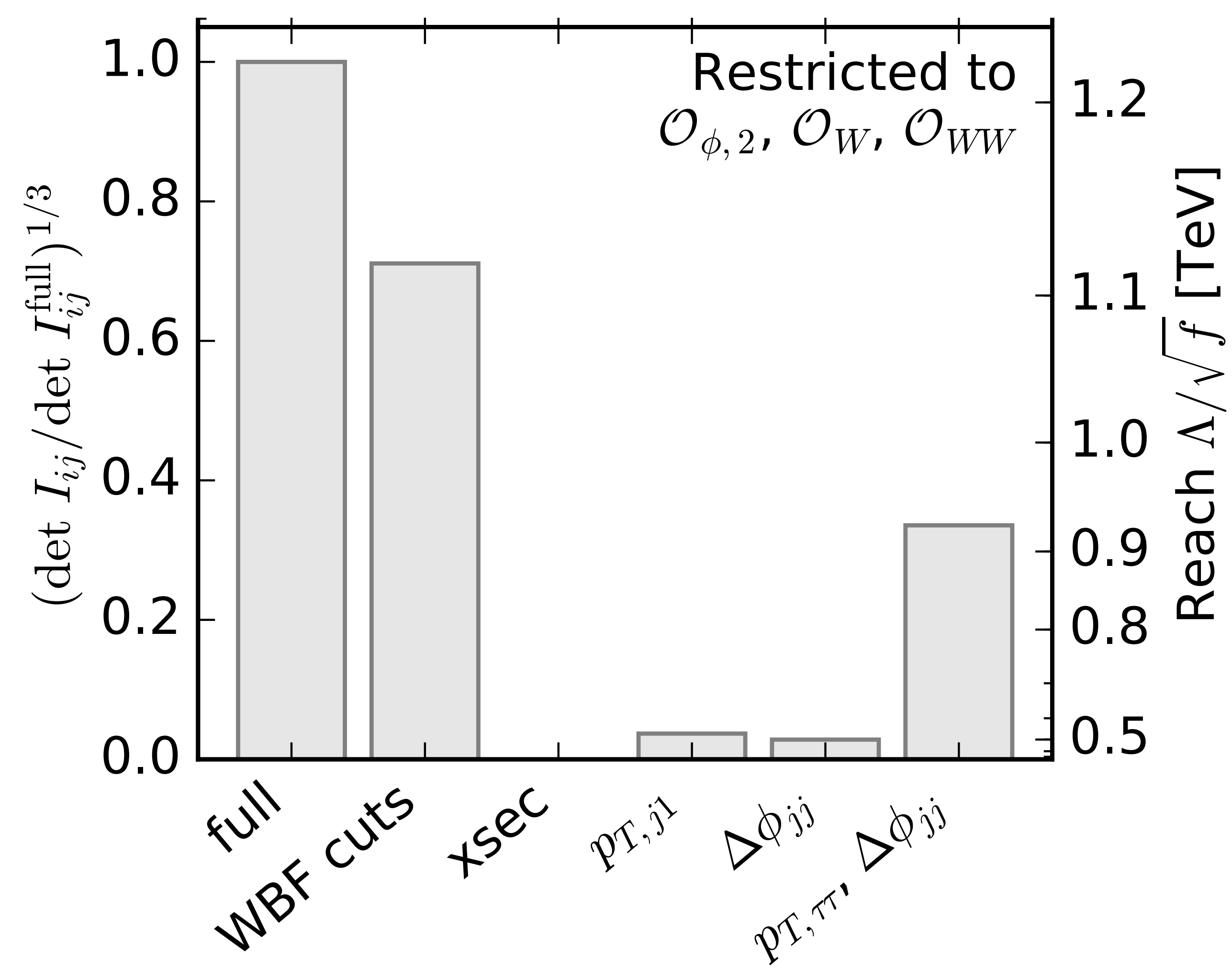
"Better Higgs Measurements Through Information Geometry"
[arXiv:1612.05261]

Compared to just using standard kinematic variables, the fully differential cross-section has the potential to dramatically improve sensitivity



Equivalent to 3x more data!



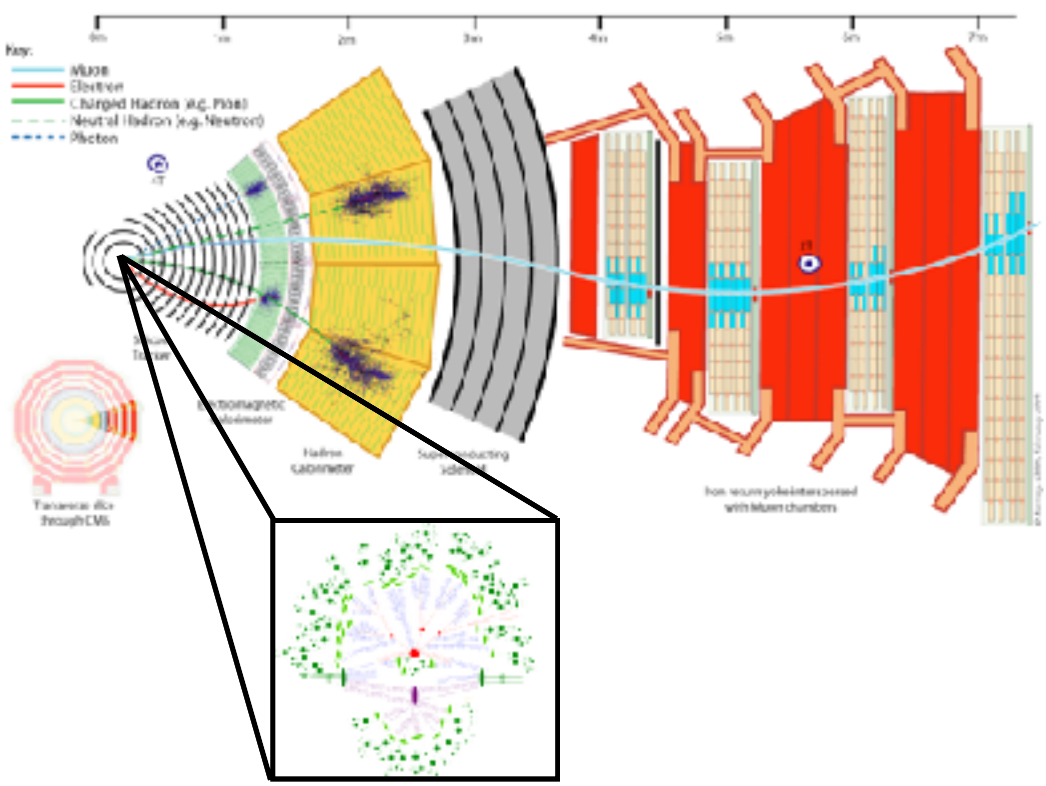


There is a lot to gain by exploiting differential information

How do we do it when there is a detector in the way?

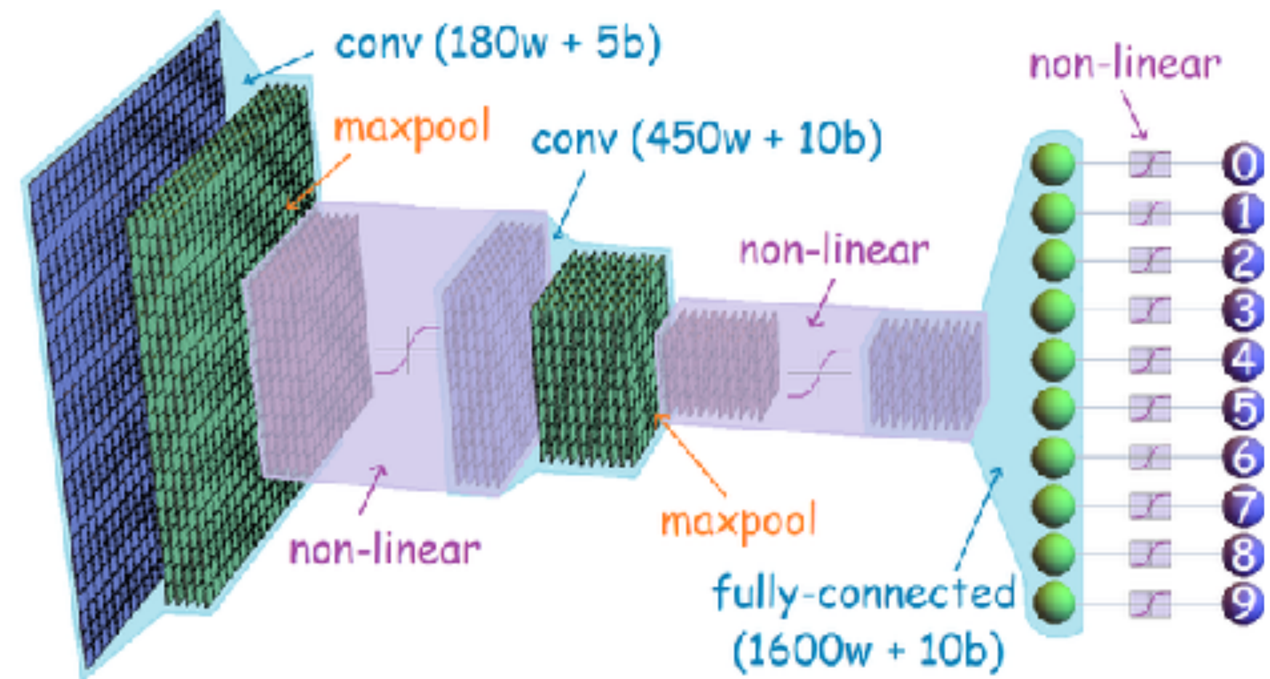
TWO APPROACHES TO SIMULATION-BASED INFERENCE

Use simulator
(much more efficiently)



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

Learn simulator
(with deep learning)

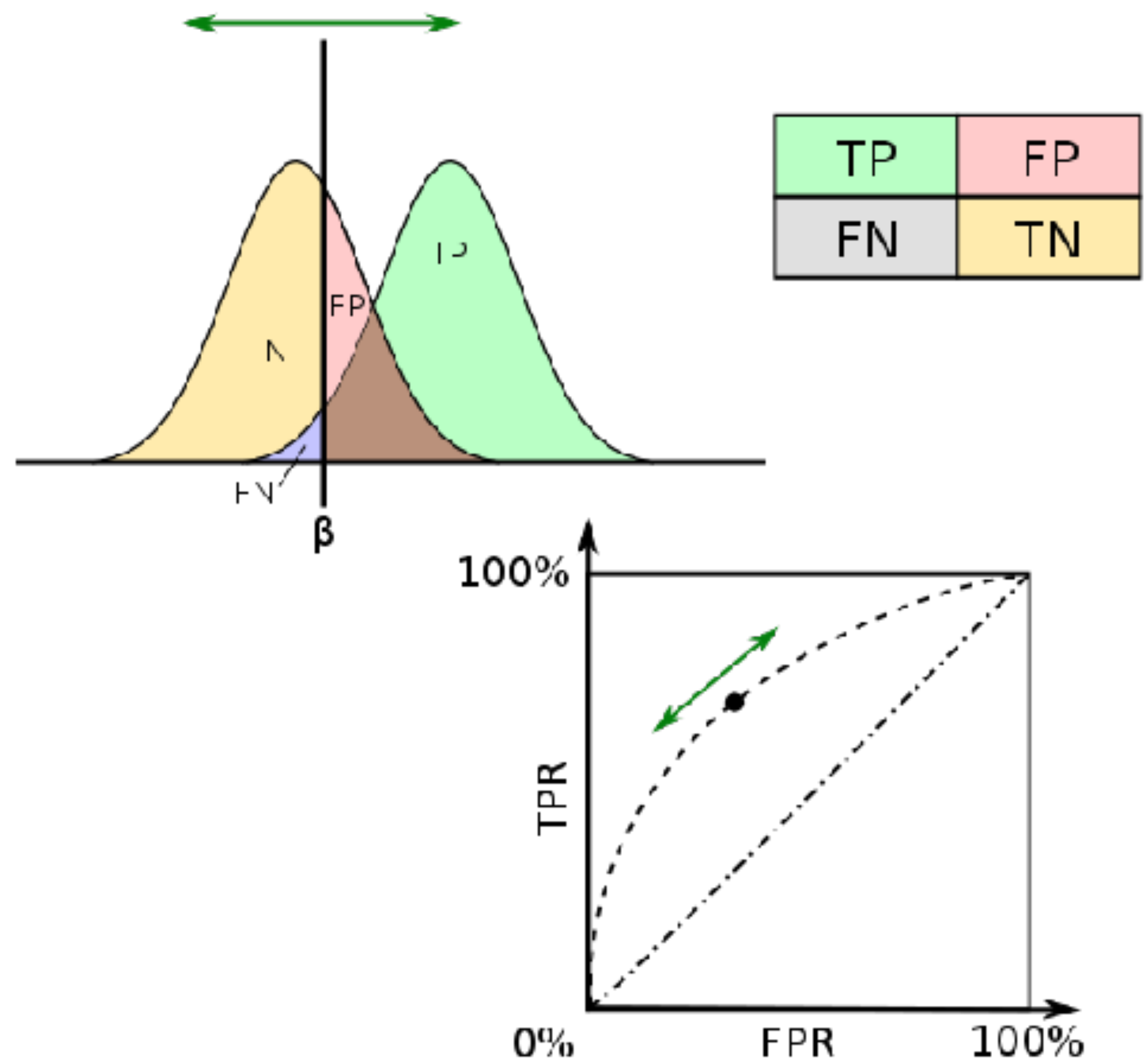


- Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAE)
- Likelihood ratio from classifiers (CARL)
- Autogregressive models, Normalizing Flows

Hypothesis Testing
&
Classification

Classical hypothesis testing typically framed in terms of true/false : positive/negative

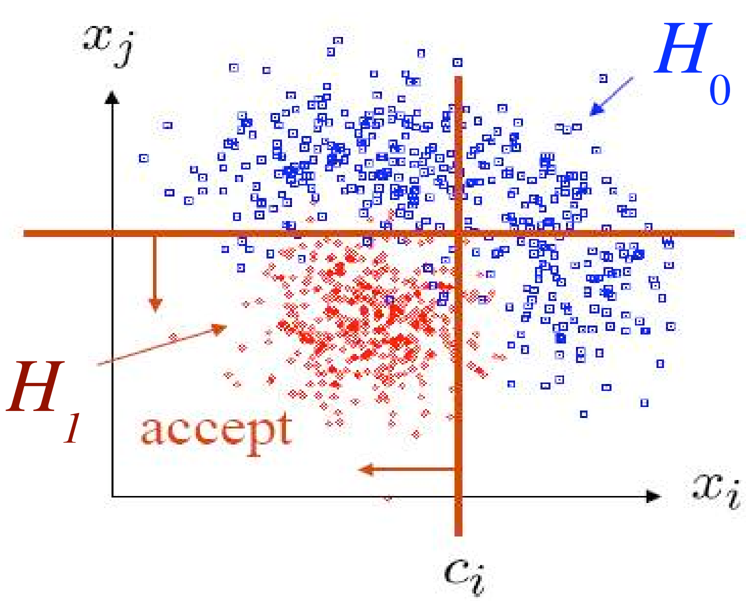
		Actual condition	
		Guilty	**Not guilty**
Decision	**Verdict of 'guilty'**	True Positive **power**	False Positive (i.e. guilt reported unfairly) **Type I error**
	Verdict of 'not guilty'	False Negative (i.e. guilt not detected) **Type II error**	True Negative

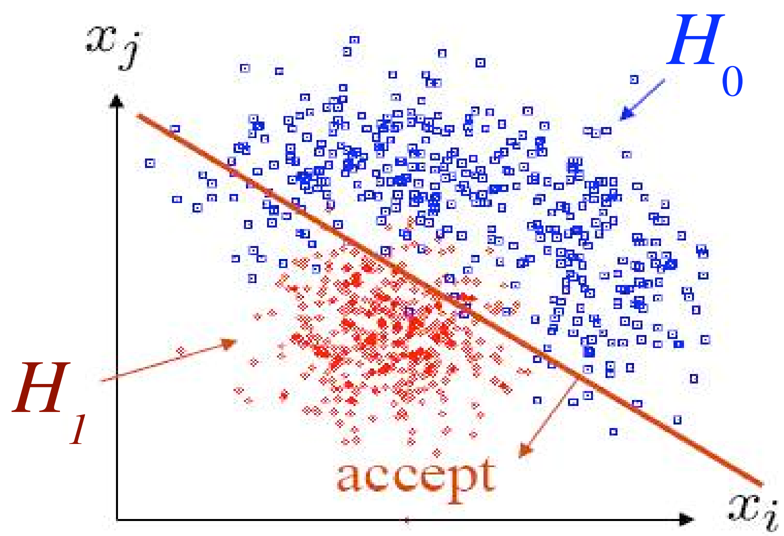


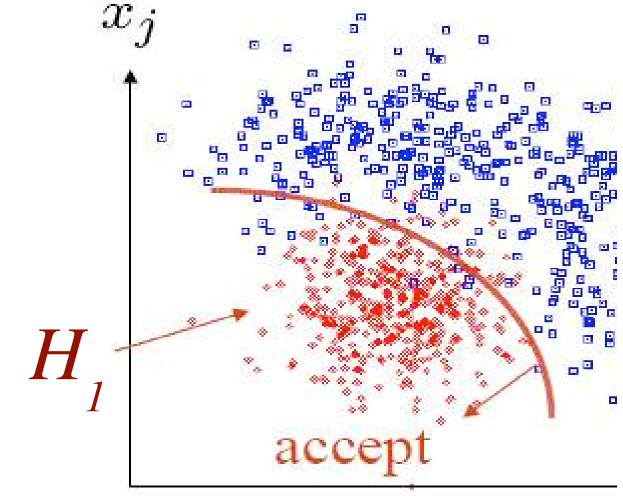
actually guilty ↔ new physics

verdict guilty ↔ claim discovery

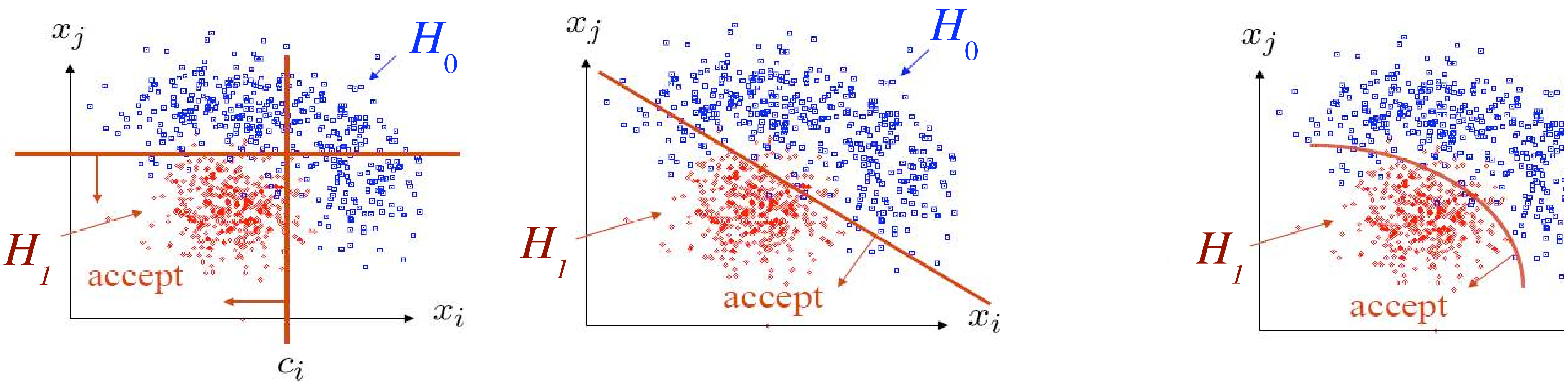
If the data are high-dimensional, it's not obvious how to draw the boundary between accept/reject the null hypothesis

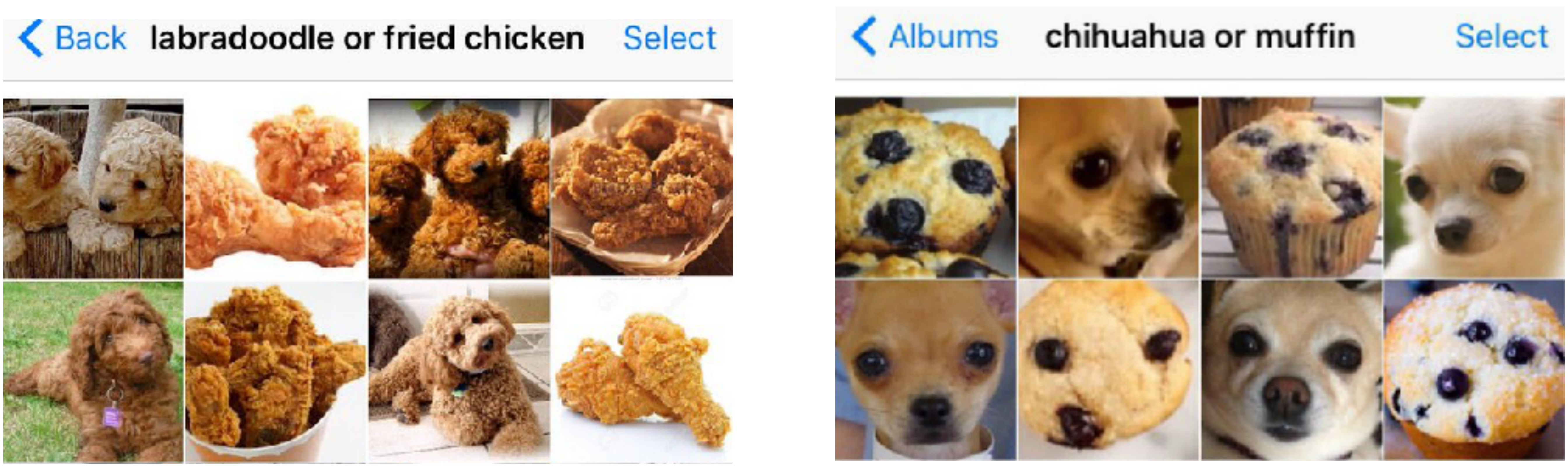






If the data are high-dimensional, it's not obvious how to draw the boundary between accept/reject the null hypothesis





THE NEYMAN-PEARSON LEMMA

In 1928-1938 Neyman & Pearson developed a theory in which one must consider competing Hypotheses:

- the Null Hypothesis H_0 (background only)
- the Alternate Hypothesis H_1 (signal-plus-background)

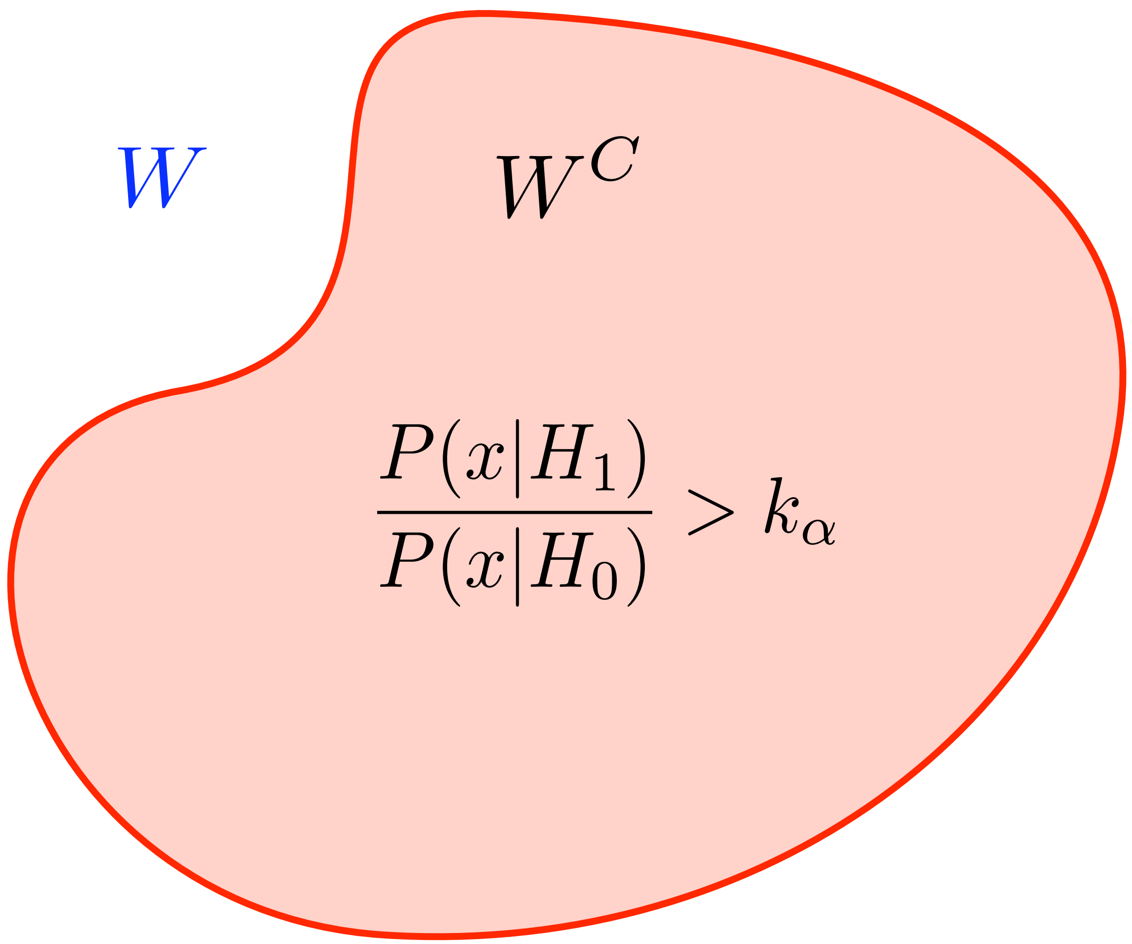
Given some probability that we wrongly reject the Null Hypothesis

$$\alpha = P(x \notin W | H_0)$$

(Convention: if data falls in W then we accept H_0)

Find the region W such that we minimize the probability of wrongly accepting the H_0 (when H_1 is true)

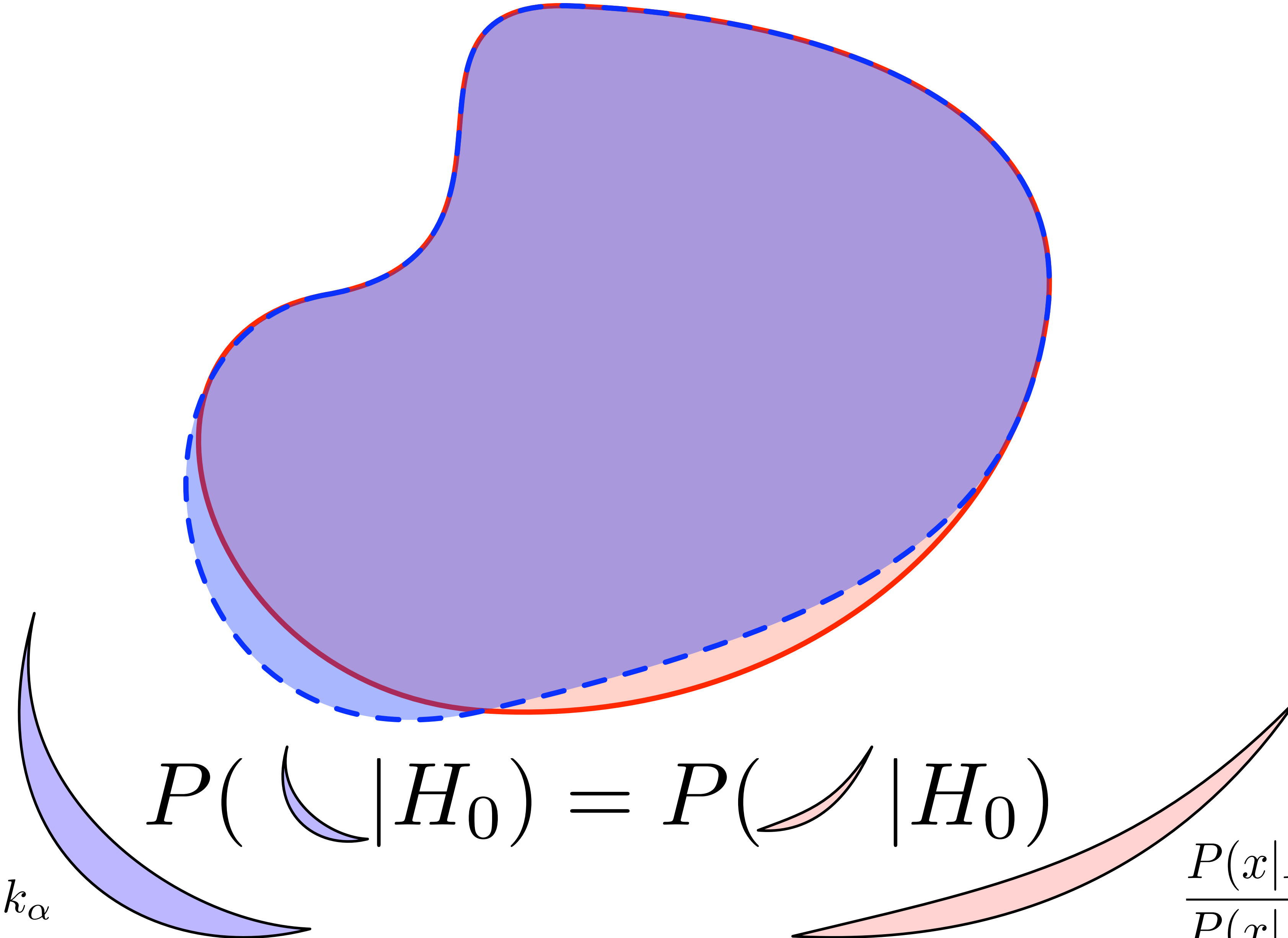
$$\beta = P(x \in W | H_1)$$



The region W that minimizes the probability of wrongly accepting H_0 is just a contour of the Likelihood Ratio

Any other region of the same size will have less power

A SHORT PROOF OF NEYMAN-PEARSON



$$P(\ \ |H_0) = P(\ \ |H_0)$$

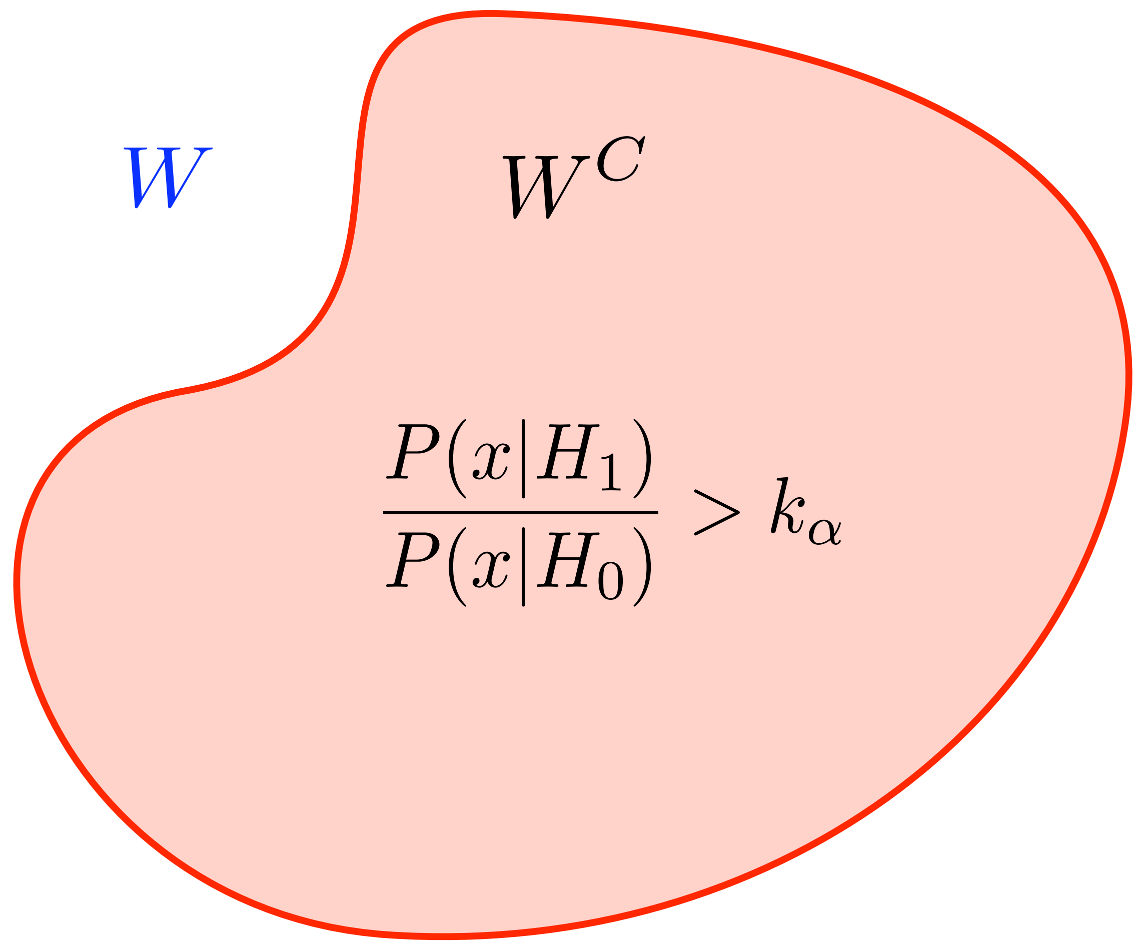
$$\frac{P(x|H_1)}{P(x|H_0)} < k_\alpha \qquad \frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

$$P(\ \ |H_1) < P(\ \ |H_0)k_\alpha \qquad P(\ \ |H_1) > P(\ \ |H_0)k_\alpha$$

$$\boxed{P(\ \ |H_1) < P(\ \ |H_1)}$$

The new region region has less power.

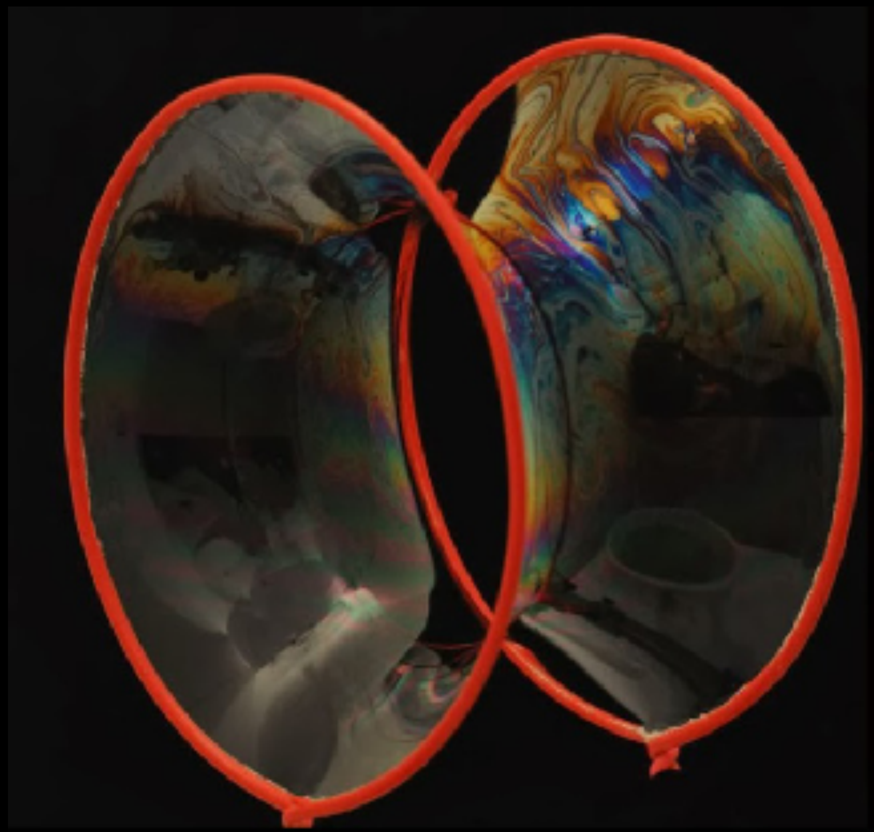
PROBLEM WITH NEYMAN-PEARSON

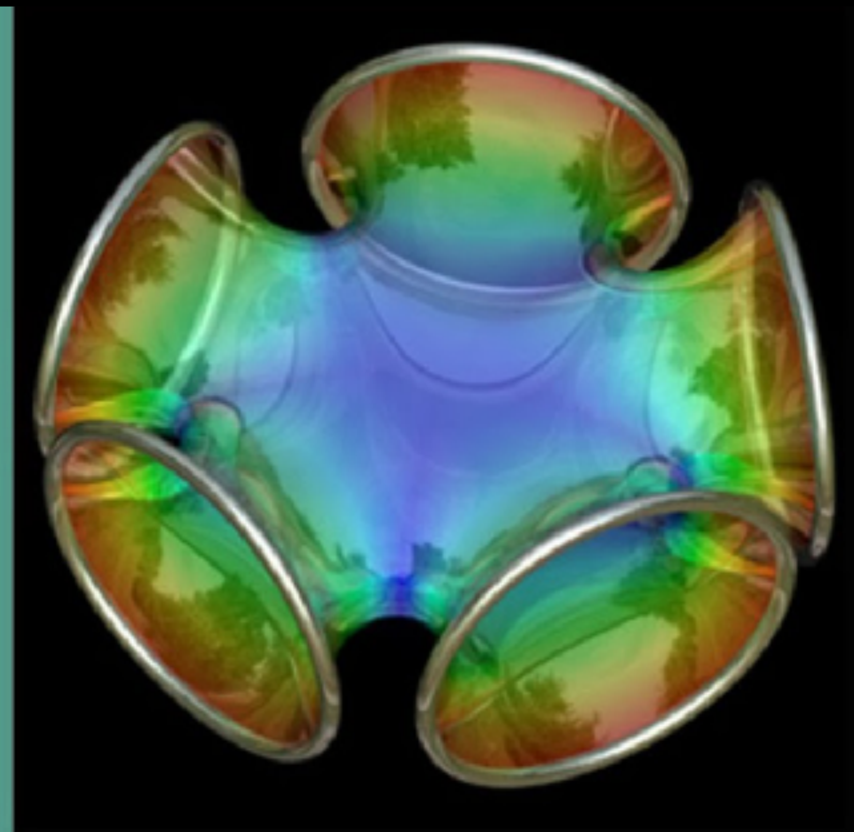


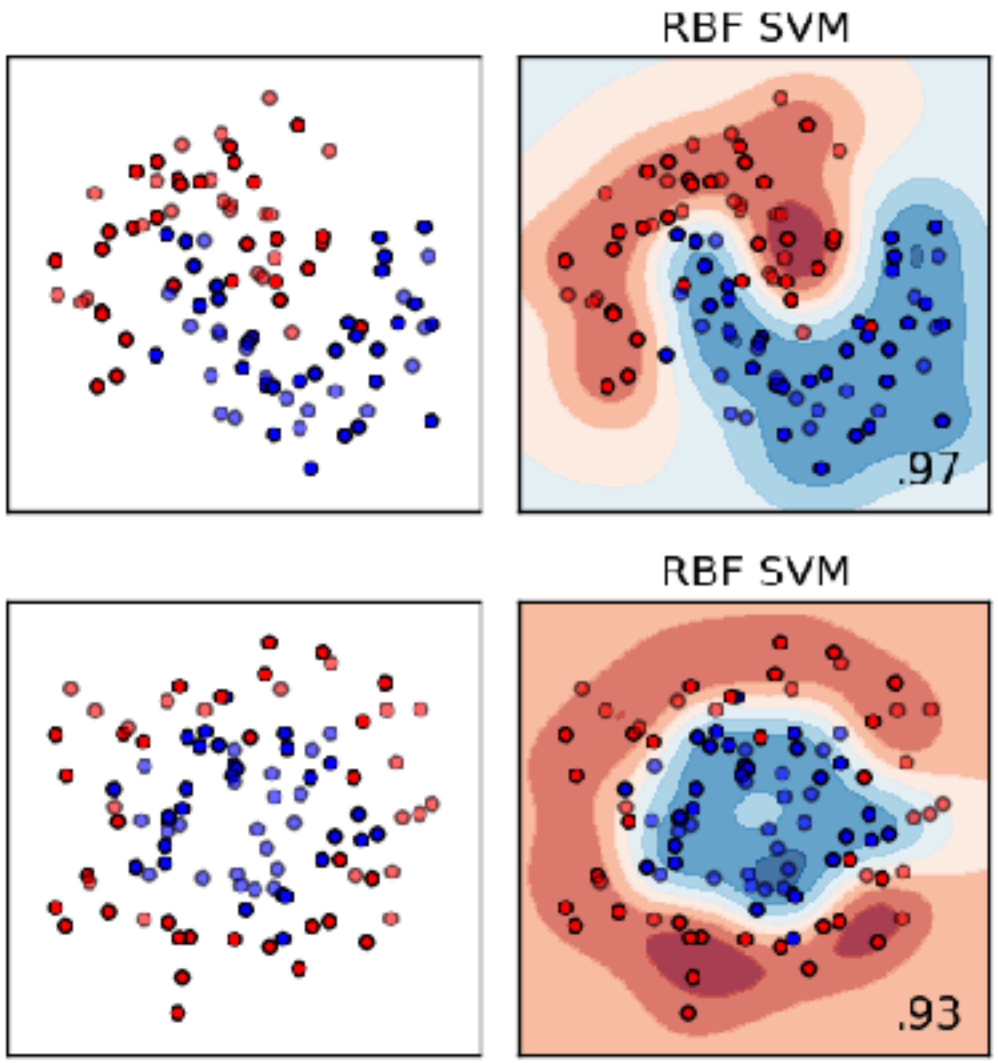
But, if I don't know $P(x|H_1)$ and $P(x|H_0)$

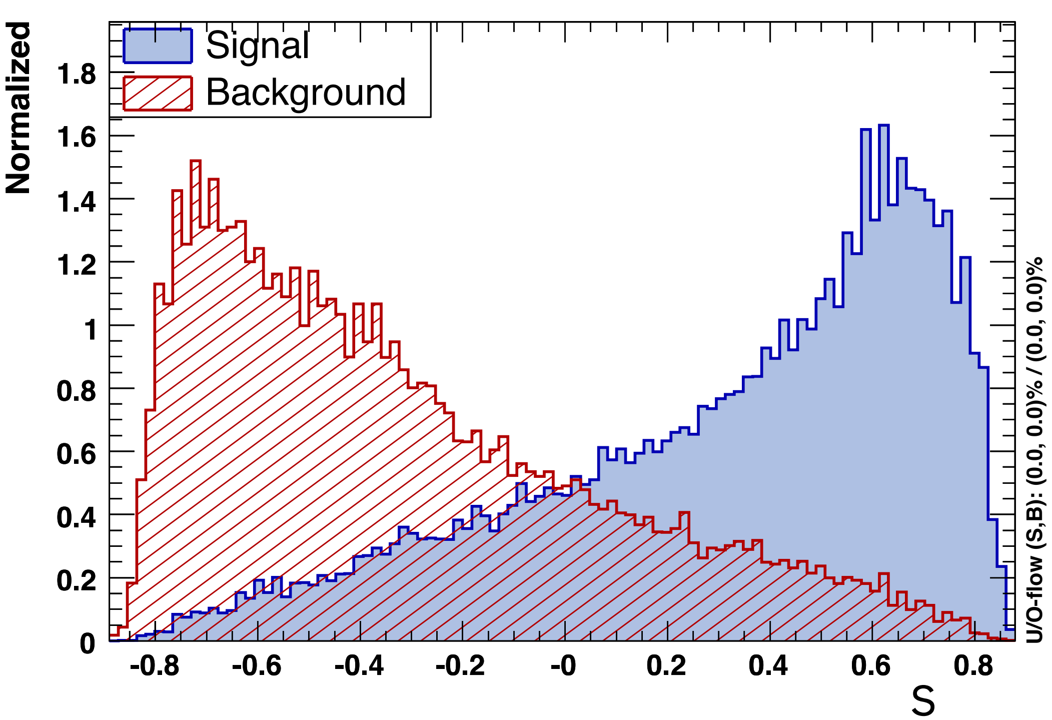
I can't evaluate this likelihood ratio!

Machine Learning = Applied Calculus of Variations







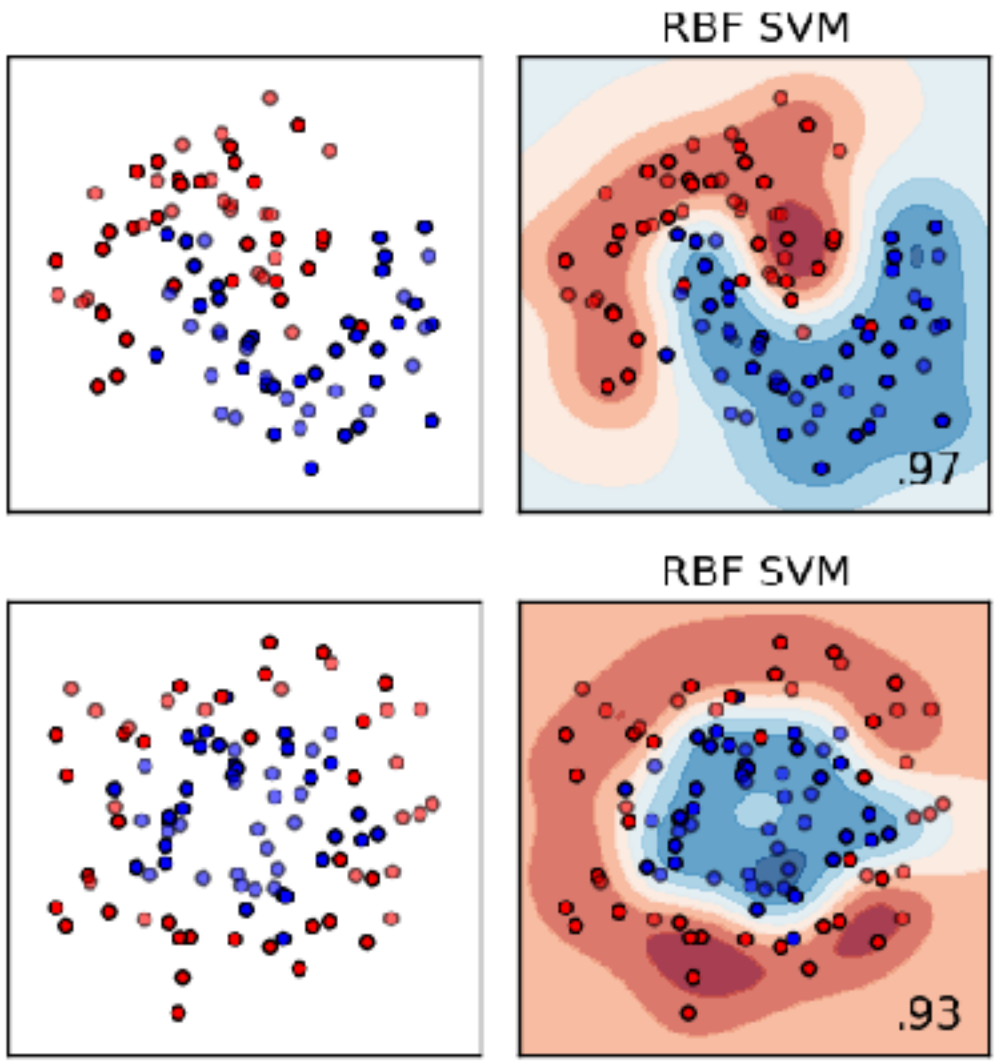


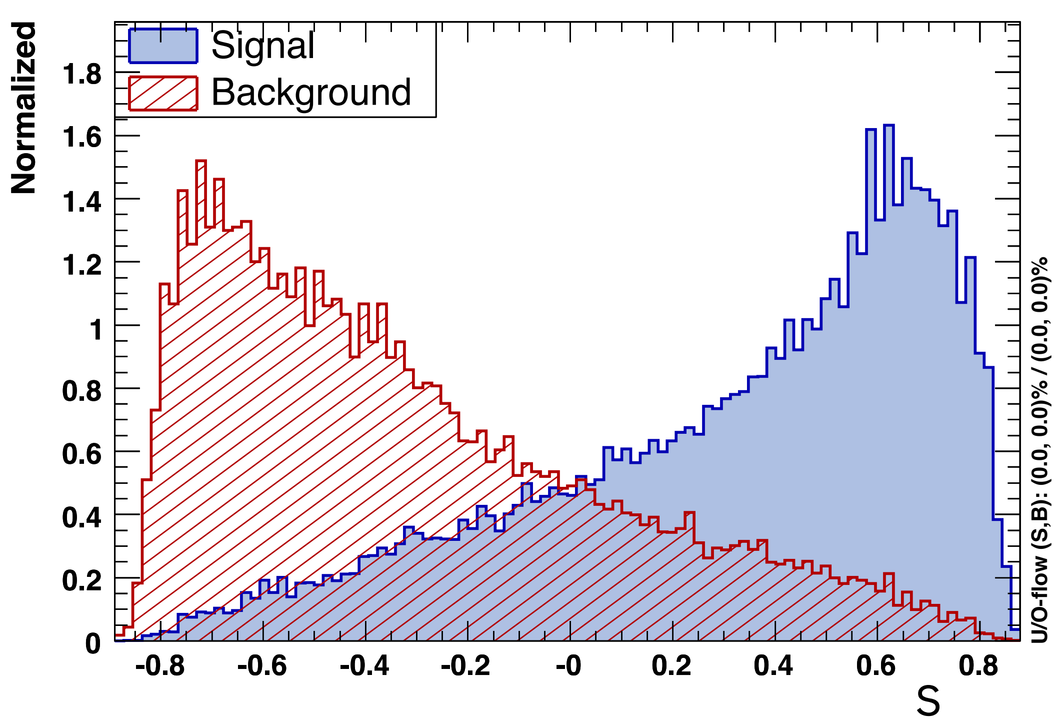
Common to use machine learning classifiers to separate signal (H_1) vs. background (H_0)

- want a function s: X→ Y that maps signal to y=1 and background to y=0
- **calculus of variations**: find function s(x) that minimizes ***loss***:

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

MACHINE LEARNING: CLASSIFIERS





- **applied calculus of variations**: find function s(x) that minimizes

loss:

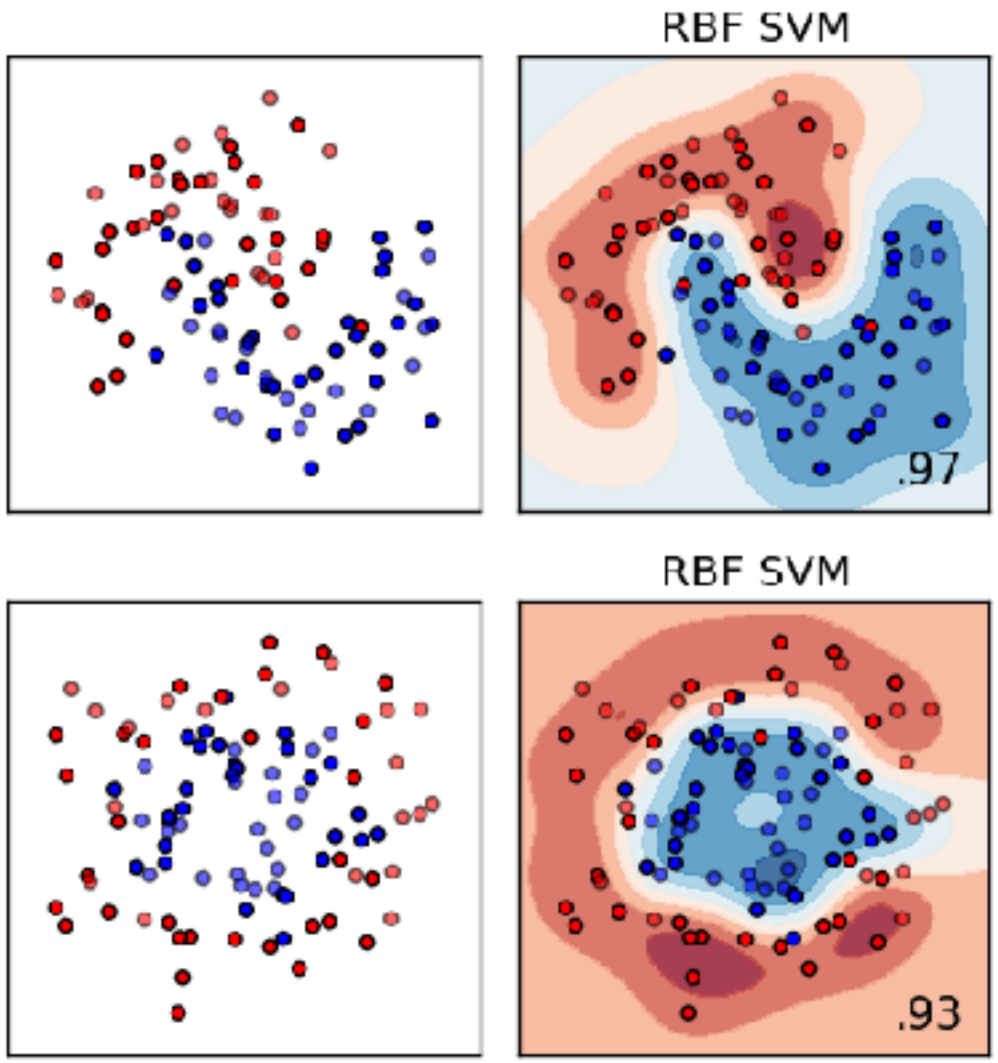
$$L[s] = \int p(x|H_0)\,(0 - s(x))^2\,dx + \int p(x|H_1)\,(1 - s(x))^2 dx$$

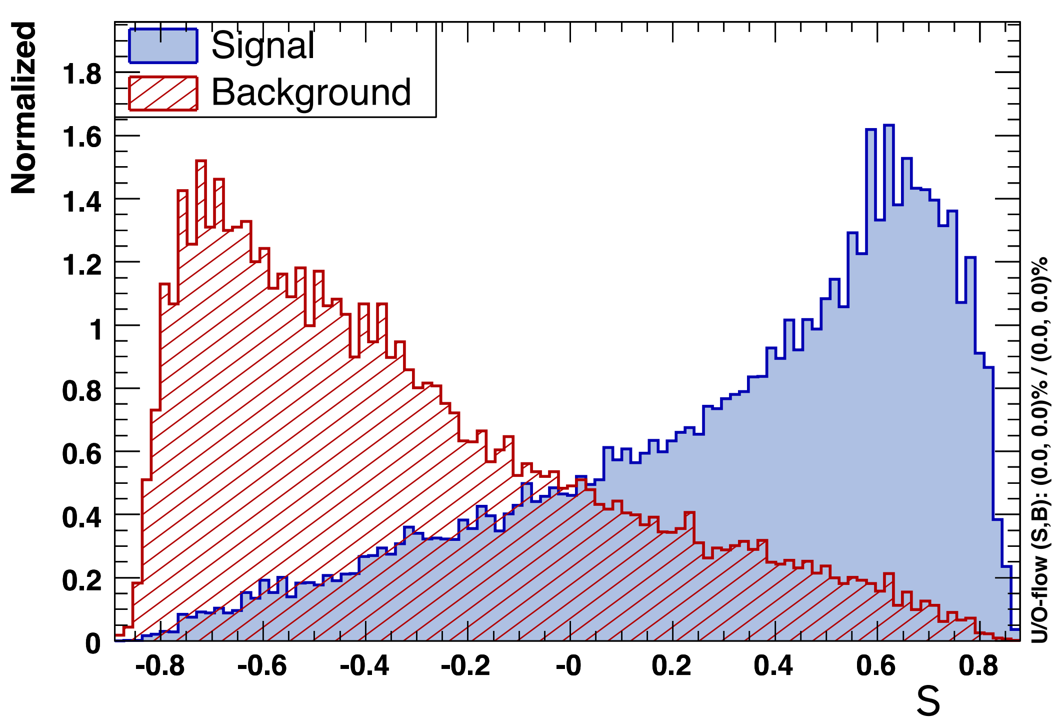
- i.e. approximate the optimal classifier

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

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

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





- **applied calculus of variations**: 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 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

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

NN = A HIGHLY FLEXIBLE FAMILY OF FUNCTIONS

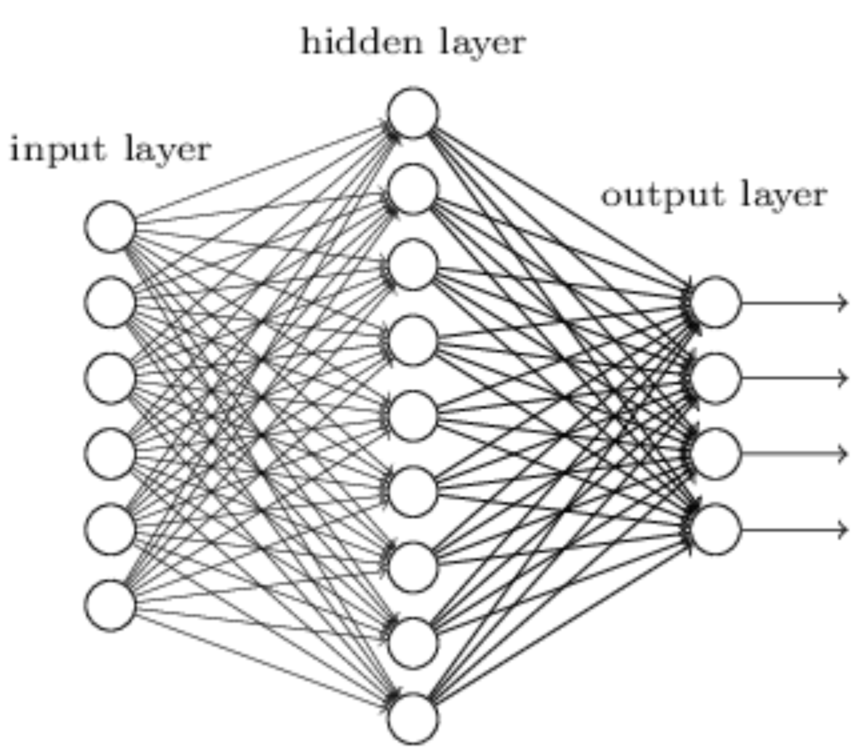
In calculus of variations, the optimization is over all functions: $\hat{s} = \underset{s}{\operatorname{argmin}}\, L[s]$

- In applied calculus of variations, we consider a highly flexible family of functions s_ϕ and optimize

- Think of neural networks as a highly flexible family of functions

- Machine learning also includes non-convex optimization algorithms that are affective even with millions of parameters!

Shallow neural network



Deep neural network

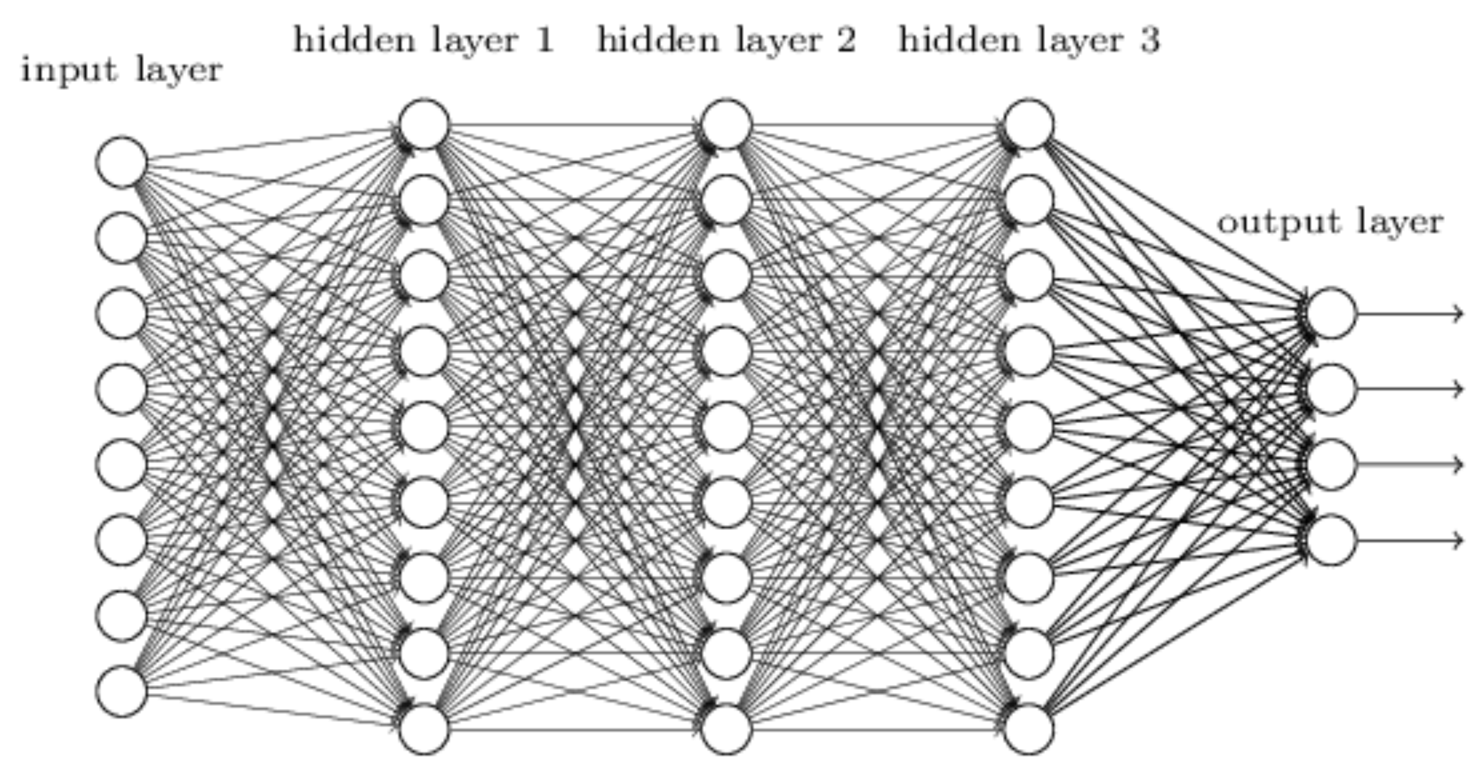


image credit: Michael Nielsen

Machine Learning for Effective Field Theory

Stronger Bounds on Higgs EFTs

Let θ denote the coefficients of higher dimensional operators in the Lagrangian and x be high-dimensional data associated to an event

- we want to compare any two points in EFT parameter space

- **goal**: estimate the true likelihood ratio $\dfrac{p(x|\theta_0)}{p(x|\theta_1)}$

- Theory language: dimension-6 operators of SM EFT, $\mathcal{L} \supset \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i$

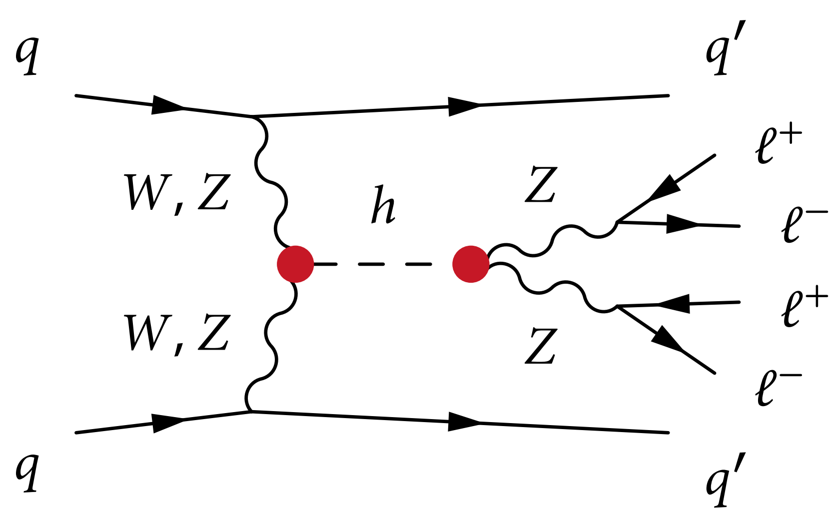
  [W. Buchmuller, D. Wyler 85; K. Hagiwara, S. Ishihara, S. R. Szalapski, D. Zeppenfeld 93; B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

  - Total rate: $\mathcal{O}_{\phi,2} = \frac{1}{2} \partial^\mu (\phi^\dagger \phi)\, \partial_\mu (\phi^\dagger \phi)$
  - New kinematic structures:

$$\mathcal{O}_B = \mathrm{i}\frac{g}{2} (D^\mu \phi^\dagger)(D^\nu \phi)\, B_{\mu\nu} \qquad \mathcal{O}_W = \mathrm{i}\frac{g}{2} (D^\mu \phi)^\dagger \sigma^k (D^\nu \phi)\, W^k_{\mu\nu}$$

$$\mathcal{O}_{BB} = -\frac{g'^2}{4} (\phi^\dagger \phi)\, B_{\mu\nu} B^{\mu\nu} \qquad \mathcal{O}_{WW} = -\frac{g^2}{4} (\phi^\dagger \phi)\, W^k_{\mu\nu} W^{\mu\nu\, k}$$

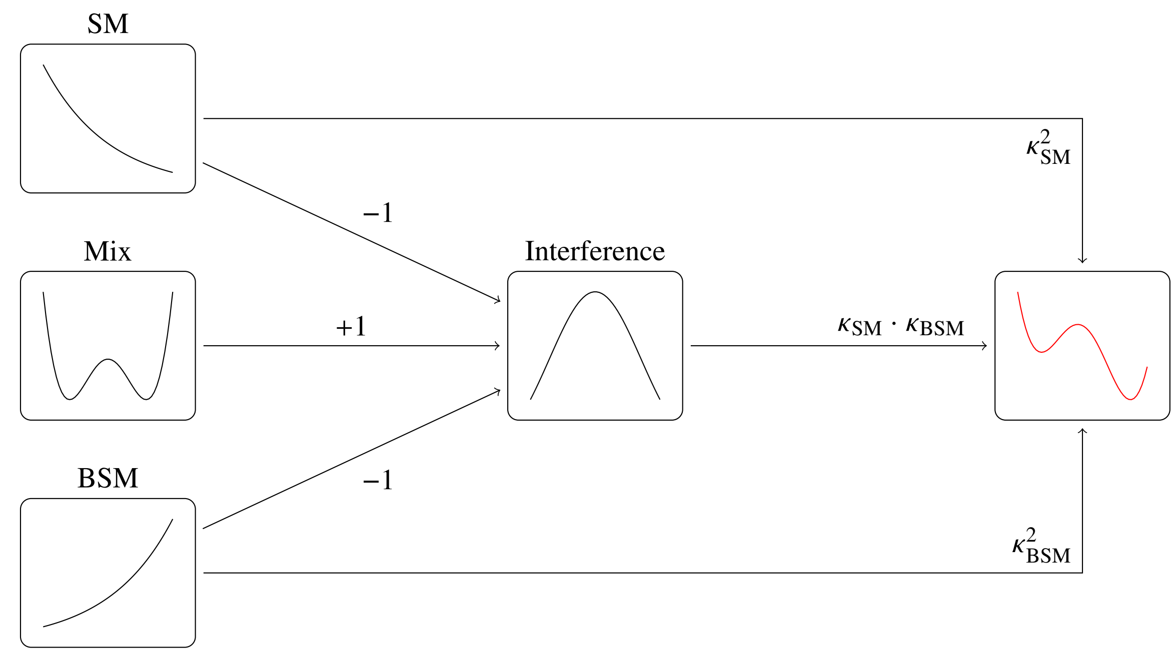
  - CP violation: $\mathcal{O}_{W\widetilde{W}} = -\frac{g^2}{4} (\phi^\dagger \phi)\, W^k_{\mu\nu} \widetilde{W}^{\mu\nu\, k}$
  - Others strongly constrained by EWPD or redundant



EFT EMBEDDED IN A VECTOR SPACE

ATL-PHYS-PUB-2015-047

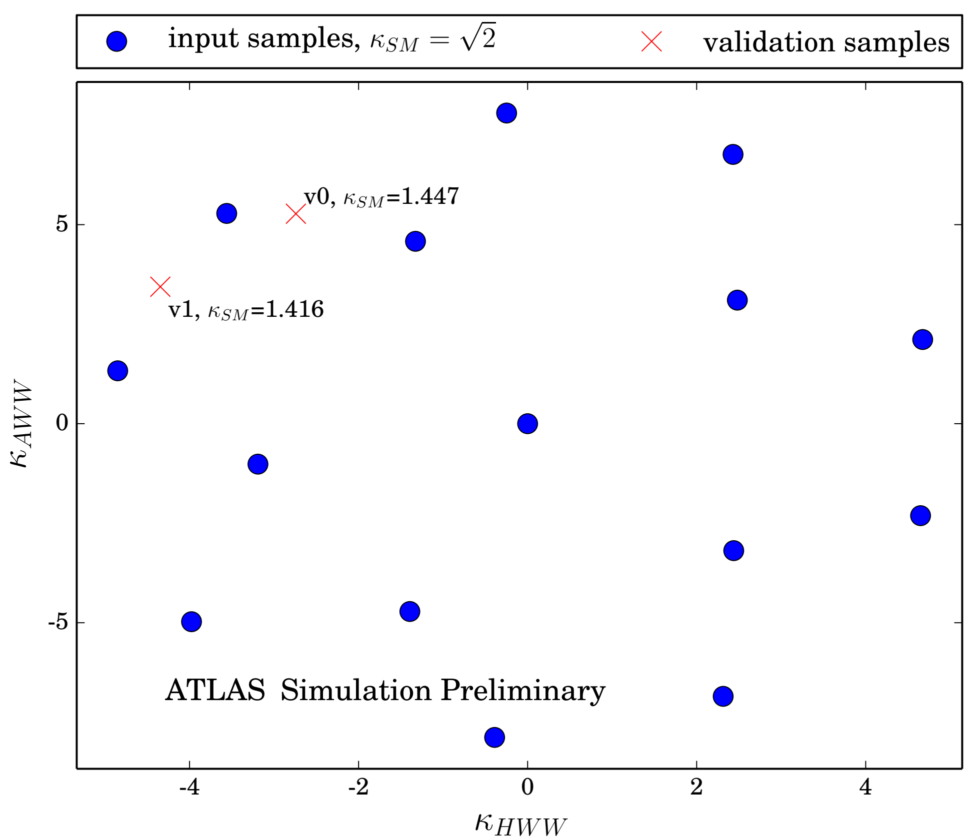
Difficulty is that one changes the parameters of the EFT, the distributions p(x|θ) change due to interference. But there is a trick:

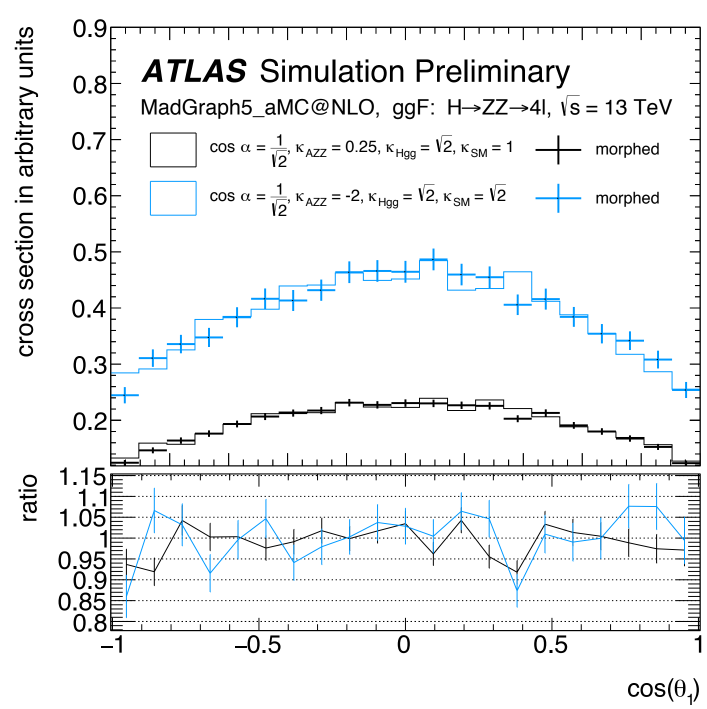


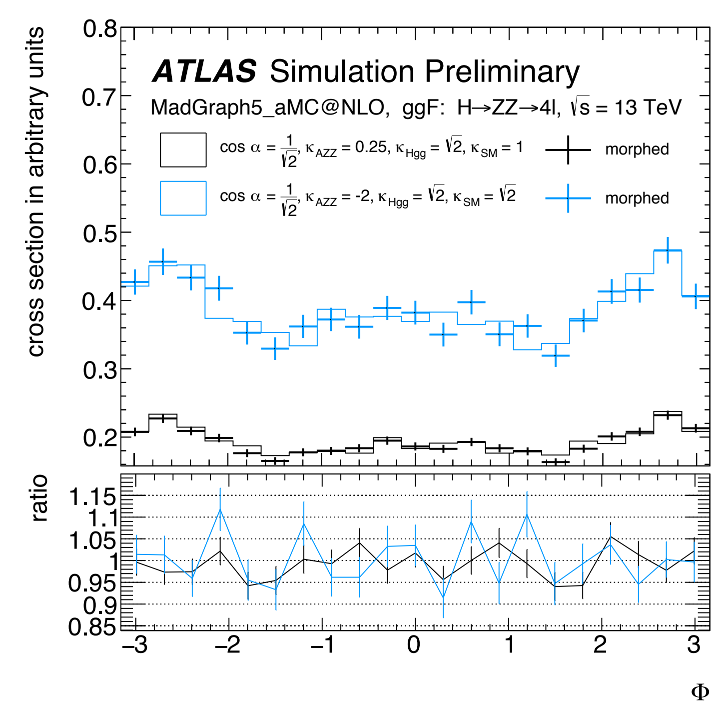
Simple example: $|g_1 M_{SM} + g_2 M_{BSM}|^2 = g_1^2 |M_{SM}|^2 + 2 g_1 g_2 Re\left[M_{SM}^* M_{BSM}\right] + g_2^2 |M_{BSM}|^2$

3-d vector space, any point in this space is linear mixture of 3 basis samples!

(real examples need more basis samples)







EFT DECOMPOSITION

$$d\sigma \propto \left| \overset{\text{production}}{\boxed{\left(\mathcal{M}^p_{\text{SM}} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{M}^p_i \right)}} \overset{\text{decay}}{\boxed{\left(\mathcal{M}^d_{\text{SM}} + \sum_j \frac{f_j}{\Lambda^2} \mathcal{M}^d_j \right)}} \right|^2$$

Express EFT as a mixture:

$$p(x|\theta) = \sum_c w_c(\theta) p_c(x)$$

Process	Number of components for n operators					
	$\mathcal{O}(\Lambda^0)$	$\mathcal{O}(\Lambda^{-2})$	$\mathcal{O}(\Lambda^{-4})$	$\mathcal{O}(\Lambda^{-6})$	$\mathcal{O}(\Lambda^{-8})$	$\sum$
hV / WBF production	1	n	$\frac{n(n+1)}{2}$			$\frac{(n+1)(n+2)}{2}$
$h \to VV$ decay	1	n	$\frac{n(n+1)}{2}$			$\frac{(n+1)(n+2)}{2}$
Production + decay	1	n	$\frac{n(n+1)}{2}$	$\binom{n+2}{3}$	$\binom{n+3}{4}$	$\binom{n+4}{4}$

Table 1: Number of components c as given in Eq. (6) for different processes, sorted by their suppression by the EFT cutoff scale Λ.

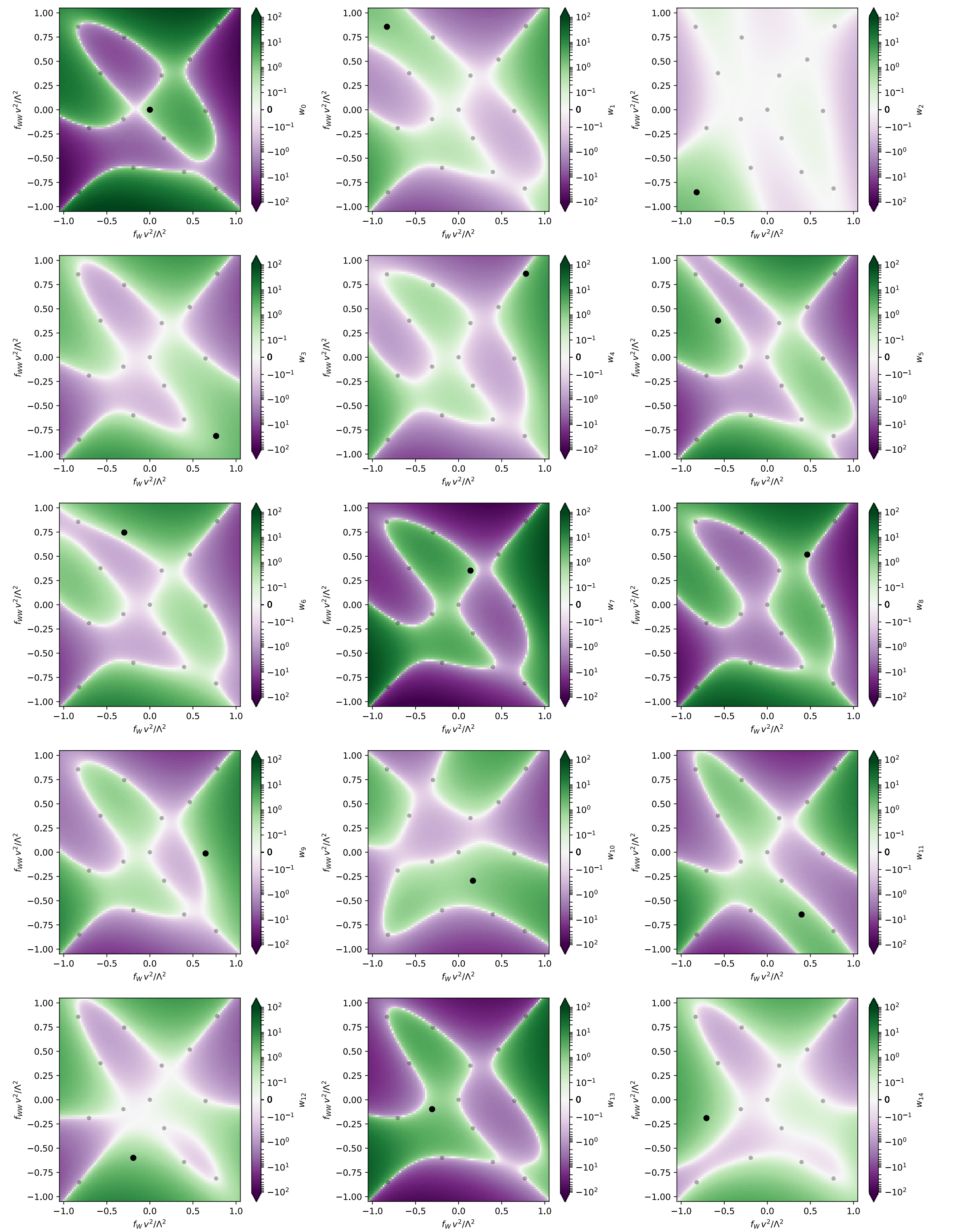


Figure 13: Morphing weights $w_i(\boldsymbol{\theta})$ for basis points distributed over the full relevant parameter space.

For 2 BSM operators affecting VBF Higgs production and decay, we need a 15-D vector space

For 5 BSM operators we need 126-D vector space

A binary classifier approximates

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

Which is one-to-one with the likelihood ratio

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

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]

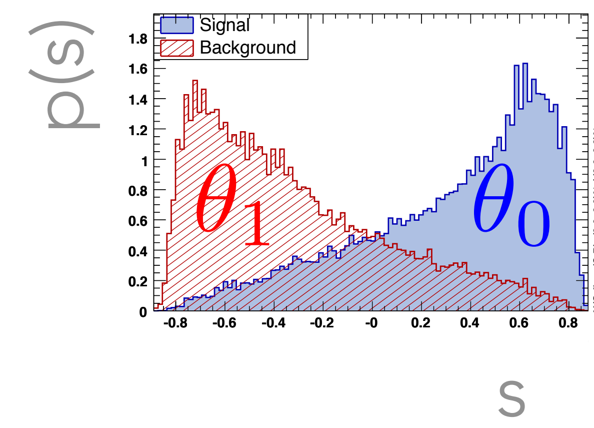
CALIBRATING THE LIKELIHOOD-RATIO TRICK

The intractable likelihood ratio based on high-dimensional features x is:

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

We can show that an **equivalent test** can be made from 1-D projection

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



if the scalar map s: X → $\mathbb{R}$ has the same level sets as the likelihood ratio

$$s(x;\theta_0;\theta_1) = \text{monotonic}[\ p(x|\theta_0)/p(x|\theta_1)\ ]$$

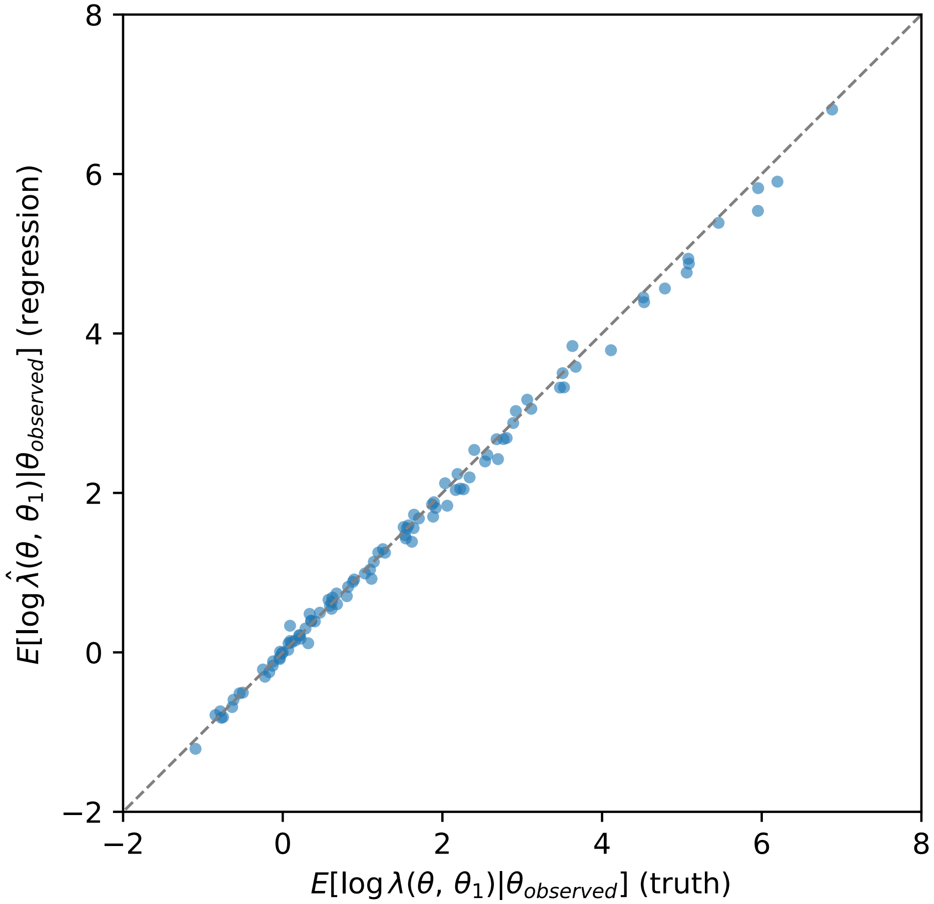
Estimating the density of $s(x;\theta_0,\theta_1)$ via the simulator calibrates the ratio.

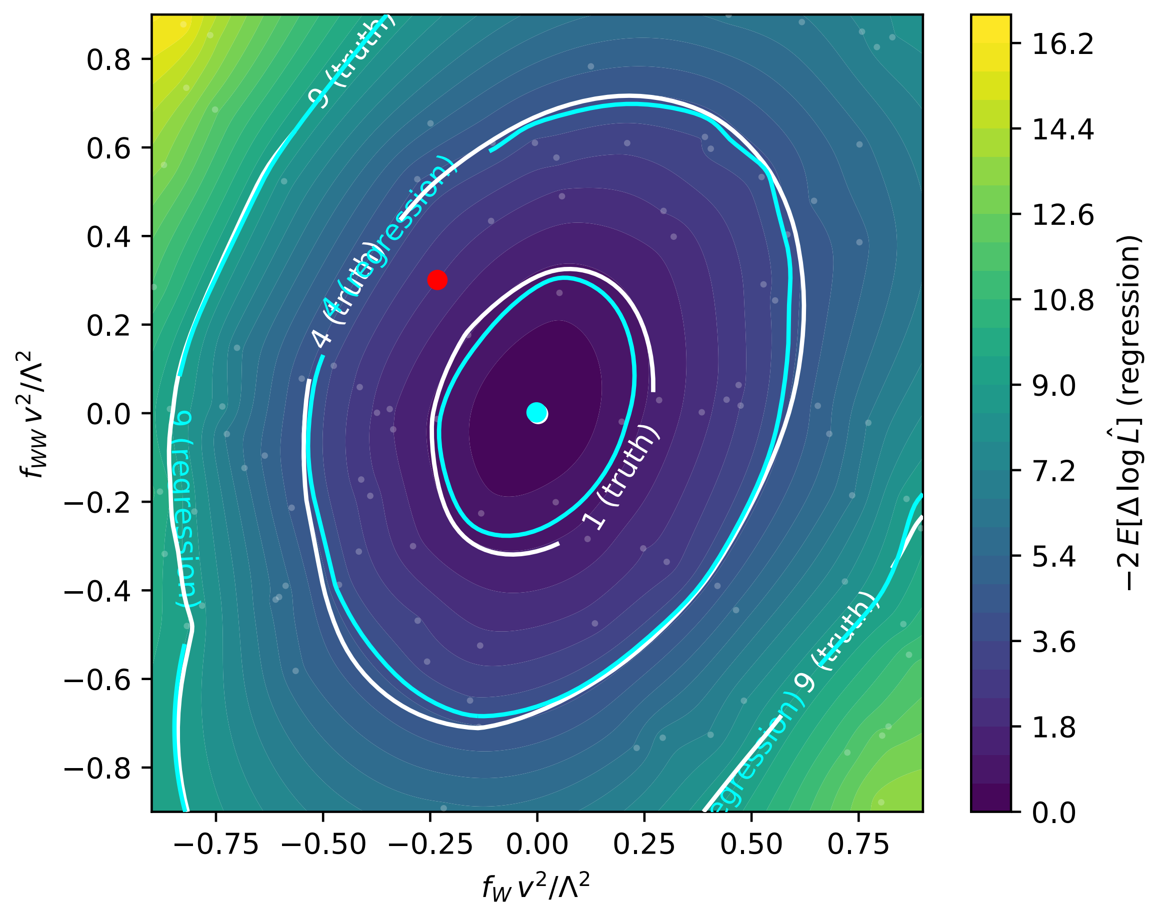
K.C., G. Louppe, J. Pavez: Approximating Likelihood Ratios with Calibrated Discriminative Classifiers [arXiv:1506.02169]

MACHINE LEARNING THE HIGGS EFFECTIVE FIELD THEORY

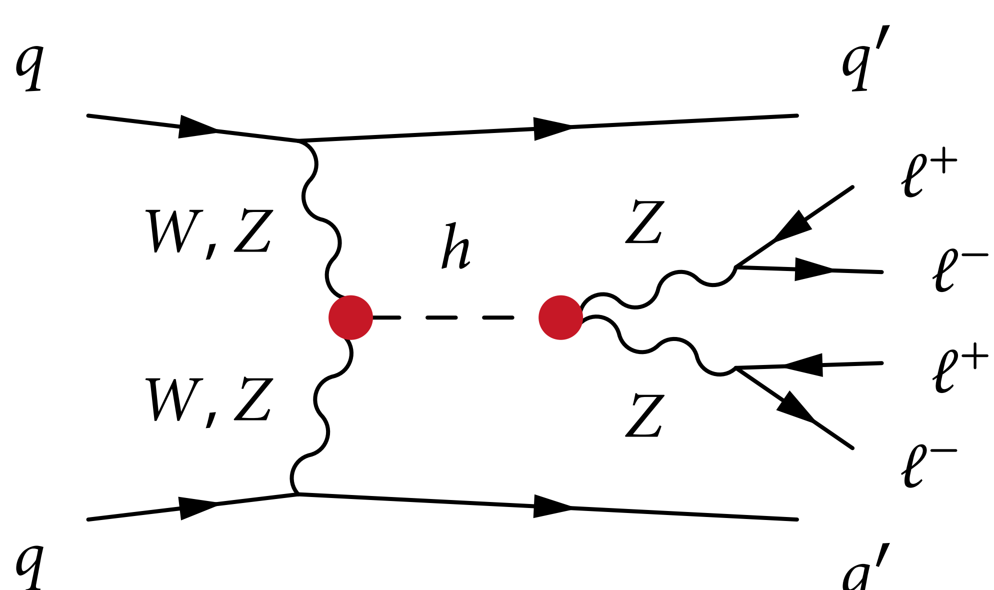
(based on a 16-Dim observation $\boldsymbol{x}$)

Estimated likelihood





True likelihood

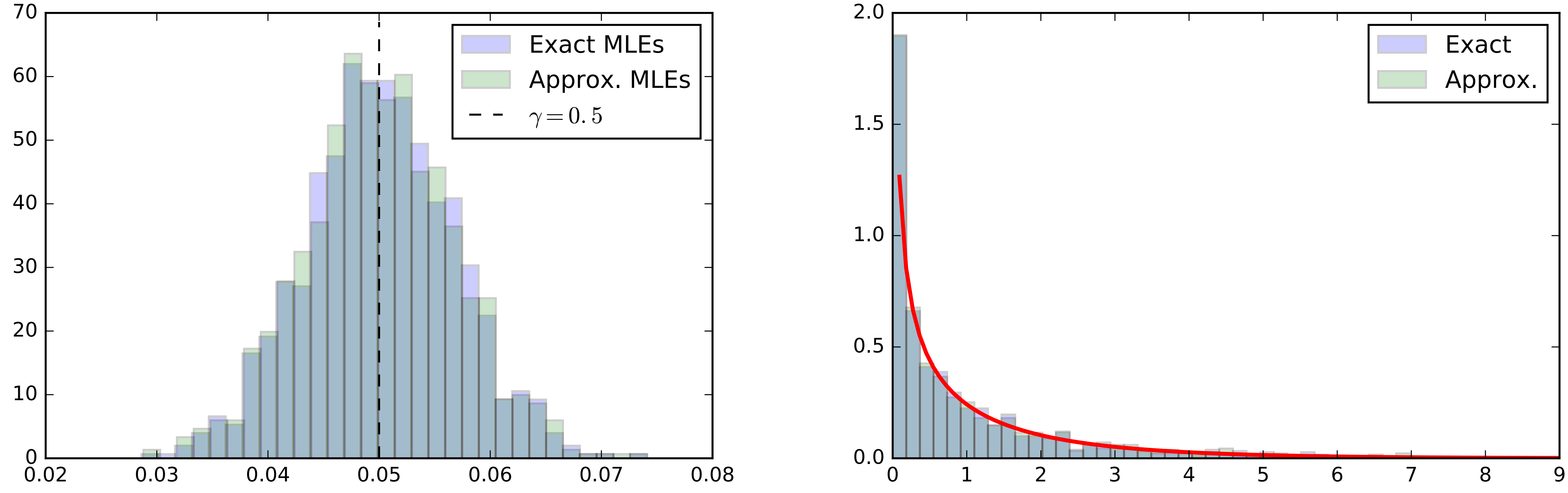


work with Juan Pavez, Gilles Louppe, Cyril Becot, and Lukas Heinrich; Johann Brehmer, Felix Kling, and Tilman Plehn
"Better Higgs Measurements Through Information Geometry" [arXiv:1612.05261] & CARL [arxiv:1506.02169]

AMORTIZED LIKELIHOOD-FREE INFERENCE

Once we've learned the function s(x; θ_0,θ_1) to approximate the likelihood, we can apply it to any data x.

- unlike MCMC, we pay biggest computational costs up front
- Here we repeat inference thousands of times & check asymptotic statistical theory

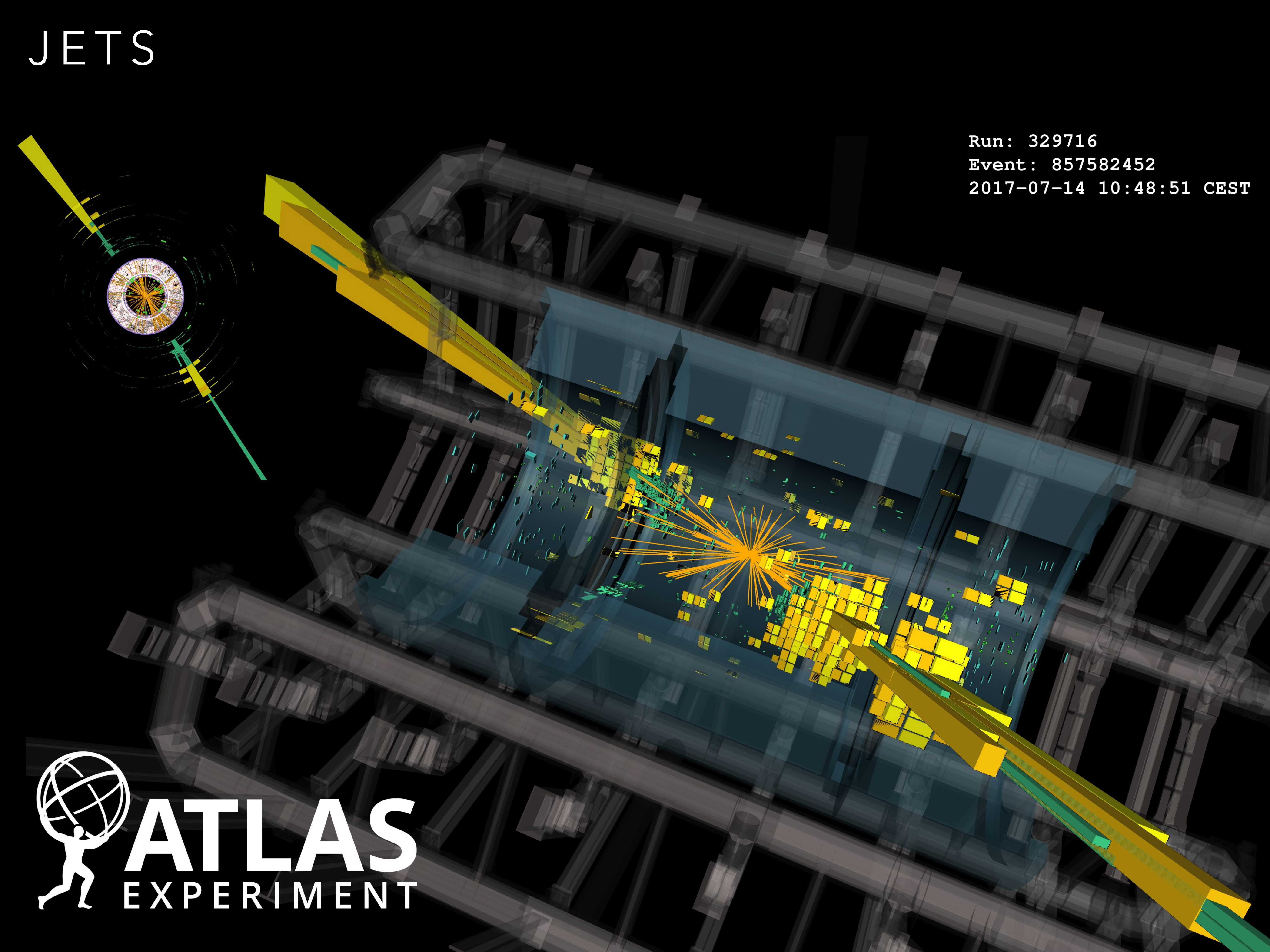


(a) Exact vs. approximated MLEs.

(b) $p(-2\log\Lambda(\gamma = 0.05)\,|\,\gamma = 0.05)$

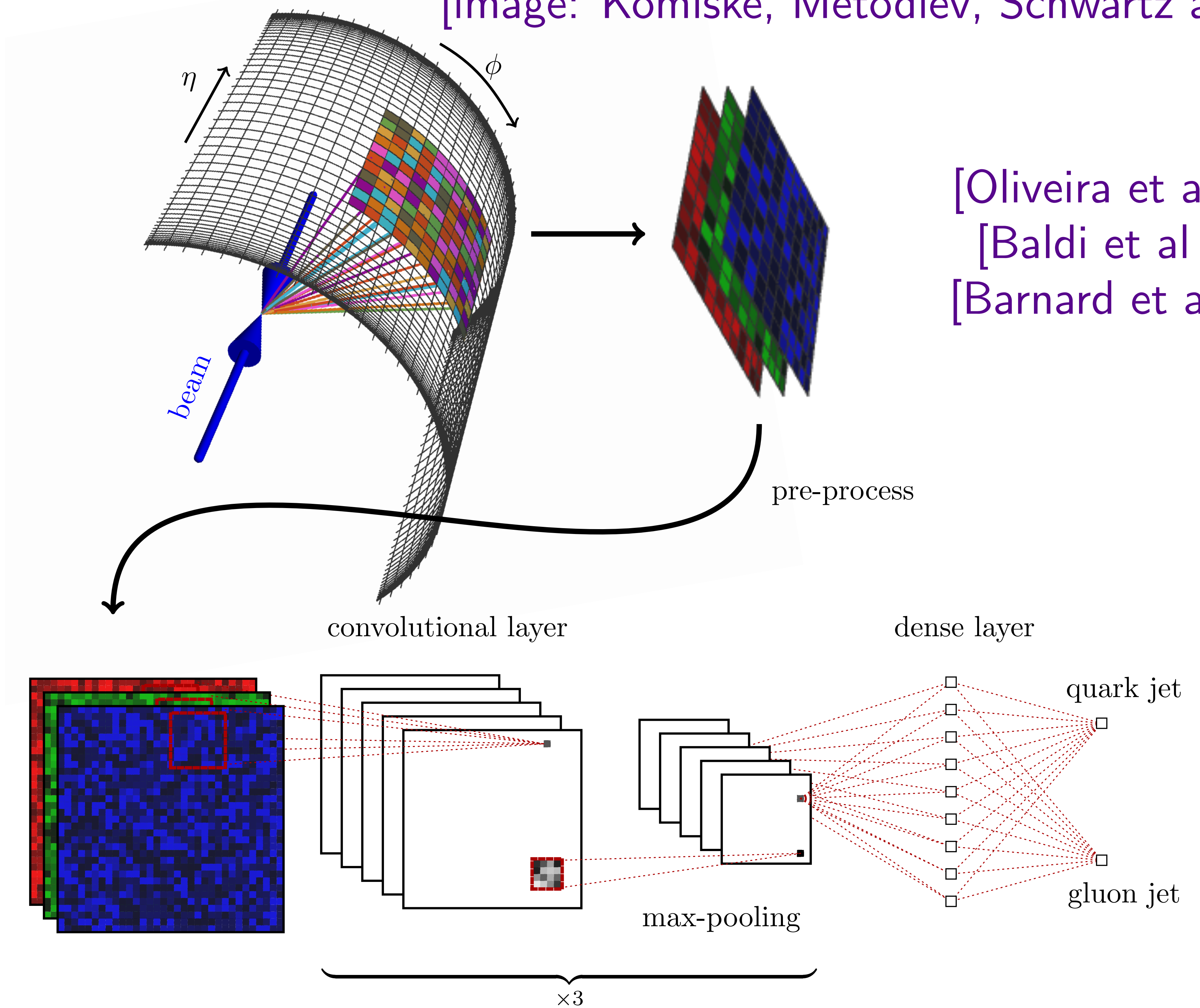
Example:

Jet Physics


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

JET IMAGES

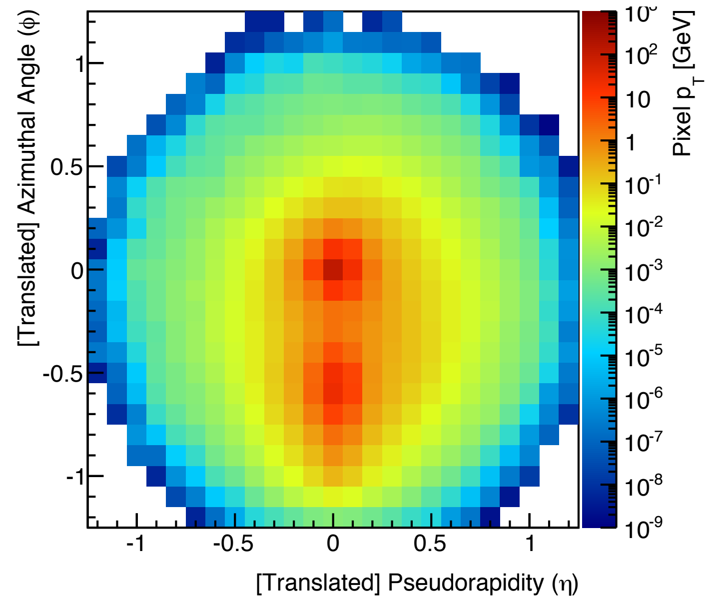
[image: Komiske, Metodiev, Schwartz arxiv:1612.01551]



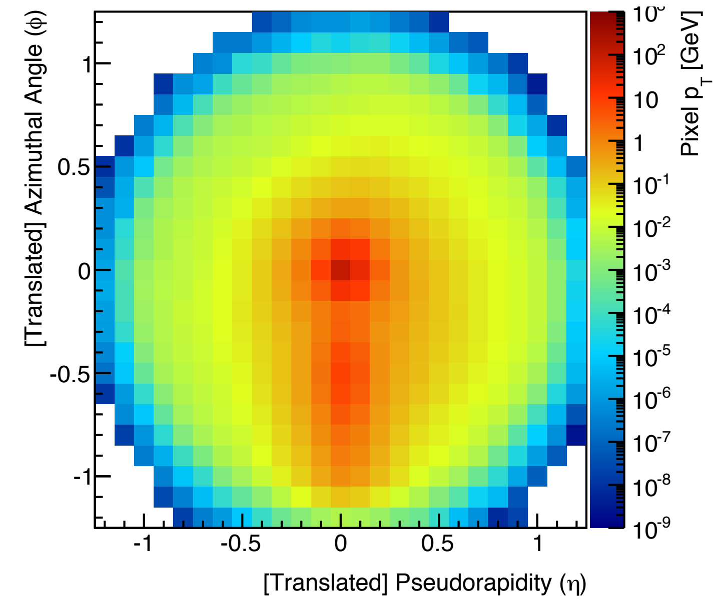
[Oliveira et al arXiv:1511.05190]
[Baldi et al arXiv:1603.09349]
[Barnard et al arXiv:1609.00607]

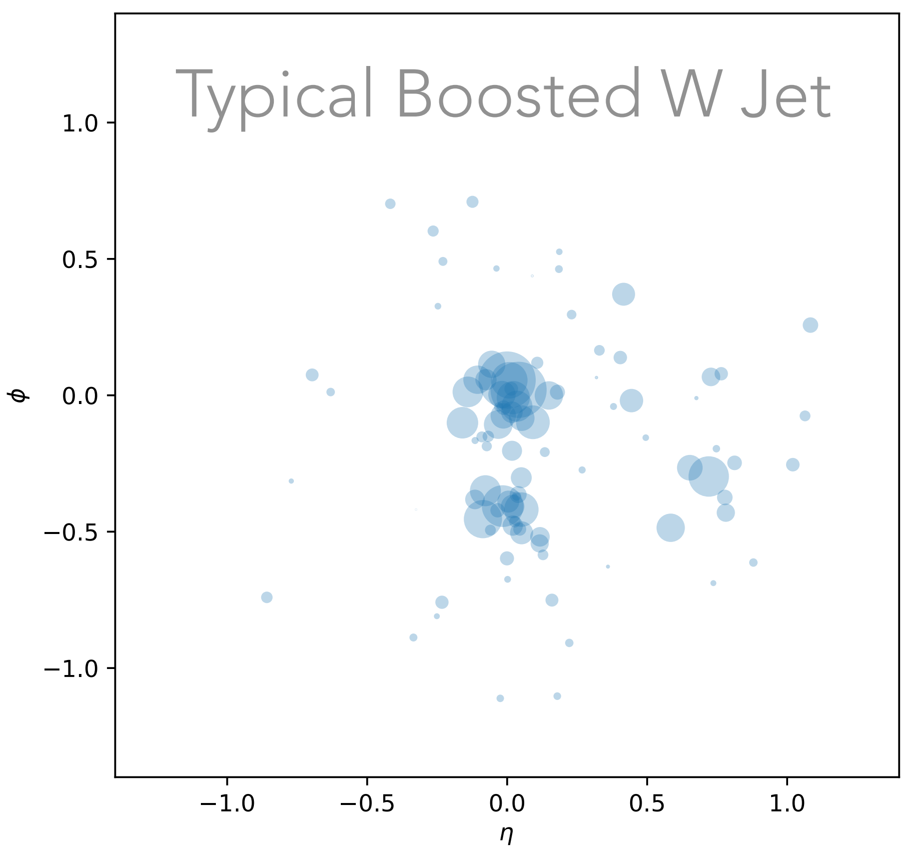
BOOSTED JET TAGGING

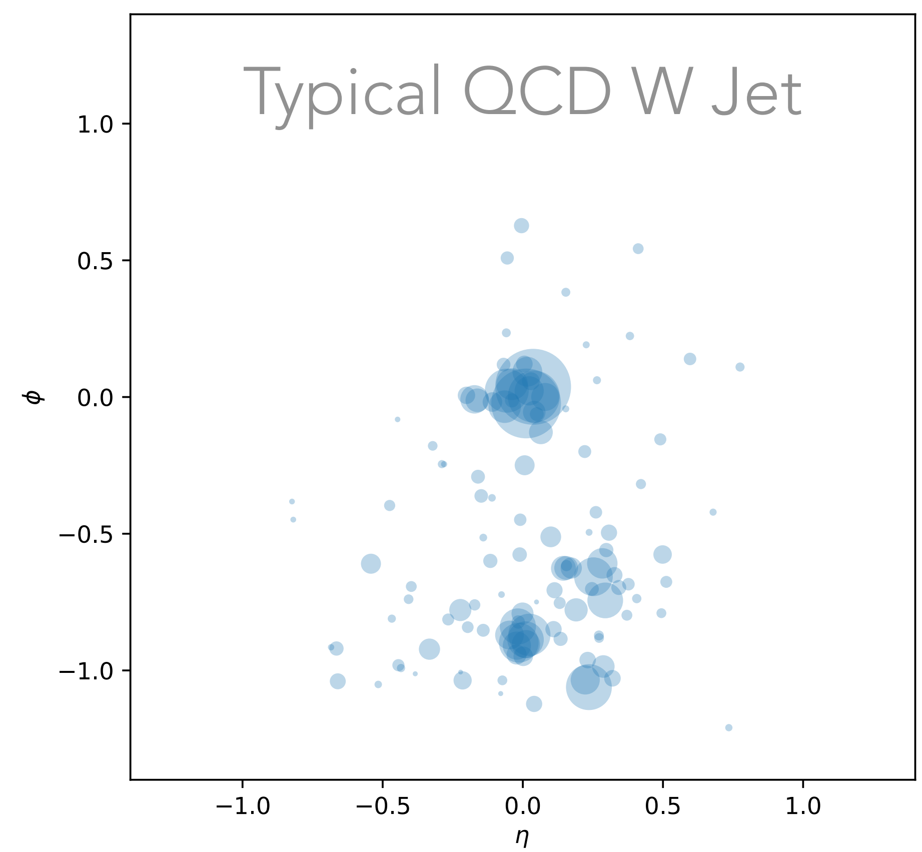
Average Boosted W Jet



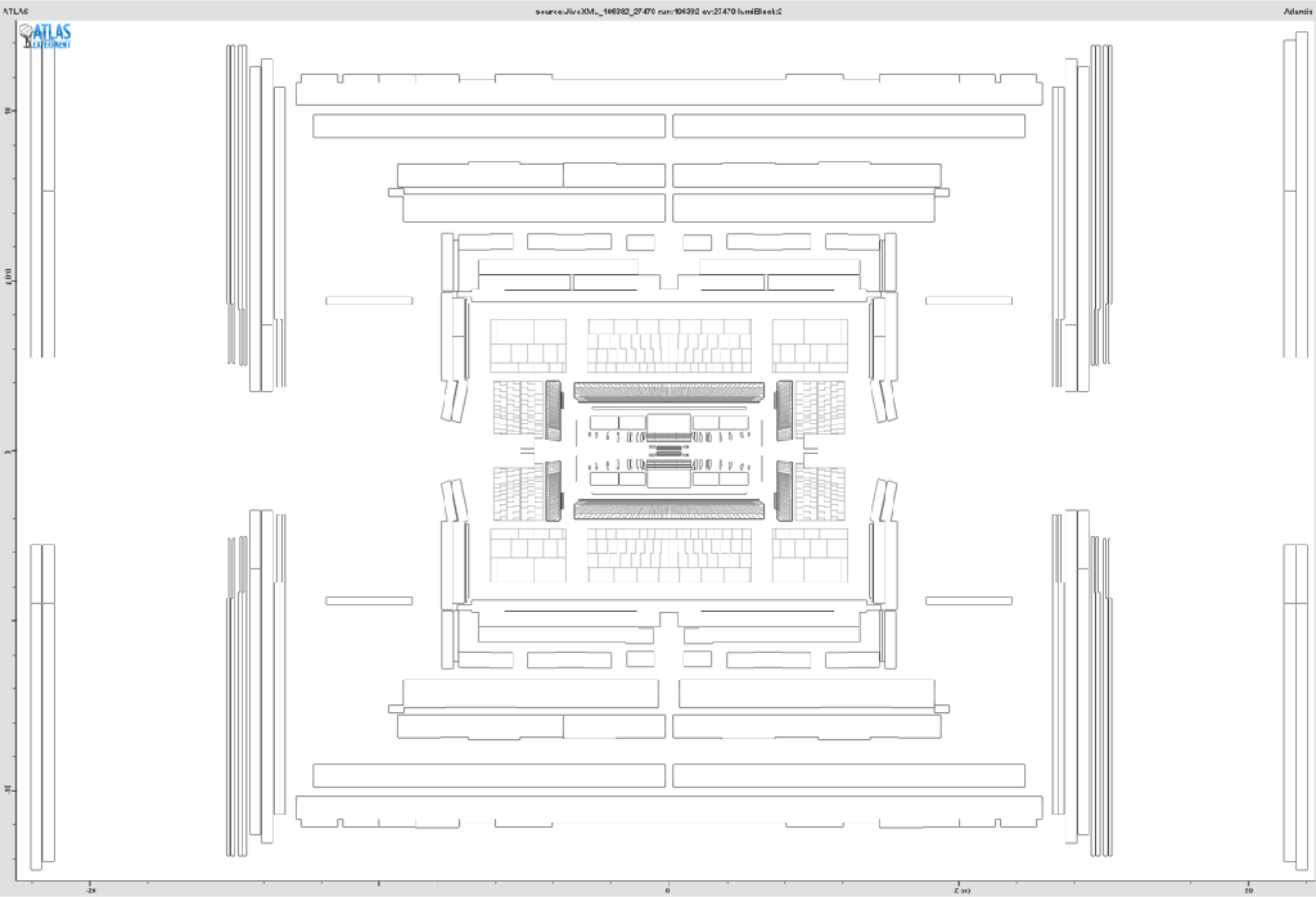
Average QCD Jet



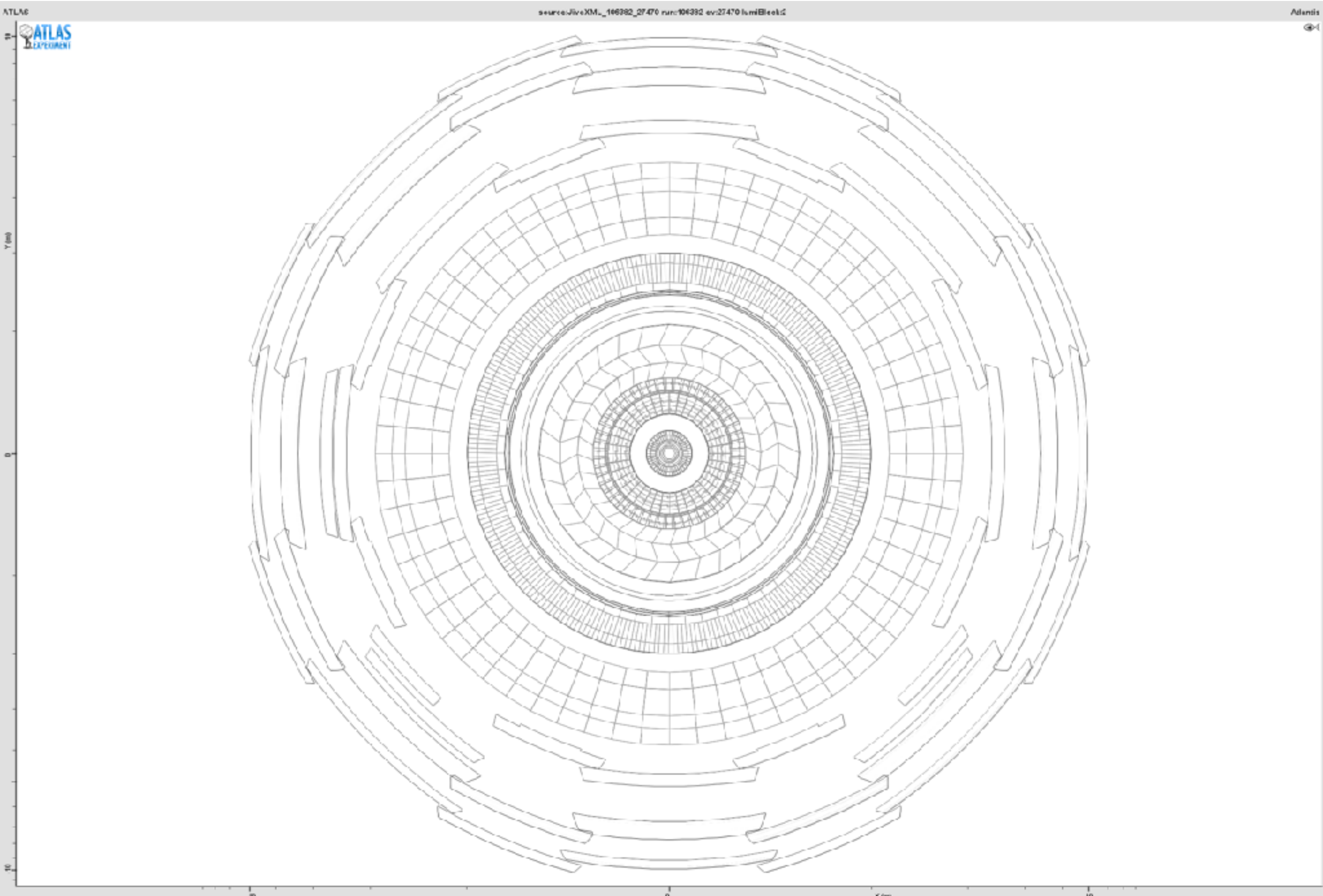




NON-UNIFORM GEOMETRY

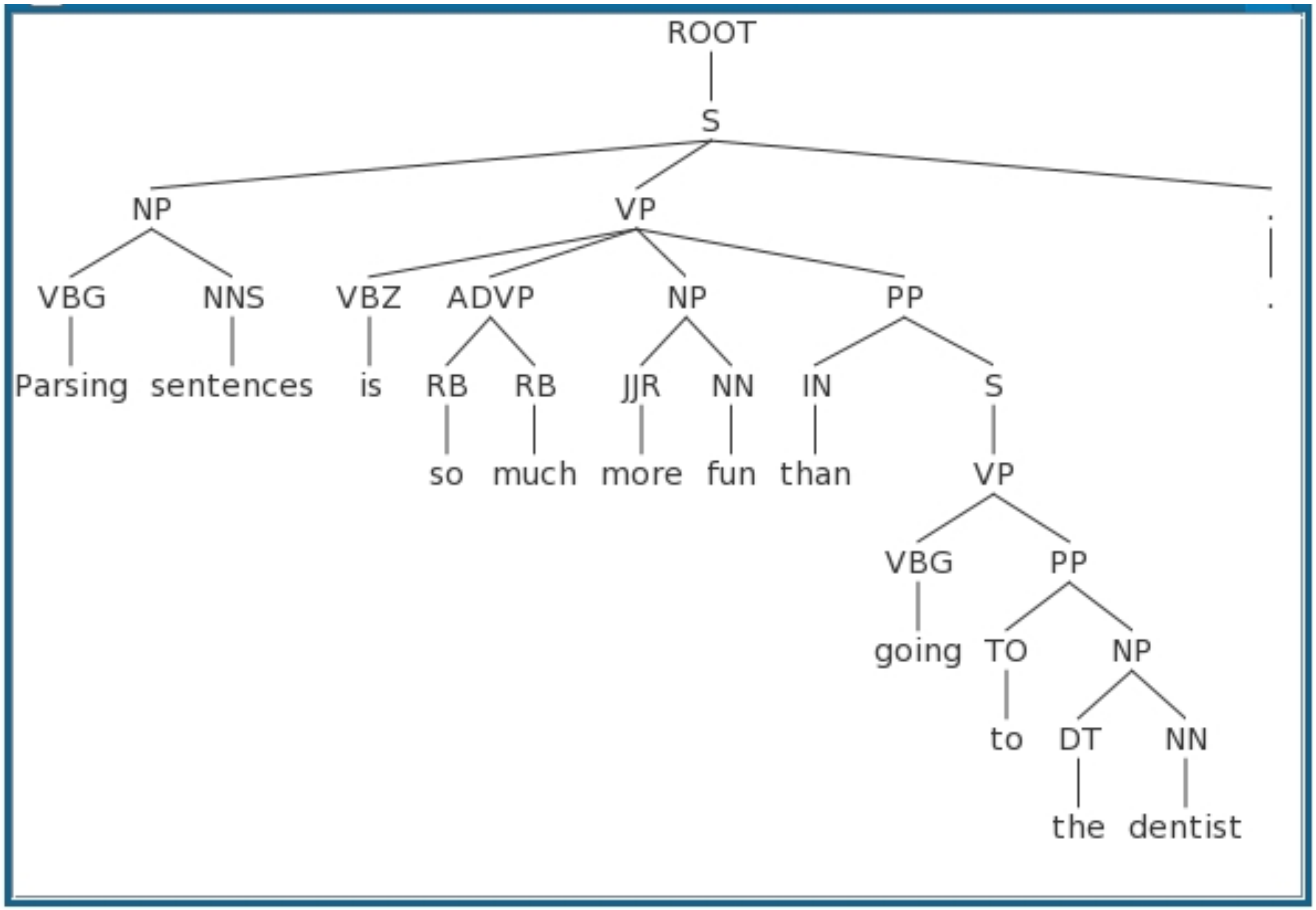


NON-UNIFORM GEOMETRY



Recursive Neural Networks showing great performance for Natural Language Processing tasks

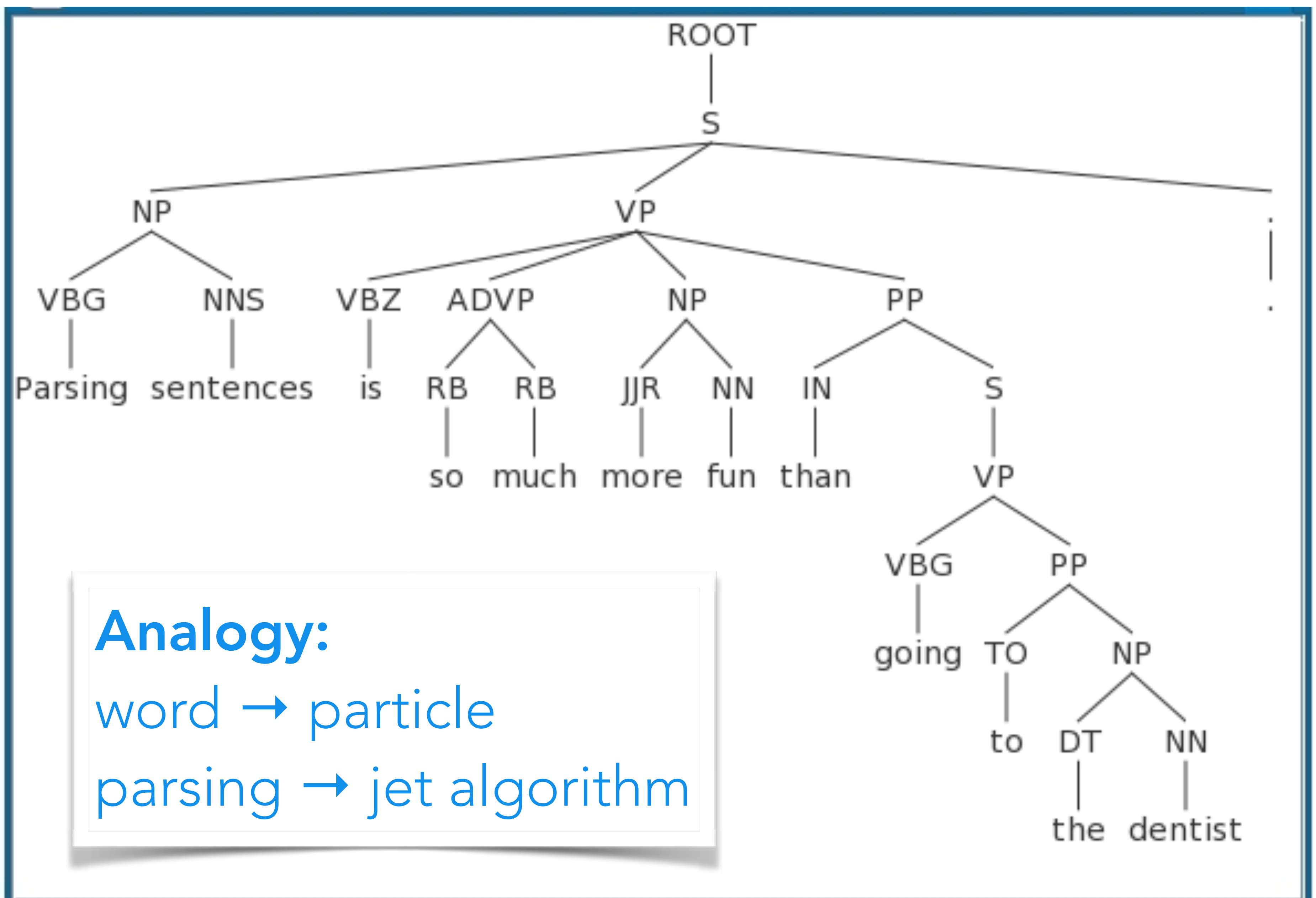
- neural network's topology given by parsing of sentence!



FROM IMAGES TO SENTENCES

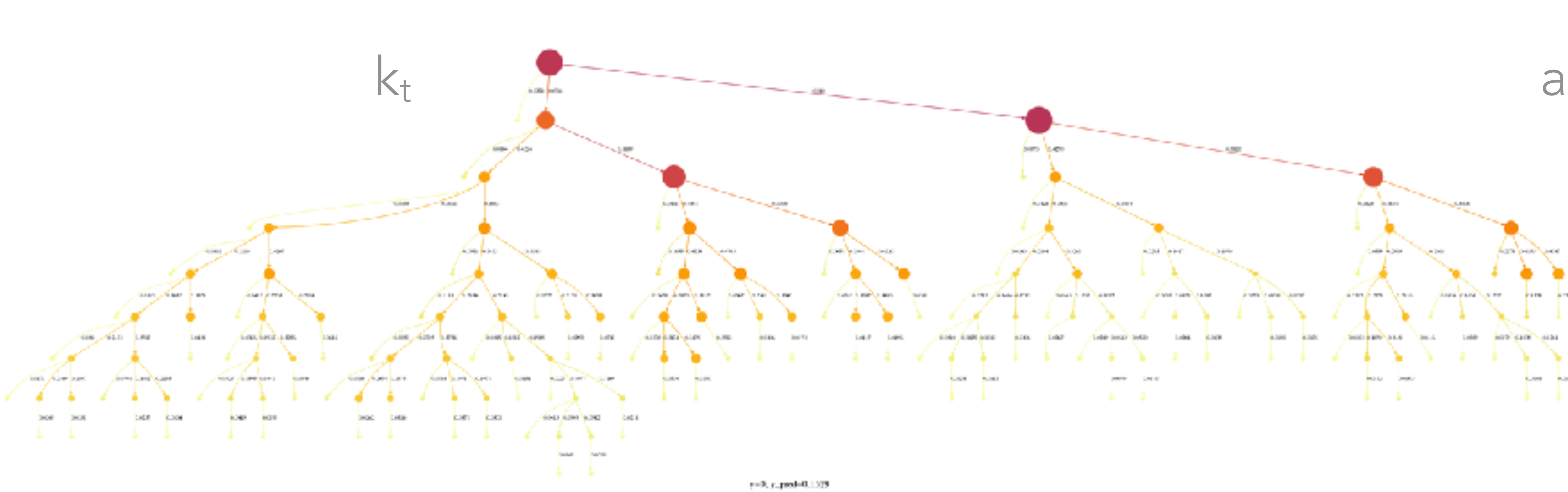
Recursive Neural Networks showing great performance for Natural Language Processing tasks

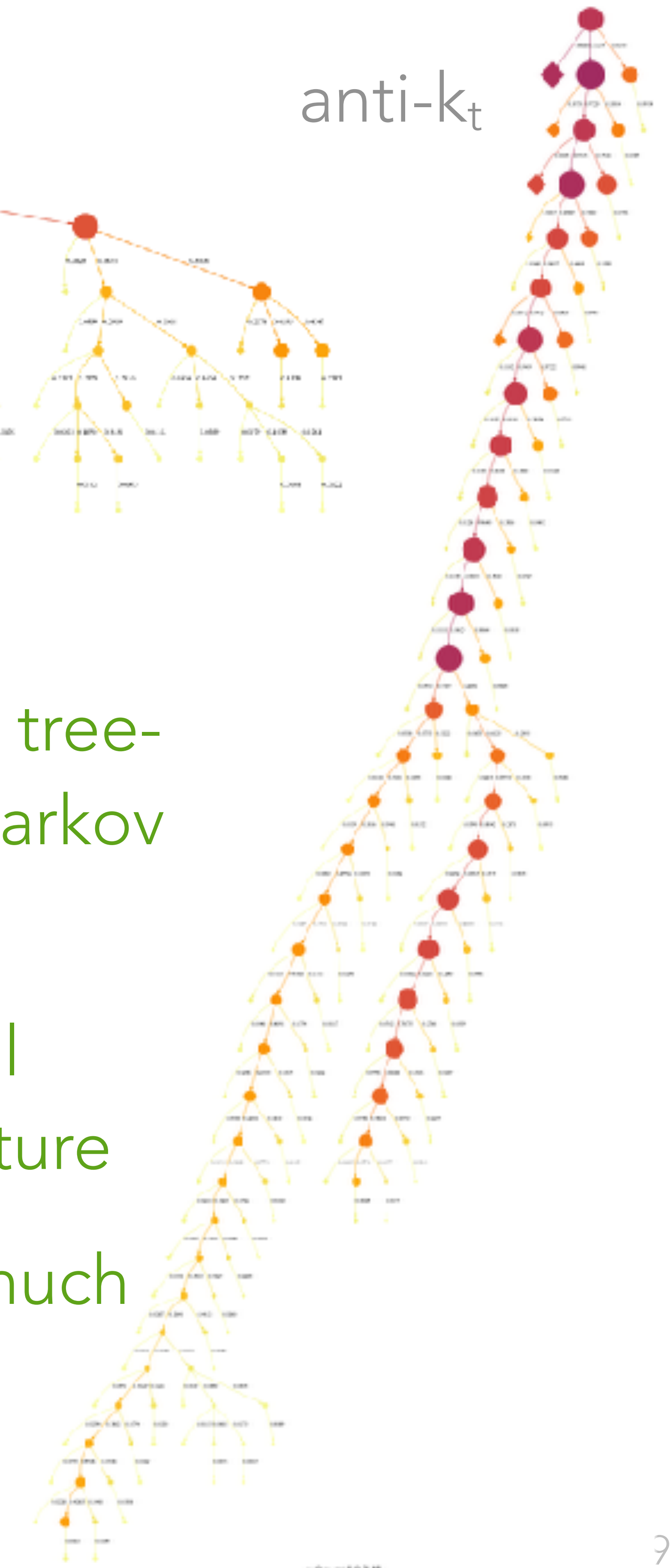
- neural network's topology given by parsing of sentence!

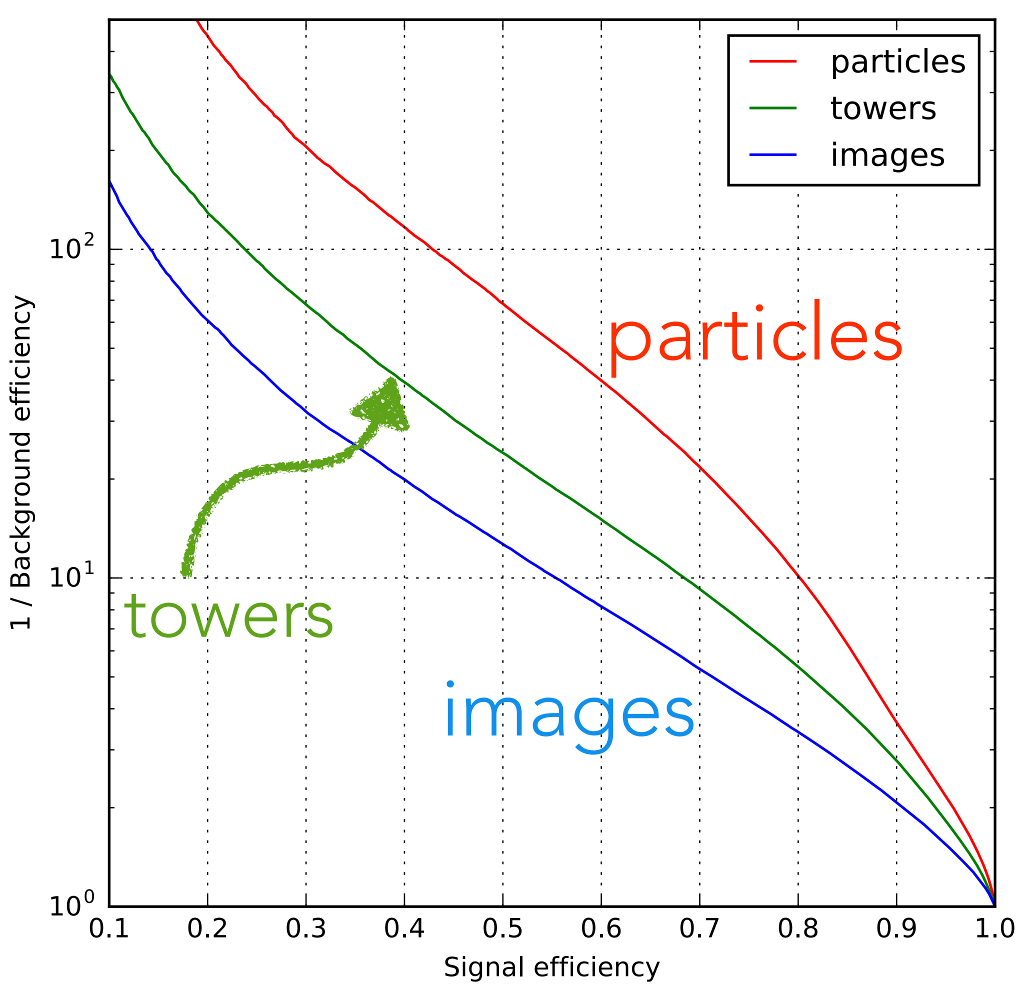


Analogy:
word → particle
parsing → jet algorithm

QCD-INSPIRED RECURSIVE NEURAL NETWORKS





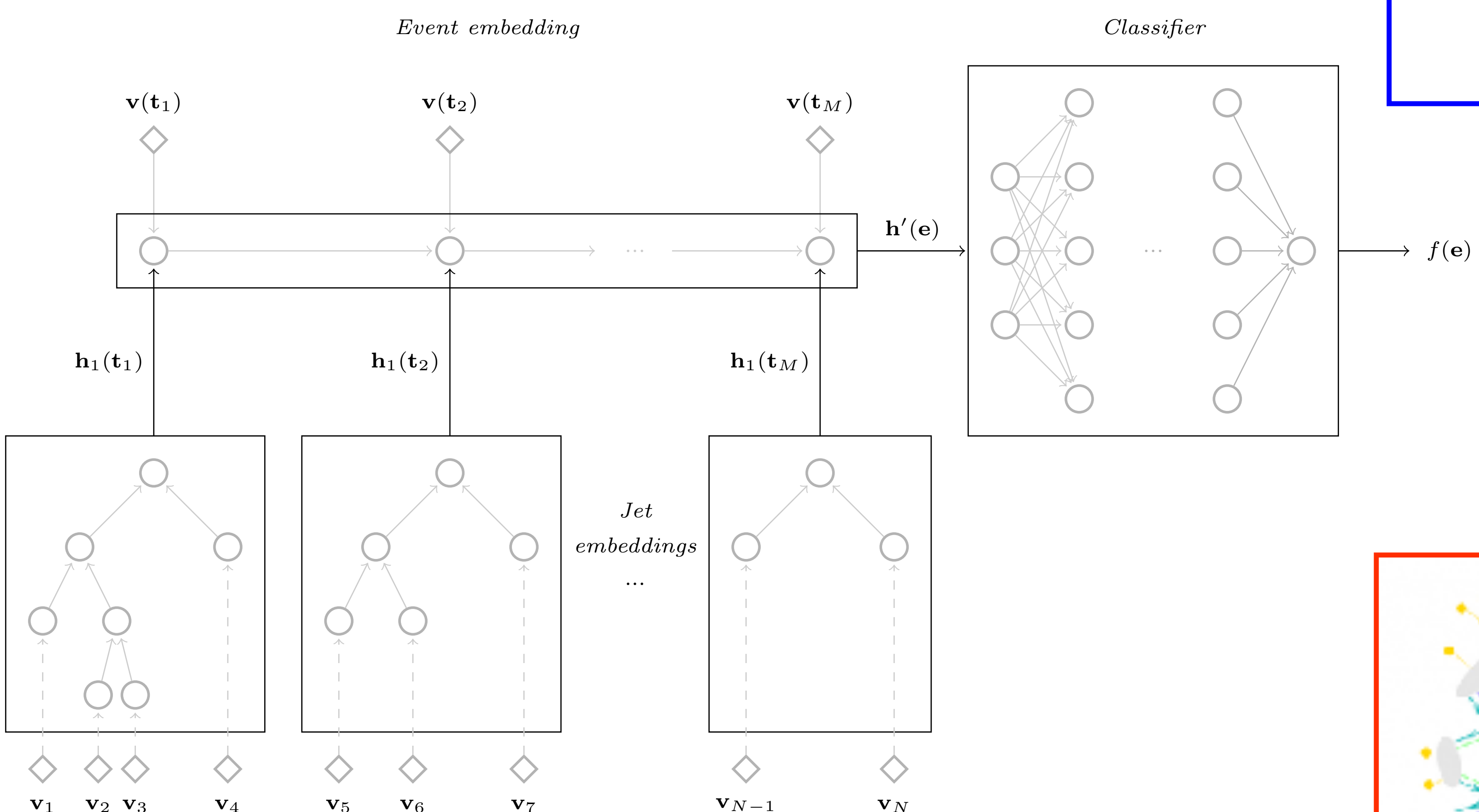


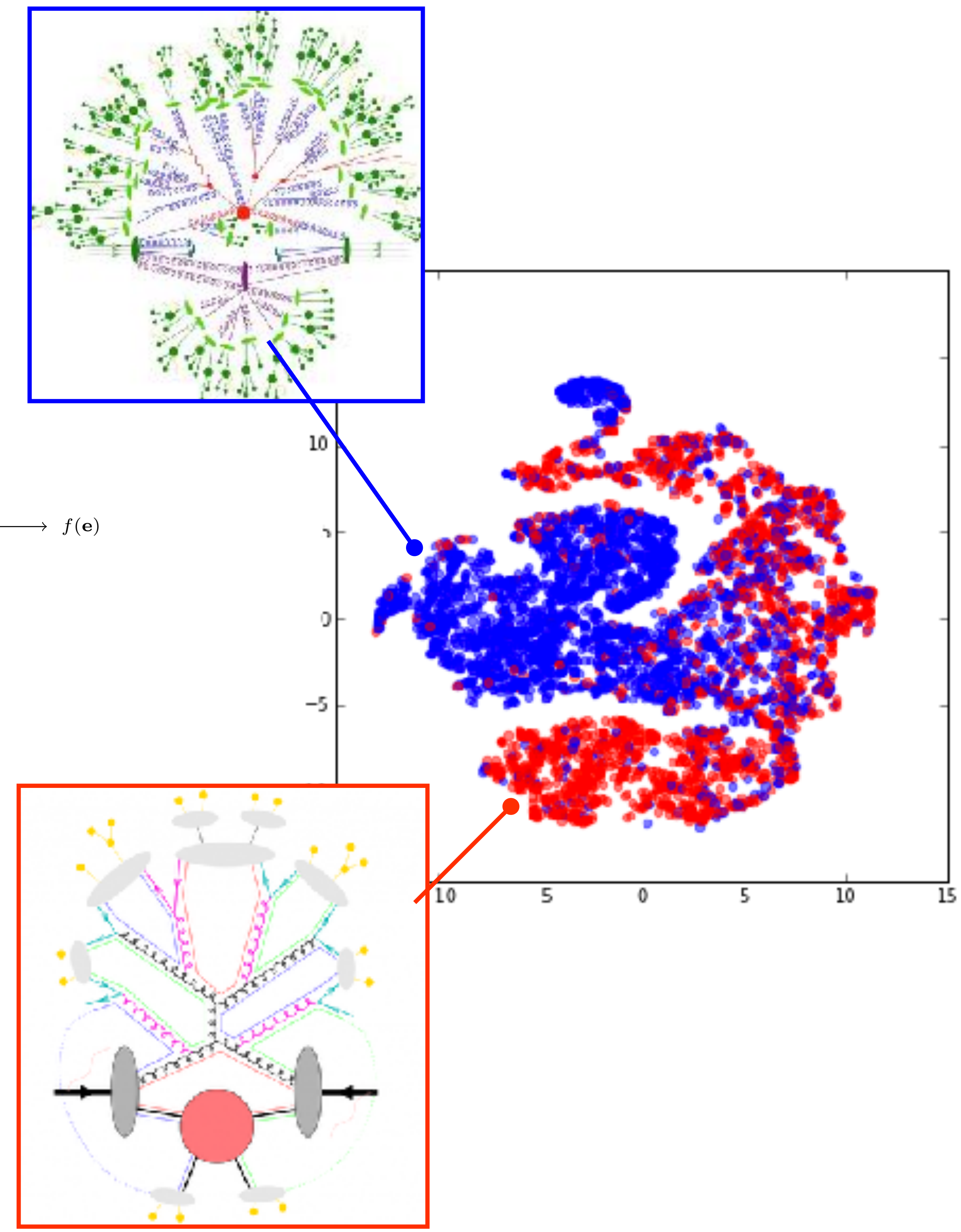
- Parton Shower is a tree-like, ~stationary Markov Process
- Neural network will leverage this structure
- Tree-RNN needs much less data to train!

EVENT EMBEDDINGS

Jointly optimize jet embedding → event embedding → classifier

It scales!





Neural Message Passing for Jet Physics

Isaac Henrion, Johann Brehmer, Joan Bruna, Kyunghyun Cho, Kyle Cranmer, Gilles Louppe, Gaspar Rochette

Courant Institute & Center for Data Science

NYU

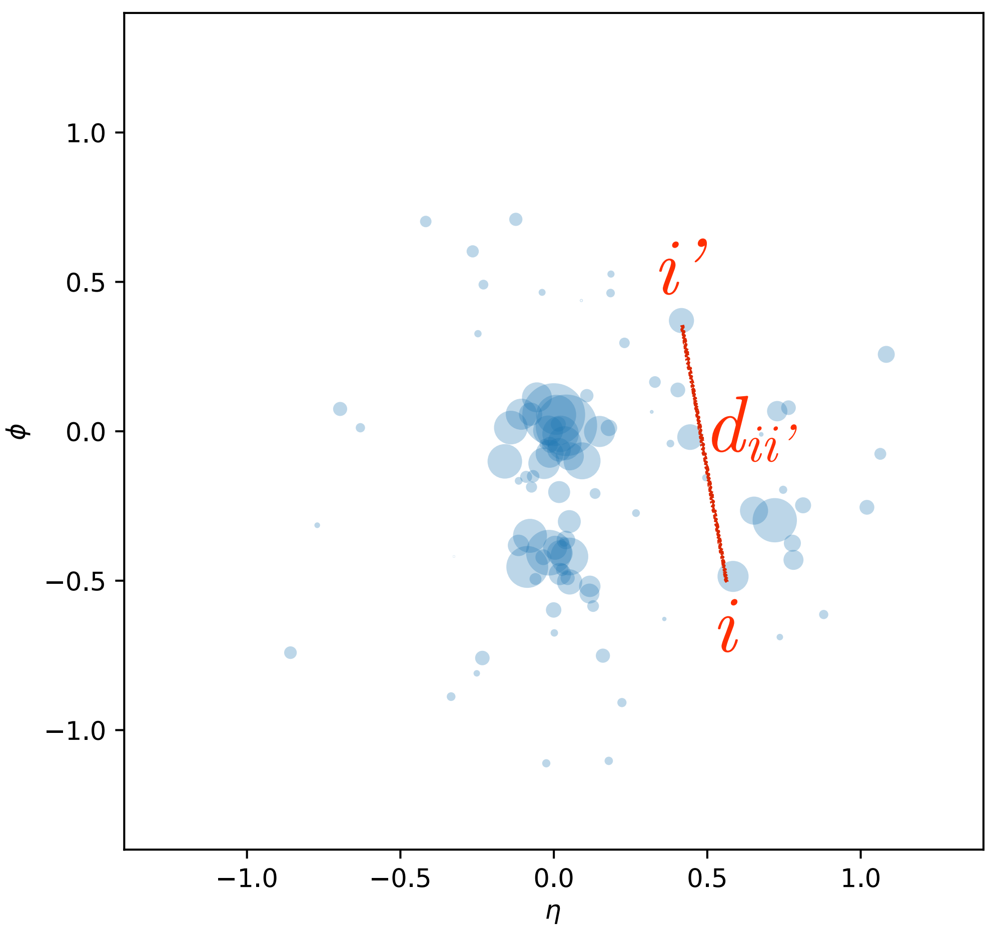
Isaac Henrion

Paper: https://dl4physicalsciences.github.io/files/nips_dlps_2017_29.pdf
Talk: https://dl4physicalsciences.github.io/files/nips_dlps_2017_slides_henrion.pdf

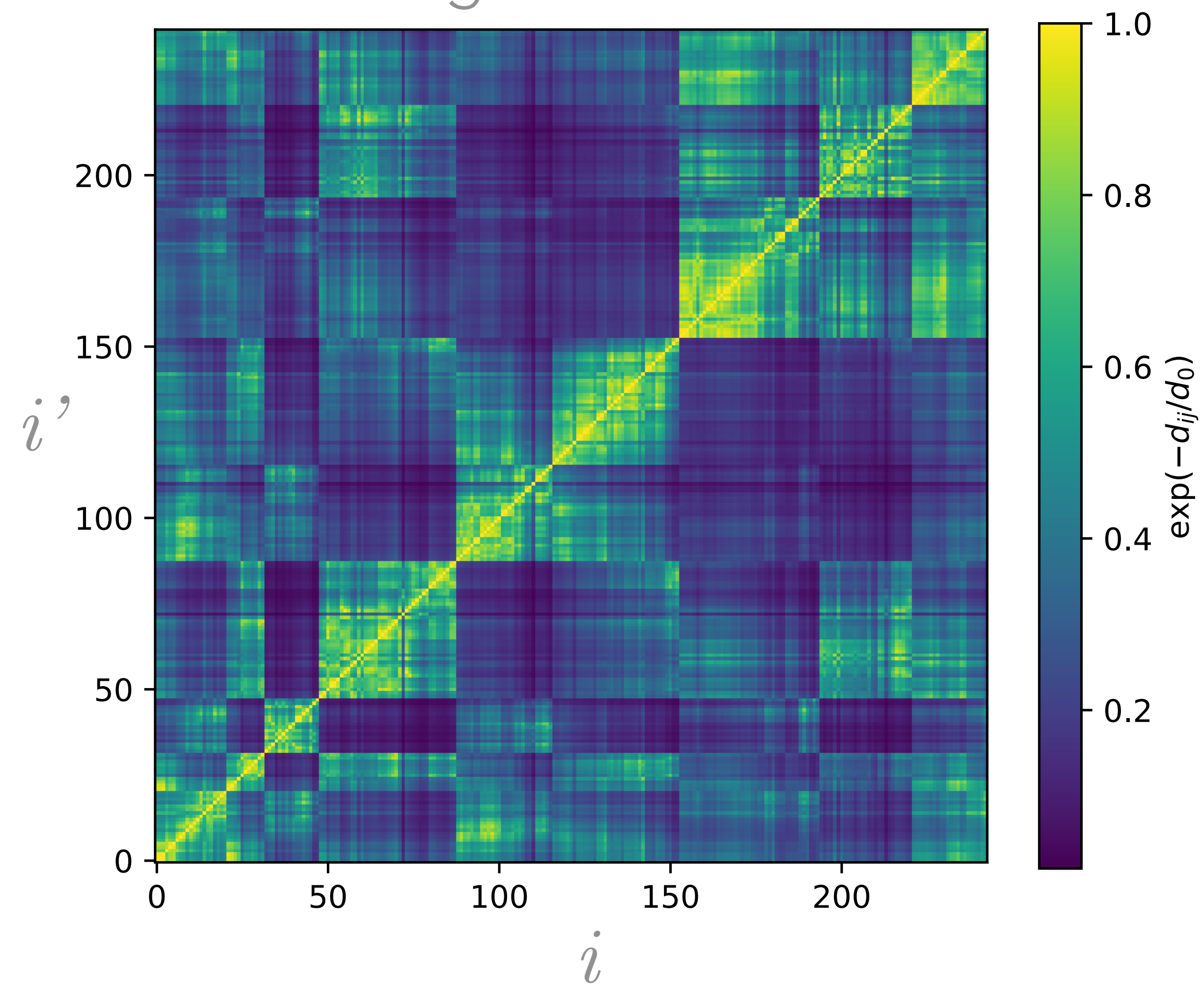
JETS AS A GRAPH

Using message passing neural networks over a fully connected graph on the particles

- Two approaches for adjacency matrix for edges
  - **import** physics knowledge by using metric of jet algorithms $d^{\alpha}_{ii'} = \min(p^{2\alpha}_{ti}, p^{2\alpha}_{ti'}) \frac{\Delta R^2_{ii'}}{R^2}$
  - learn adjacency matrix and **export** new jet algorithm



C/A algorithm with α=0



QCD Jet rejection @ 50% W-jet tagging efficiency

Model	Iterations	$R_{\epsilon=50\%}$
Rec-NN (no gating)	1	70.4 ± 3.6
Rec-NN (gating)	1	$\mathbf{83.3 \pm 3.1}$
MPNN (directed)	1	89.4 ± 3.5
MPNN (directed)	2	$\mathbf{98.3 \pm 4.3}$
MPNN (directed)	3	85.9 ± 8.5
MPNN (identity)	3	74.5 ± 5.2
Relation Network	1	67.7 ± 6.8

Significant improvement on W vs. QCD jet classification!

This is with a learned adjacency matrix

- what did it learn? Is that adjacency matrix useful?
- we are working MPNN with QCD-motivated adjacency matrix

PHYSICS-AWARE MACHINE LEARNING

We can **inject** our knowledge of physics into the machine learning models!

We can **extract** knowledge learned from the data!

Physics-aware Gaussian Processes

arXiv:1709.05681

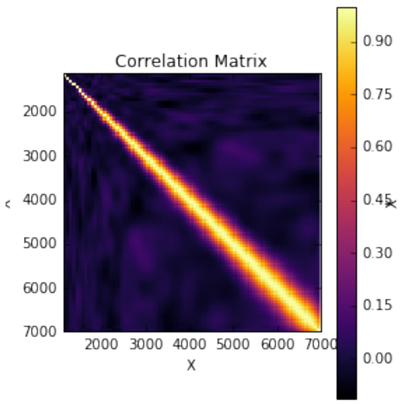
Final Kernel =

Poisson fluctuations

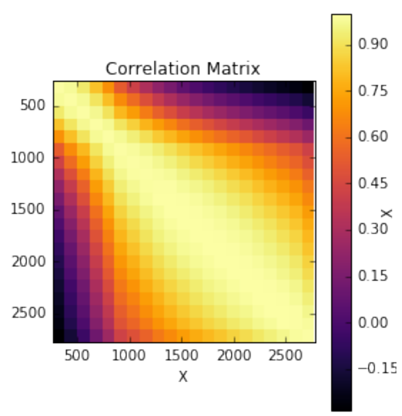
+ Mass Resolution

+ Parton Density Functions

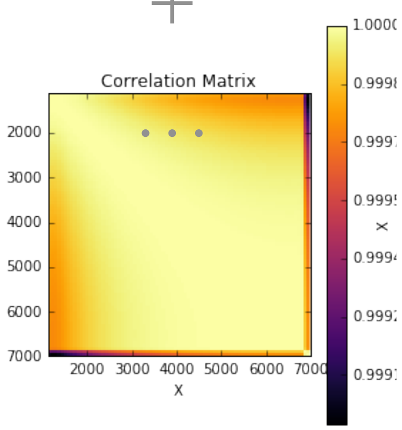
+ Jet Energy Scale



=



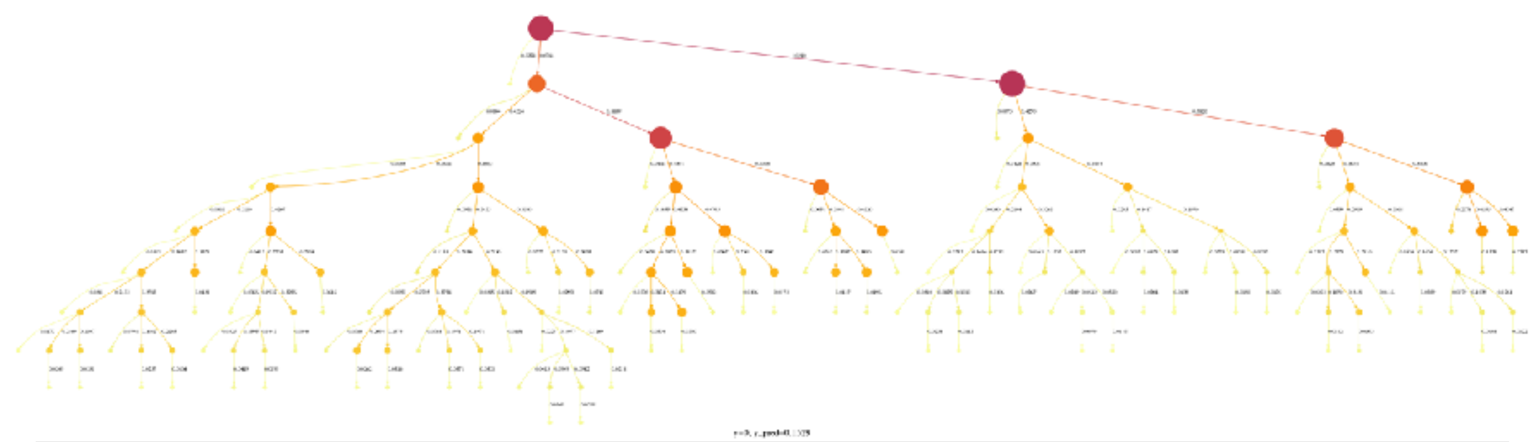
+



+

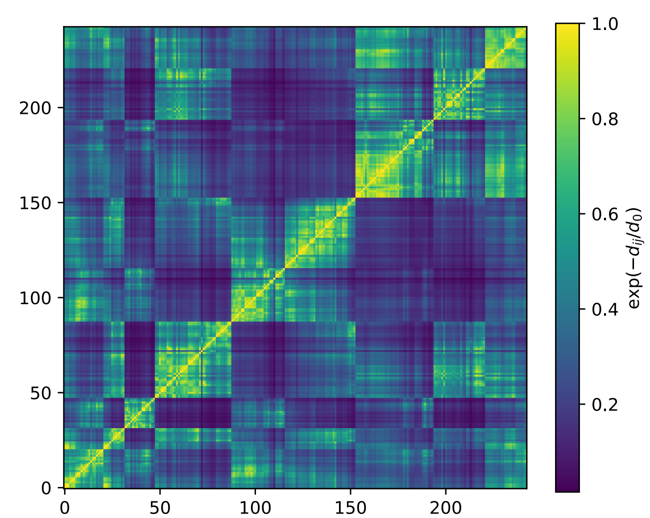
QCD-Aware recursive neural networks

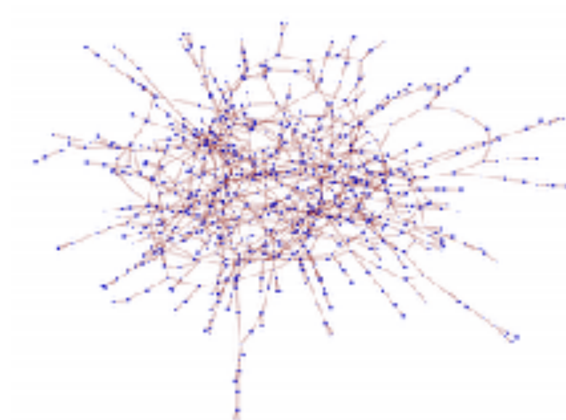
arXiv:1702.00748



QCD-Aware graph convolutional neural networks

NIPS2017 workshop [http://bit.ly/2AkwYRG]





$$d^{\alpha}_{ii'} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha}) \frac{\Delta R^2_{ii'}}{R^2}$$

THE PLAYERS

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

$p(x, z \mid \theta, \nu)$

z
latent variables
Monte Carlo truth

x
observed data
simulated data

ν
nuisance parameters

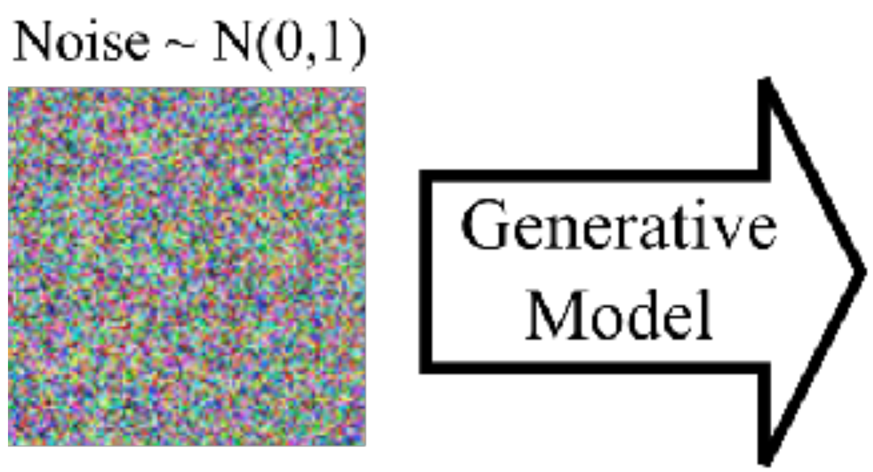
INFERENCE

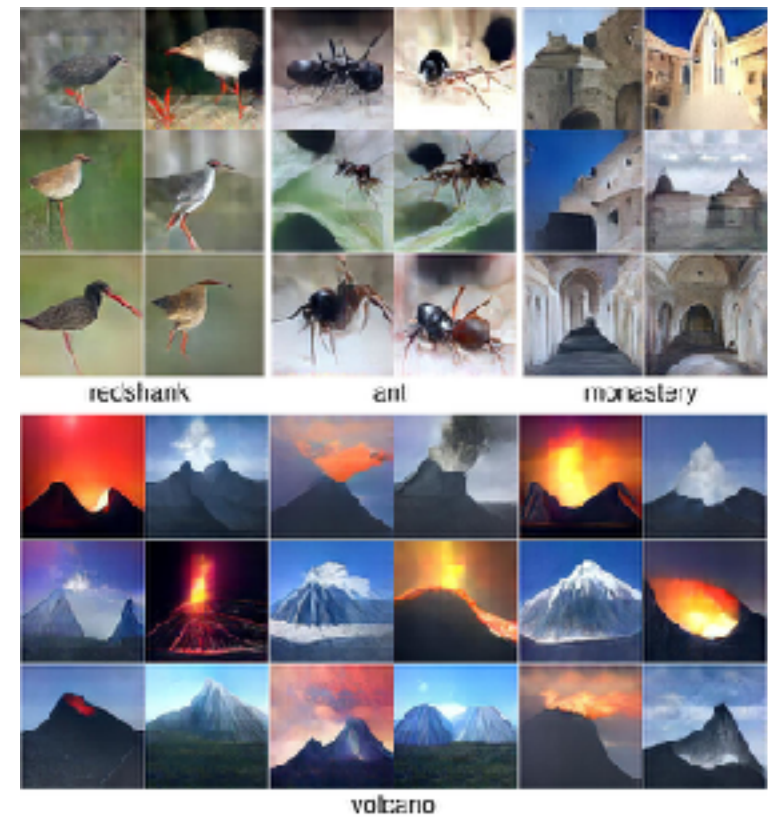
inverse problem
measurement
parameter estimation

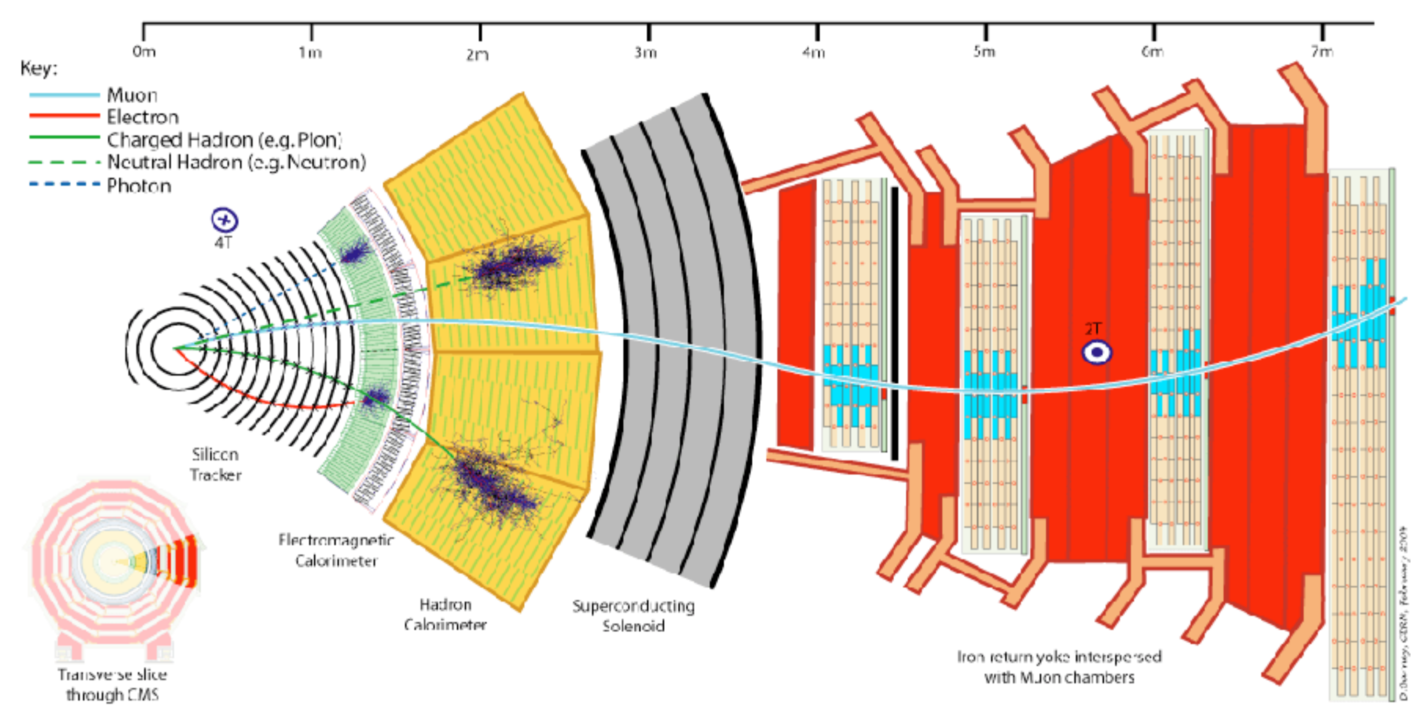
LEARNING THE GENERATIVE MODEL

Z

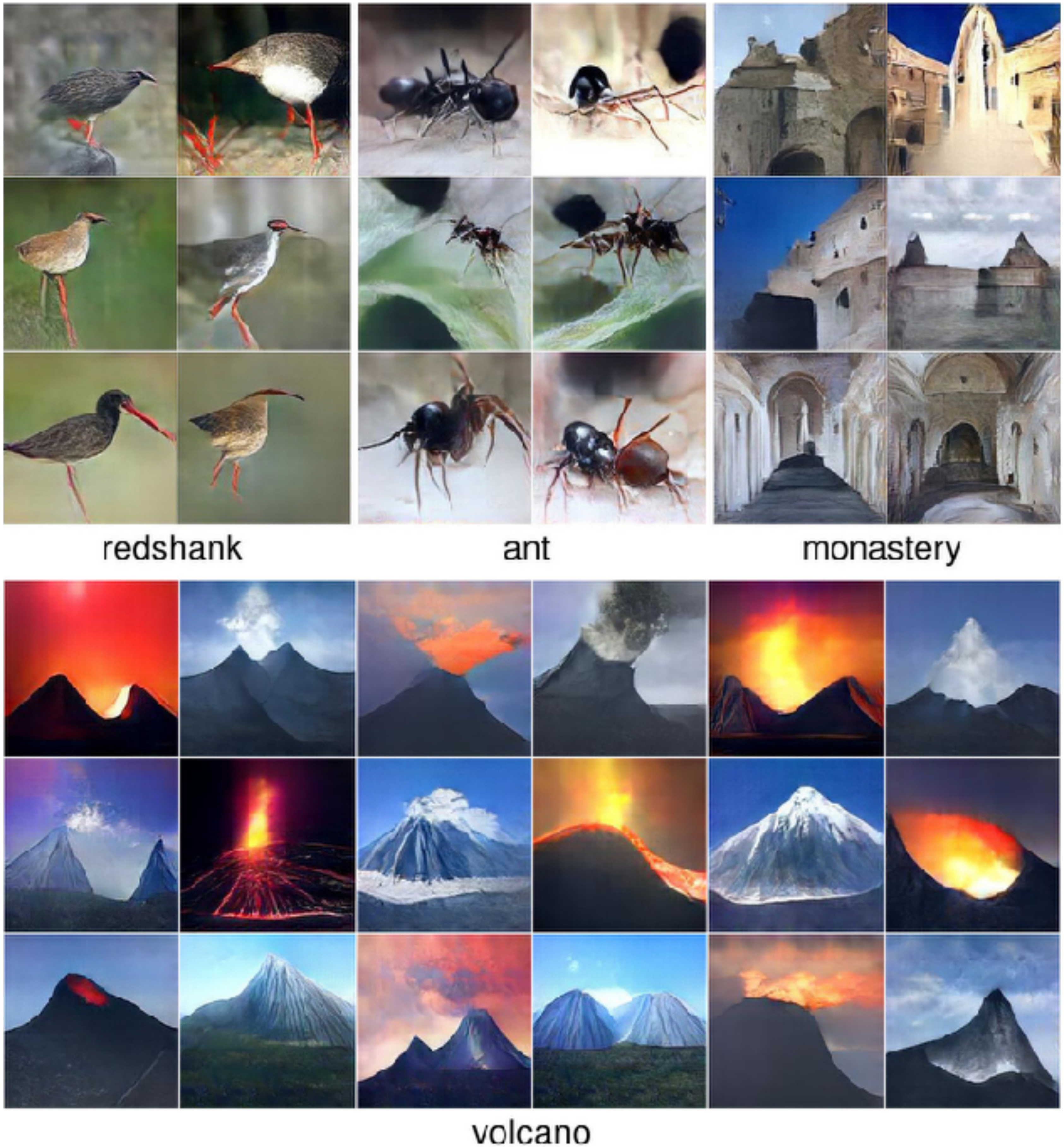
X







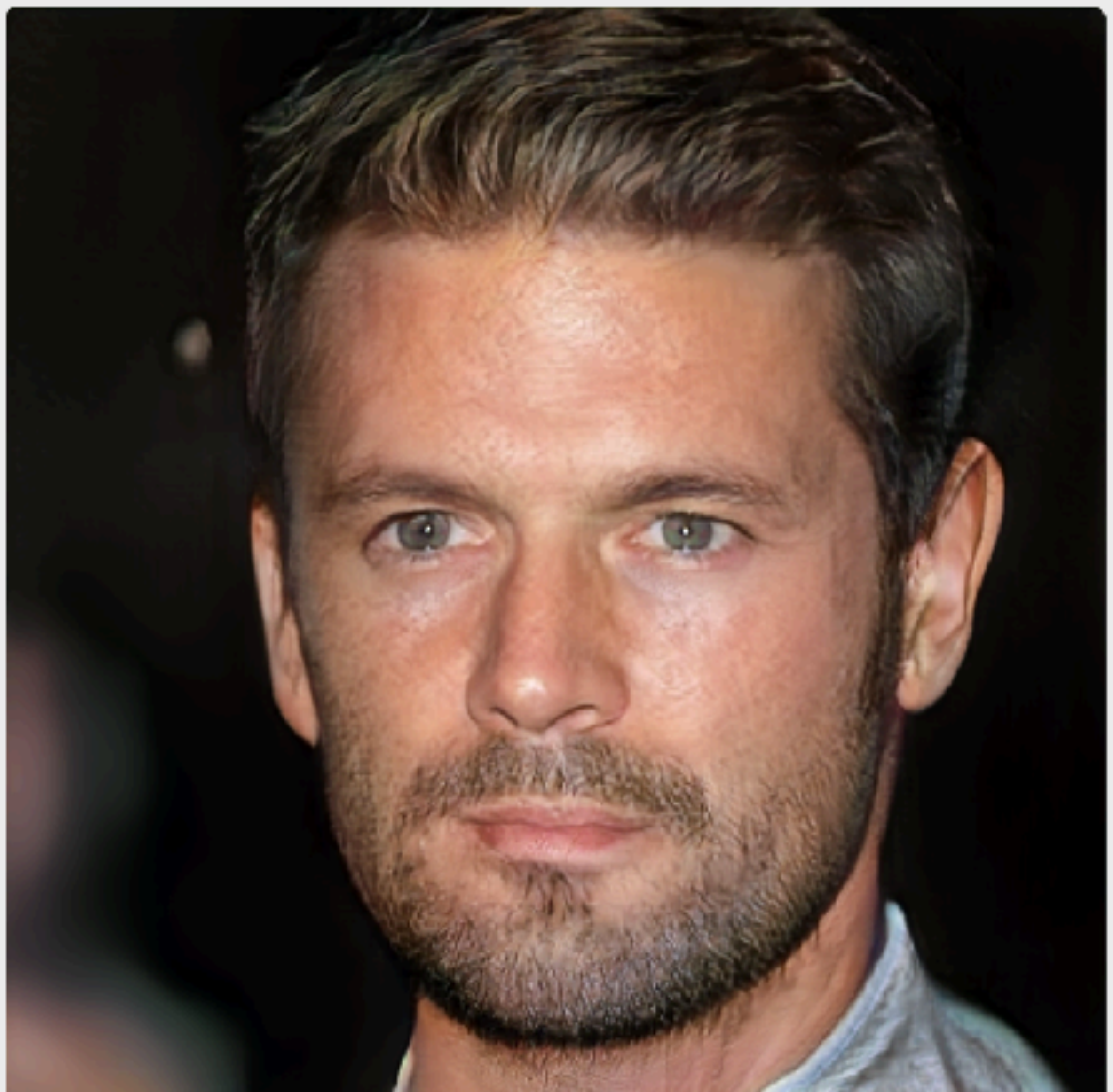
GENERATIVE MODEL FOR IMAGES



GENERATIVE MODEL FOR IMAGES

How an A.I. 'Cat-and-Mouse Game' Generates Believable Fake Photos

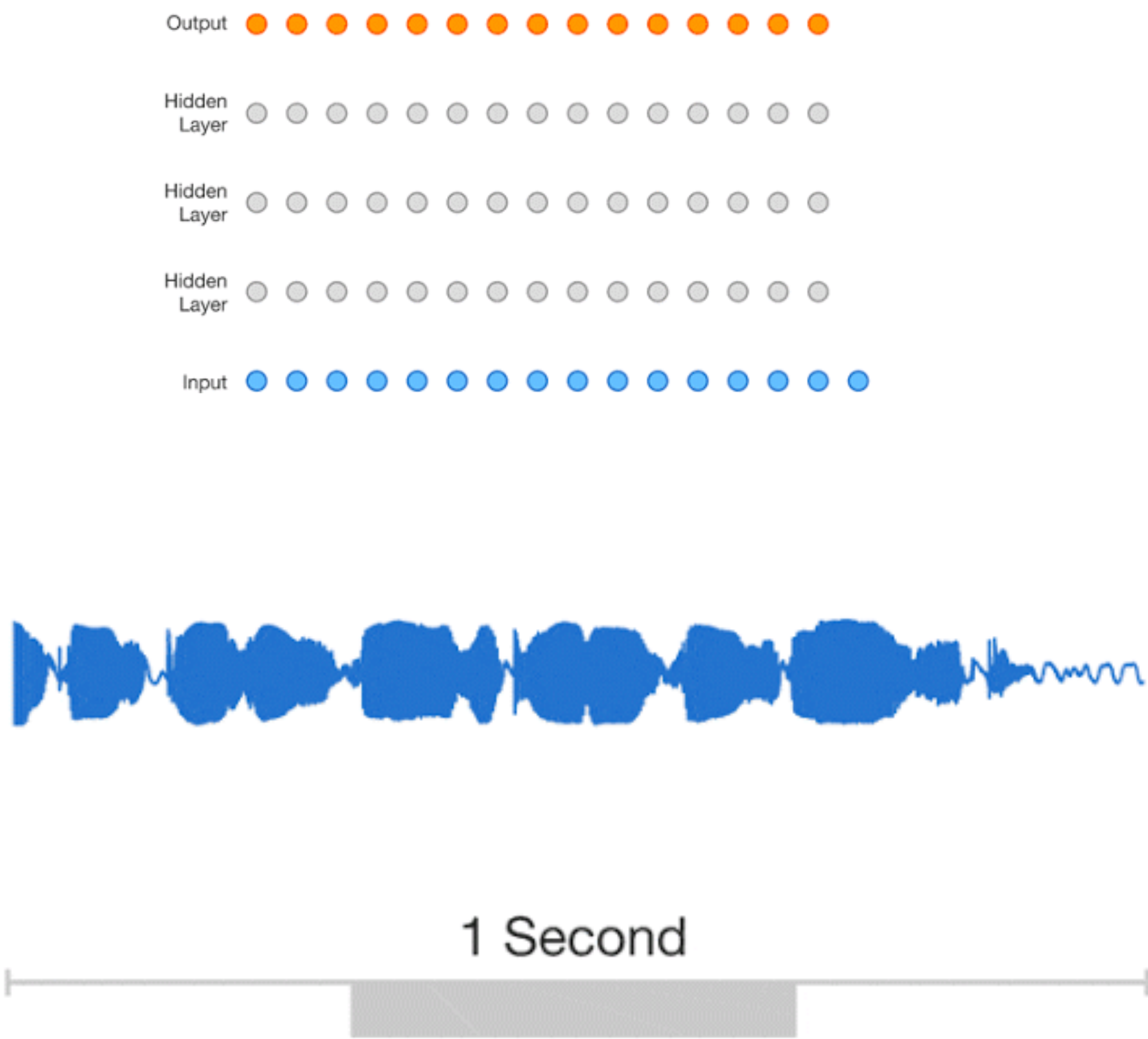
By CADE METZ and KEITH COLLINS JAN. 2, 2018



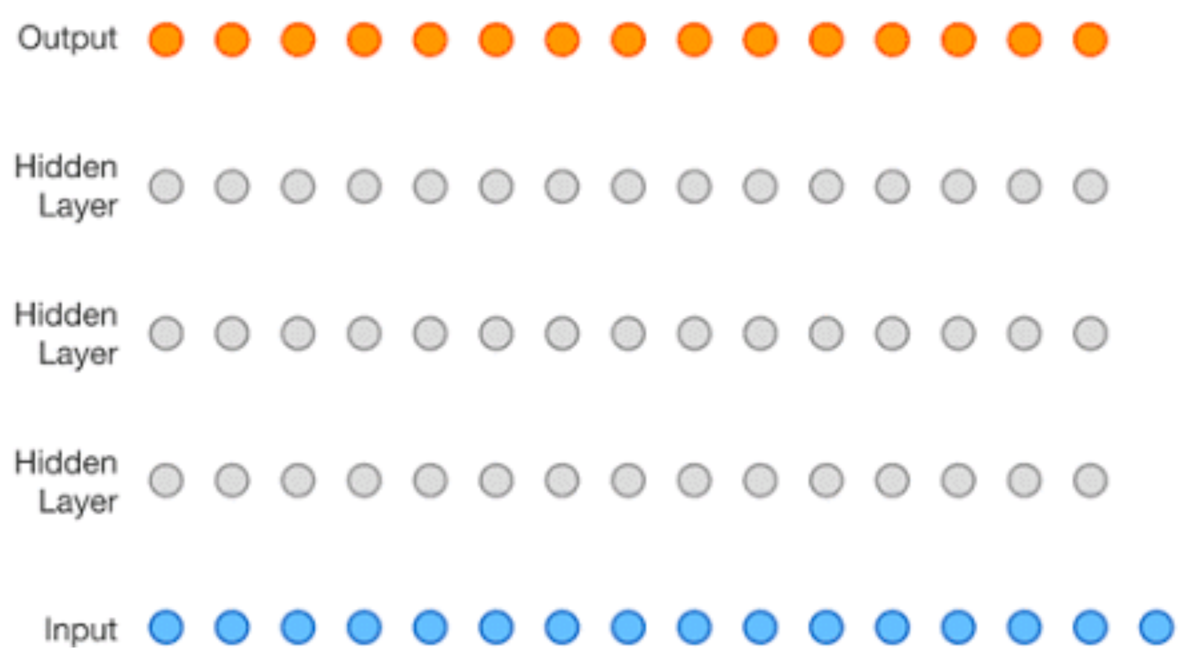
This one is computer-generated

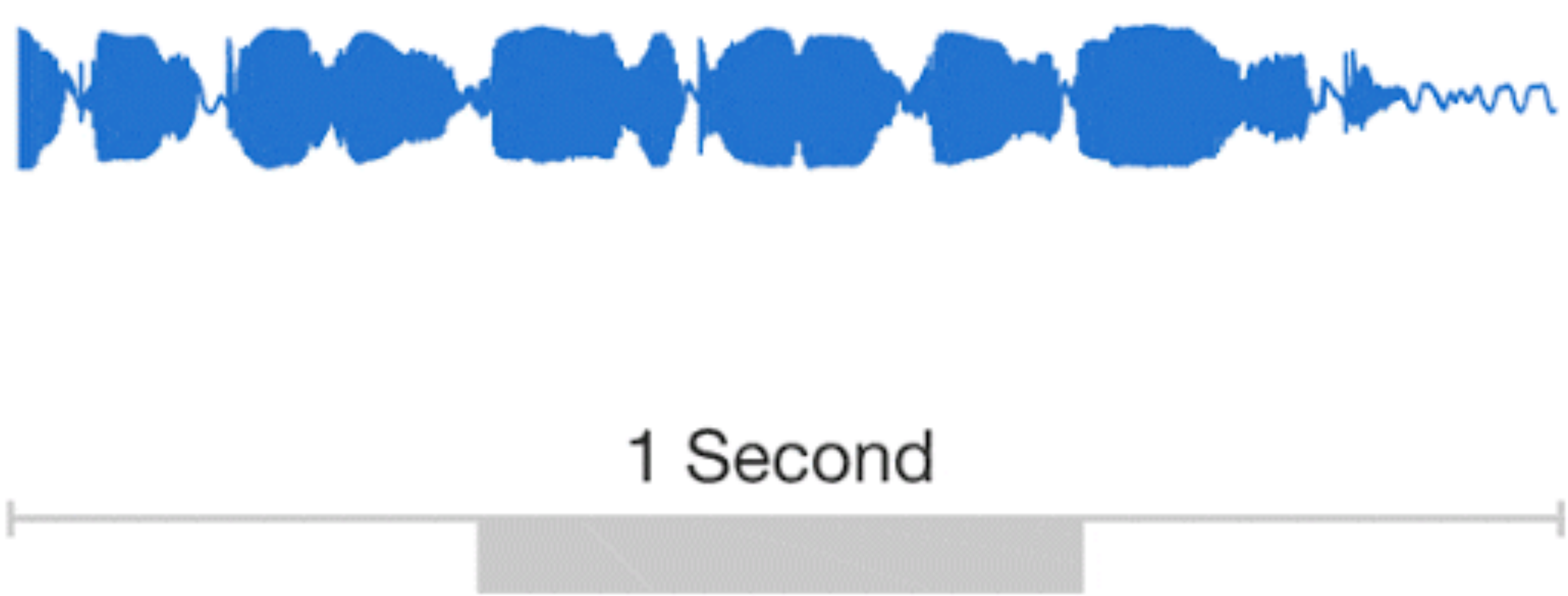
This one is also computer-generated

WAVENET: A GENERATIVE MODEL FOR RAW AUDIO

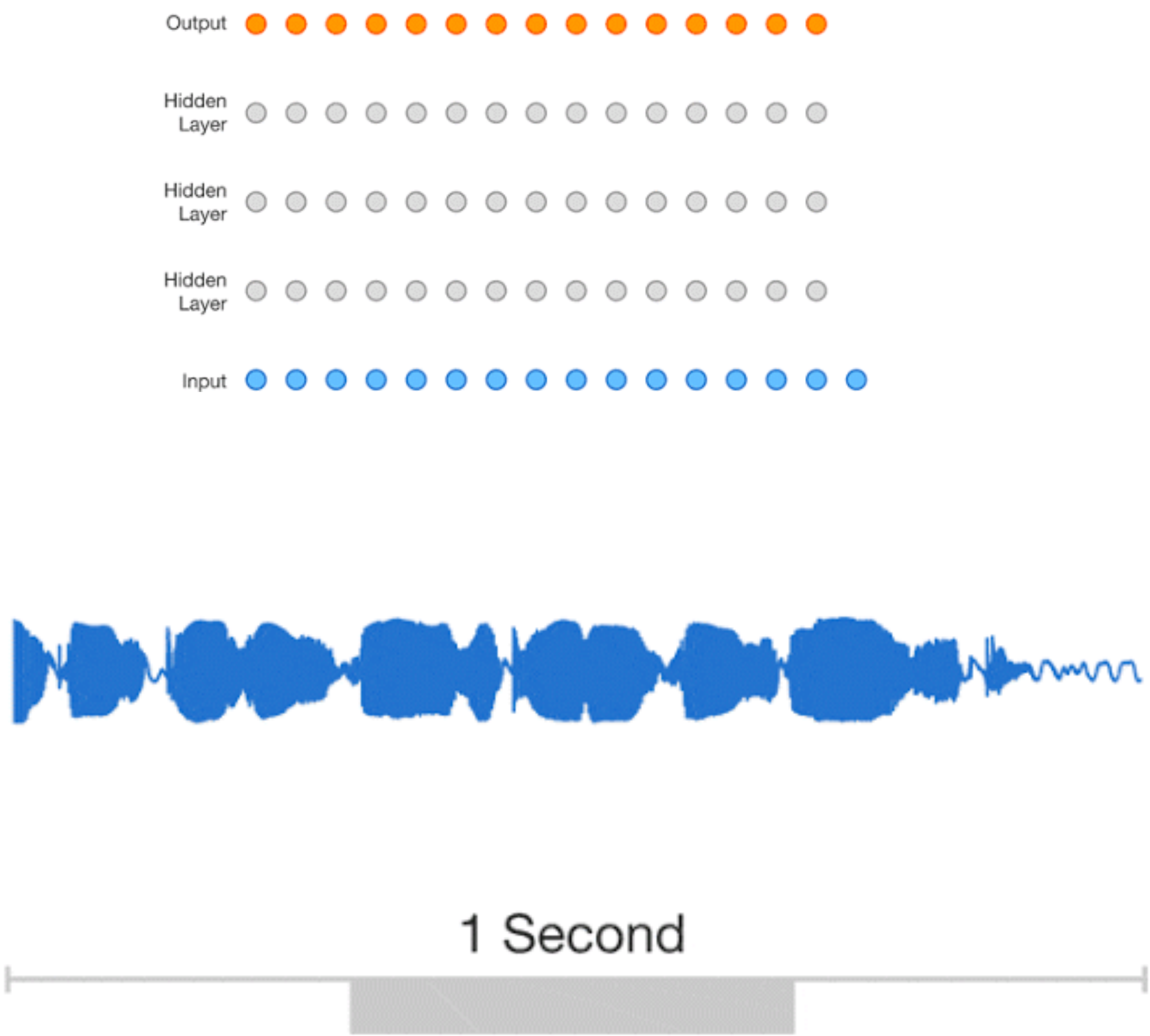


WAVENET: A GENERATIVE MODEL FOR RAW AUDIO





WAVENET: A GENERATIVE MODEL FOR RAW AUDIO



NEW! AVO

G. Louppe & K.C. arXiv:1707.07113

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe[1] and **Kyle Cranmer**[1]

[1] *New York University*

Complex computer simulators are increasingly used across fields of science as generative models tying parameters of an underlying theory to experimental observations. Inference in this setup is often difficult, as simulators rarely admit a tractable density or likelihood function. We introduce Adversarial Variational Optimization (AVO), a likelihood-free inference algorithm for fitting a non-differentiable generative model incorporating ideas from empirical Bayes and variational inference. We adapt the training procedure of generative adversarial networks by replacing the differentiable generative network with a domain-specific simulator. We solve the resulting non-differentiable mini-max problem by minimizing variational upper bounds of the two adversarial objectives. Effectively, the procedure results in learning a proposal distribution over simulator parameters, such that the corresponding marginal distribution of the generated data matches the observations. We present results of the method with simulators producing both discrete and continuous data.

Leo is G Tom is D

Similar to GAN setup, but instead of using a neural network as the generator, use the actual simulation (eg. Pythia, GEANT)

Continue to use a neural network discriminator / critic.

Difficulty: the simulator isn't differentiable, but there's a **trick**!

Allows us to efficiently fit / **tune simulation** with stochastic gradient techniques!

GANs FOR PHYSICS

CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

Michela Paganini[a,b], **Luke de Oliveira**[a], **and Benjamin Nachman**[a]

[a] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA, 94720, USA*
[b] *Department of Physics, Yale University, New Haven, CT 06520, USA*

E-mail: michela.paganini@yale.edu, lukedeoliveira@lbl.gov, bnachman@cern.ch

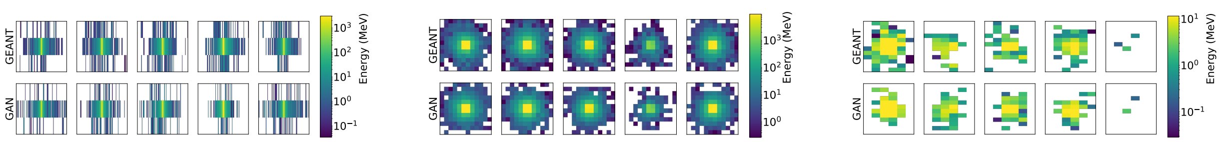


Figure 9: Five randomly selected e^+ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CaloGAN candidates.

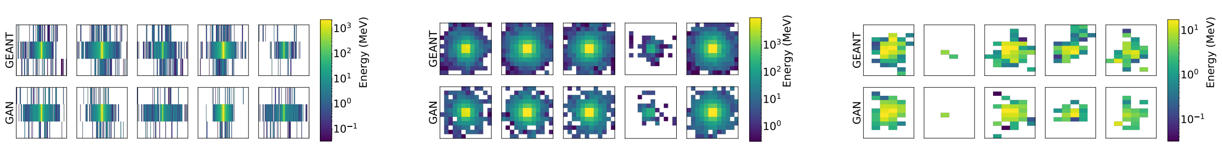


Figure 10: Five randomly selected γ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CaloGAN candidates.

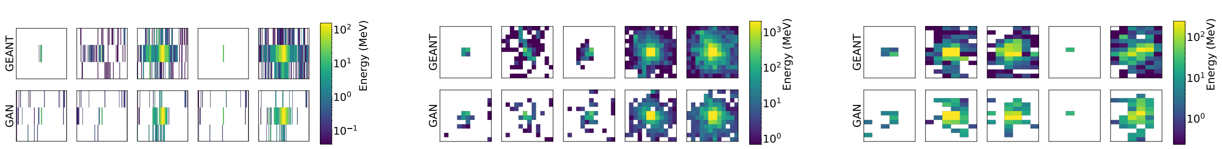
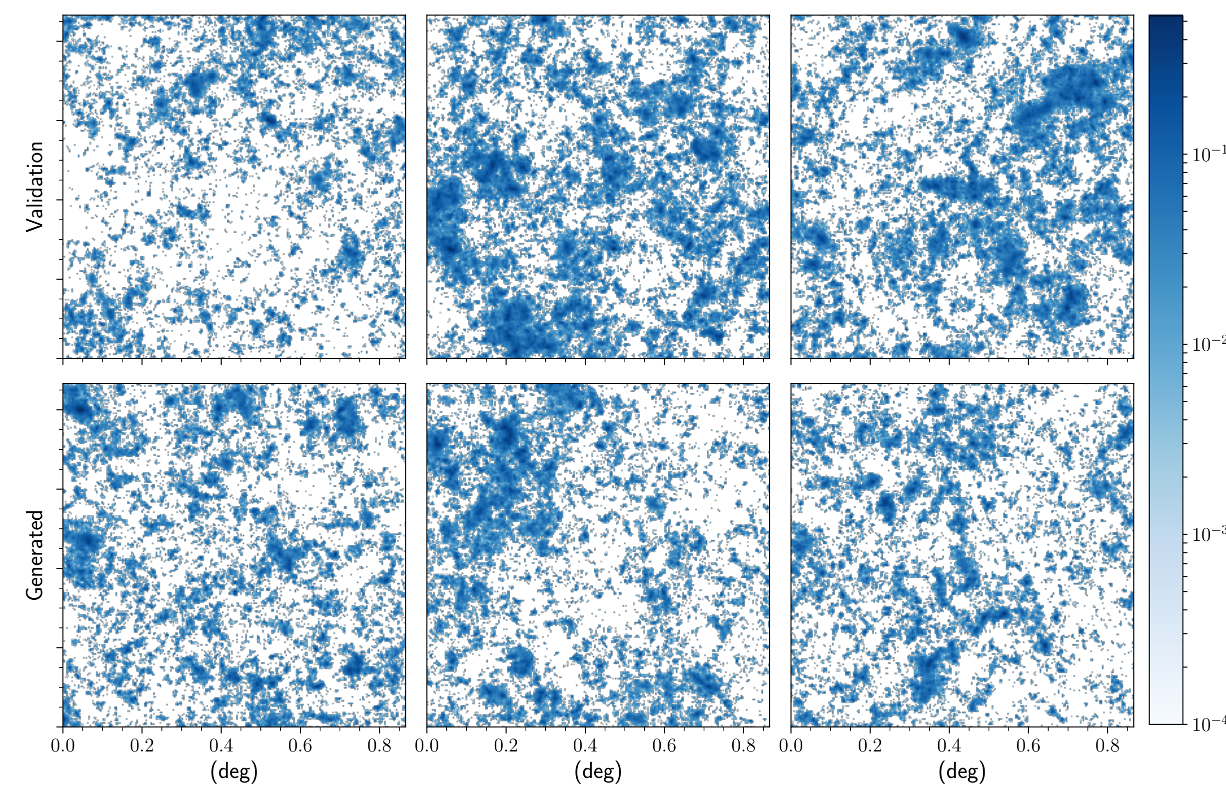


Figure 11: Five randomly selected π^+ showers per calorimeter layer from the training set (top) and the five nearest neighbors (by euclidean distance) from a set of CaloGAN candidates.

Creating Virtual Universes Using Generative Adversarial Networks

Mustafa Mustafa*[1], Deborah Bard[1], Wahid Bhimji[1], Rami Al-Rfou[2], and Zarija Lukić[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA 94720
[2]Google Research, Mountain View, CA 94043



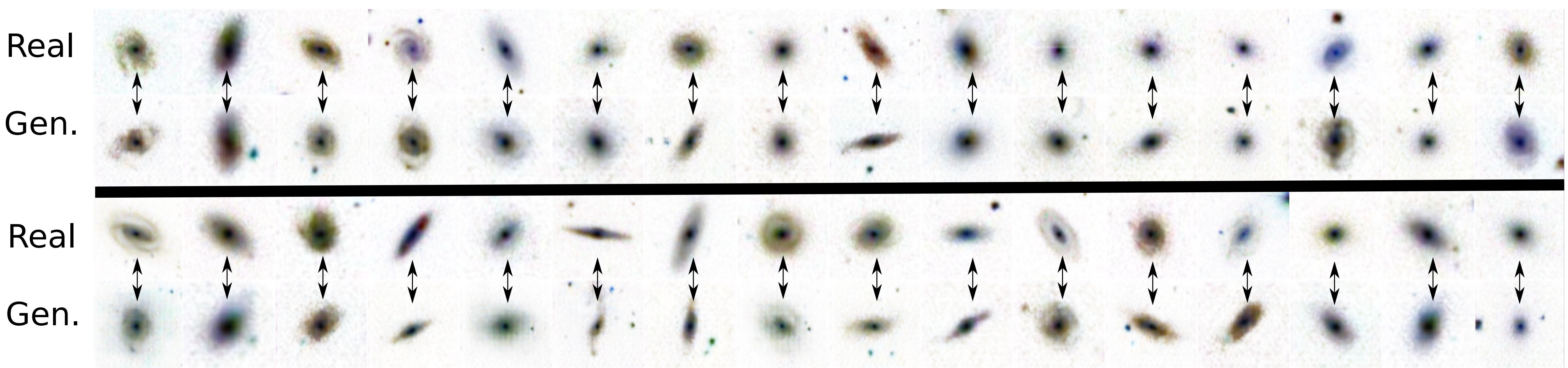
Use of generative models of galaxy images to help calibrate down-stream analysis in next-generation surveys.

Enabling Dark Energy Science with Deep Generative Models of Galaxy Images

Siamak Ravanbakhsh[1], François Lanusse[2], Rachel Mandelbaum[2], Jeff Schneider[1], and Barnabás Póczos[1]

[1]School of Computer Science, Carnegie Mellon University
[2]McWilliams Center for Cosmology, Carnegie Mellon University

***Abstract*—Understanding the nature of dark energy, the mysterious force driving the accelerated expansion of the Universe, is a major challenge of modern cosmology. The next generation of cosmological surveys, specifically designed to address this issue, rely on accurate measurements of the apparent shapes of distant galaxies. However, shape measurement methods suffer from various unavoidable biases and therefore will rely on a precise calibration to meet the accuracy requirements of the science analysis. This calibration process remains an open challenge as it requires large sets of high quality galaxy images. To this end, we study the application of deep conditional generative models in generating realistic galaxy images. In particular we consider variations on conditional variational autoencoder and introduce a new adversarial objective for training of conditional generative networks. Our results suggest a reliable alternative to the acquisition of expensive high quality observations for generating the calibration data needed by the next generation of cosmological surveys.**



CONCLUSIONS

While "Big Data" is relevant for the computing aspects of HEP, we have not had much of an impact on industry or taken much advantage of their developments

In contrast, recent developments in machine learning and AI are closely aligned with our "Big Decisions" when formulated in a statistical language

- more opportunities for HEP community to collaborate with ML community
- likelihood-free inference and generative models are two particularly exciting areas for physics

Our understanding of how to leverage our prior physics knowledge while letting machine learning do what it's good at is maturing.

- ability to inject and extract physics knowledge from models

Harnessing the full potential of these techniques will require deep integration into our scientific workflow

COLLABORATORS


Gilles Louppe
U. Liège


Kyunghyun Cho


Joan Bruna


Brenden Lake


Meghan Frate


Juan Pavez


Tilman Plehn


Johann Brehmer


Isaac Henrion


Lukas Heinrich


Heiko Müller


Tim Head


Michael Kagan


David Rousseau


Peter Sadowski


Daniel Whiteson


Pierre Baldi


Lezcano Casado


Atılım Güneş Baydin
University of Oxford

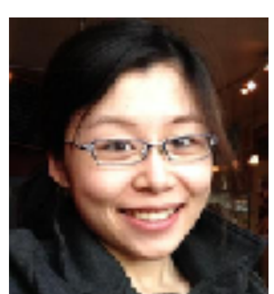

Prabhat
NERSC, Berkeley Lab


Wahid Bhimji
NERSC, Berkeley Lab

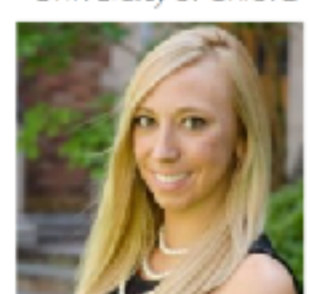

Frank Wood
University of Oxford

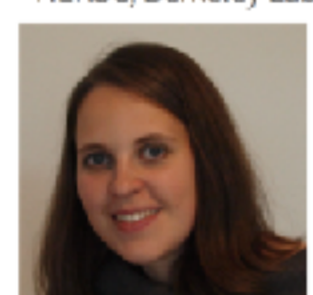

Phiala Shanahan

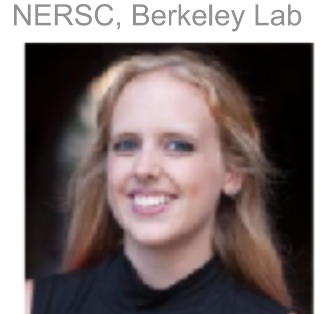

William Detmold


Karen Ng


Tuan Anh Le

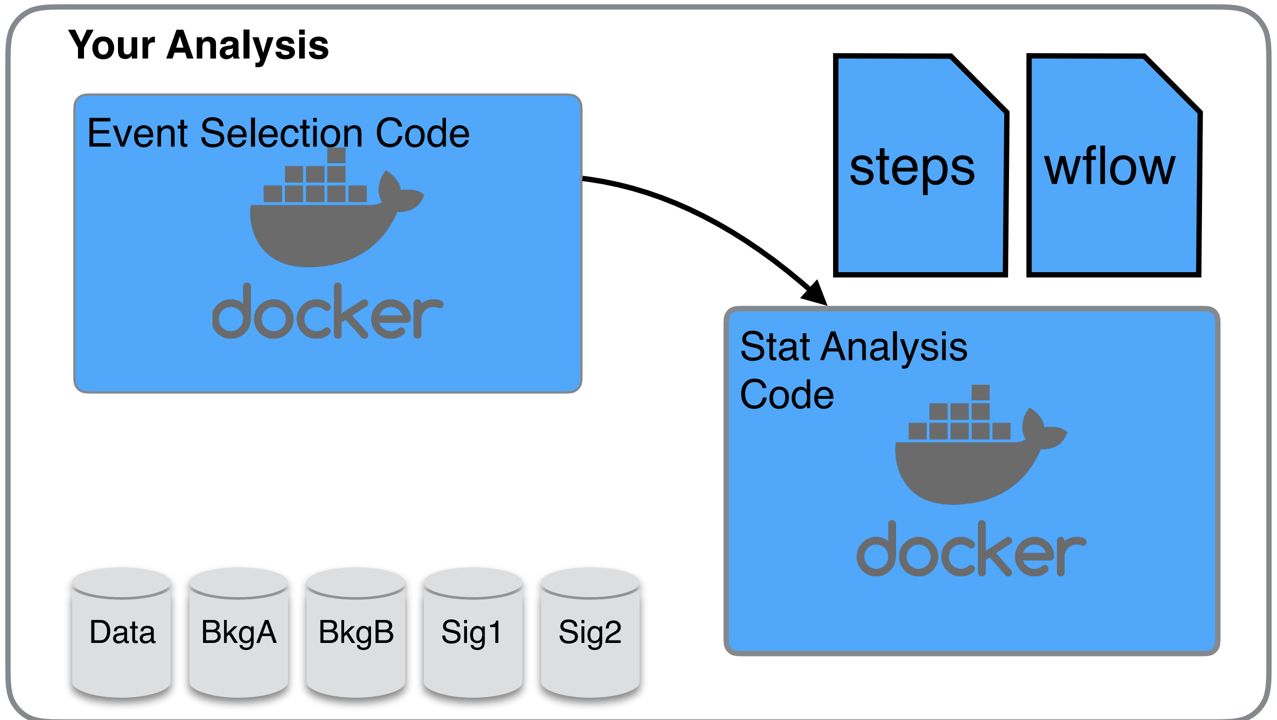

Michela Paganini
Yale University

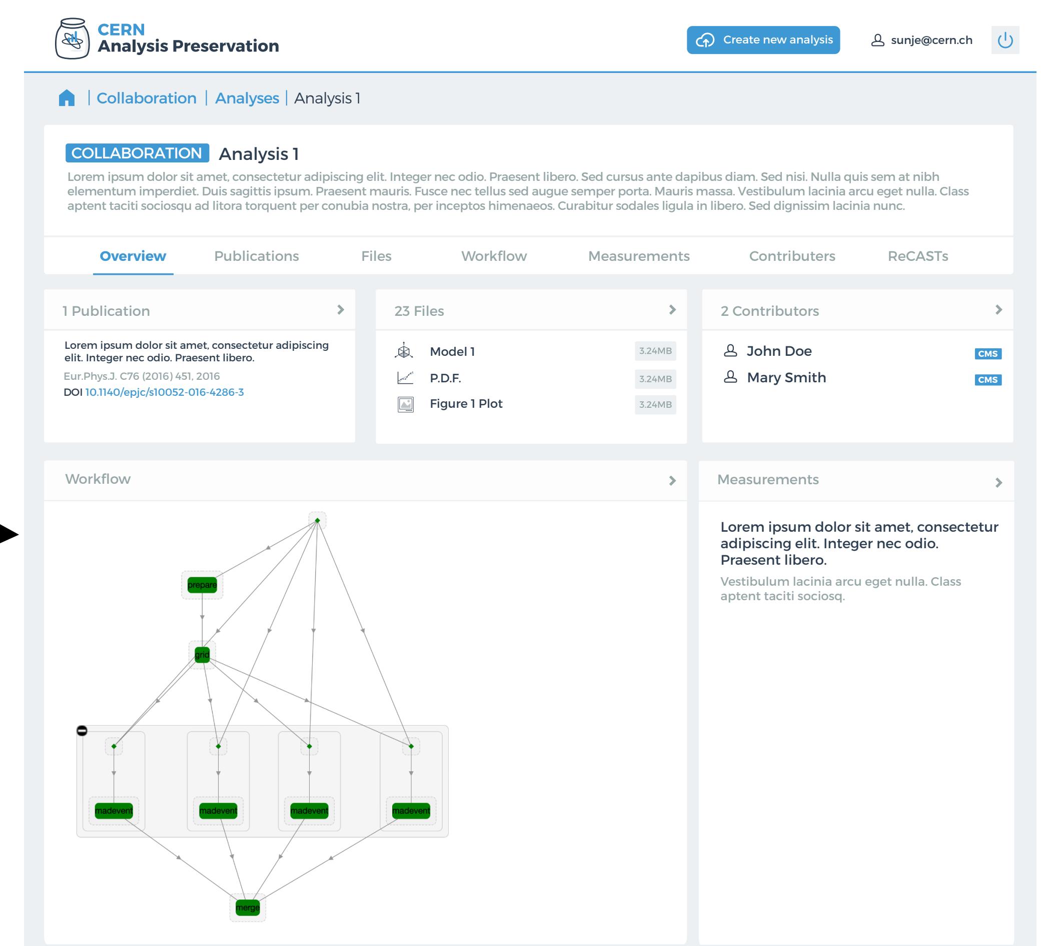

Daniela Huppenkothen
New York University


Savannah Thais
Yale University


Ruth Angus
Columbia University

Integration with CERN Analysis Preservation





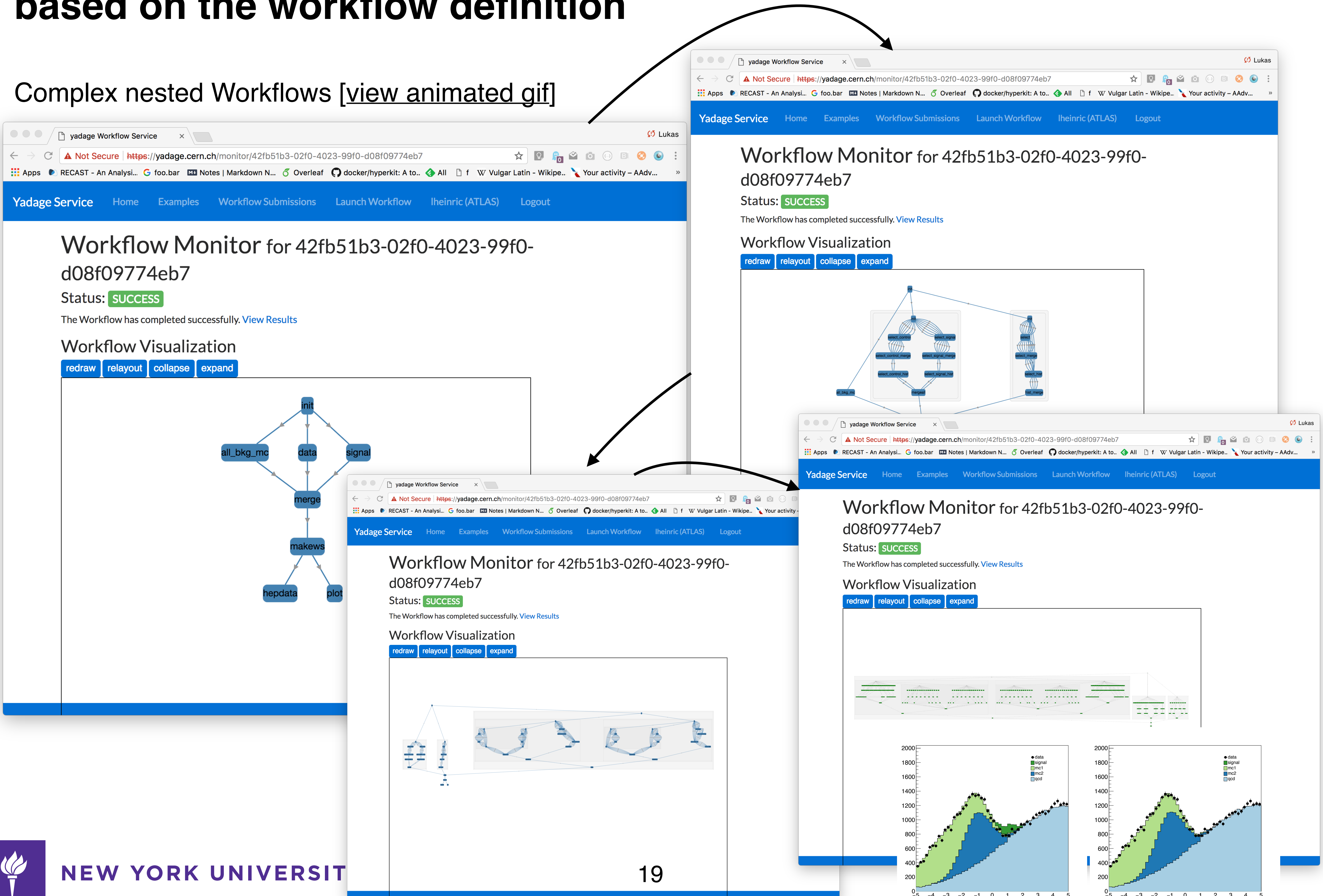
Automated import into CAP from your repos

- copy docker images
- copy workflow info
- copy data (if needed)

CERN Analysis Preservation

We provide infrastructure to re-run the analysis based on the workflow definition

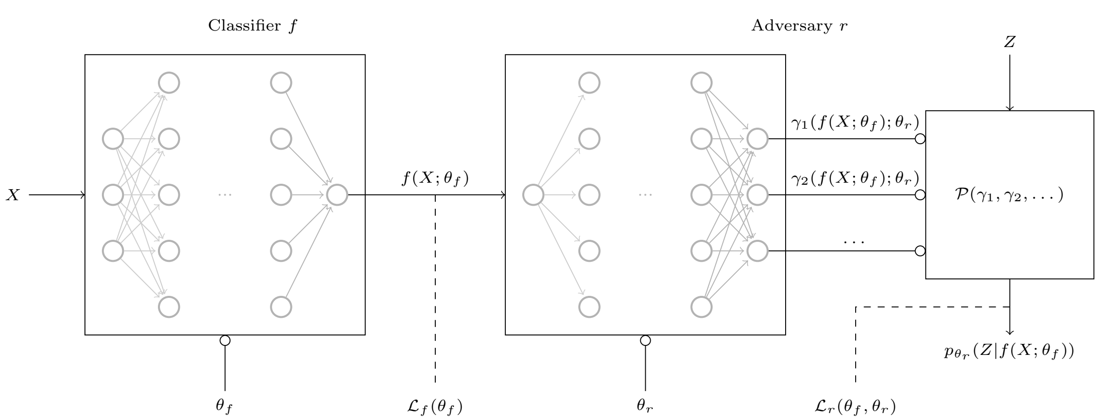
Complex nested Workflows [view animated gif]

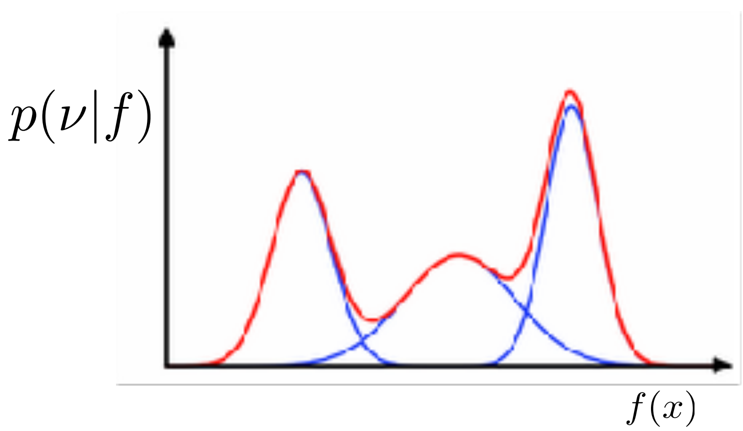


LEARNING TO PIVOT WITH ADVERSARIAL NETWORKS

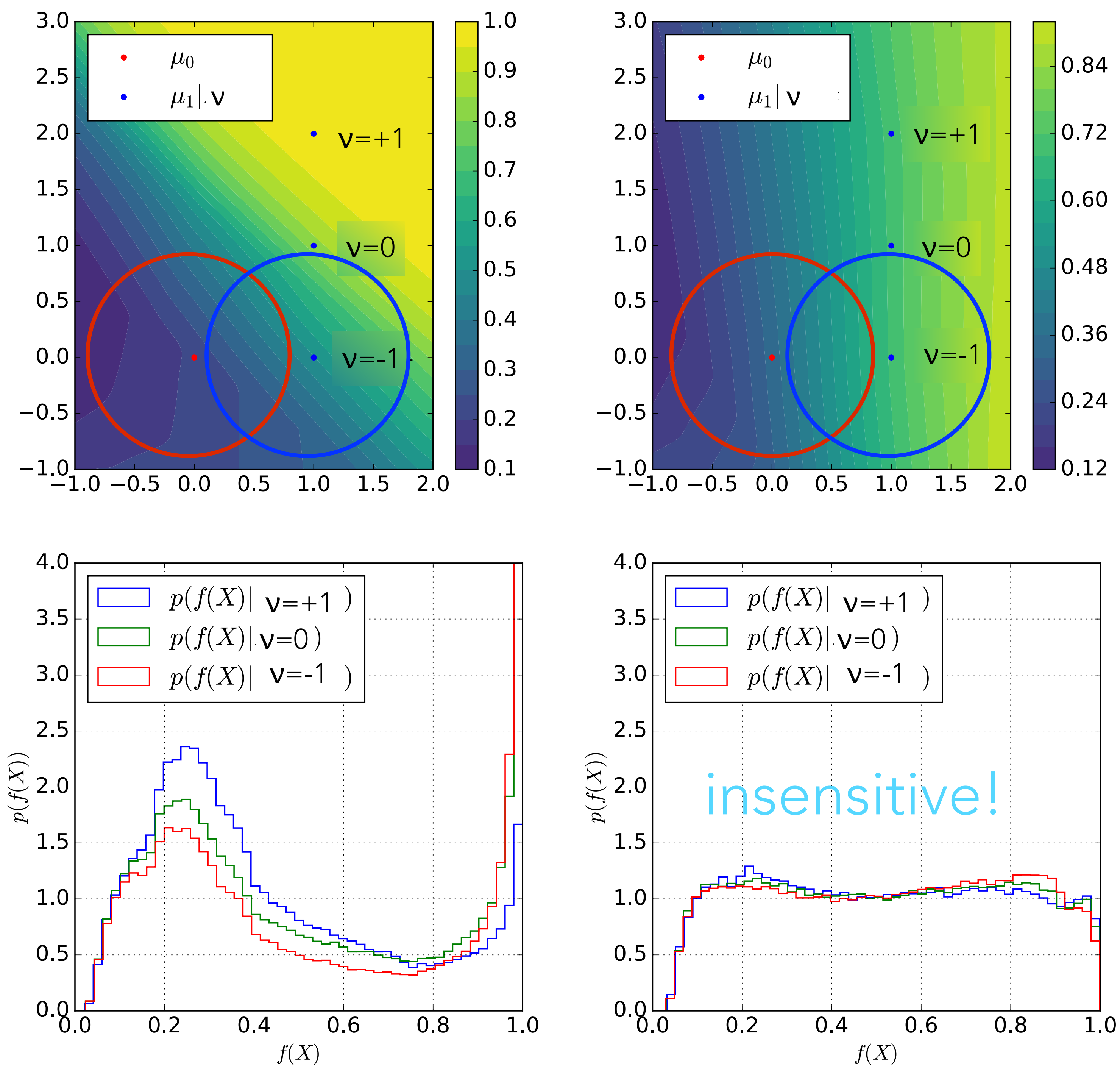
Typically classifier **f(x)** trained to minimize loss $\mathbf{L_f}$.

- want classifier output to be insensitive to systematics (nuisance parameter $\boldsymbol{\nu}$)
- introduce an **adversary r** that tries to predict ν based on f.
- setup as a minimax game:





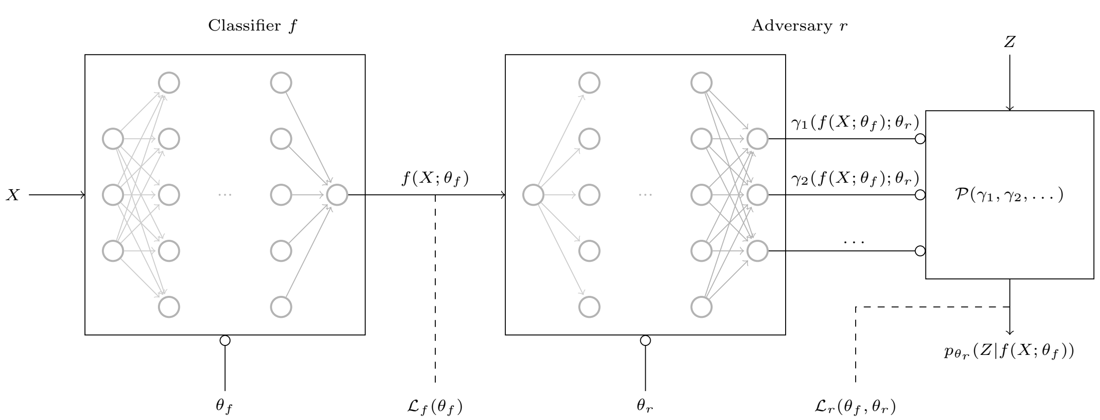
normal training | adversarial training

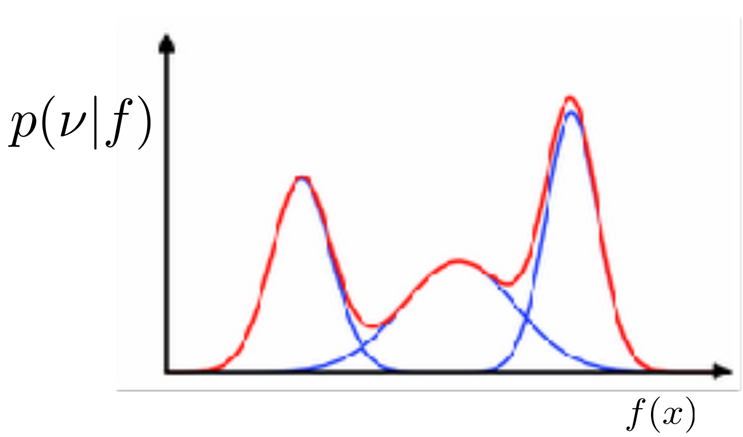


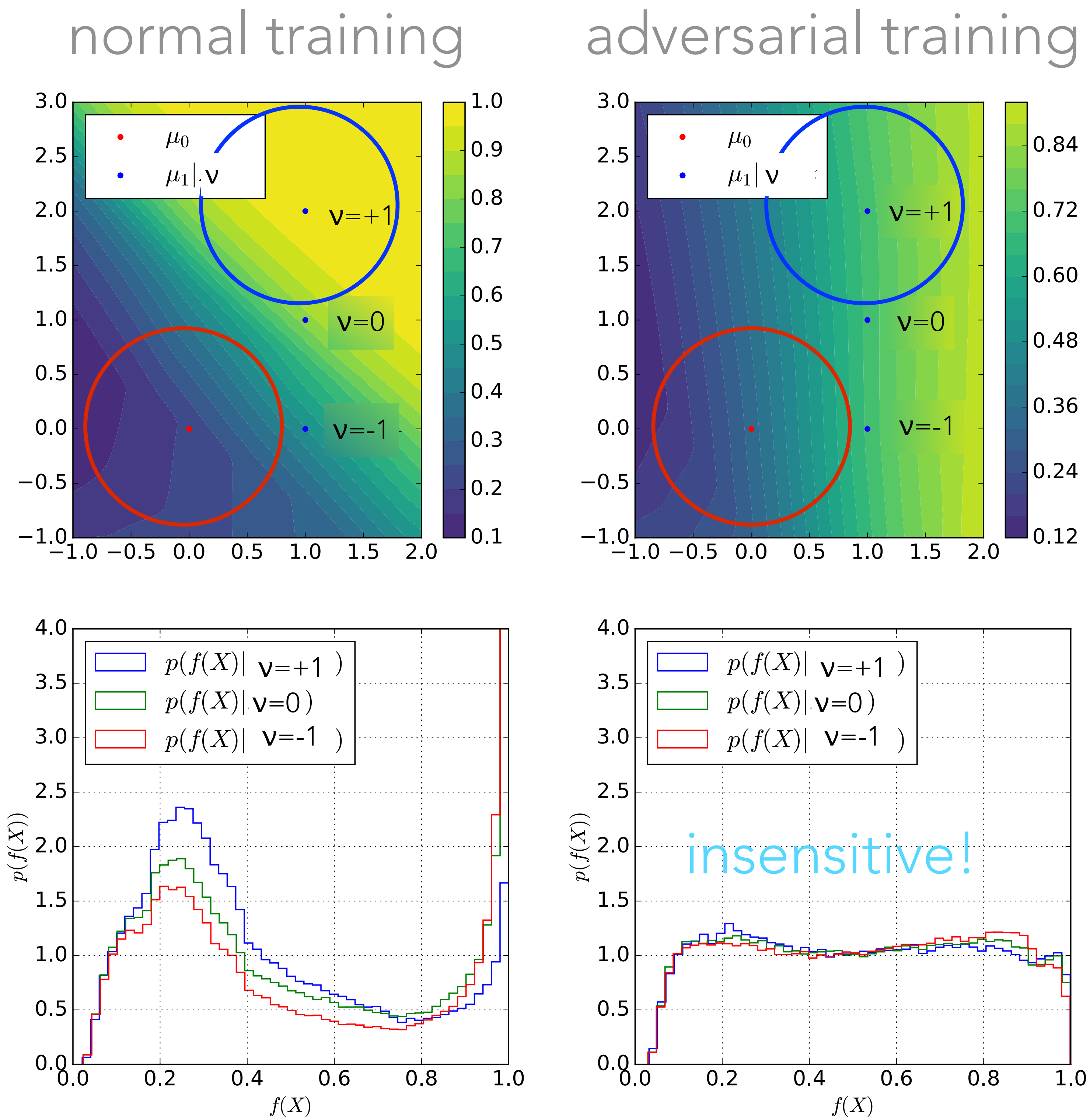
G. Louppe, M. Kagan, K. Cranmer, Learning to Pivot with Adversarial Networks [arXiv:1611.01046]

Typically classifier **f(x)** trained to minimize loss **L_f**.

- want classifier output to be insensitive to systematics (nuisance parameter **ν**)
- introduce an **adversary r** that tries to predict ν based on f.
- setup as a minimax game:

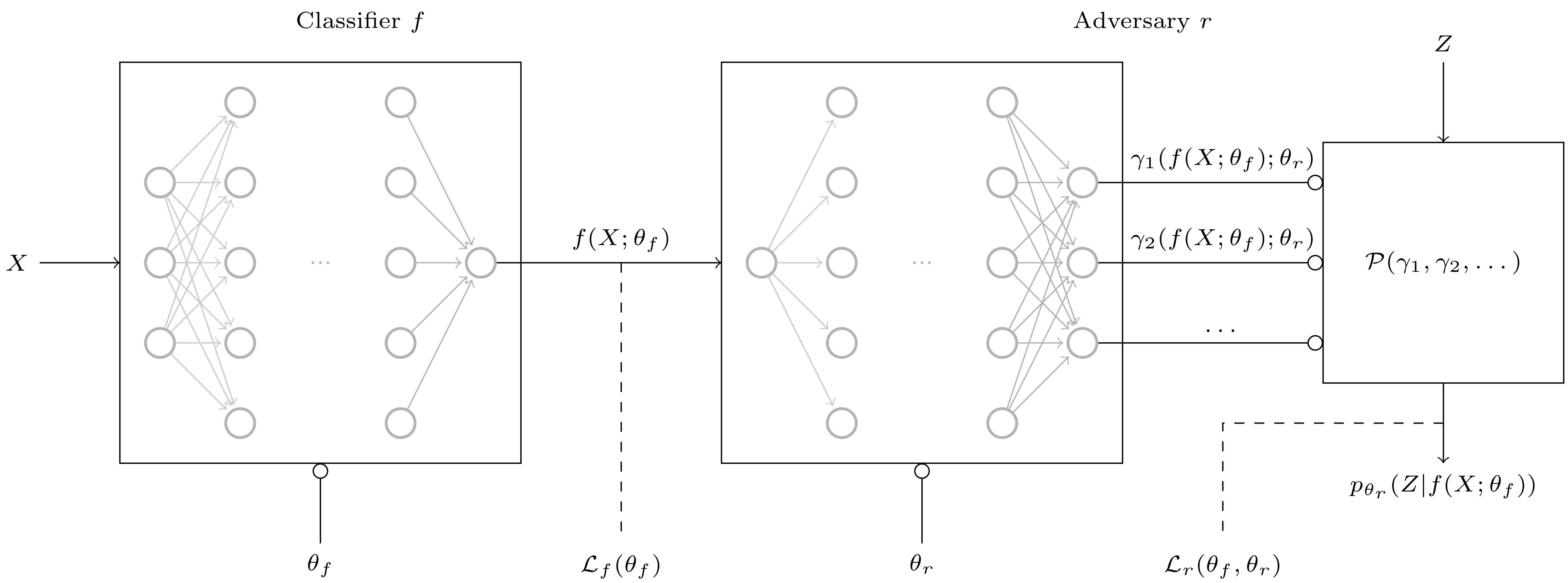






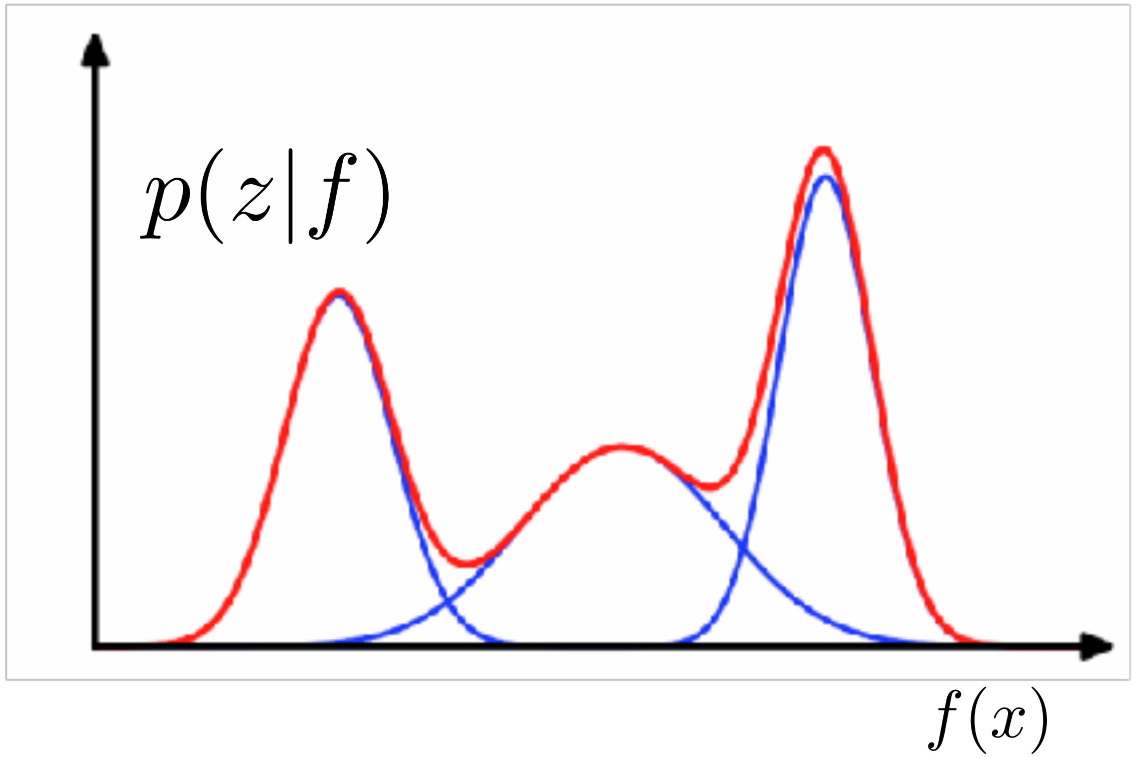
G. Louppe, M. Kagan, K. Cranmer, Learning to Pivot with Adversarial Networks [arXiv:1611.01046]

THE ADVERSARIAL MODEL



the γ_1, γ_2, … are the mean, standard deviation, and amplitude for the Gaussian Mixture Model.

- the neural network takes in f and predicts γ_1, γ_2, …

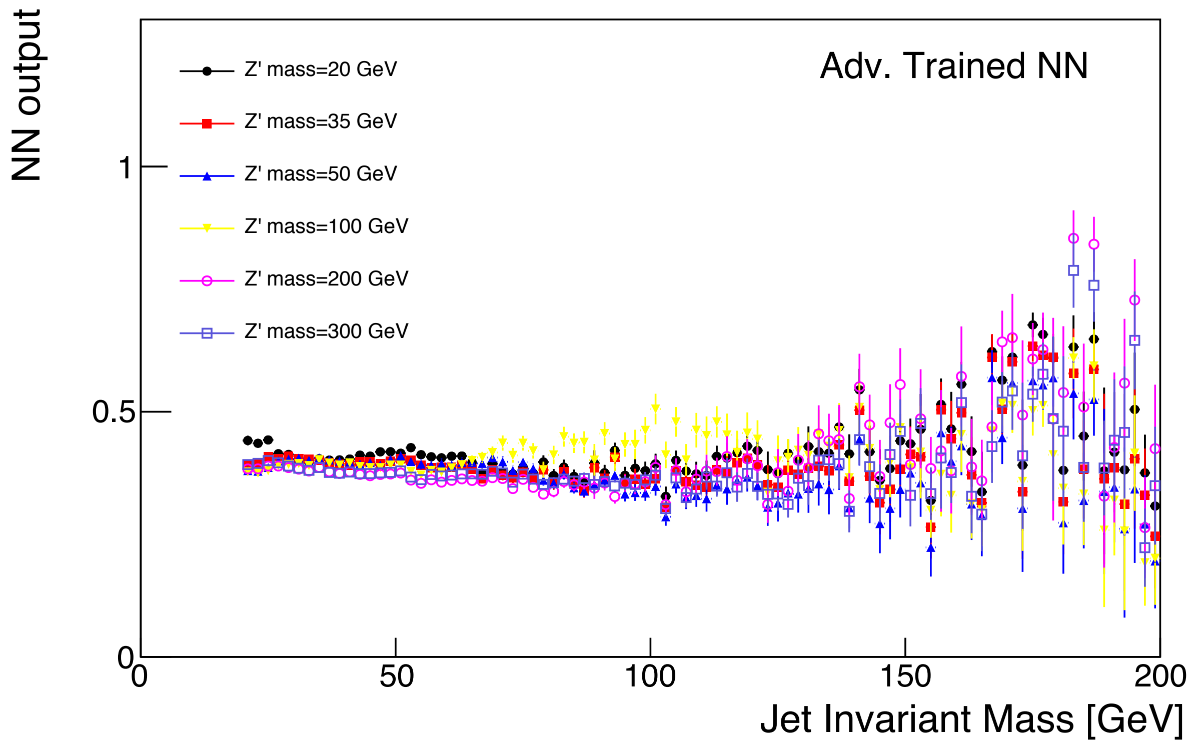


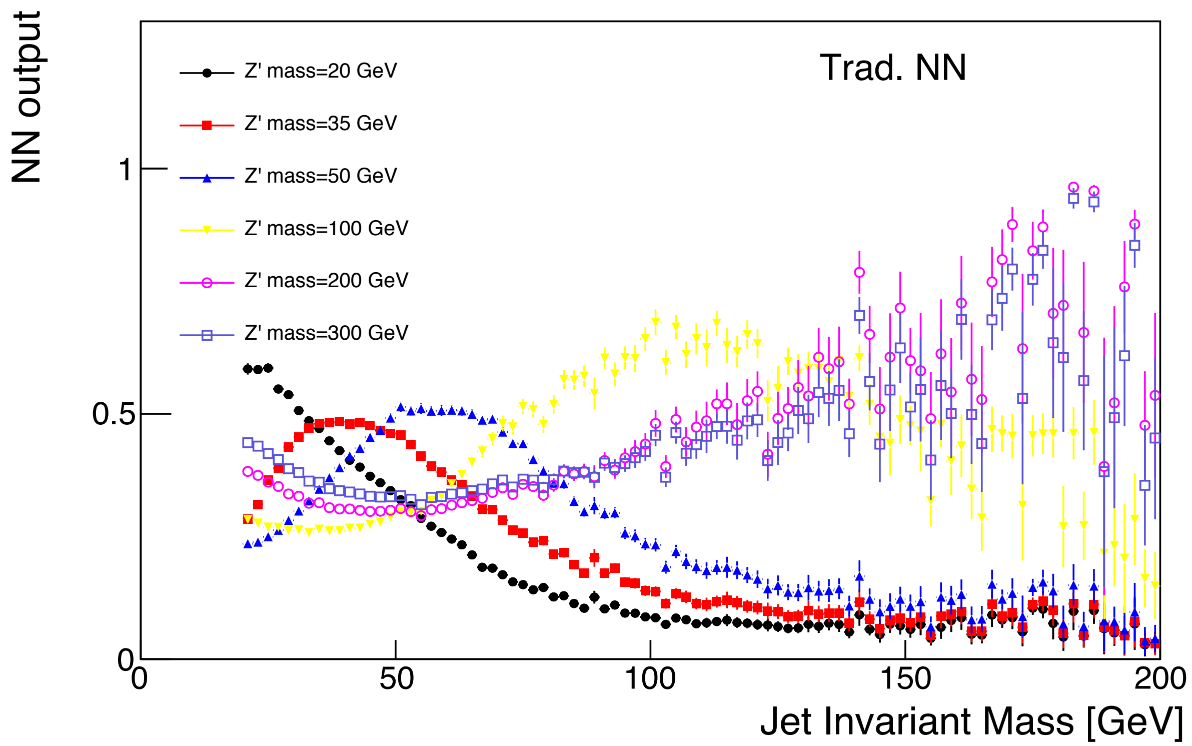
FAIR CLASSIFIERS

K.C, J. Pavez, and G. Louppe, arXiv:1506.02169
P. Baldi, K.C, T. Faucett, P. Sadowski, D. Whiteson arXiv:1601.07913
G. Louppe, M. Kagan, K.C, arXiv:1611.01046
Shimmin, et. al. arXiv:1703.03507

Adversarial approach of "Learning to Pivot" can also be used to train a classifier that is independent from some other continuous variable.

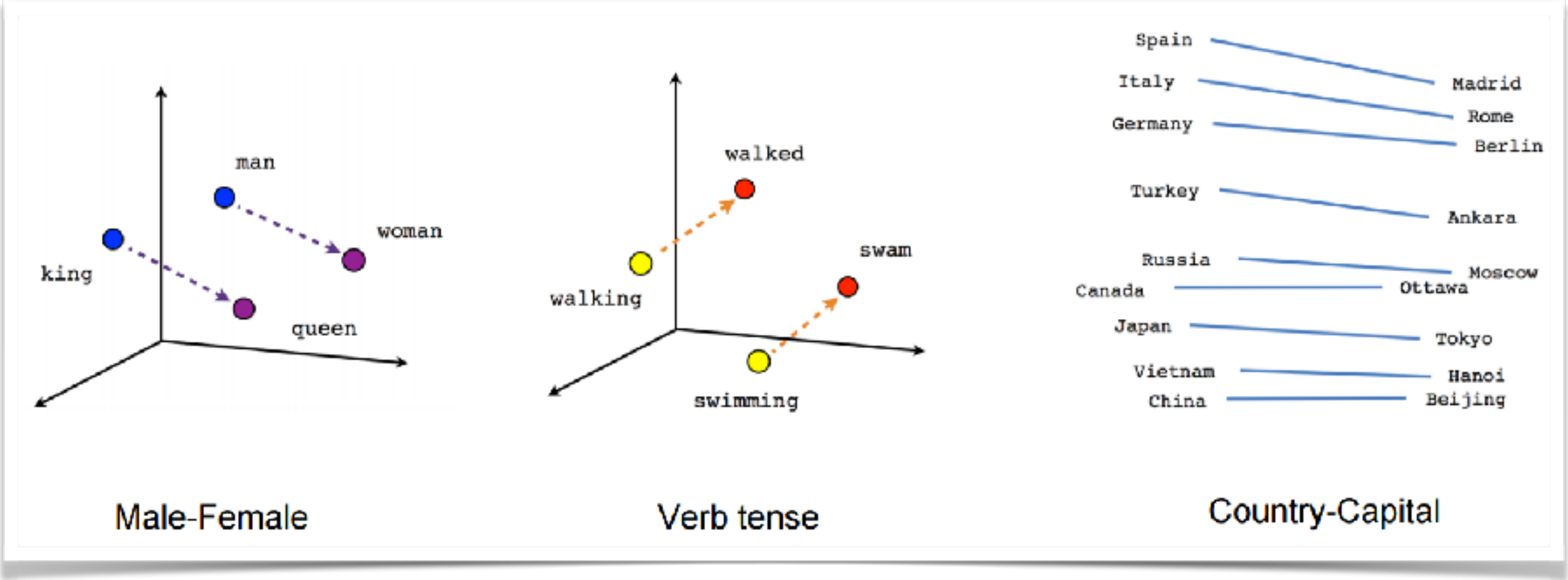
- fairness to continuous attribute
- motivation for doing this is related to robustnesss to uncertainties and interpretability

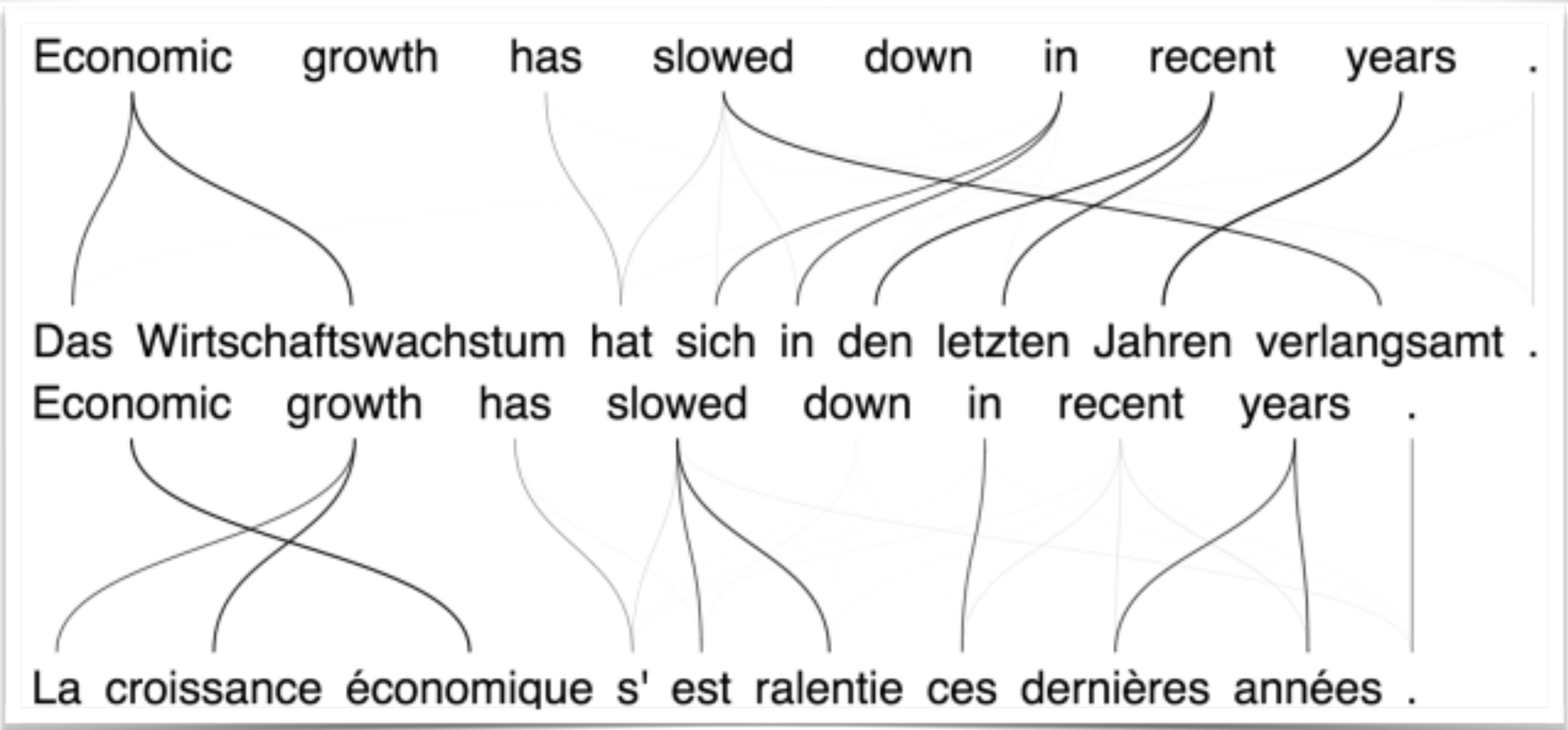




Deep Learning:
A Revolution in AI

WORD EMBEDDINGS & TRANSLATION



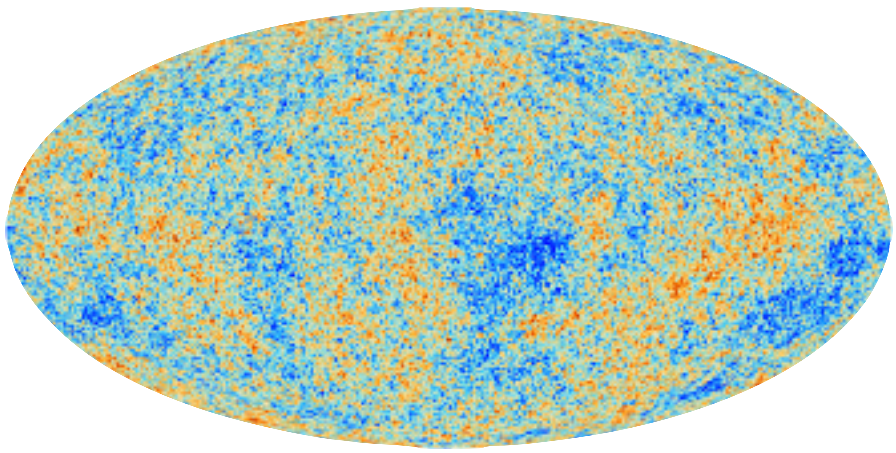


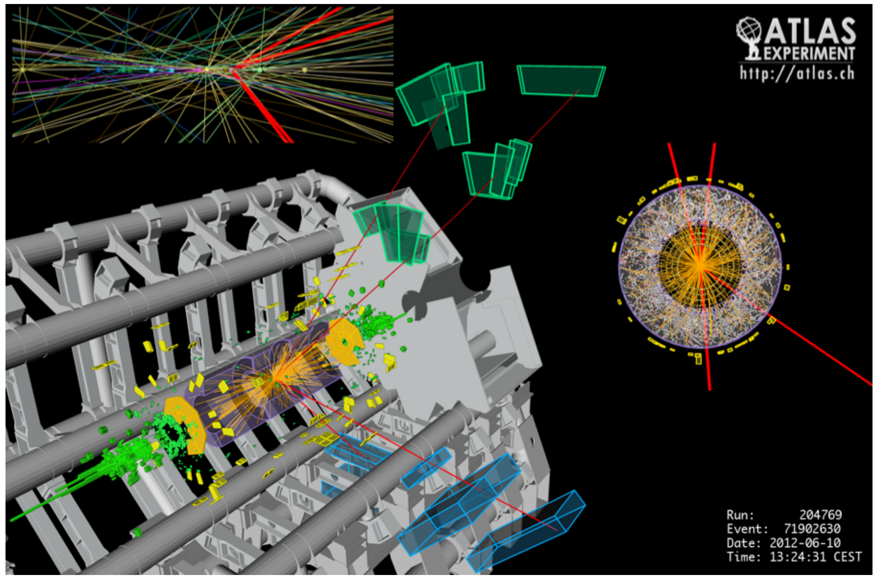
Deep Learning for Physical Sciences

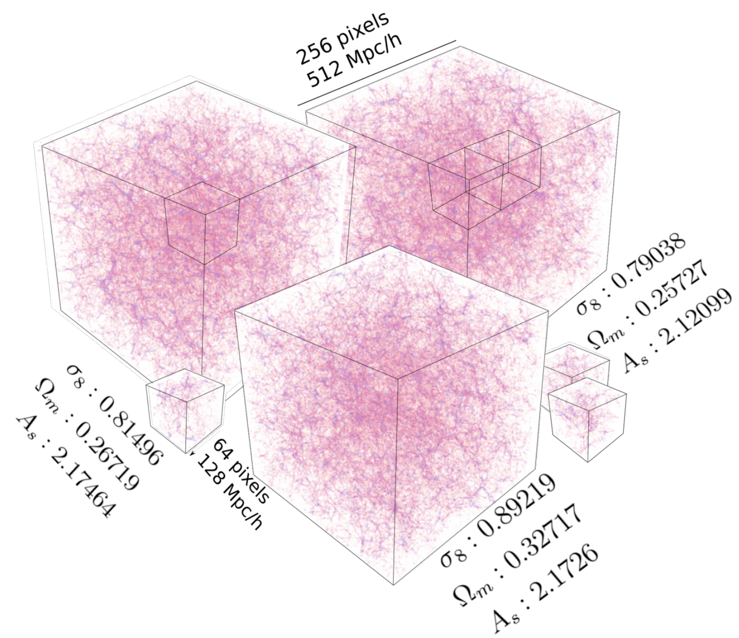
Workshop at the 31st Conference on Neural Information Processing Systems (NIPS)

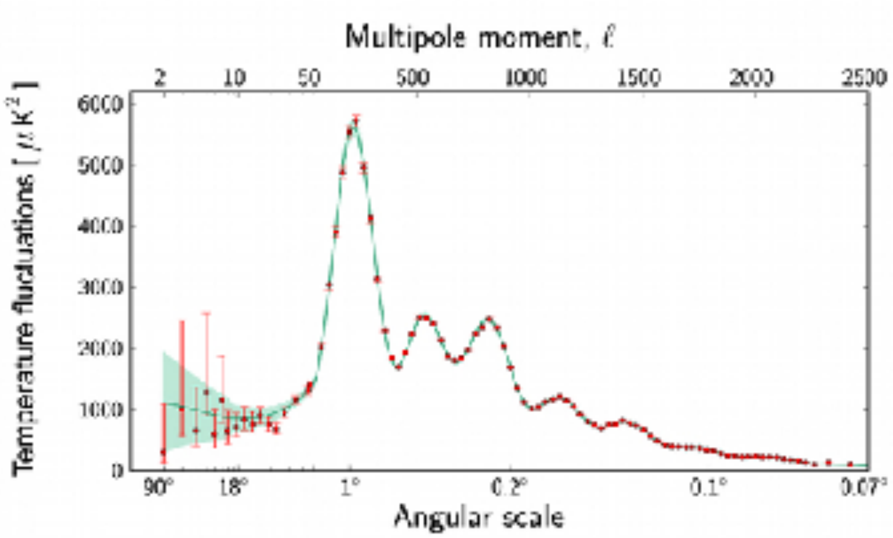
December 8, 2017

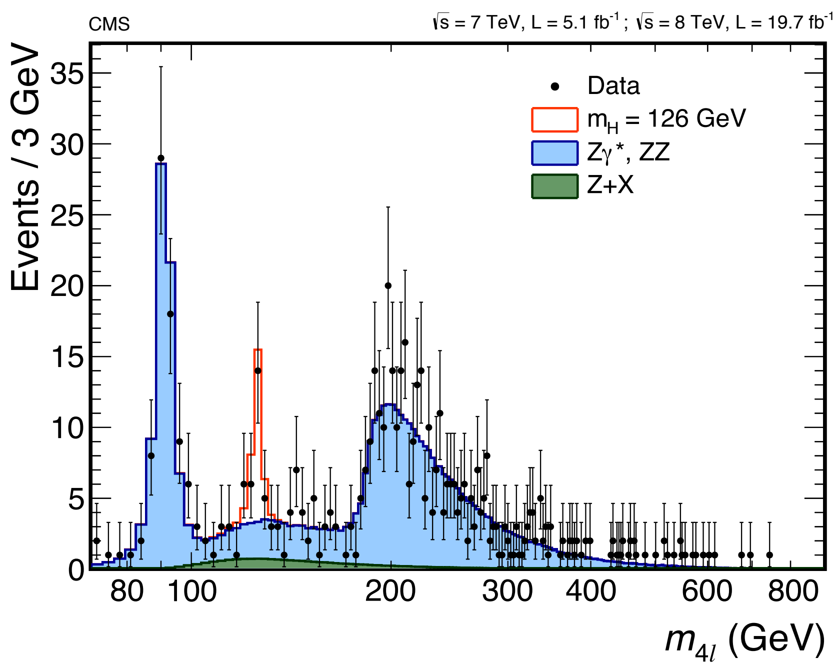
PREDICTION: THE FORWARD MODEL

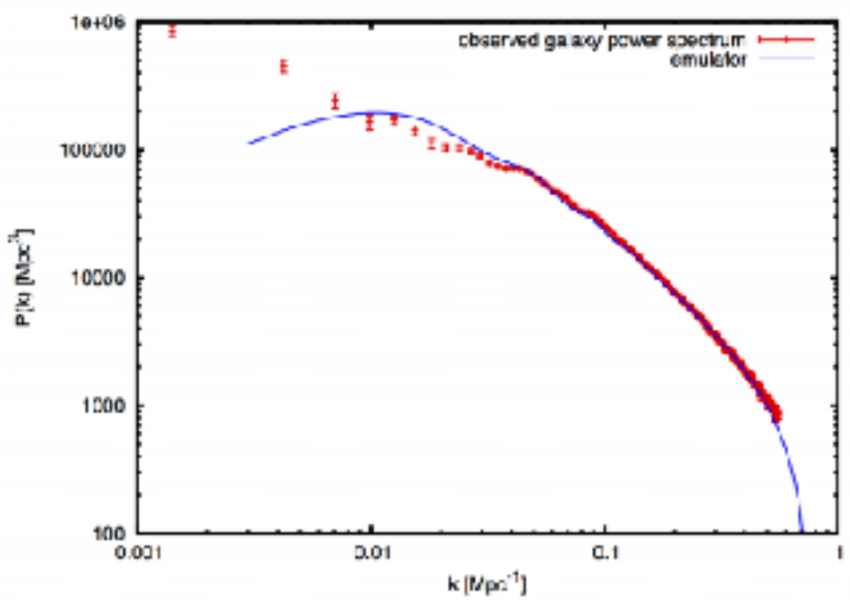






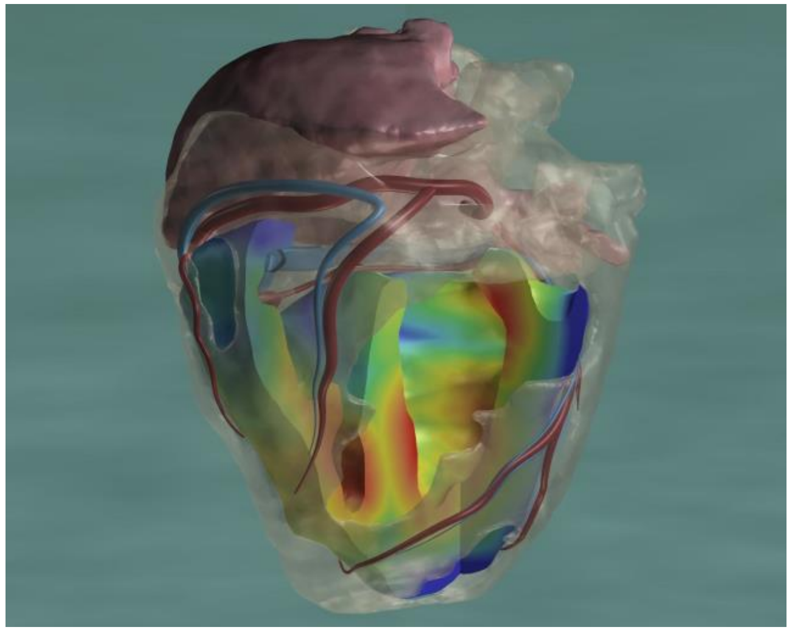


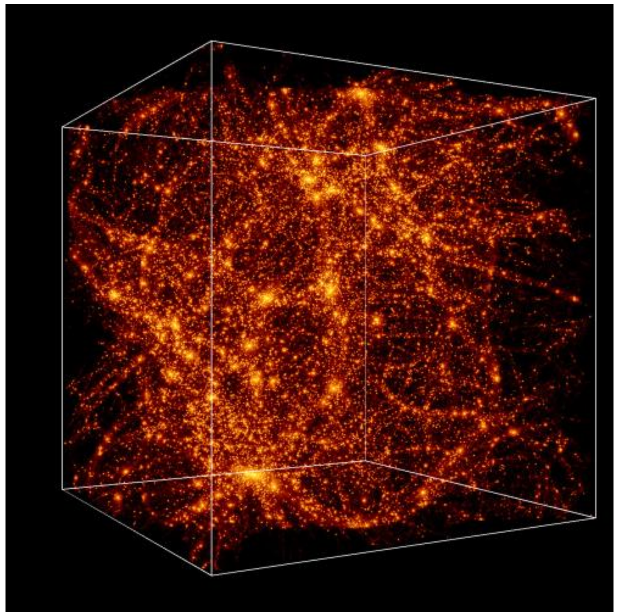


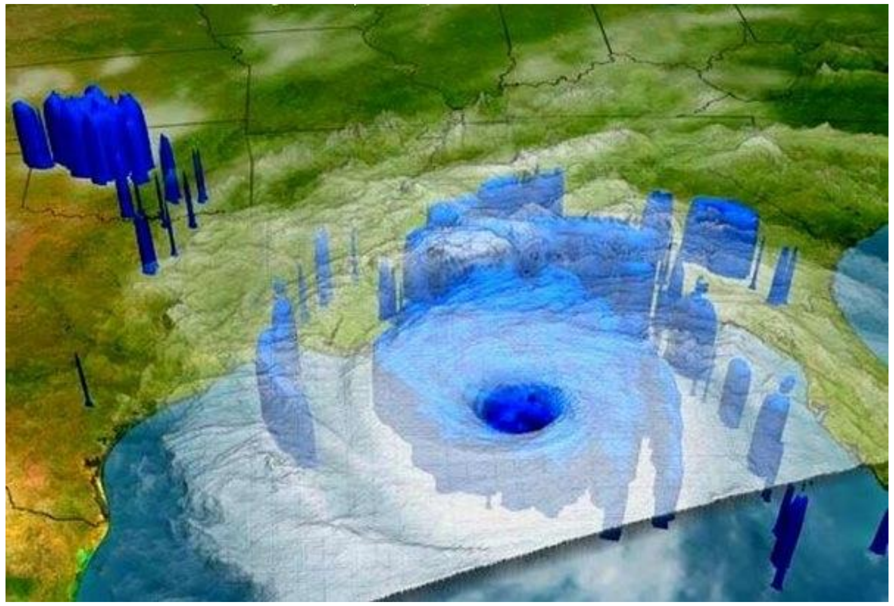


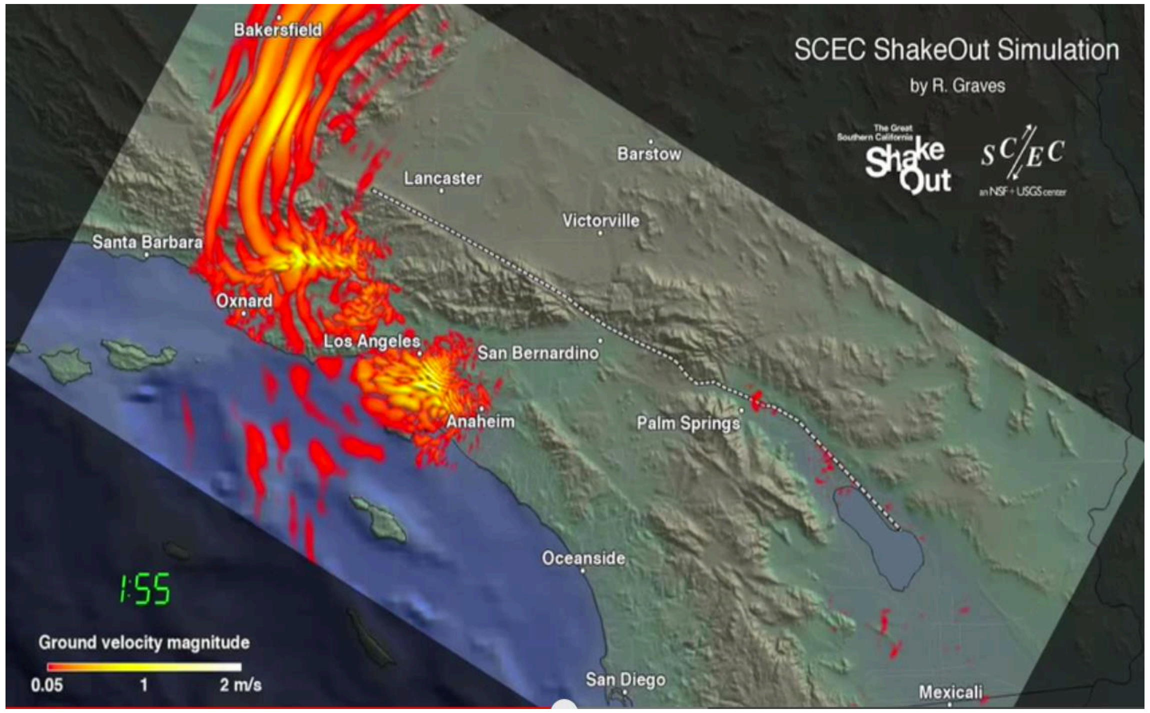
Generative Models: Simulators

Max Welling

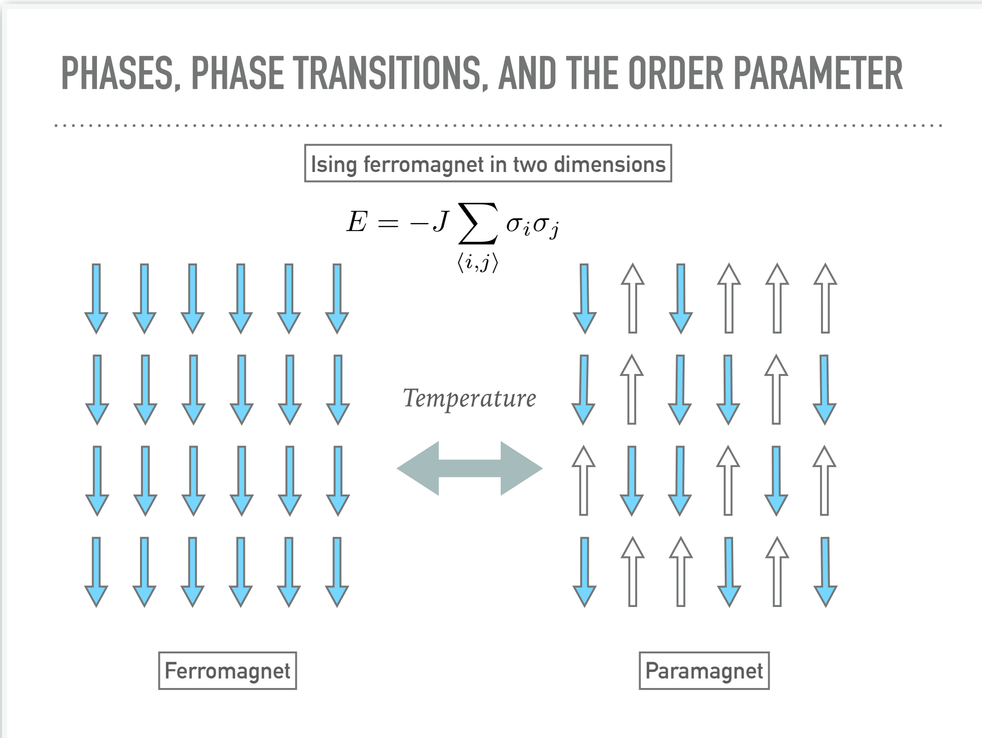


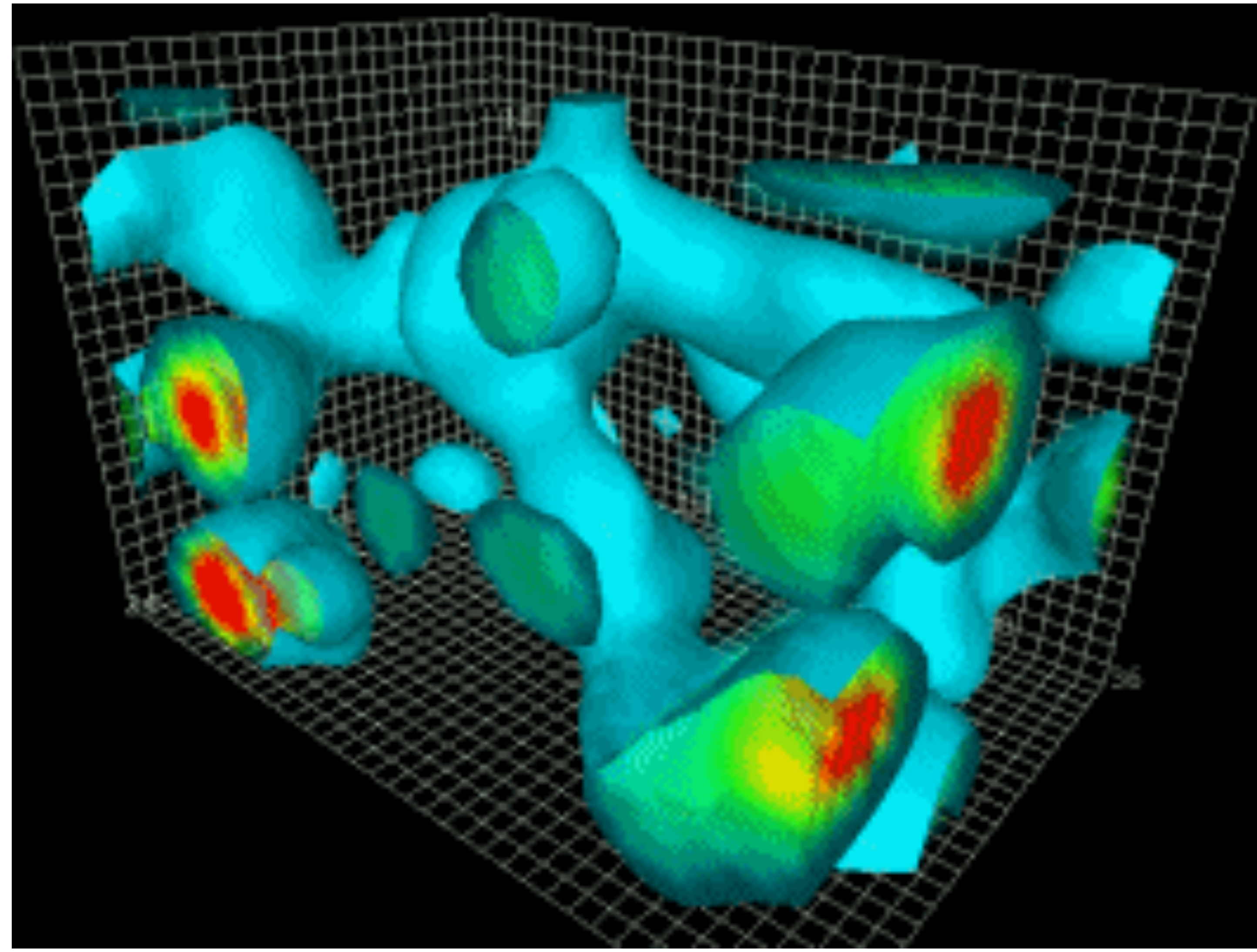






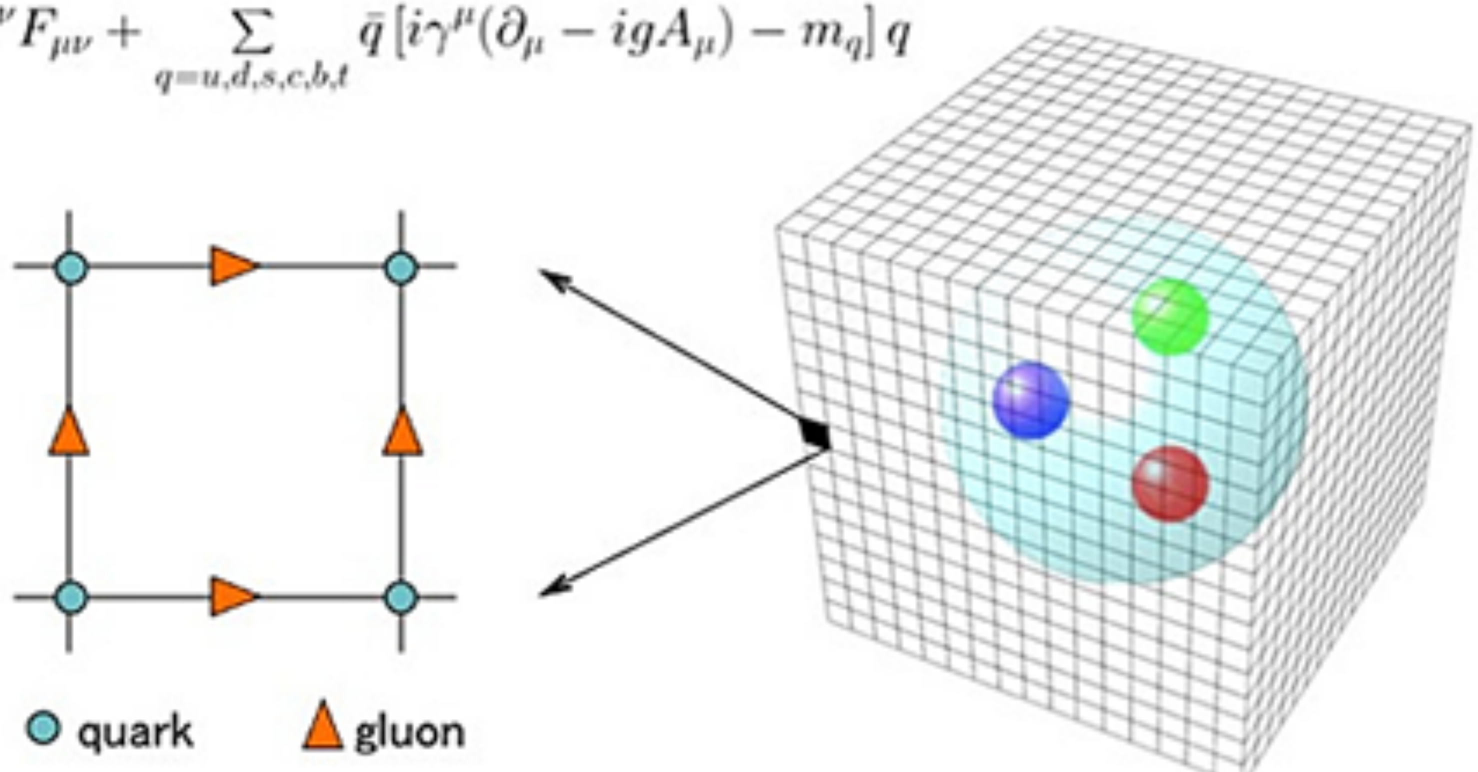
LATTICE FIELD THEORY



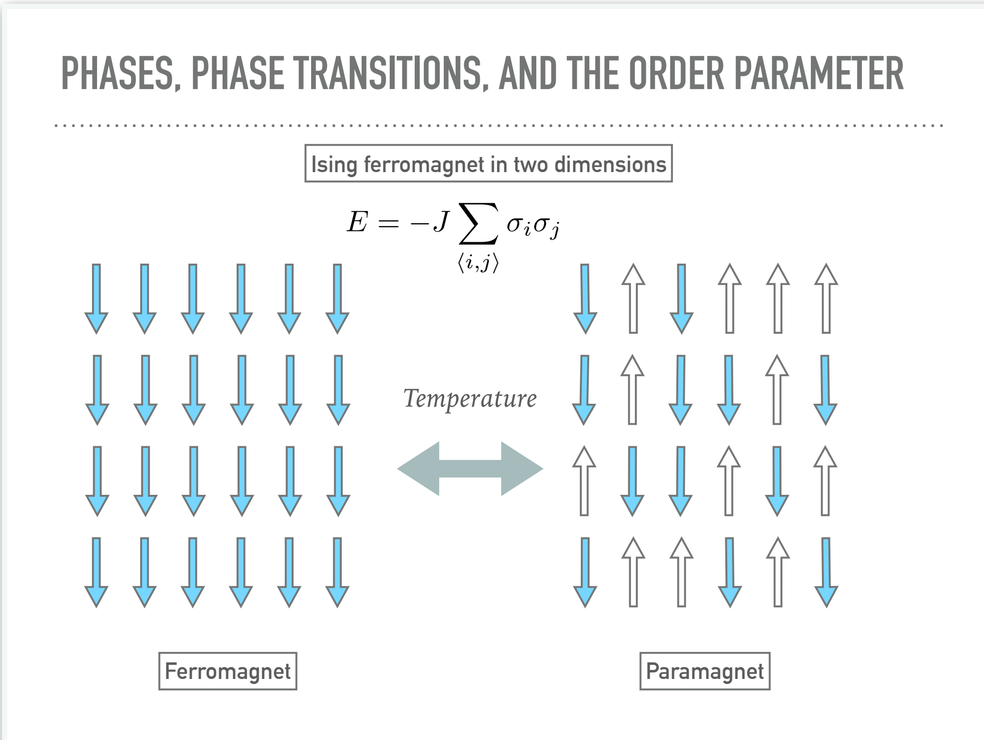


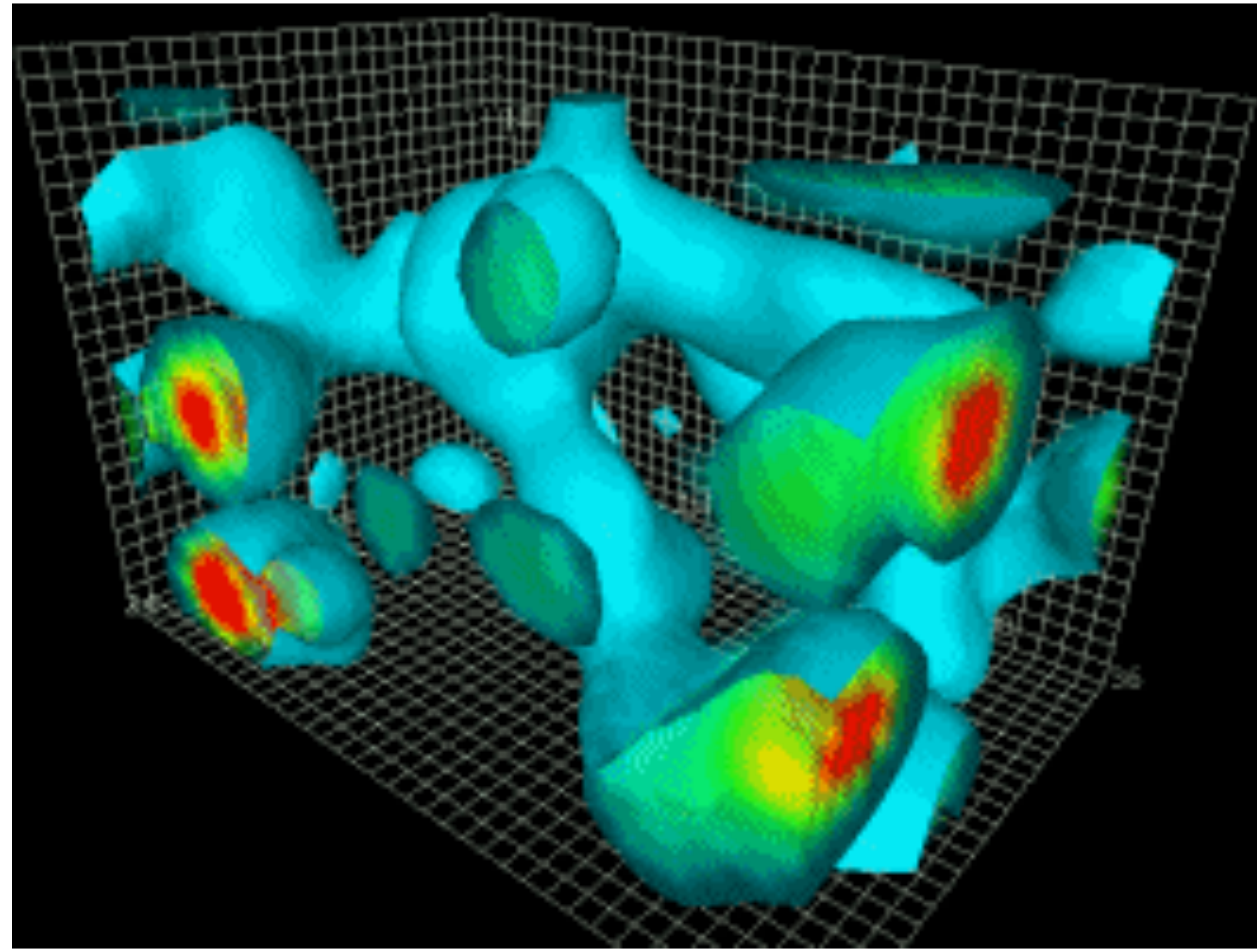
QCD Lagrangian

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



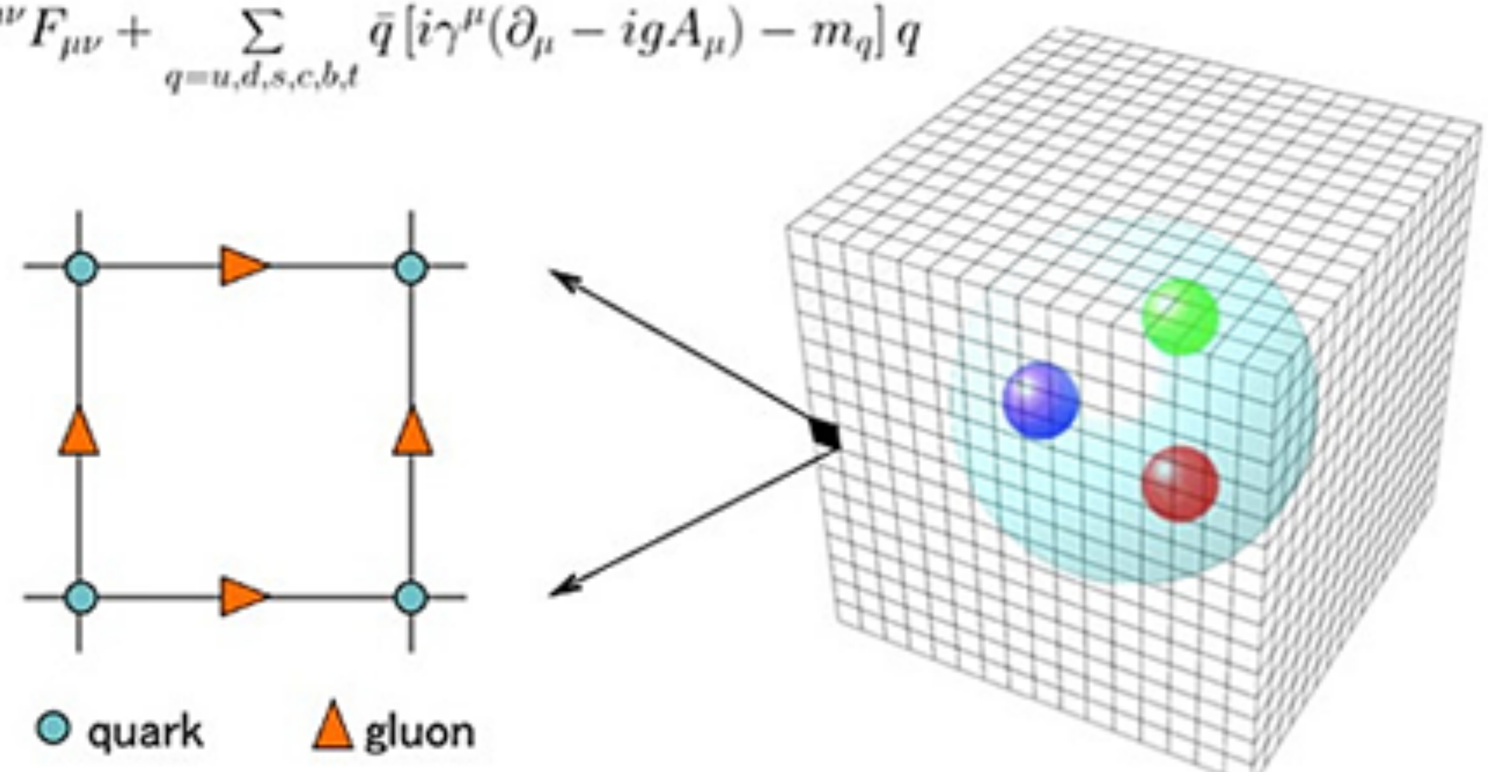
LATTICE FIELD THEORY





QCD Lagrangian

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



WHY WE SHOULD CARE

Many areas of science have simulations based on some well-motivated mechanistic model.

However, the aggregate effect of many interactions between these low-level components leads to an intractable inverse problem.

The developments in machine learning and AI go way beyond improved classifiers and have the potential to effectively bridge the microscopic - macroscopic divide & aid in the inverse problem.

- they can provide effective statistical models that describe macroscopic phenomena that are tied back to the low-level microscopic (reductionist) model
- generative models and likelihood-free inference are two particularly exciting areas

A COMMON THEME

ABC

resources on approximate Bayesian computational methods

Search

Home

Home

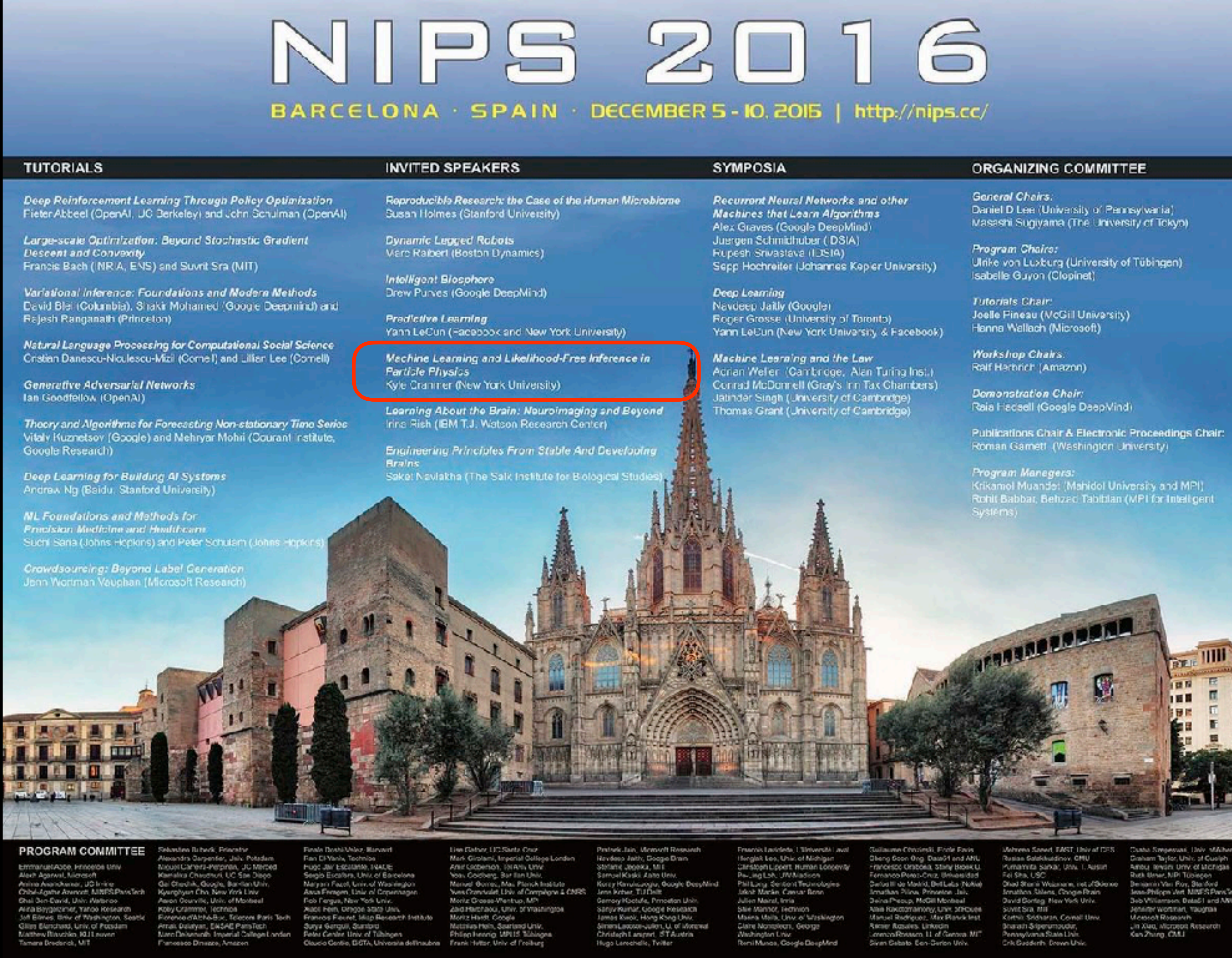
This website keeps track of developments in approximate Bayesian computation (ABC) (a.k.a. likelihood-free), a class of computational statistical methods for Bayesian inference under intractable likelihoods. The site is meant to be a resource both for biologists and statisticians who want to learn more 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.



NIPS 2016
BARCELONA · SPAIN · DECEMBER 5 - 10, 2016 | http://nips.cc/
TUTORIALS
Deep Reinforcement Learning Through Policy Optimization
Pieter Abbeel (OpenAI, UC Berkeley) and John Schulman (OpenAI)
Large-scale Optimization: Beyond Stochastic Gradient Descent and Convexity
Francis Bach (INRIA, ENS) and Suvrit Sra (MIT)
Variational Inference: Foundations and Modern Methods
David Blei (Columbia), Shakir Mohamed (Google Deepmind) and Rajesh Ranganath (Princeton)
Natural Language Processing for Computational Social Science
Cristian Danescu-Niculescu-Mizil (Cornell) and Lillian Lee (Cornell)
Generative Adversarial Networks
Ian Goodfellow (OpenAI)
Theory and Algorithms for Forecasting Non-stationary Time Series
Vitaly Kuznetsov (Google) and Mehryar Mohri (Courant Institute, Google Research)
Deep Learning for Building AI Systems
Andrew Ng (Baidu, Stanford University)
ML Foundations and Methods for Precision Medicine and Healthcare
Suchi Saria (Johns Hopkins) and Peter Schulam (Johns Hopkins)
Crowdsourcing: Beyond Label Generation
Jenn Wortman Vaughan (Microsoft Research)
INVITED SPEAKERS
Reproducible Research: the Case of the Human Microbiome
Susan Holmes (Stanford University)
Dynamic Legged Robots
Marc Raibert (Boston Dynamics)
Intelligent Biosphere
Drew Purves (Google DeepMind)
Predictive Learning
Yann LeCun (Facebook and New York University)
Machine Learning and Likelihood-Free Inference in Particle Physics
Kyle Cranmer (New York University)
Learning About the Brain: Neuroimaging and Beyond
Irina Rish (IBM T.J. Watson Research Center)
Engineering Principles From Stable And Developing Brains
Saket Navlakha (The Salk Institute for Biological Studies)
SYMPOSIA
Recurrent Neural Networks and other Machines that Learn Algorithms
Alex Graves (Google DeepMind)
Juergen Schmidhuber (IDSIA)
Rupesh Srivastava (IDSIA)
Sepp Hochreiter (Johannes Kepler University)
Deep Learning
Navdeep Jaitly (Google)
Roger Grosse (University of Toronto)
Yann LeCun (New York University & Facebook)
Machine Learning and the Law
Adrian Weller (Cambridge, Alan Turing Inst.)
Conrad McDonnell (Gray's Inn Tax Chambers)
Jatinder Singh (University of Cambridge)
Thomas Grant (University of Cambridge)
ORGANIZING COMMITTEE
General Chairs:
Daniel D Lee (University of Pennsylvania)
Masashi Sugiyama (The University of Tokyo)
Program Chairs:
Ulrike von Luxburg (University of Tübingen)
Isabelle Guyon (Clopinet)
Tutorials Chair:
Joelle Pineau (McGill University)
Hanna Wallach (Microsoft)
Workshop Chairs:
Ralf Herbrich (Amazon)
Demonstration Chair:
Raia Hadsell (Google DeepMind)
Publications Chair & Electronic Proceedings Chair:
Roman Garnett (Washington University)
Program Managers:
Krikamol Muandet (Mahidol University and MPI)
Rohit Babbar, Behzad Tabibian (MPI for Intelligent Systems)
PROGRAM COMMITTEE

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.

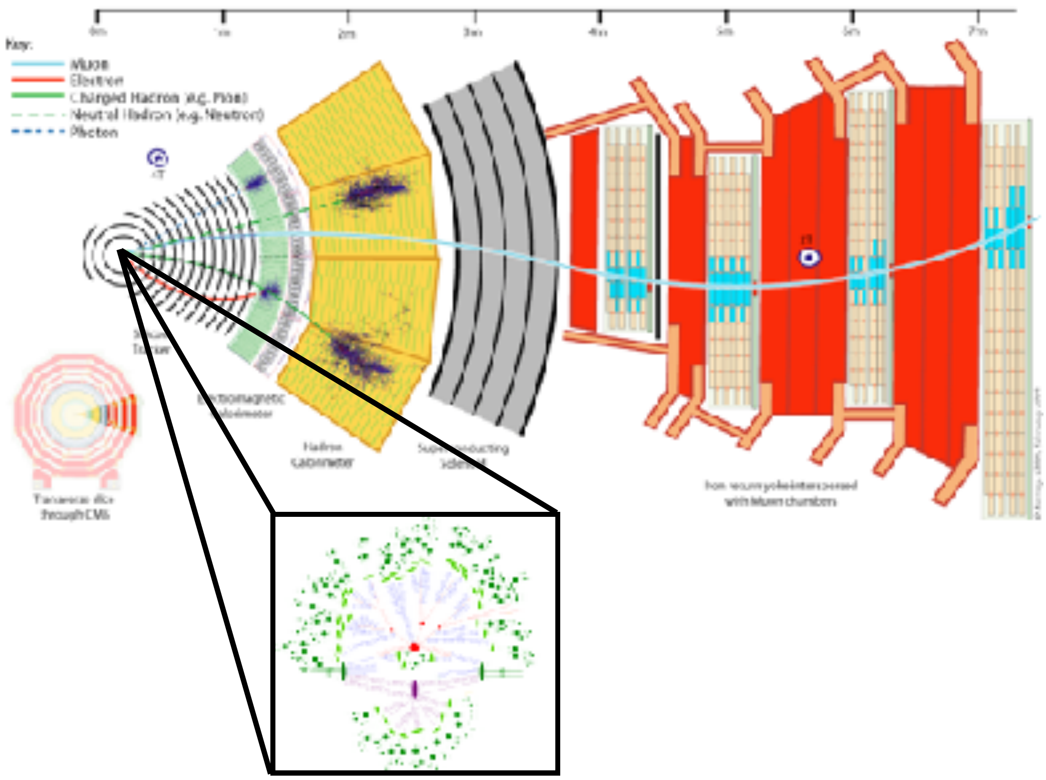
TWO APPROACHES

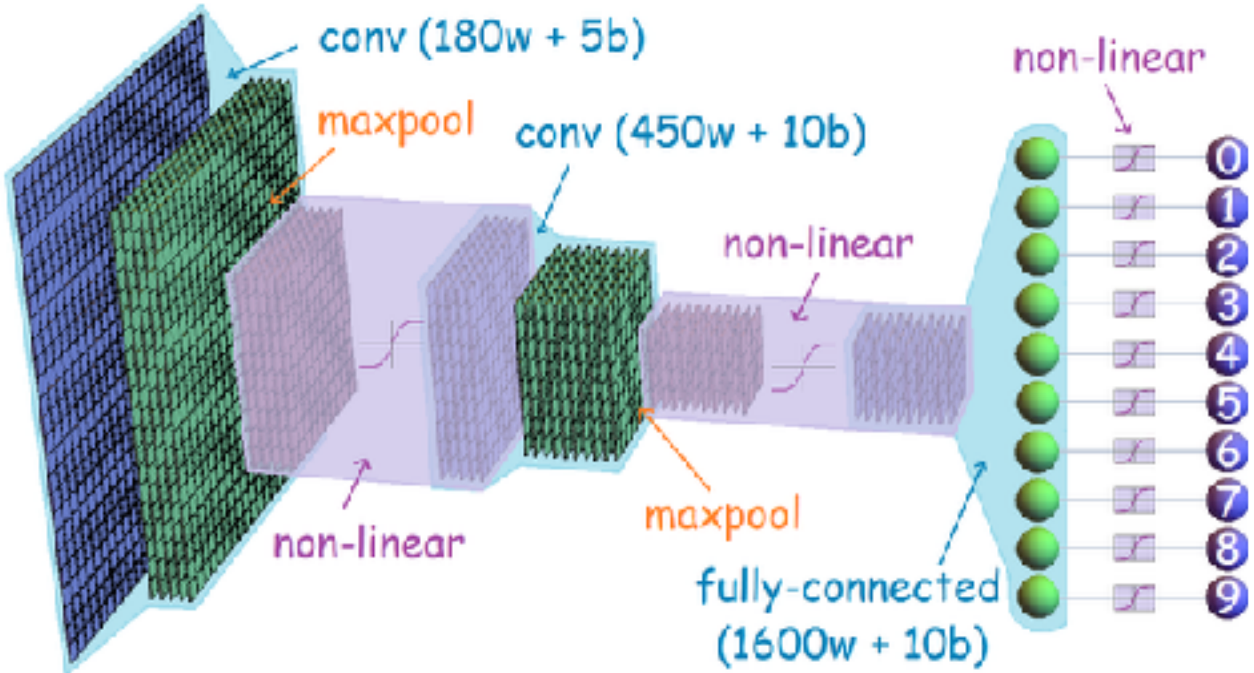
Use simulator

(much more efficiently)

Learn simulator

(with deep learning)





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

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

[image credit: A.P. Goucher]

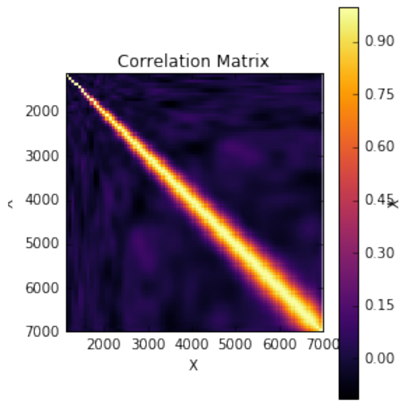
PHYSICS-AWARE MACHINE LEARNING

We can **inject** our knowledge of physics into the machine learning models!

We can **extract** knowledge learned from the data!

Physics-aware Gaussian Processes

arXiv:1709.05681

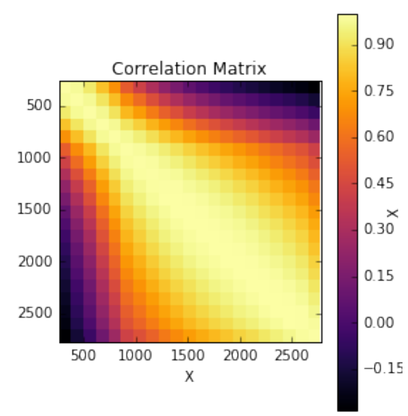


Final Kernel =

Poisson fluctuations

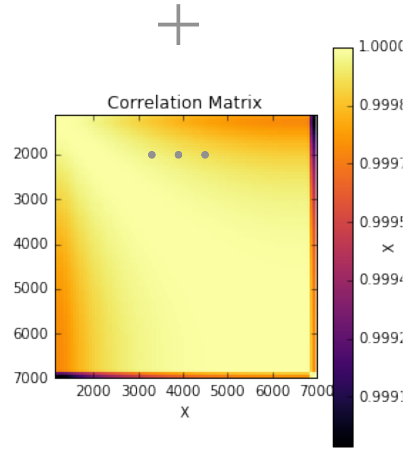
=

+ Mass Resolution



+ Parton Density Functions

+

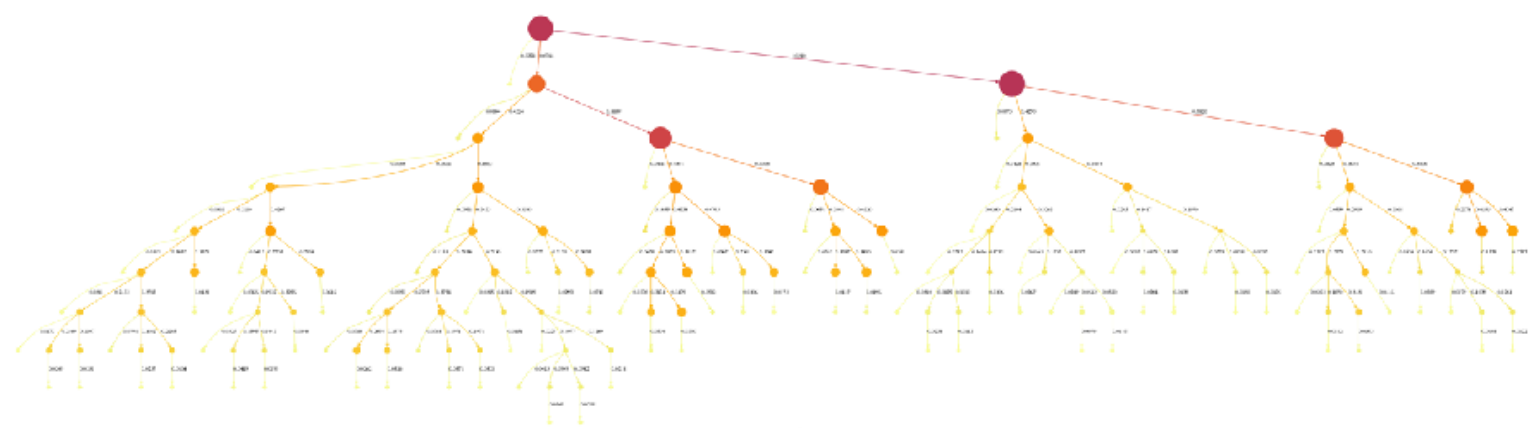


+ Jet Energy Scale

+

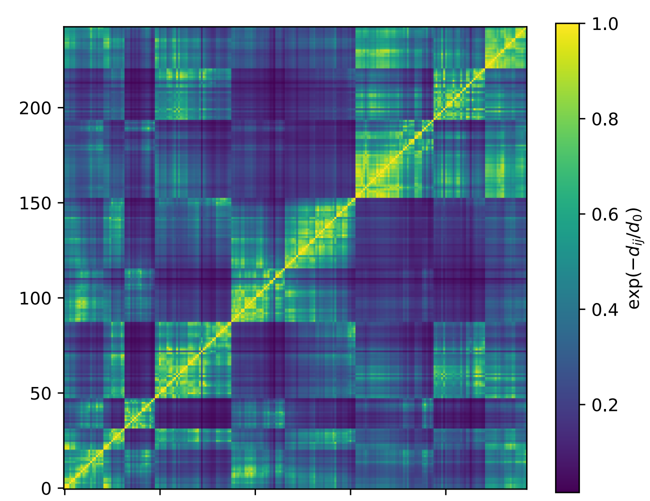
QCD-Aware recursive neural networks

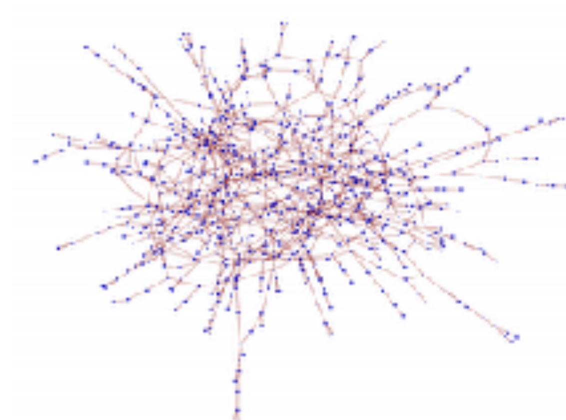
arXiv:1702.00748



QCD-Aware graph convolutional neural networks

NIPS2017 workshop [http://bit.ly/2AkwYRG]





$$d^{\alpha}_{ii'} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha}) \frac{\Delta R^2_{ii'}}{R^2}$$

THE PLAYERS

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

p(x, z | θ, ν)

x
observed data
simulated data

z
latent variables
Monte Carlo truth

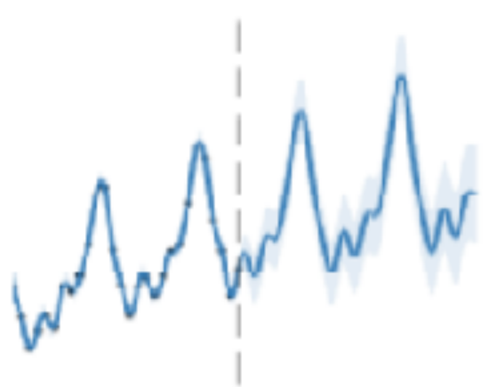
ν
nuisance parameters

INFERENCE

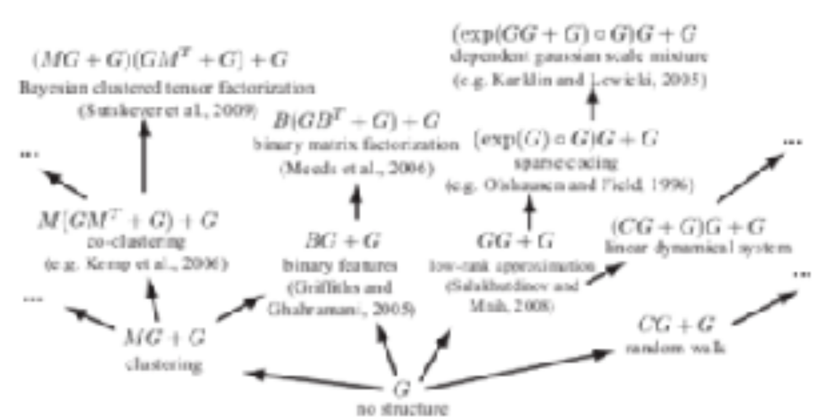
inverse problem
measurement
parameter estimation

Physics goes into the construction of a "Kernel" that defines M.L. model

- Vocabulary of kernels + grammar for composition = powerful modeling

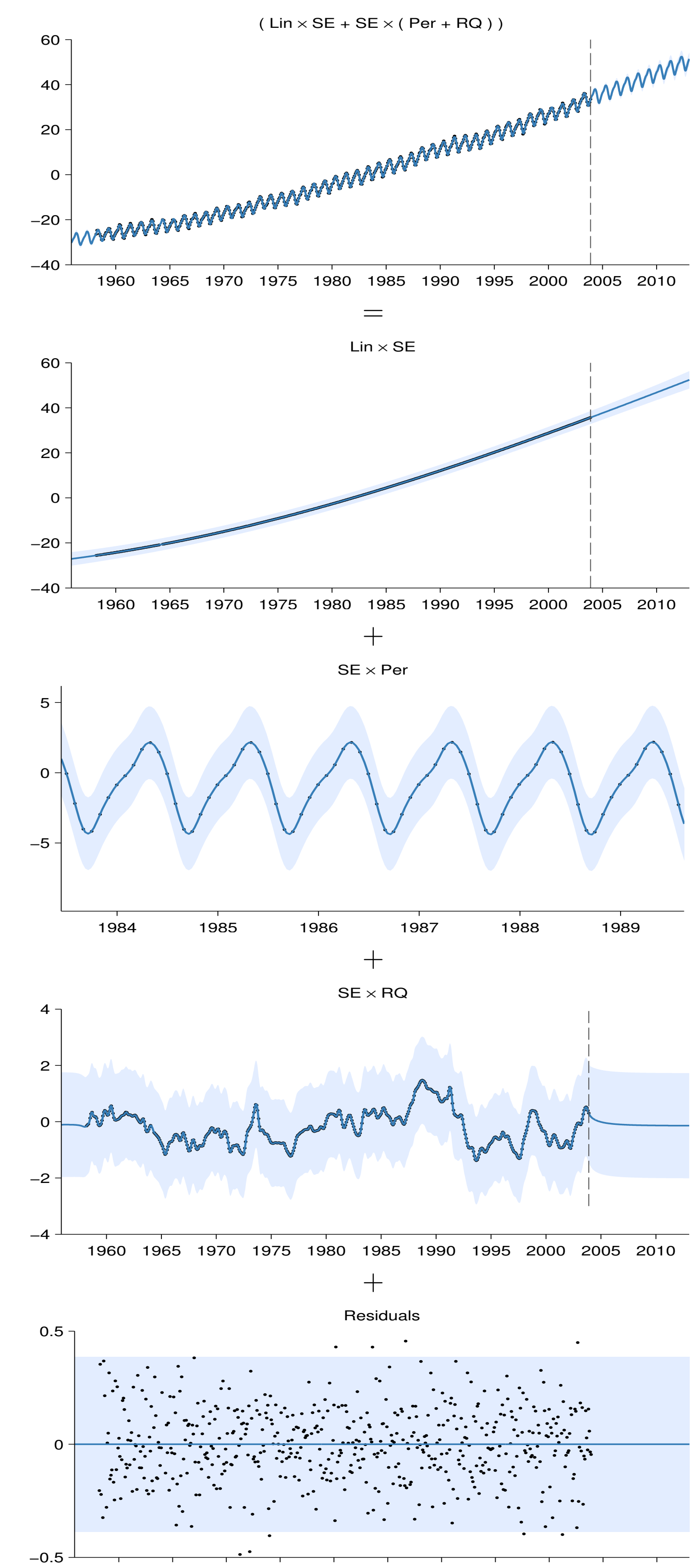


Structure Discovery in Nonparametric Regression through Compositional Kernel Search
David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, Zoubin Ghahramani
International Conference on Machine Learning, 2013
pdf | code | poster | bibtex



Exploiting compositionality to explore a large space of model structures
Roger Grosse, Ruslan Salakhutdinov, William T. Freeman, Joshua B. Tenenbaum
Conference on Uncertainty in Artificial Intelligence, 2012
pdf | code | bibtex

Mauna Loa atmospheric CO_2

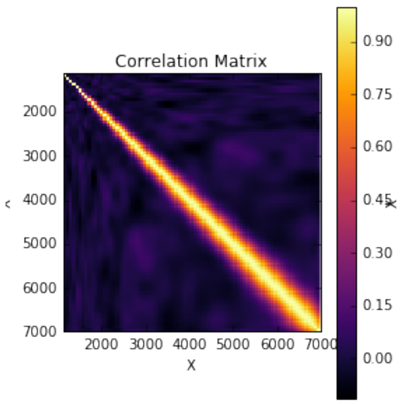


PHYSICS-AWARE MACHINE LEARNING

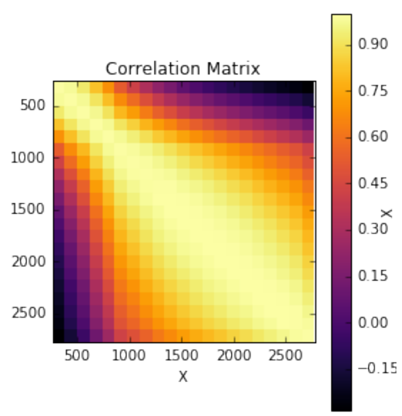
We can **inject** our knowledge of physics into the machine learning models!
We can **extract** knowledge learned from the data!

Physics-aware Gaussian Processes

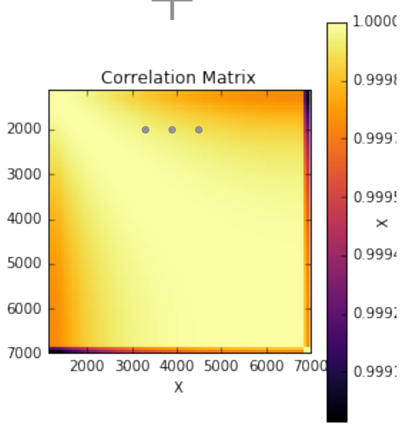
arXiv:1709.05681



=



+



+

Final Kernel =

Poisson fluctuations

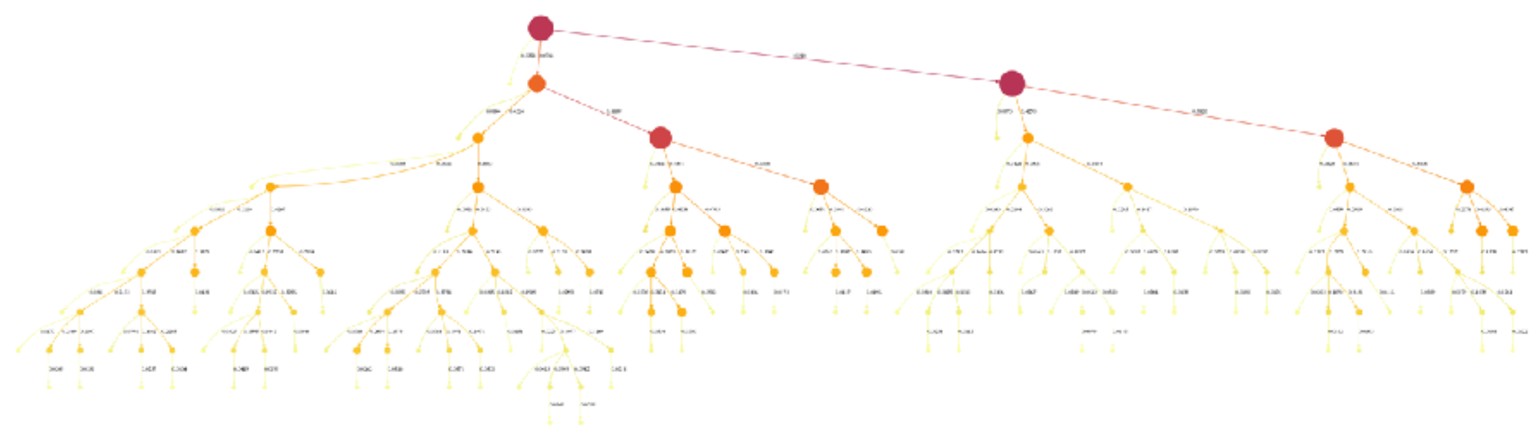
+ Mass Resolution

+ Parton Density Functions

+ Jet Energy Scale

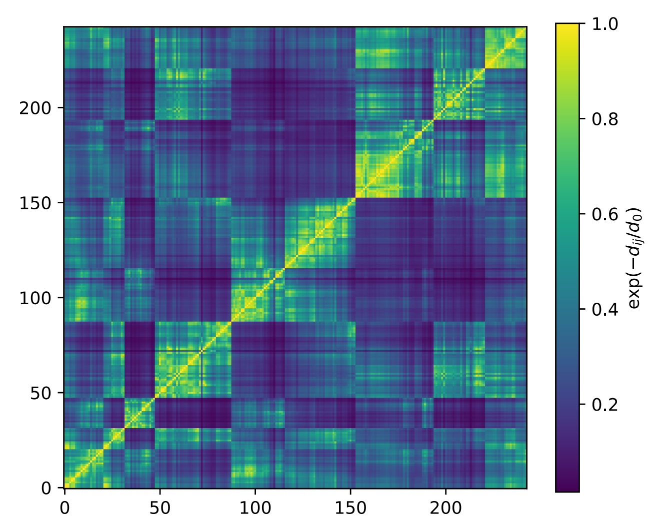
QCD-Aware recursive neural networks

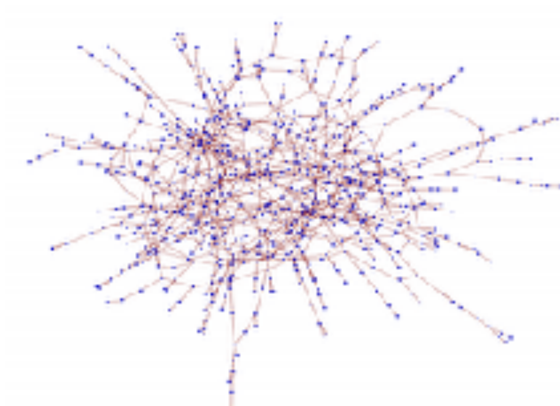
arXiv:1702.00748



QCD-Aware graph convolutional neural networks

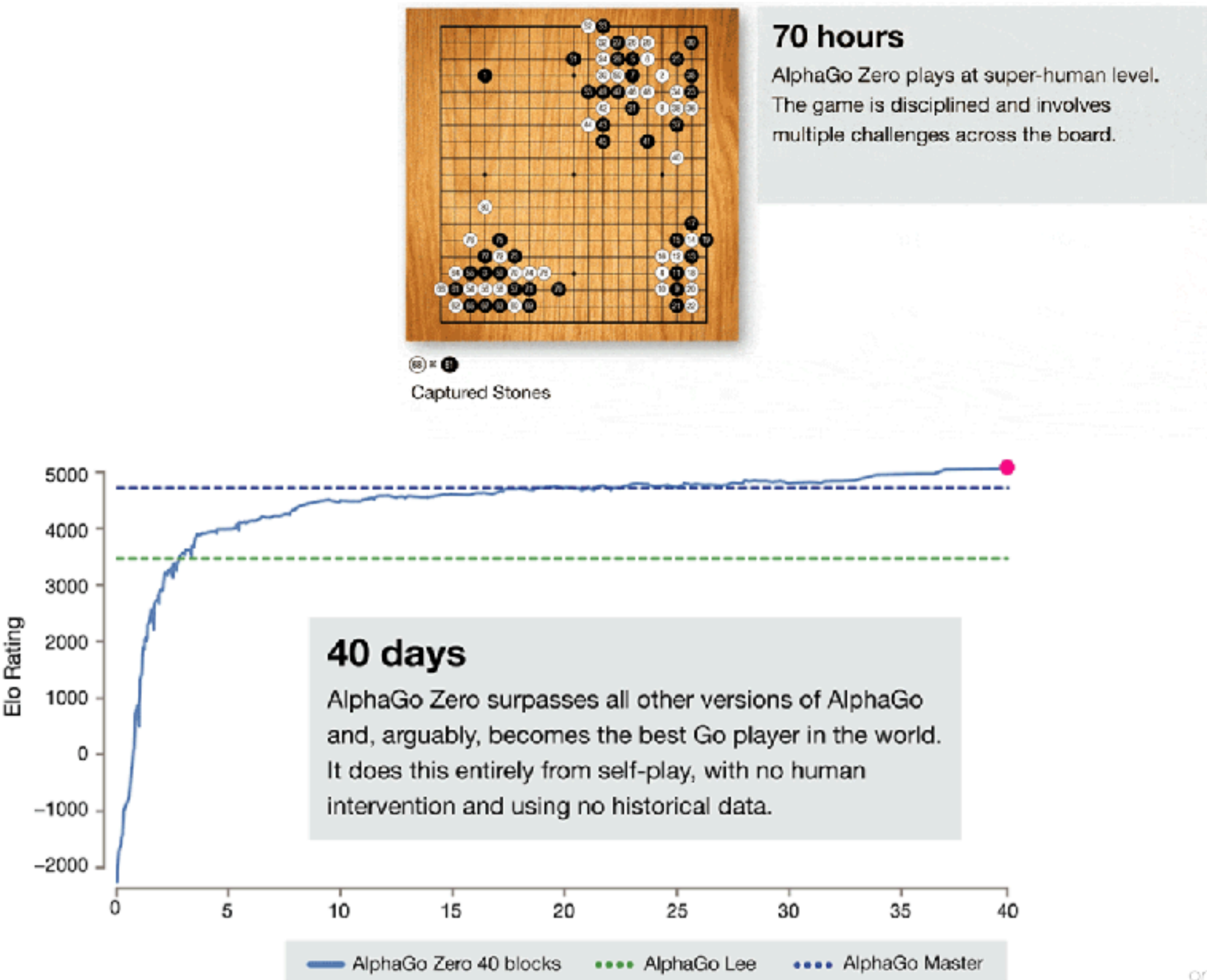
NIPS2017 workshop [http://bit.ly/2AkwYRG]



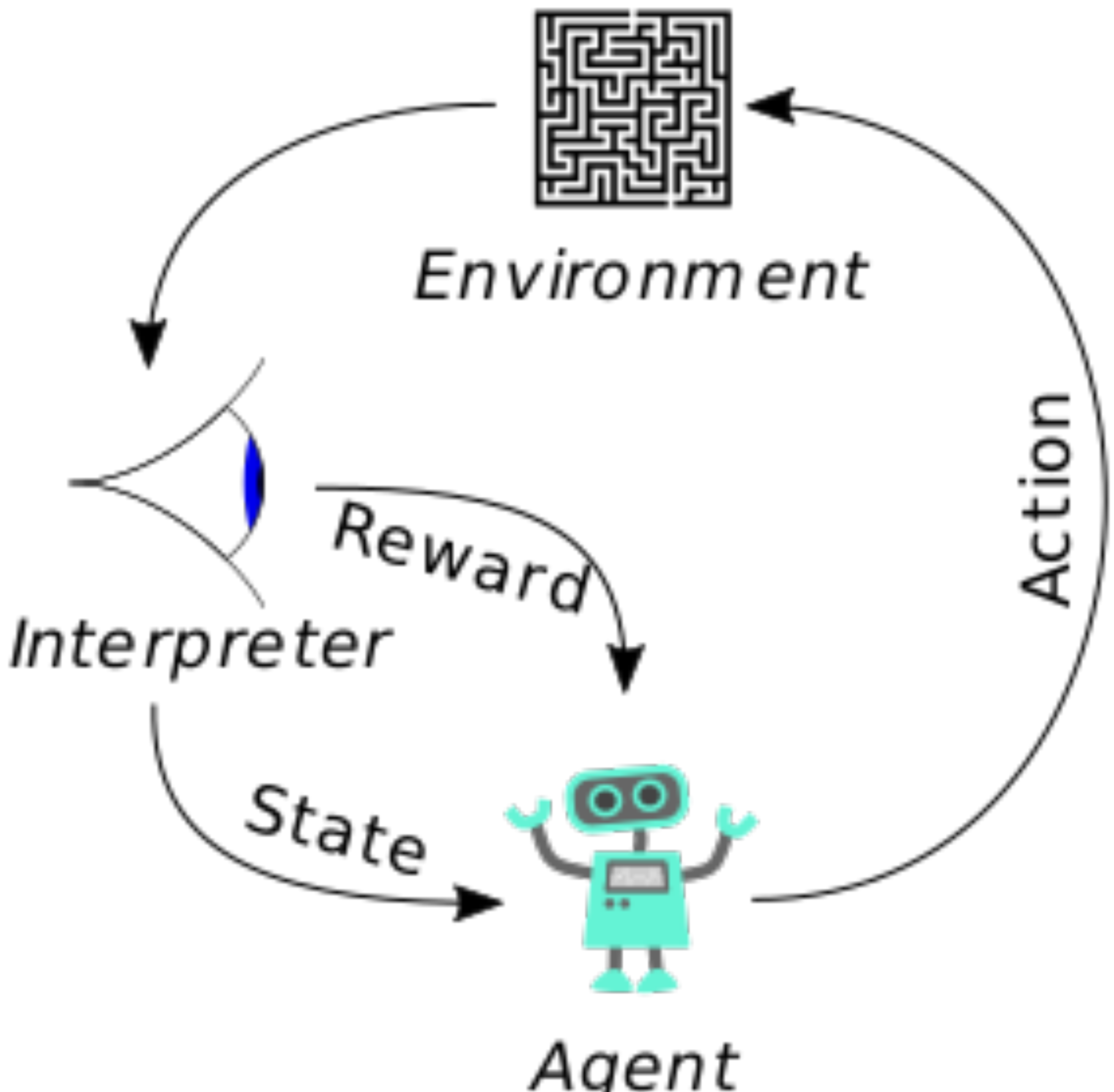


$$d^{\alpha}_{ii'} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha})\frac{\Delta R^2_{ii'}}{R^2}$$

Scientist trying to decide what experiment to do next



Scientist trying to decide what experiment to do next

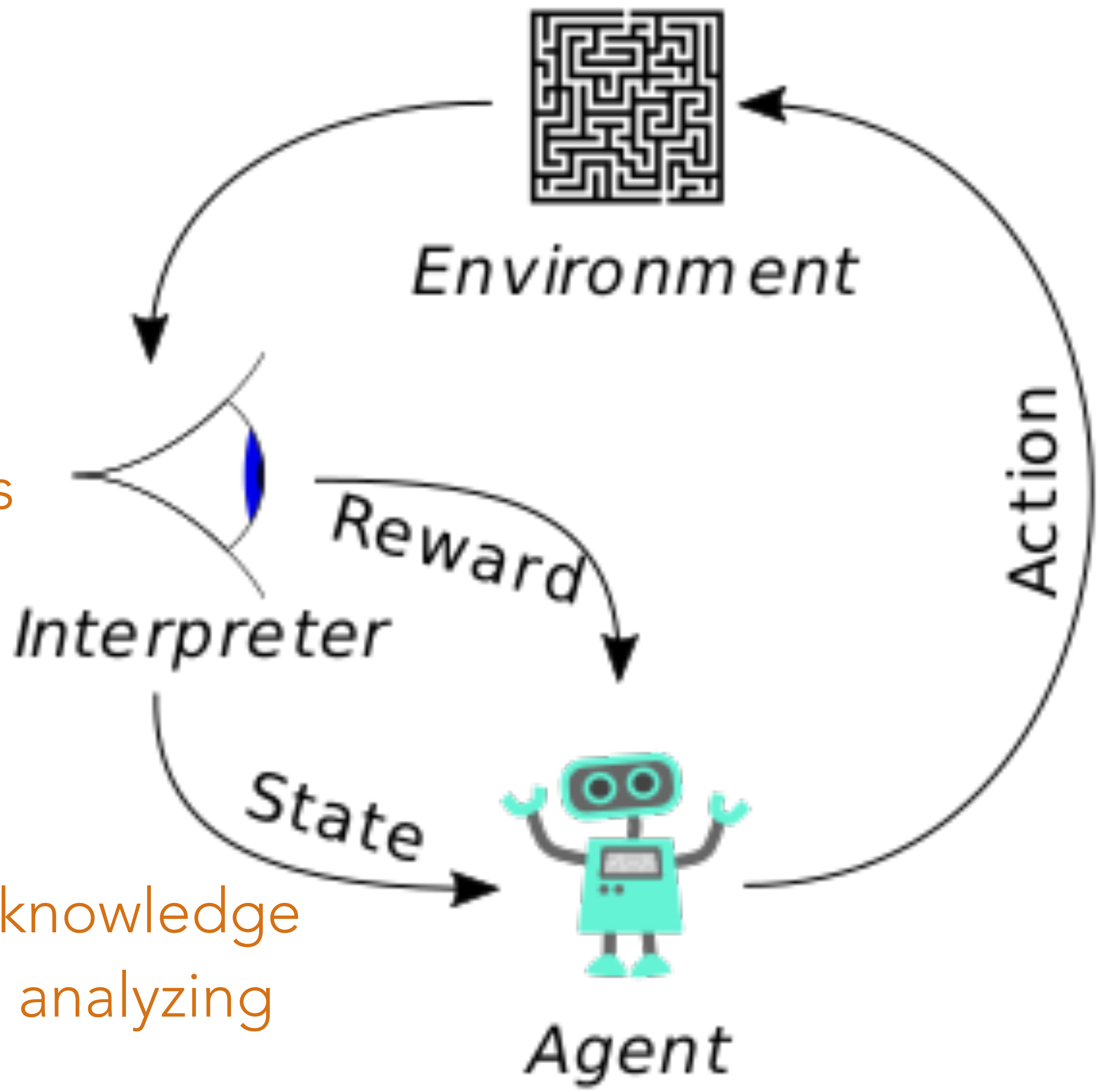


REINFORCEMENT LEARNING & SCIENTIFIC METHOD

Scientist trying to decide what experiment to do next

perform experiment,
gather data

statistical analysis

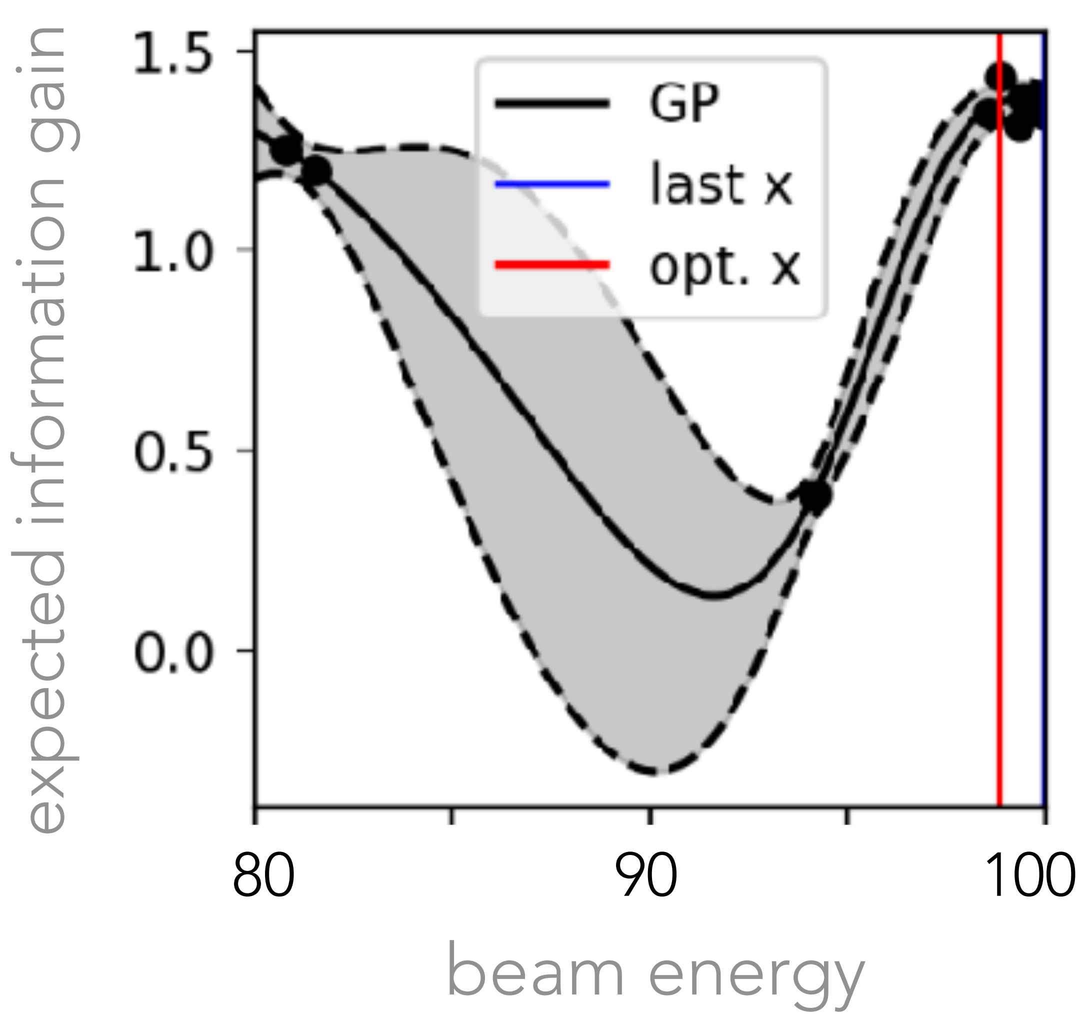


decide which
experiment to
perform

updated knowledge
based on analyzing
data

Proof-of-principle algorithm can:

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



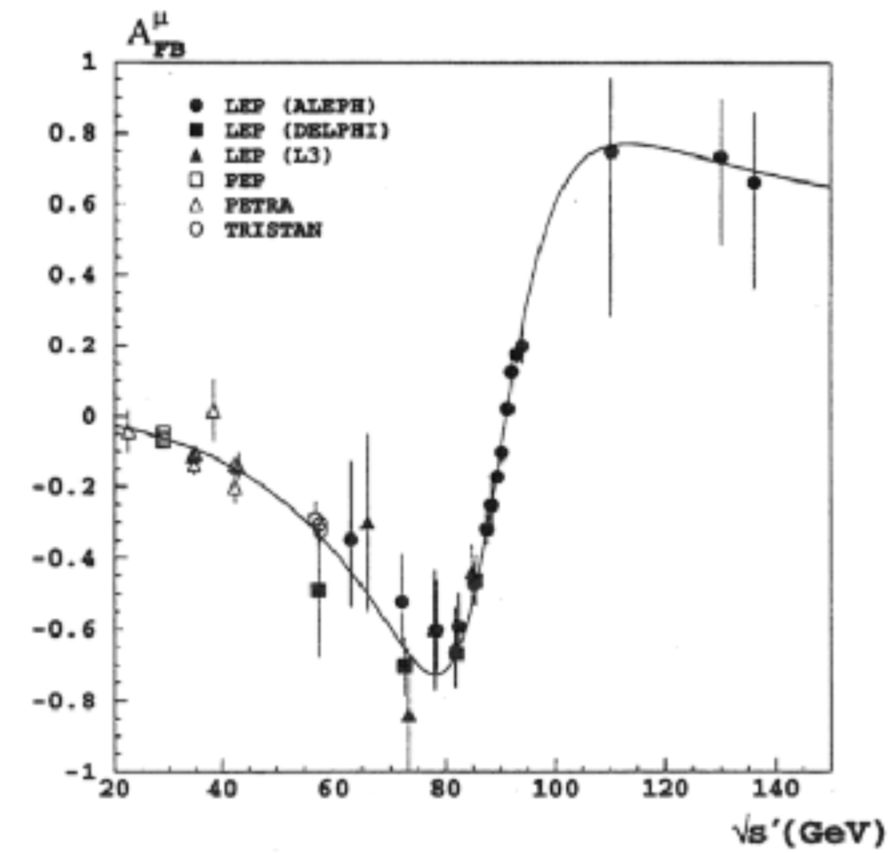


Figure 2: Measured forward-backward asymmetries of muon-pair production compared with the model independent fit results.

CONCLUSIONS

The developments in machine learning and AI go way beyond improved classifiers and have the potential to revolutionize physics

- likelihood-free inference and generative models are two particularly exciting areas

Our understanding of how to leverage our prior physics knowledge while letting machine learning do what it's good at is maturing.

- ability to inject and extract physics knowledge from models

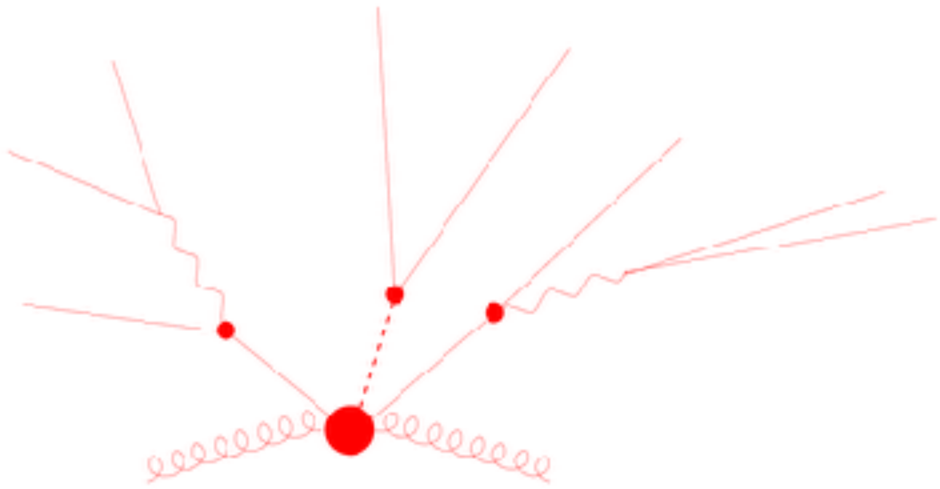
- exploit hierarchical structure of data

Harnessing the full potential of these techniques will require deep integration into our scientific workflow

SUB-ATOMIC SCALE

pencil & paper calculable from first principles

$p(z_1 \mid \theta)$

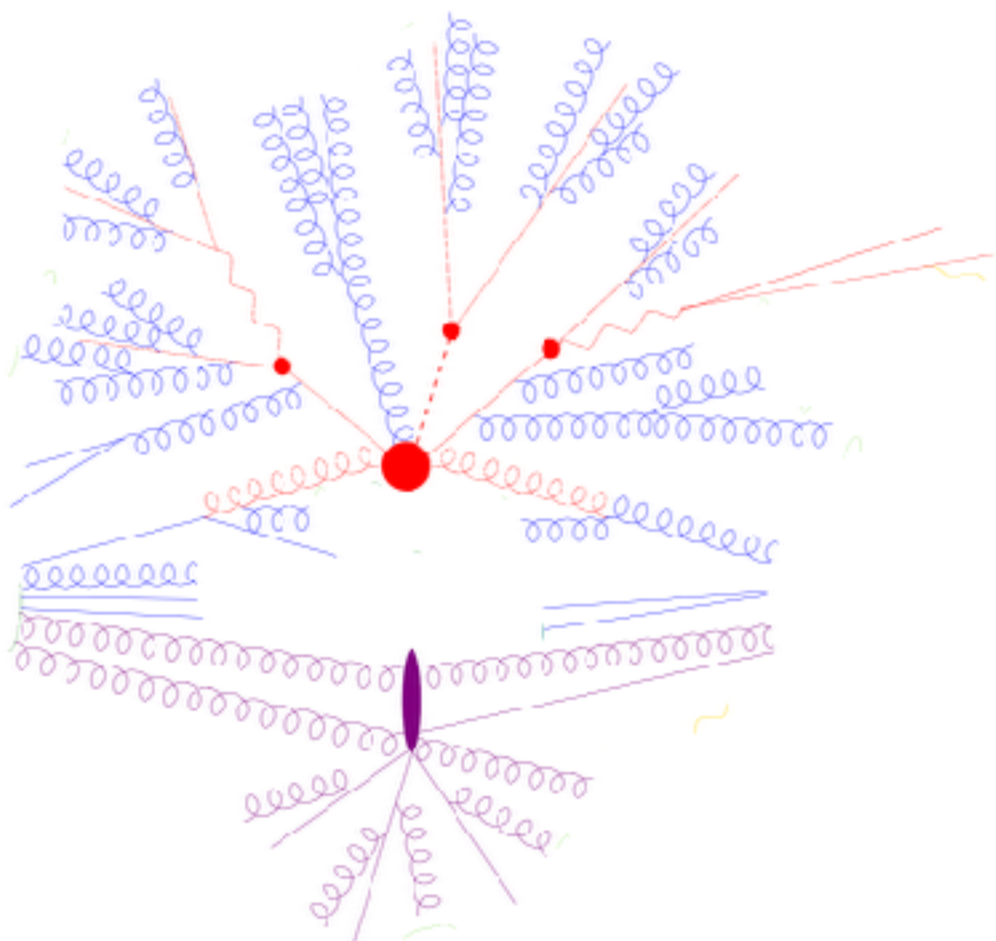


pencil & paper calculable from first principles

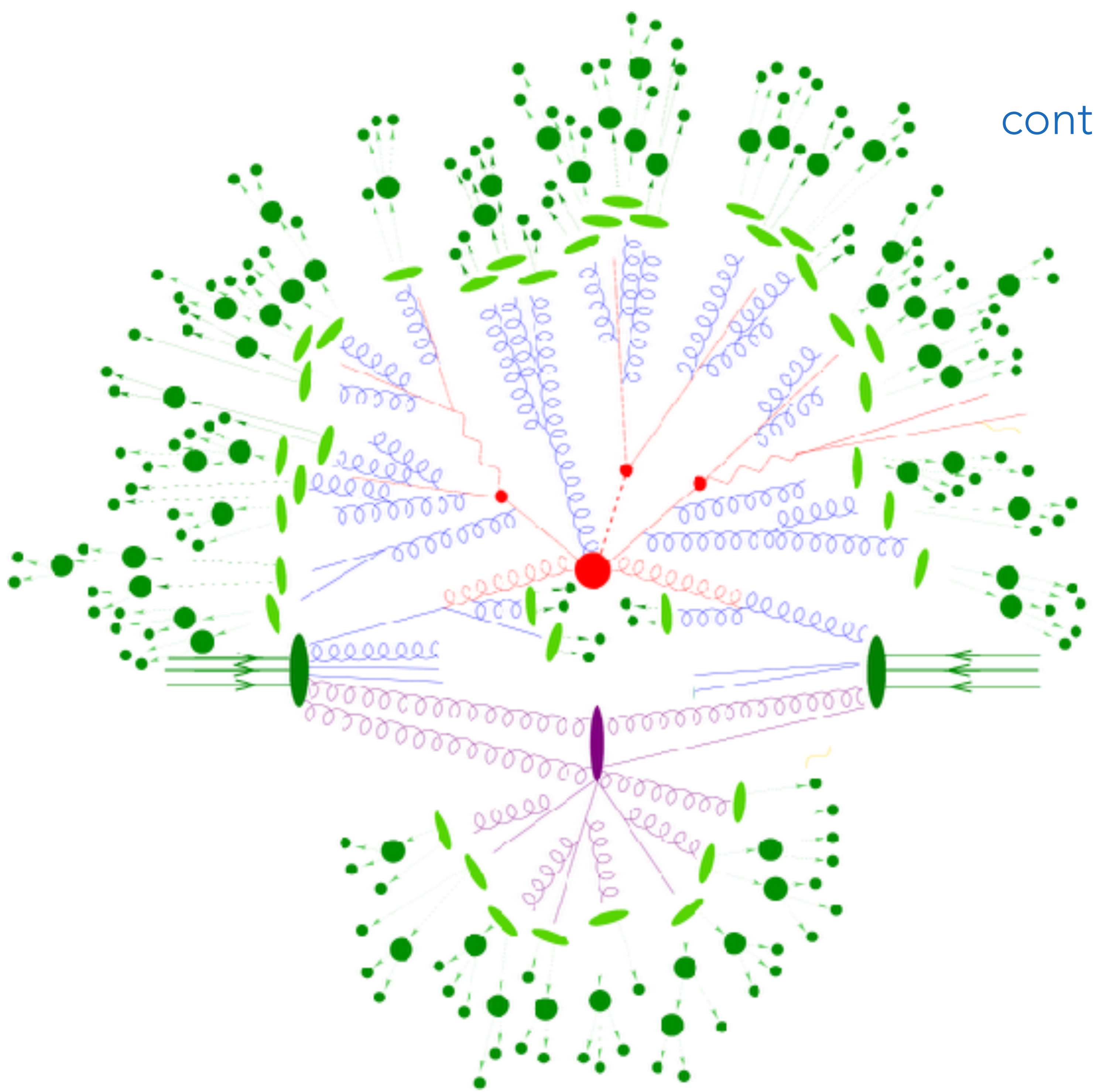
$p(z_1 \mid \theta)$

controlled approximation of first principles

$p(z_2 \mid z_1, \nu_1)$



SUB-ATOMIC SCALE



pencil & paper calculable from first principles

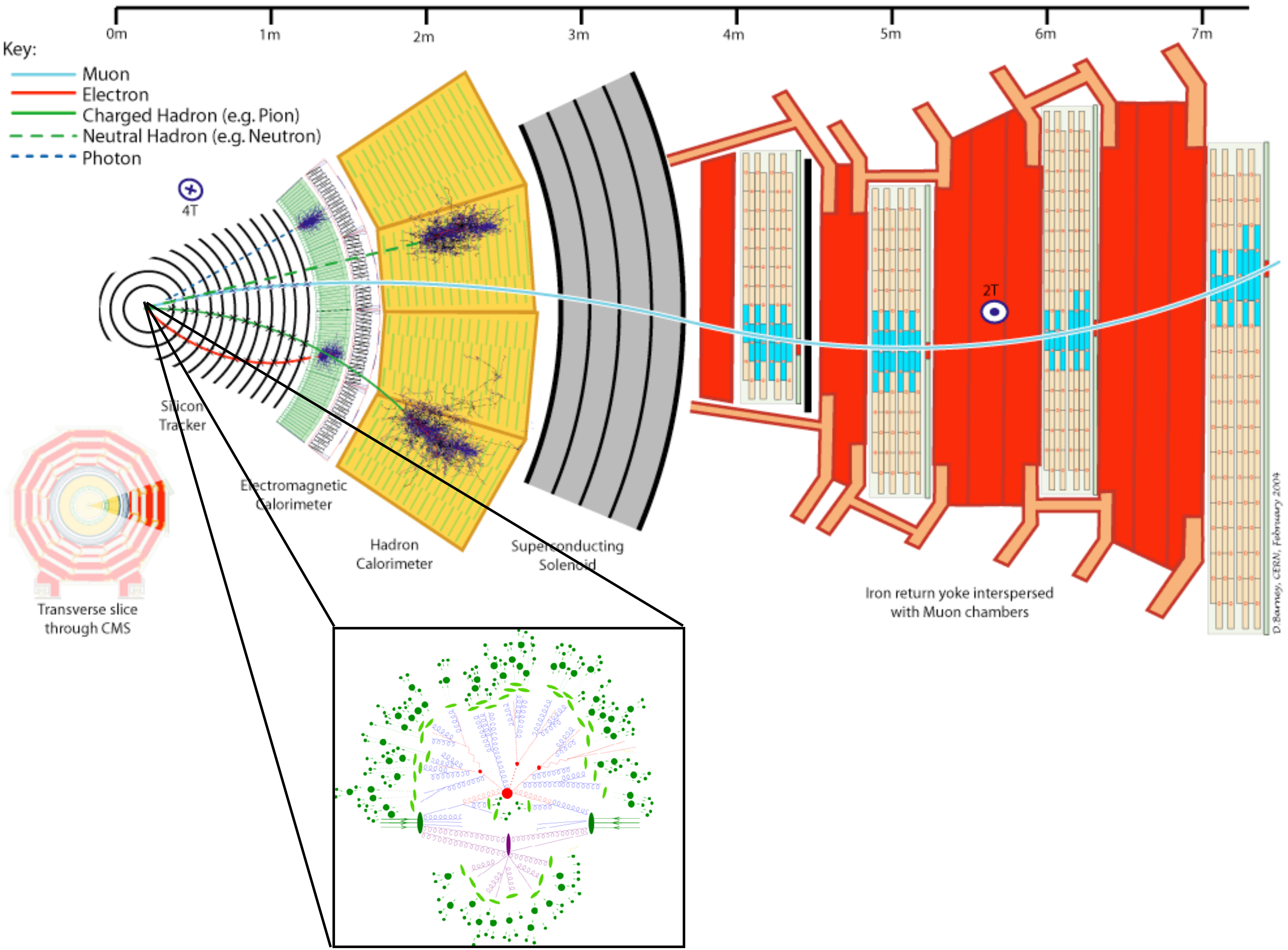
$p(z_1 \mid \theta)$

controlled approximation of first principles

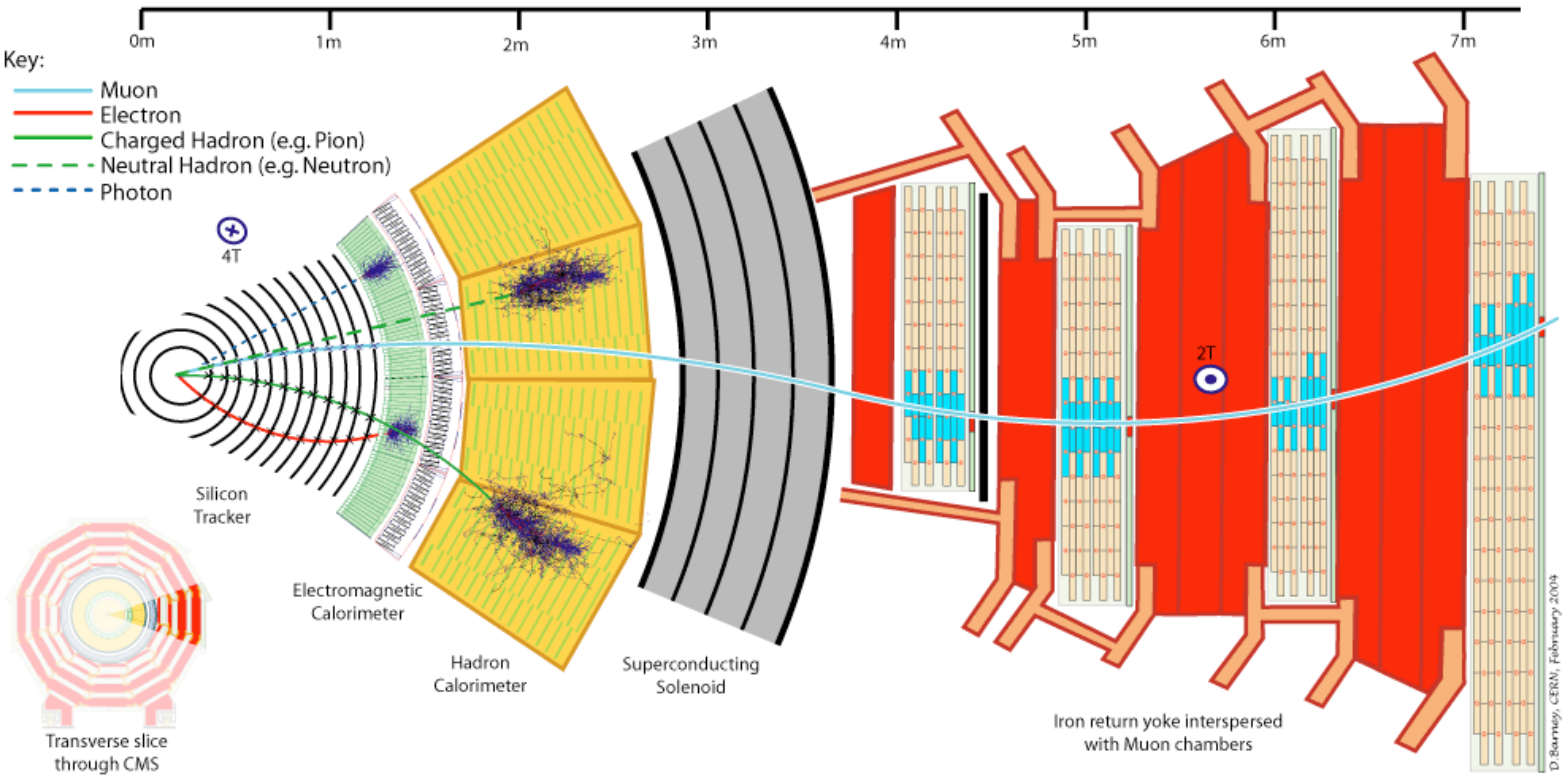
$p(z_2 \mid z_1, \nu_1)$

phenomenological model

$p(z_3 \mid z_2, \nu_2)$



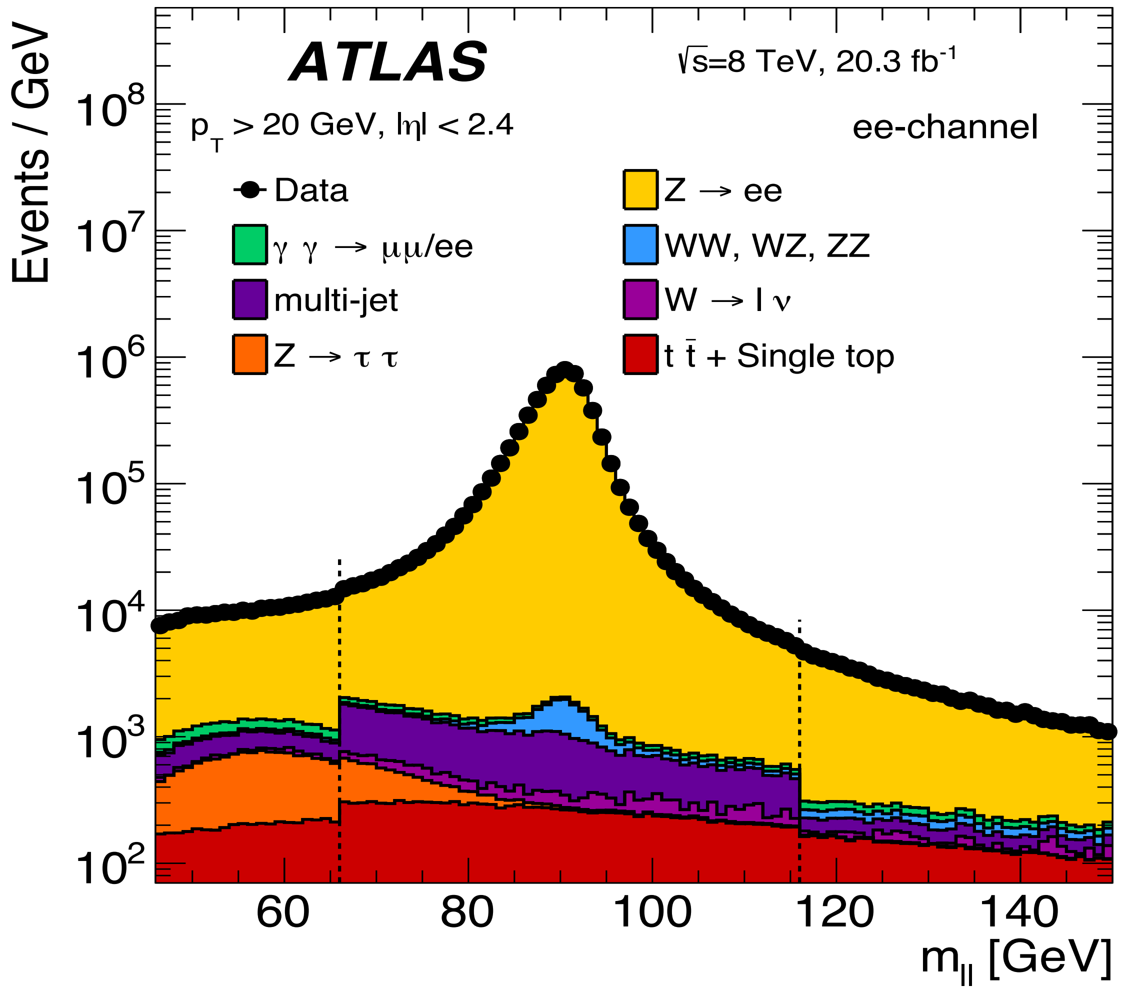
0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS
D.Barney, CERN, February 2004

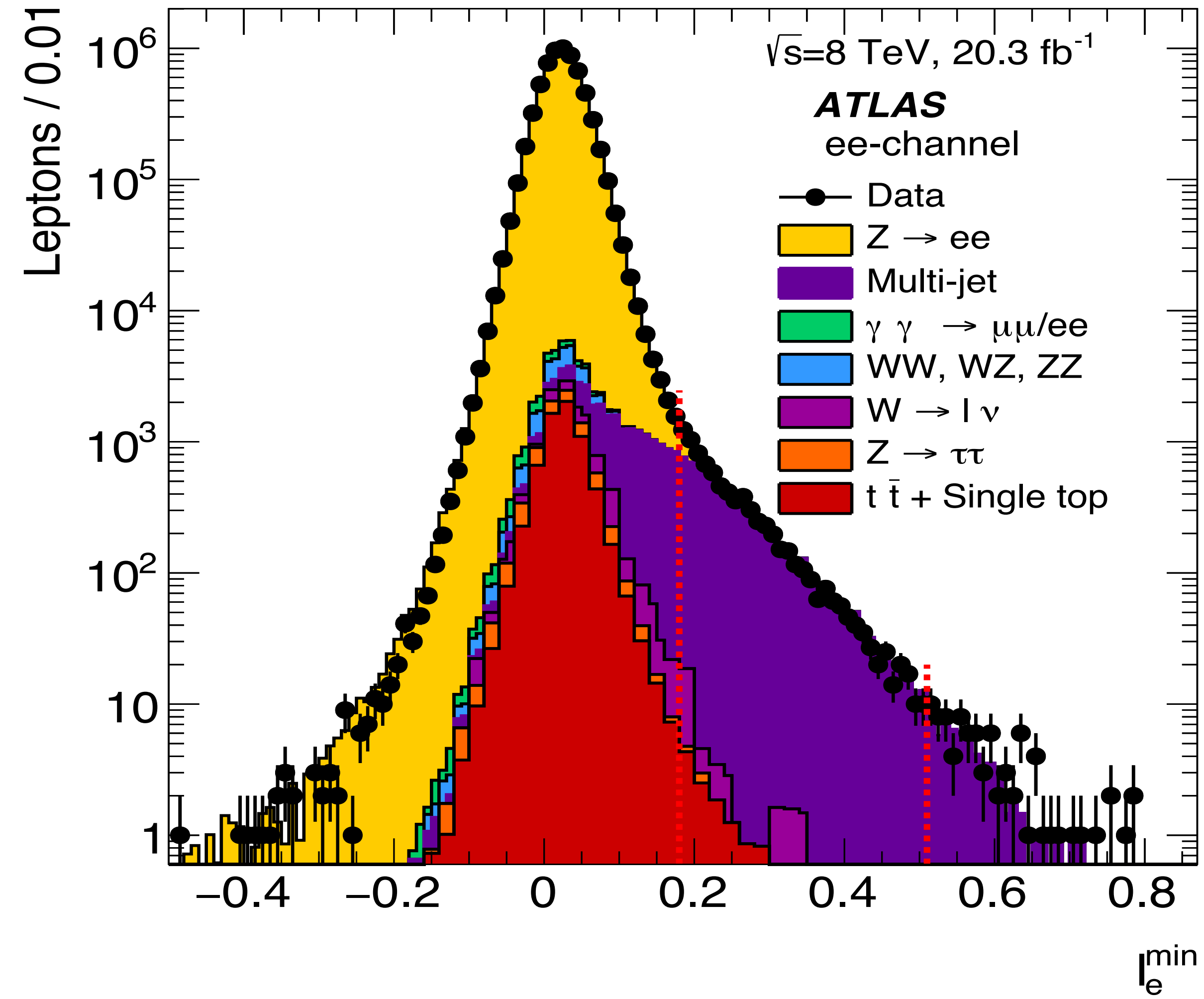


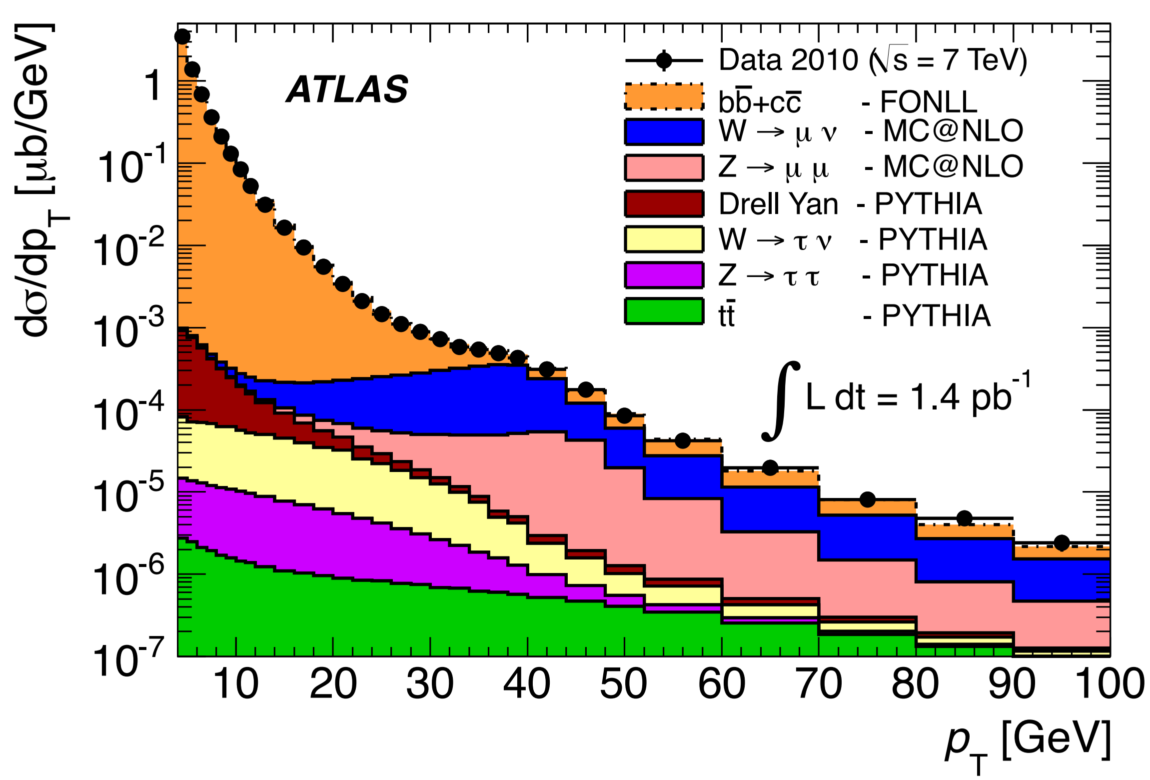
Detector Simulation **p(x | z3, ν_3):**

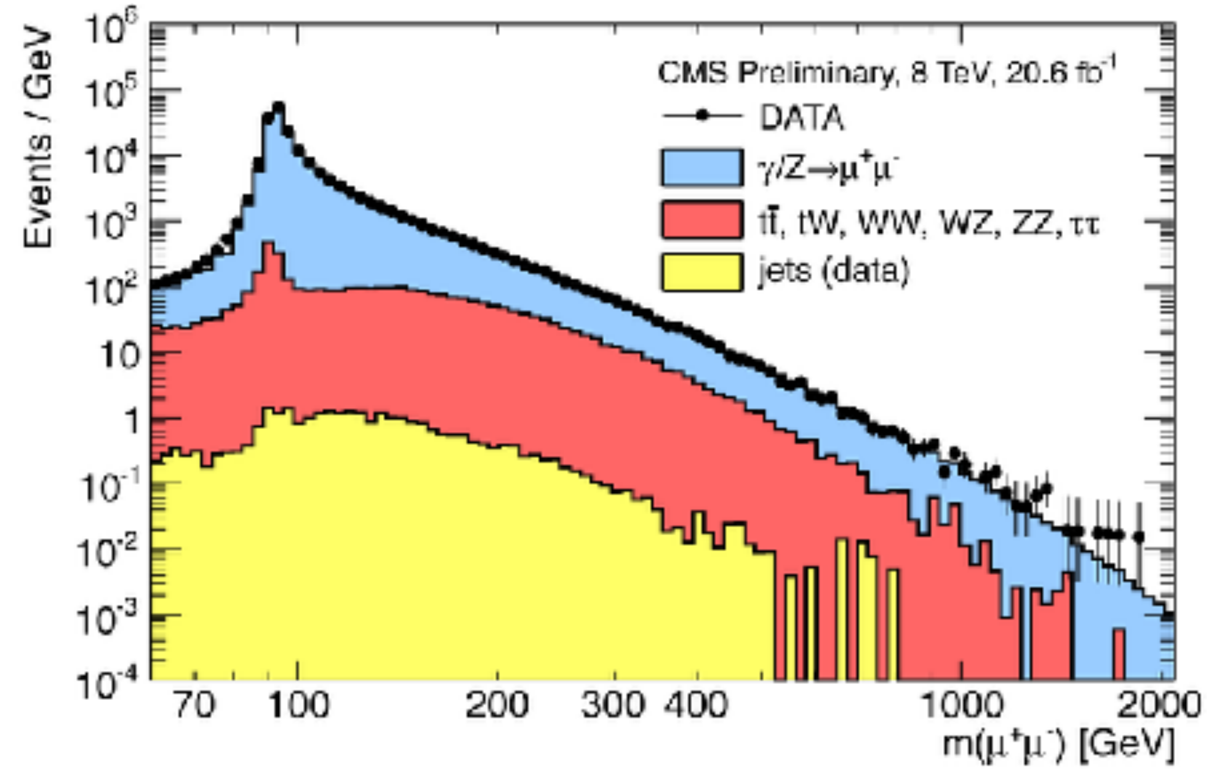
- detailed engineering (CAD)
- in situ measurements of temperature, magnetic field, alignment, calibration constants
- first-principles description of interaction of particles with matter
- measured interaction of particles with matter

HIGH FIDELITY SIMULATION

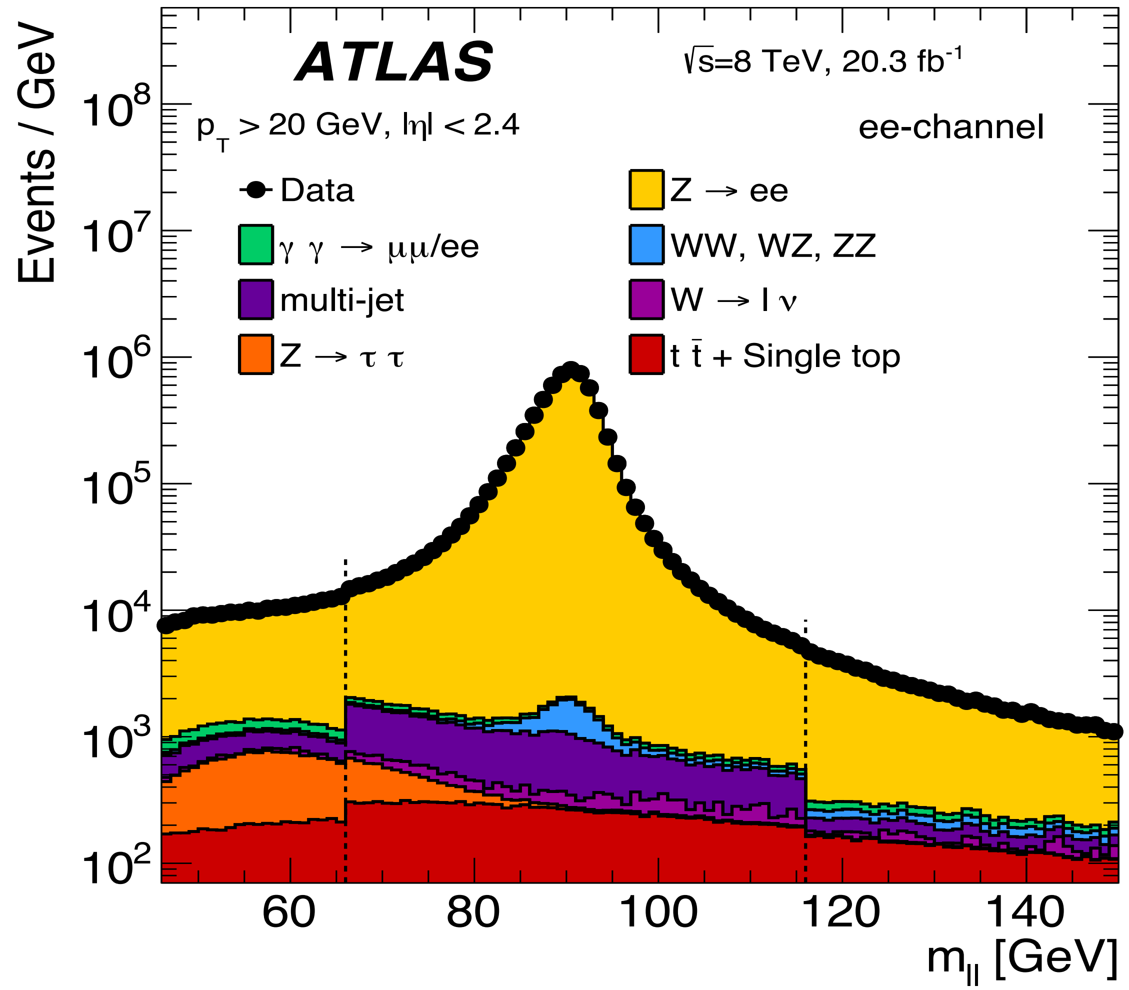


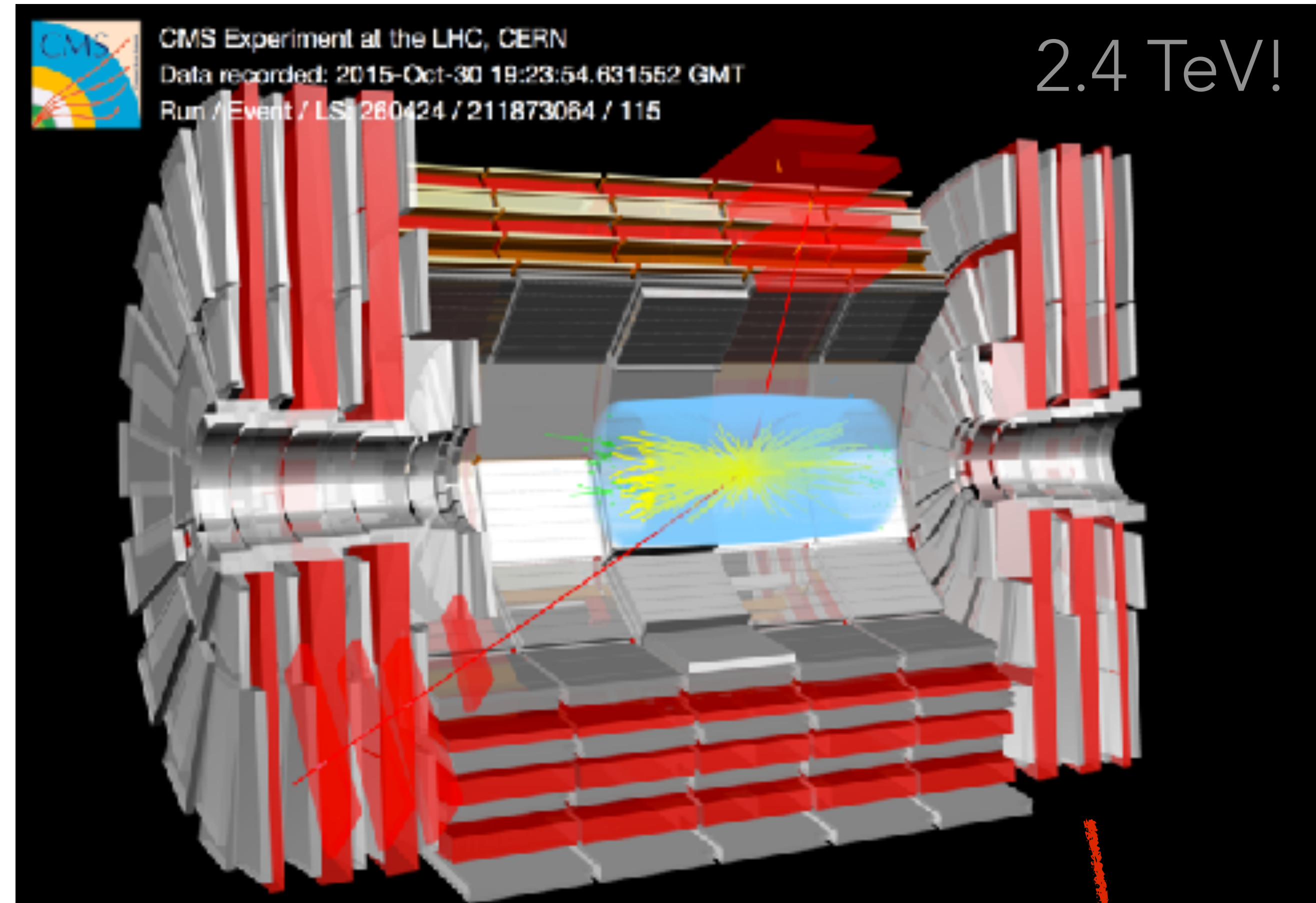


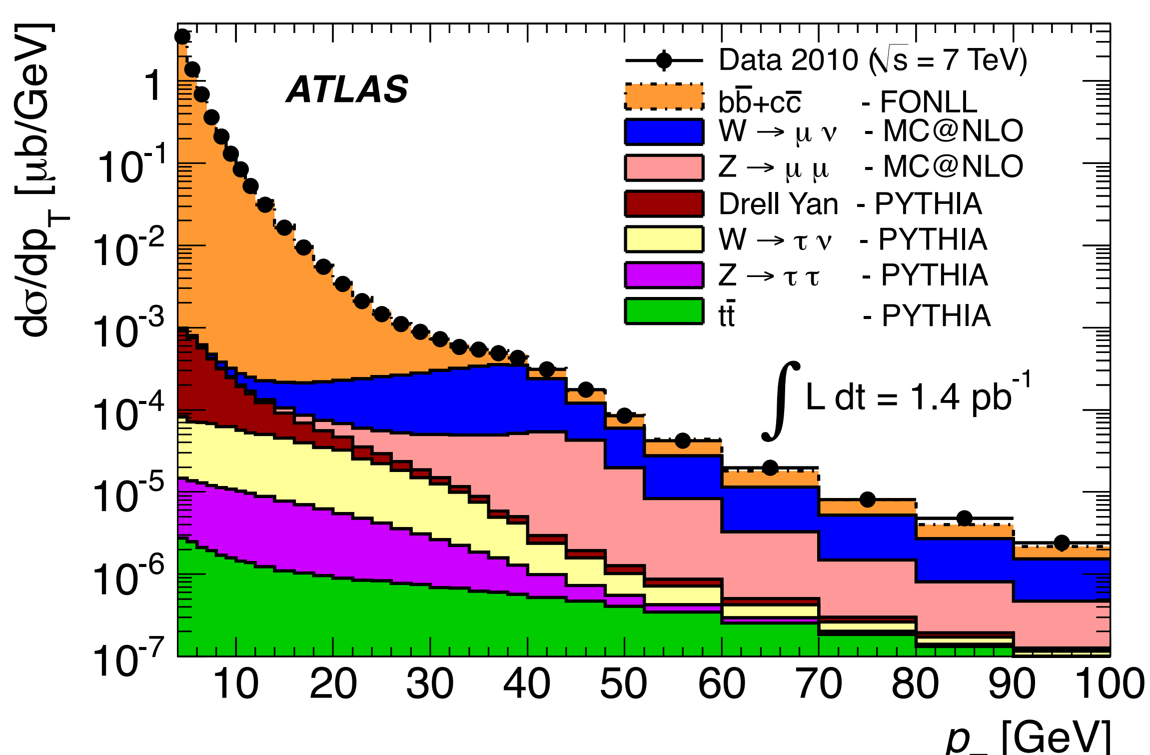


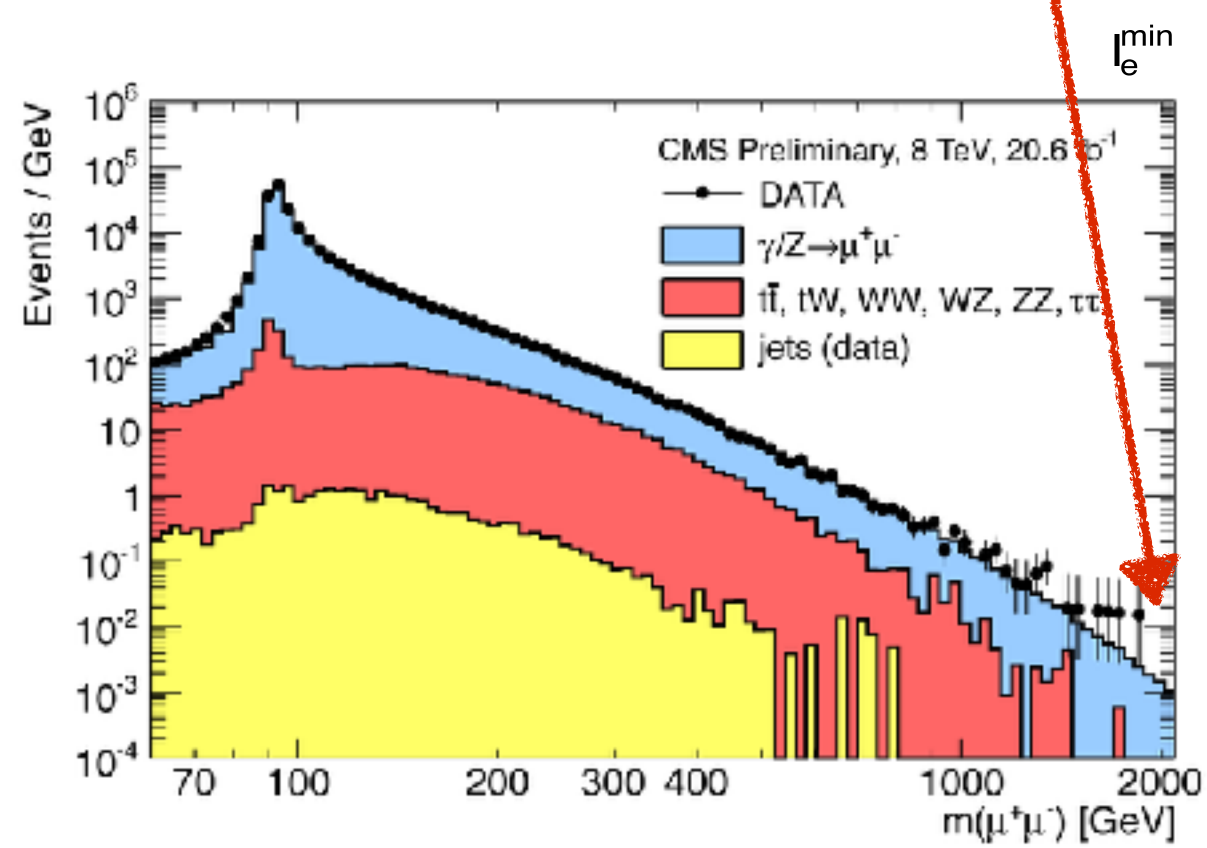


HIGH FIDELITY SIMULATION

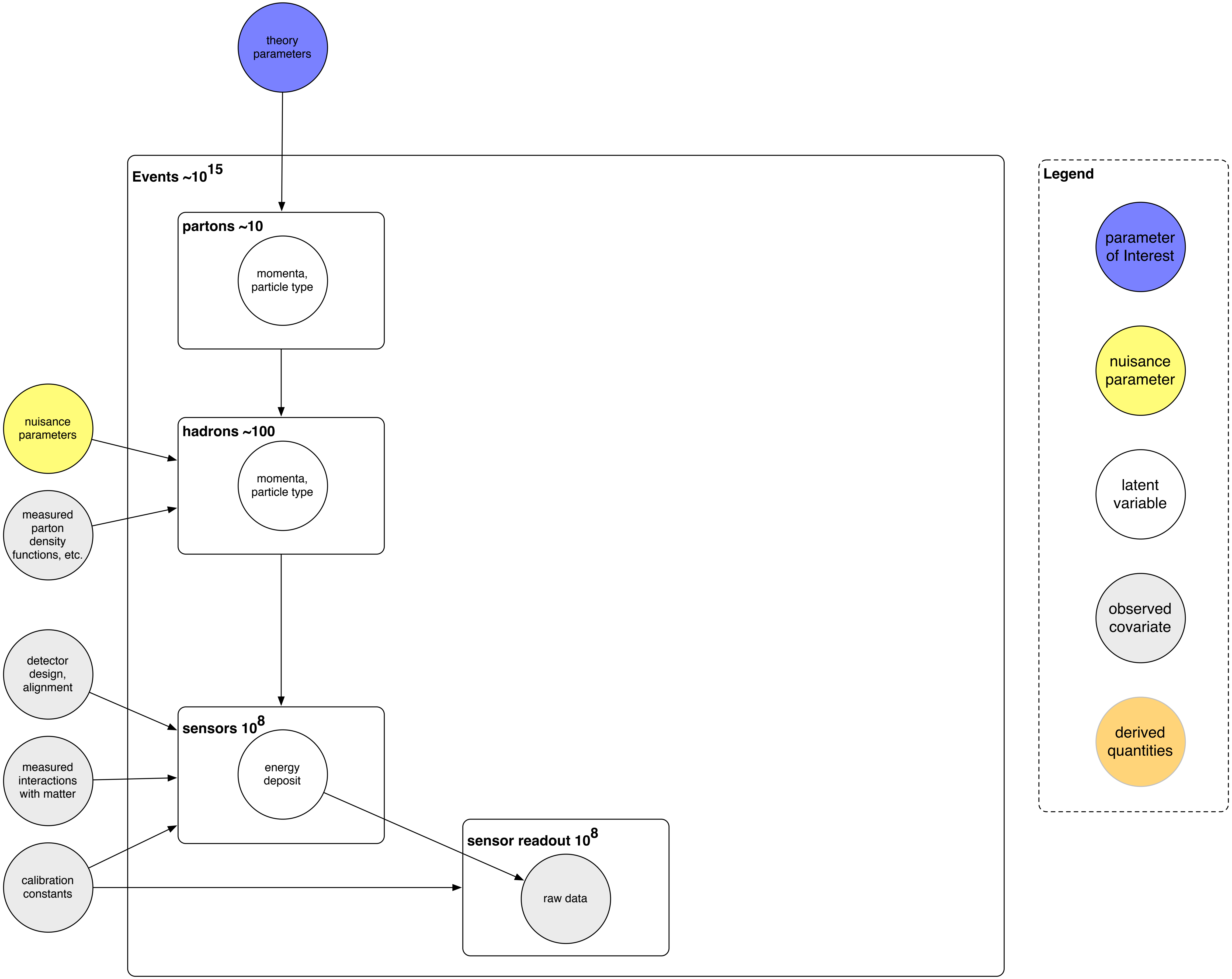






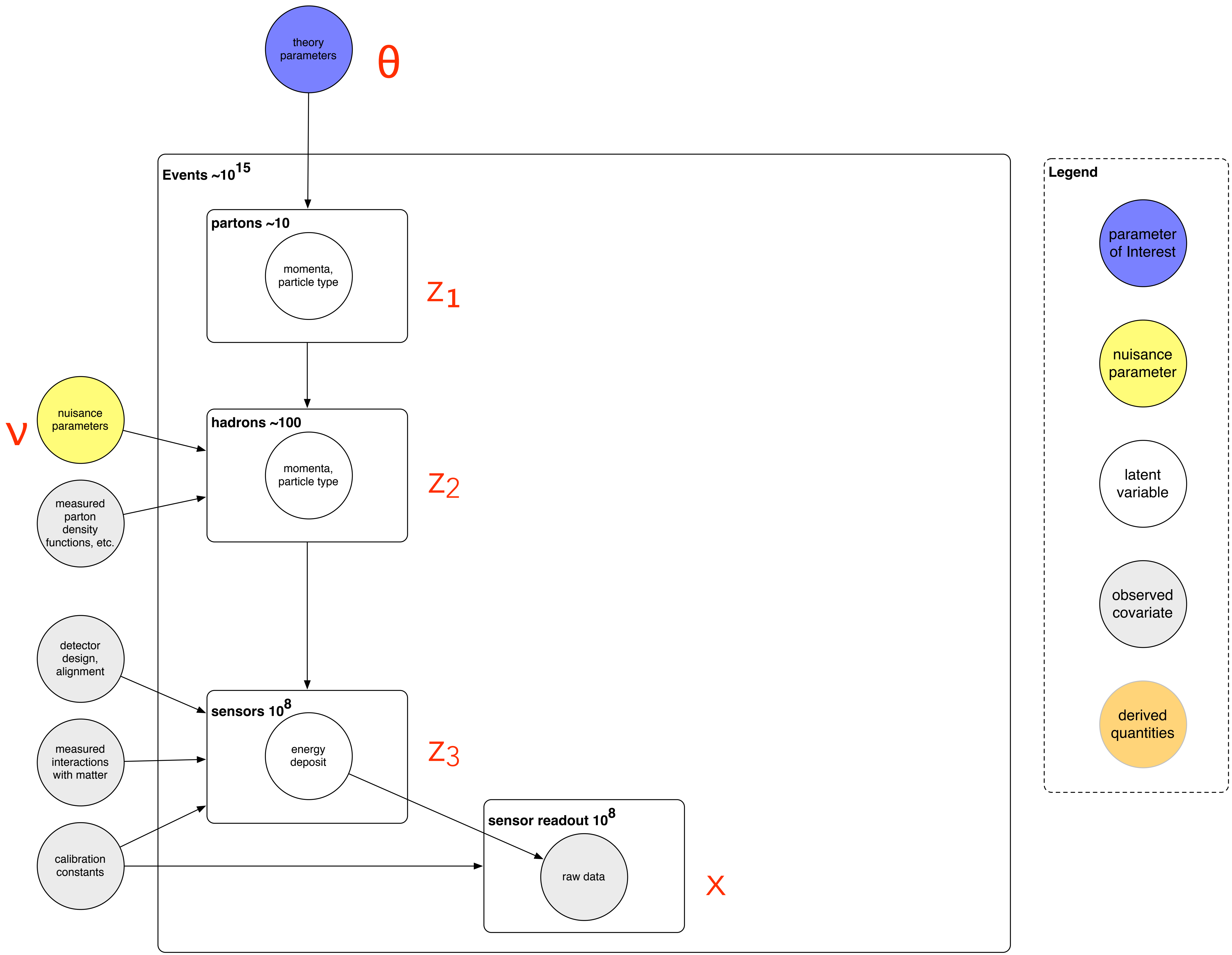


SIMULATION

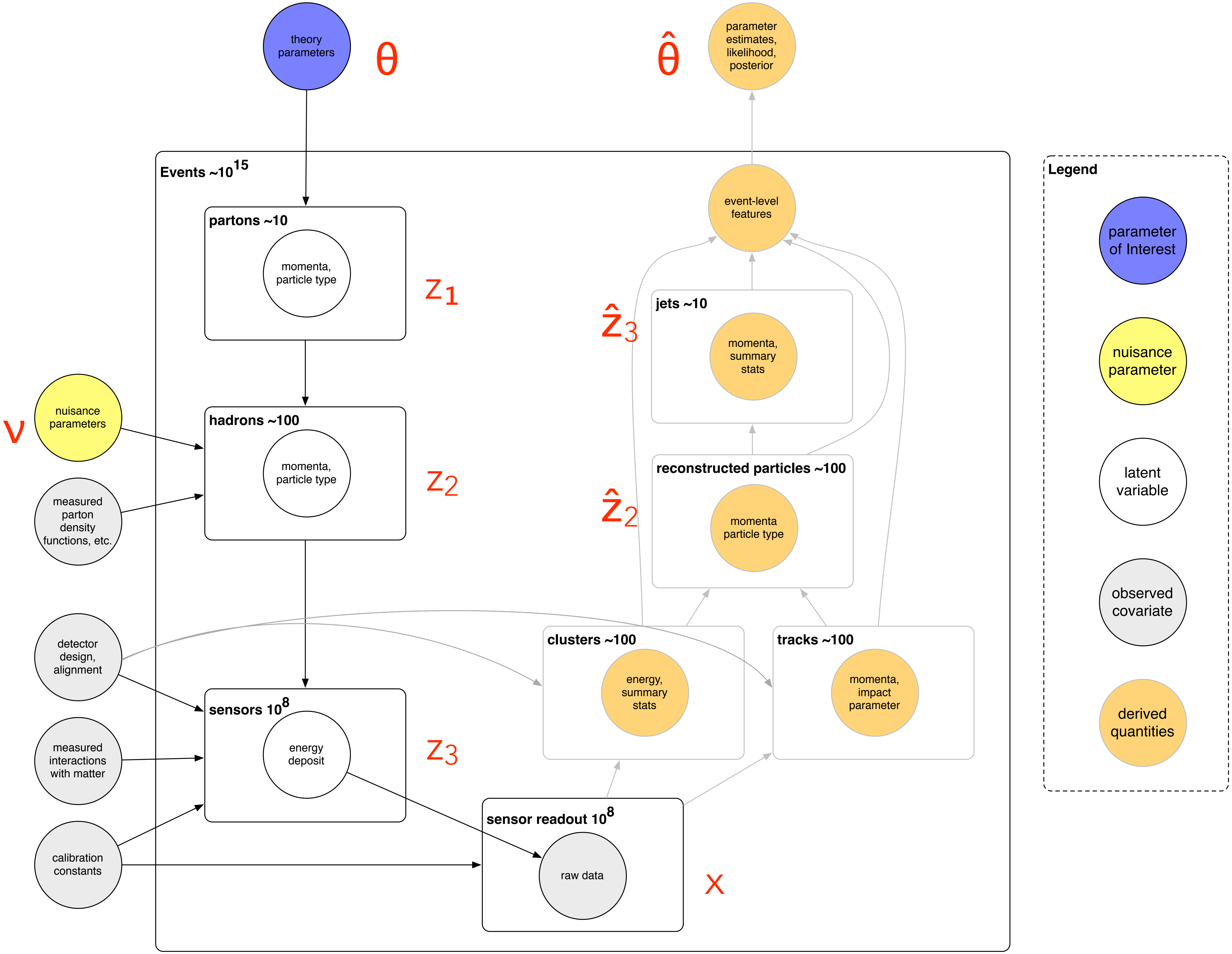


theory parameters
Events ~10^15
partons ~10
momenta, particle type
hadrons ~100
momenta, particle type
nuisance parameters
measured parton density functions, etc.
detector design, alignment
measured interactions with matter
calibration constants
sensors 10^8
energy deposit
sensor readout 10^8
raw data
Legend
parameter of Interest
nuisance parameter
latent variable
observed covariate
derived quantities

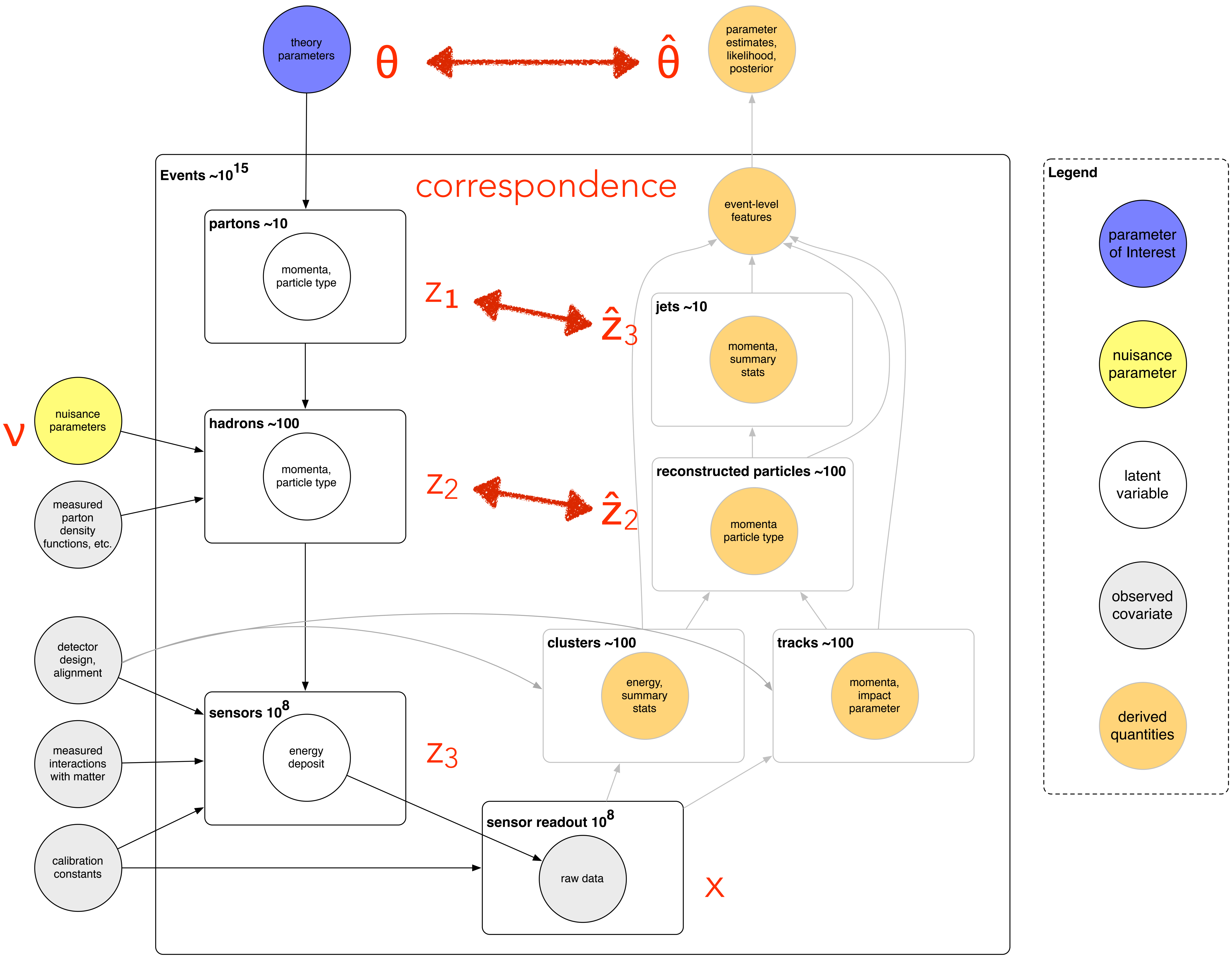
SIMULATION


theory parameters
θ
Events ~10^{15}
partons ~10
momenta, particle type
z_1
hadrons ~100
momenta, particle type
z_2
ν
nuisance parameters
measured parton density functions, etc.
detector design, alignment
measured interactions with matter
calibration constants
sensors 10^8
energy deposit
z_3
sensor readout 10^8
raw data
x
Legend
parameter of Interest
nuisance parameter
latent variable
observed covariate
derived quantities

SIMULATION + RECONSTRUCTION



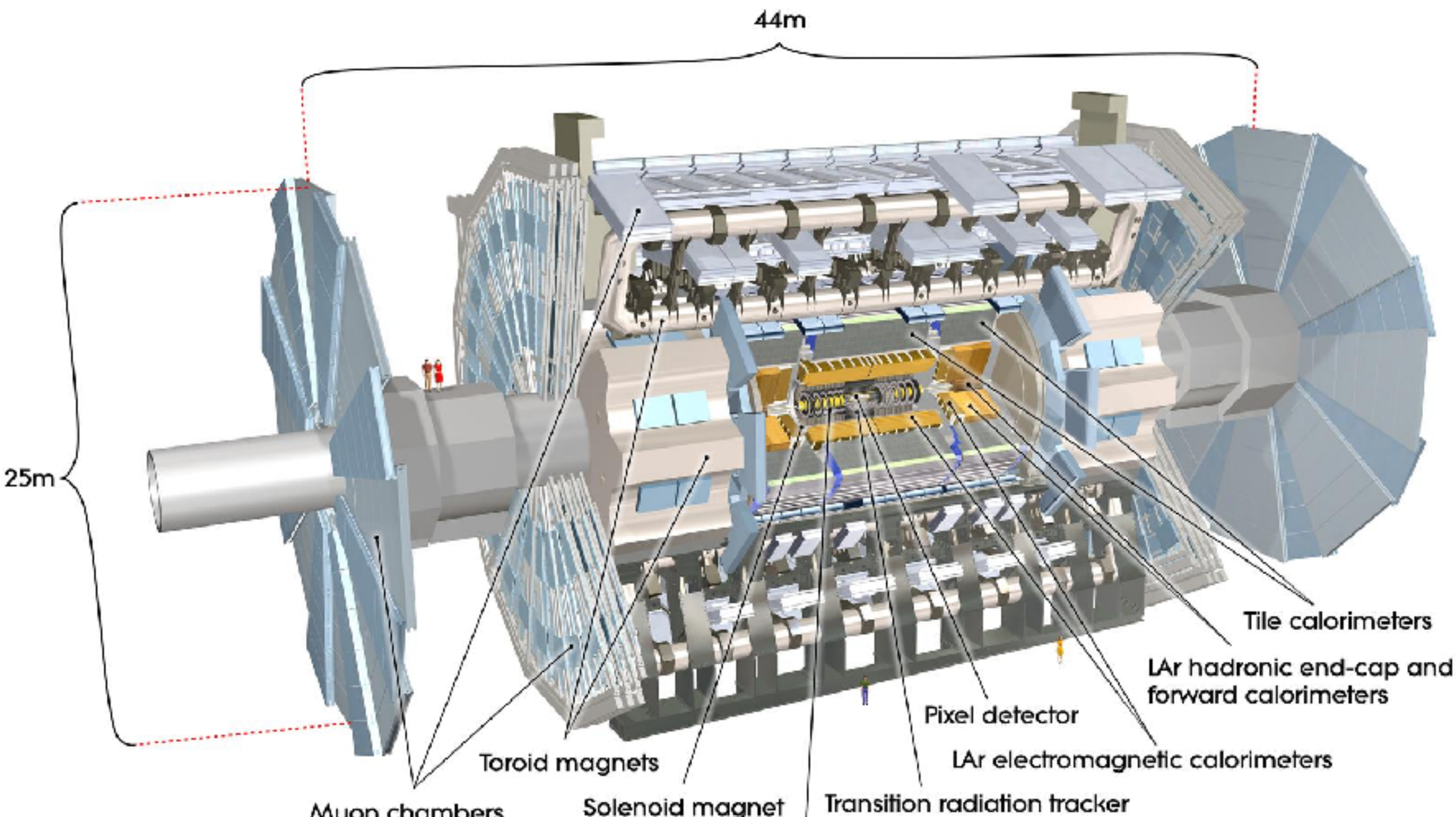
SIMULATION + RECONSTRUCTION



HIGH FIDELITY SIMULATION

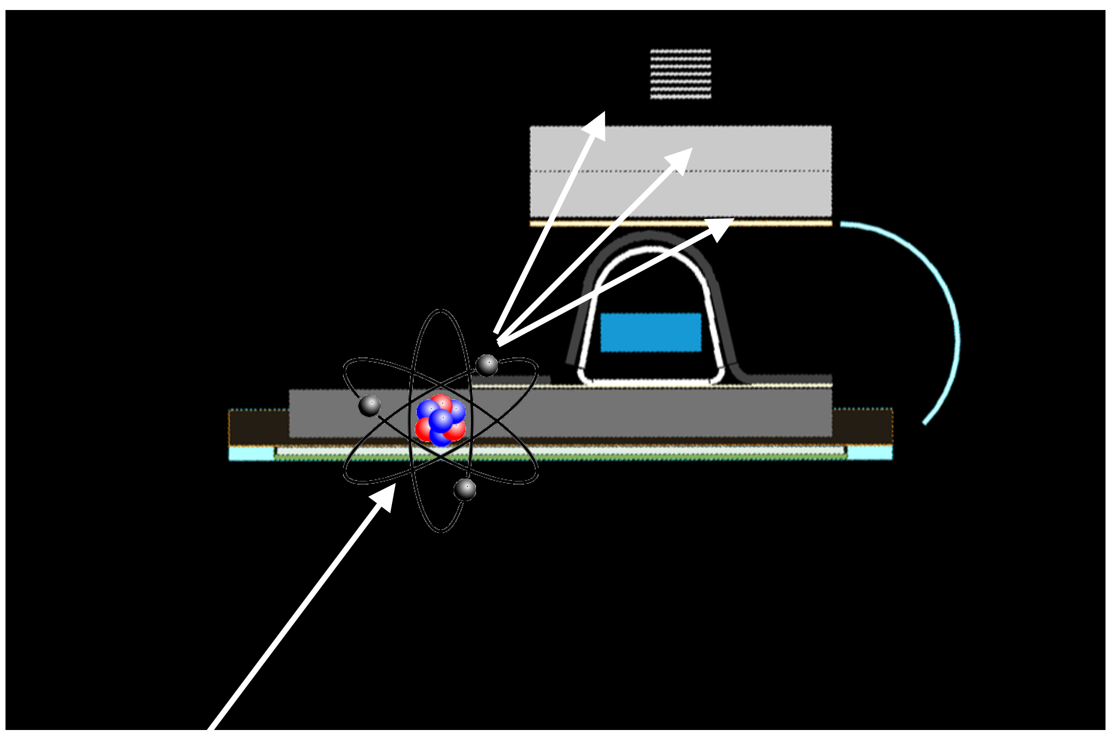
Detector is 44m long

- Detector resolves details at <mm scale; Simulation accurate!



Detector is 44m long

- Detector resolves details at <mm scale; Simulation accurate!



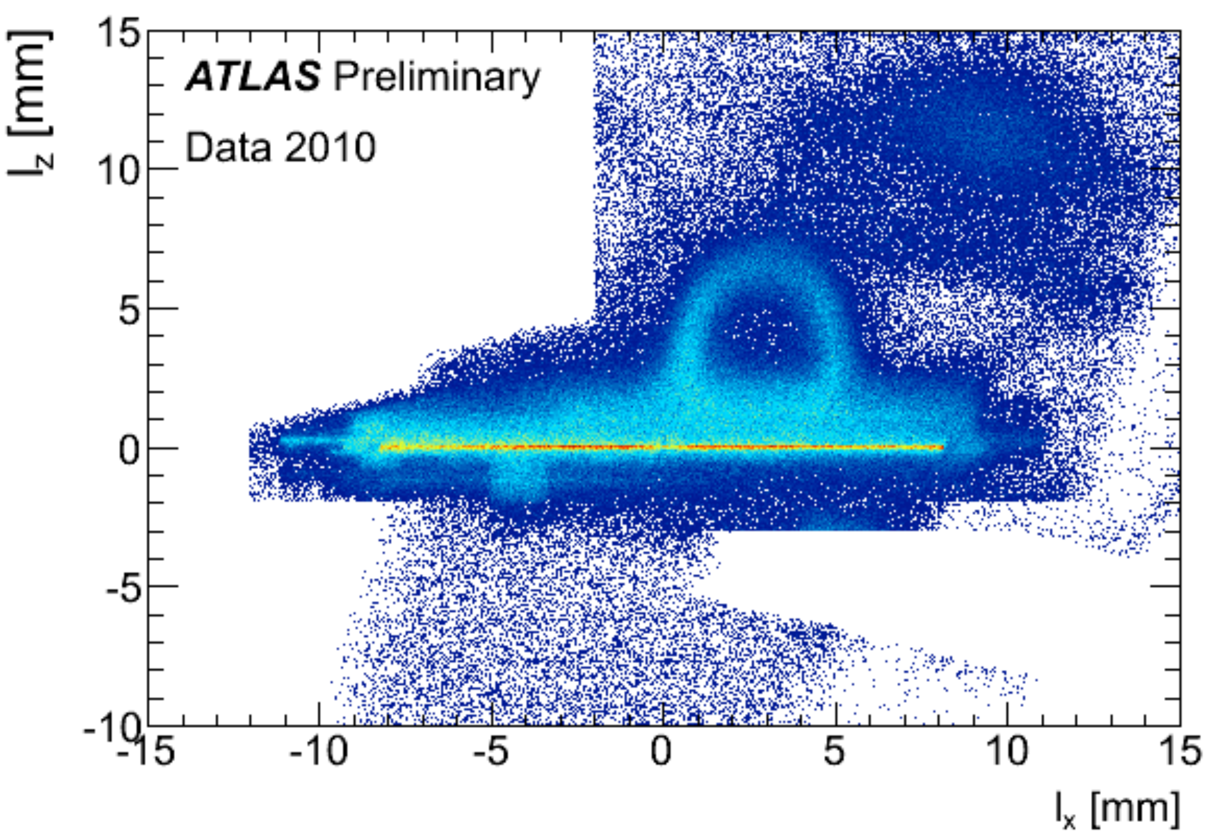


Figure:
ATLAS pixel model as described in simulation (left), tomography from vertices built from tracks for hadronic interactions (right)

Slide Credit: A. Salzburger (CERN)