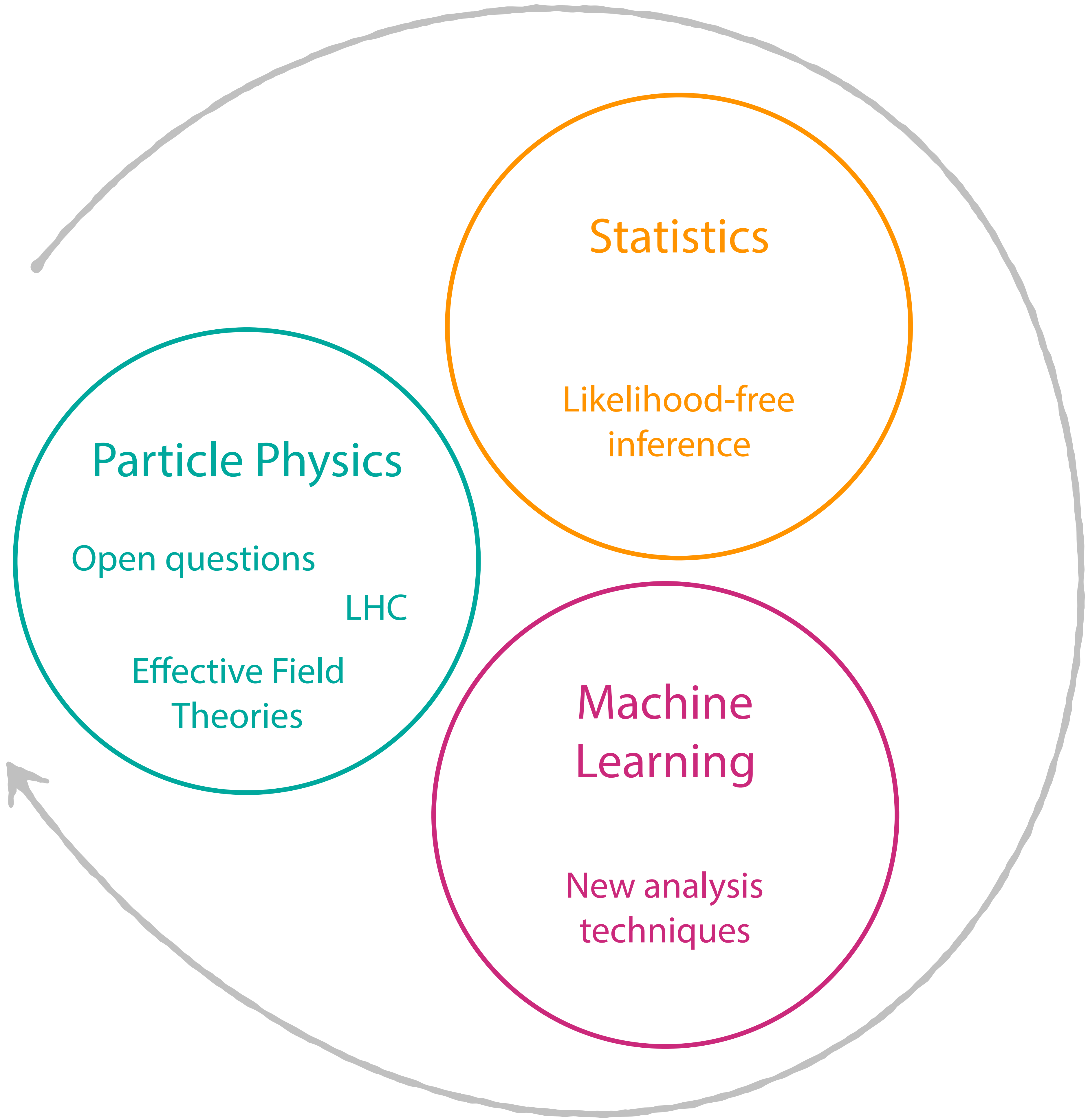
Learning from the LHC, Effectively

Johann Brehmer
Center for Cosmology and Particle Physics & Center for Data Science, NYU

CityTech physics seminar
May 10, 2018

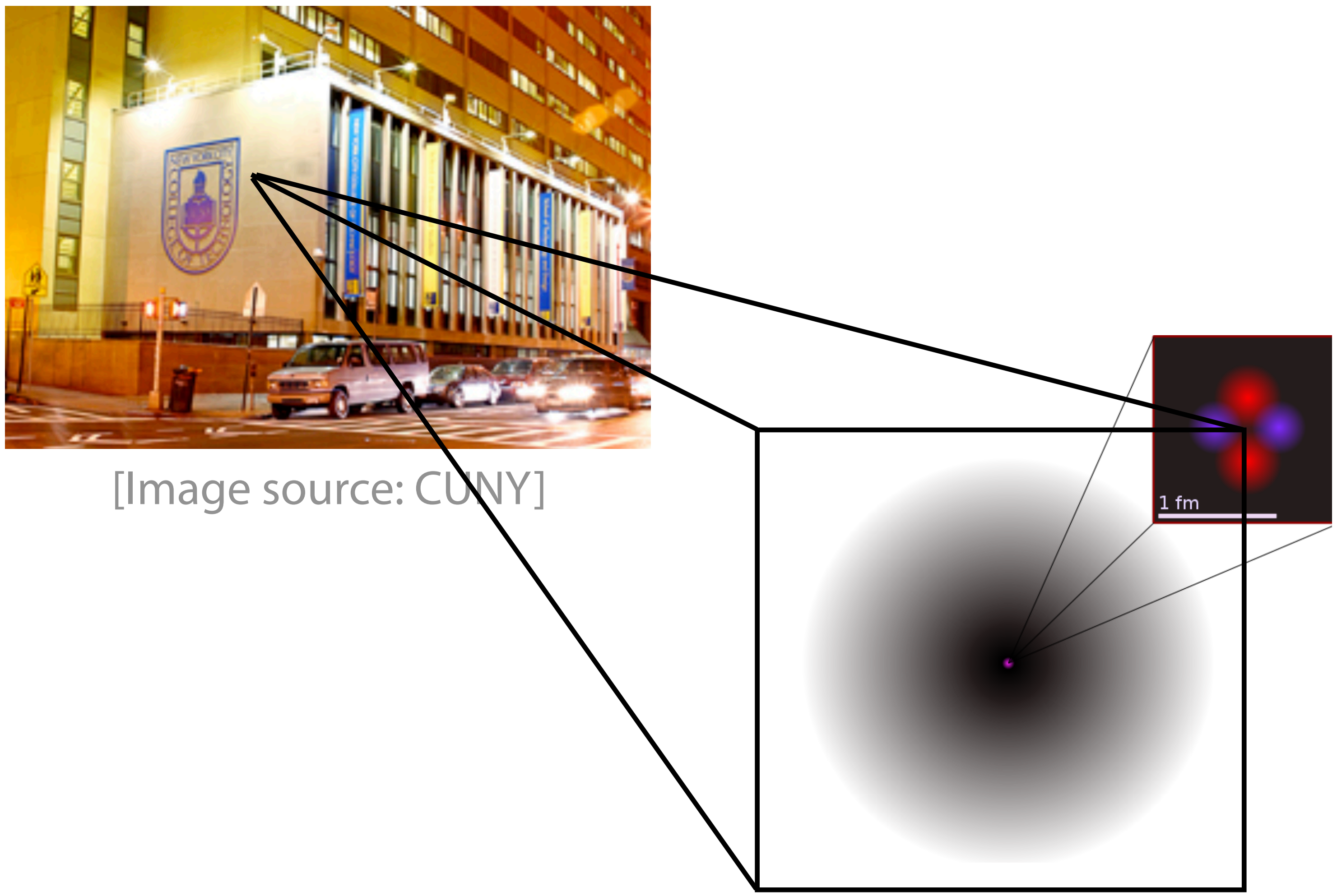

Statistics
Likelihood-free inference
Particle Physics
Open questions
LHC
Effective Field Theories
Machine Learning
New analysis techniques

Big questions for a large collider

The Standard Model of particle physics (SM)

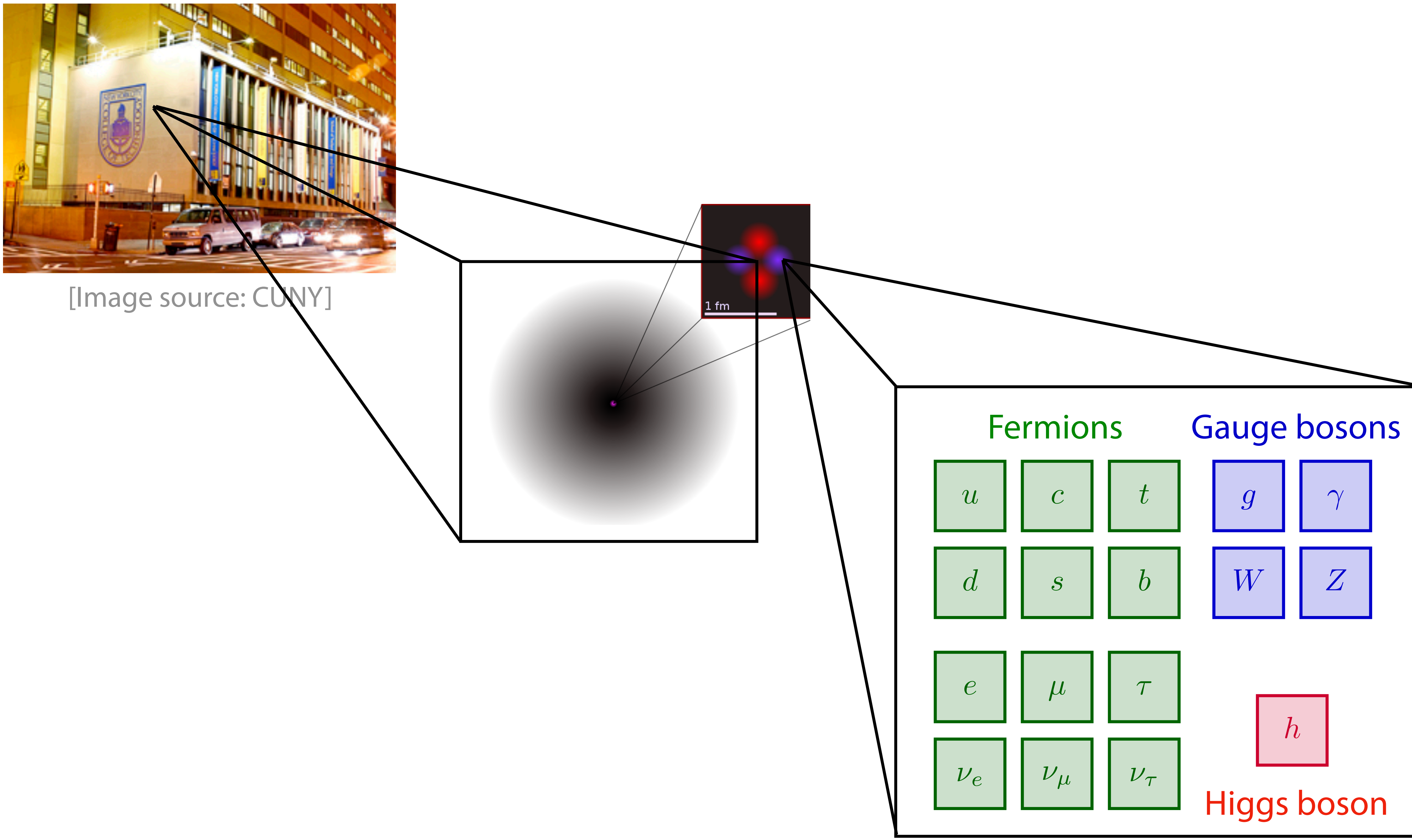
[Image source: CUNY]

The Standard Model of particle physics (SM)



[Image source: CUNY]

The Standard Model of particle physics (SM)



[Image source: CUNY]

SM track record

- ✓ All observed elementary particles
- ✓ Electromagnetism, weak force, strong force
- ✓ In agreement with virtually every experimental test, up to $\sim 10^{-10}$ precision

SM track record

- ✓ All observed elementary particles
- ✓ Electromagnetism, weak force, strong force

- ✓ In agreement with virtually every experimental test, up to $\sim 10^{-10}$ precision

- x Gravity
- x Dark matter
- x Baryon asymmetry
- x (Neutrino masses)

- x Few experimental anomalies
  (Muon g-2, proton radius, …)

- x Complicated structure, many parameters
- x Unexplained smallness of some parameters
  (Hierarchy problem, strong CP problem, cosmological constant problem, …)

SM track record

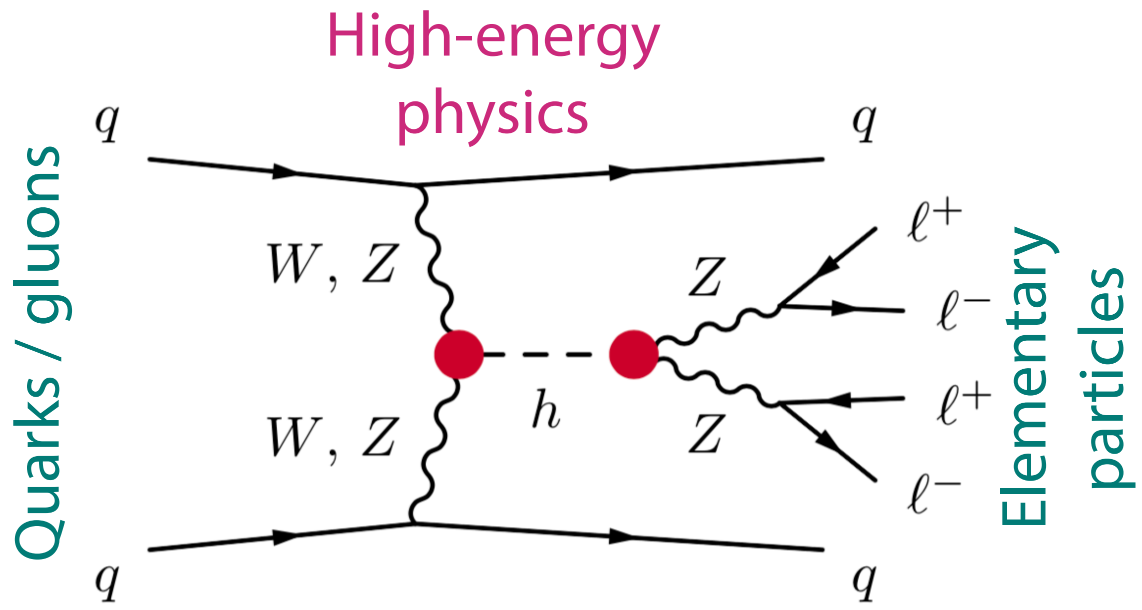
- ✓ All observed elementary particles
- ✓ Electromagnetism, weak force, strong force
- ✓ In agreement with virtually every experimental test, up to $\sim 10^{-10}$ precision

- x Gravity
- x Dark matter
- x Baryon asymmetry
- x (Neutrino masses)
- x Few experimental anomalies
  (Muon g-2, proton radius, …)
- x Complicated structure, many parameters
- x Unexplained smallness of some parameters
  (Hierarchy problem, strong CP problem, cosmological constant problem, …)

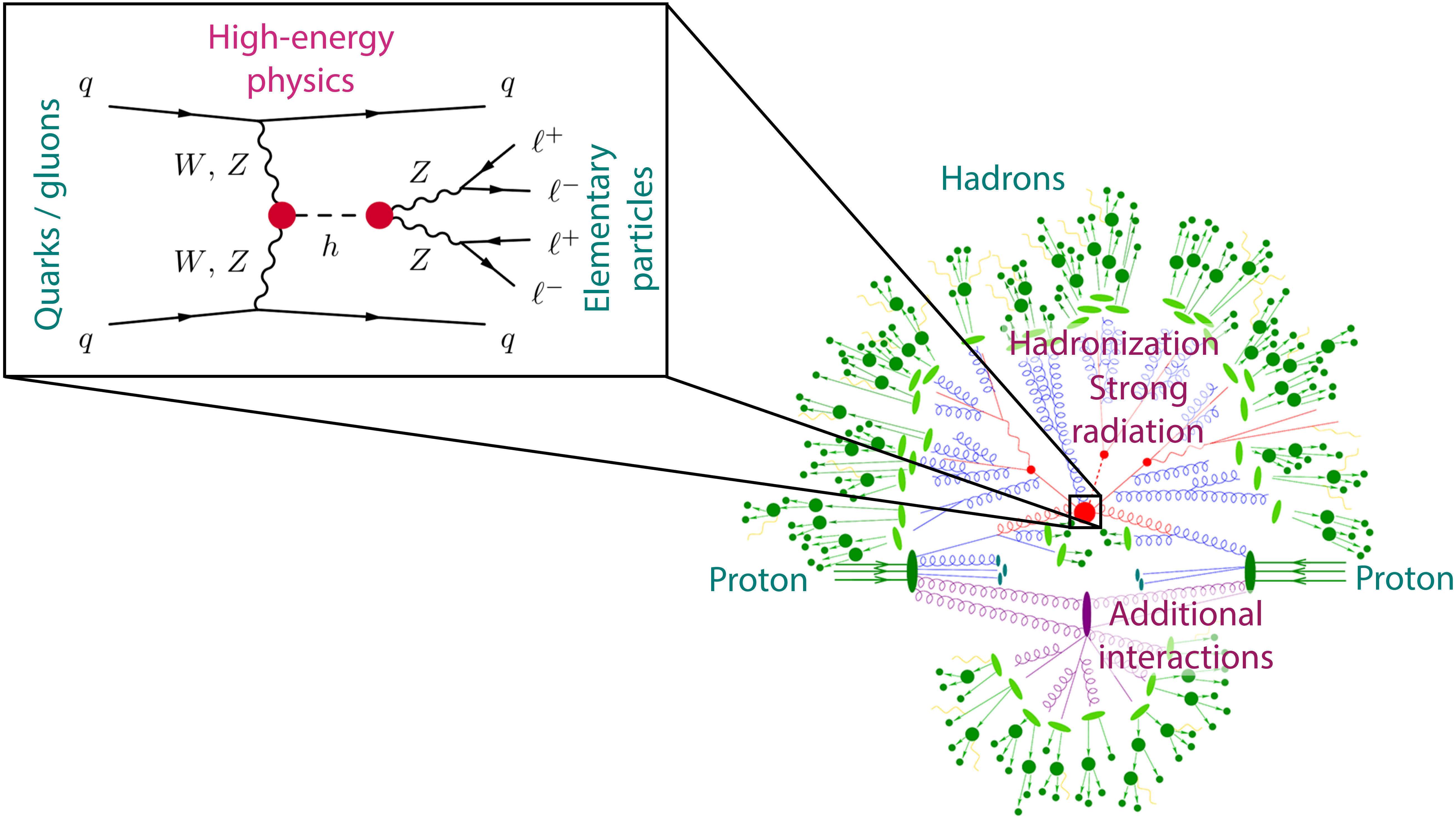
Many models of new physics to answer the open questions

$\Rightarrow$ Look for them at the LHC!

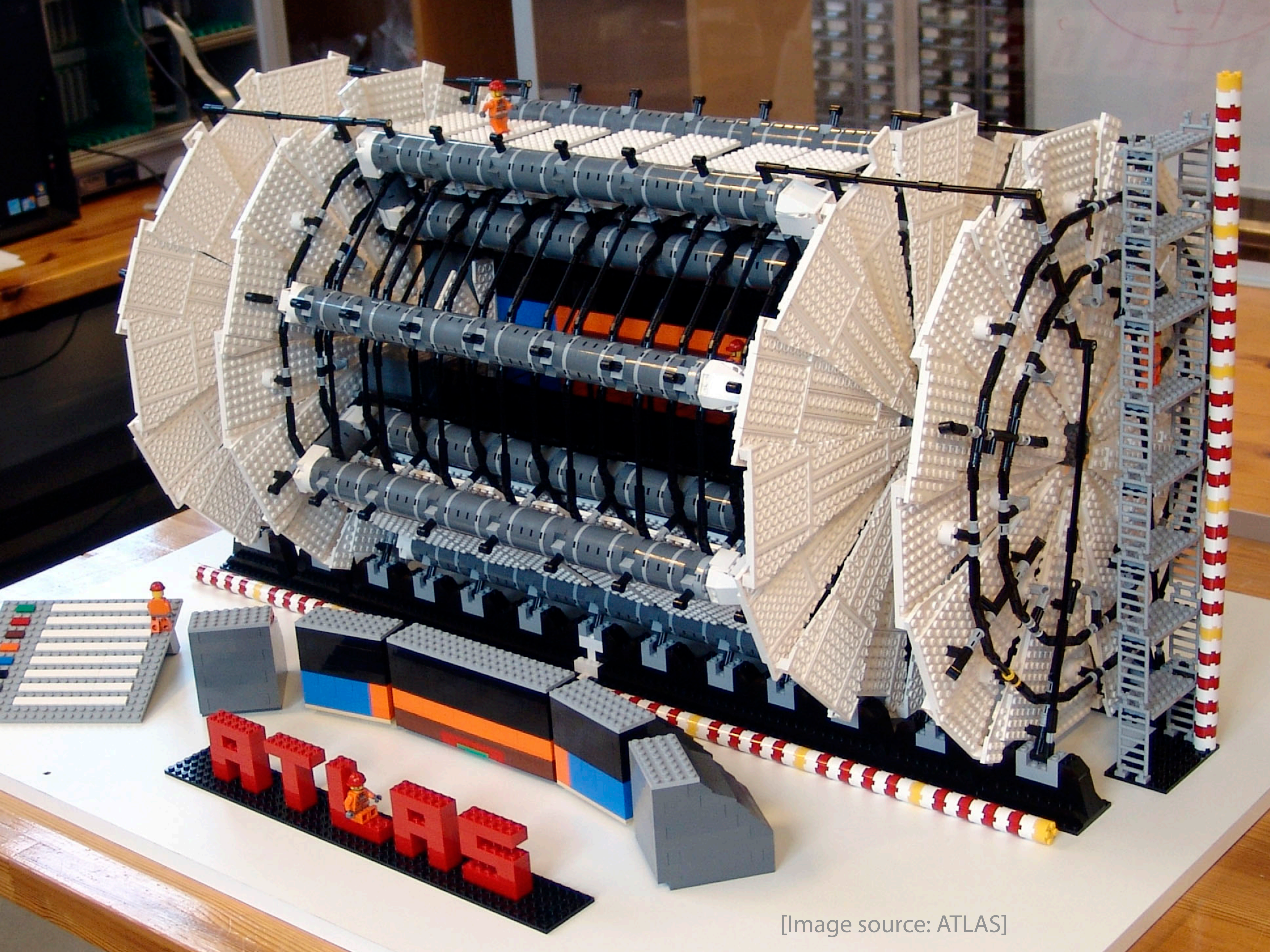
LHC processes



LHC processes

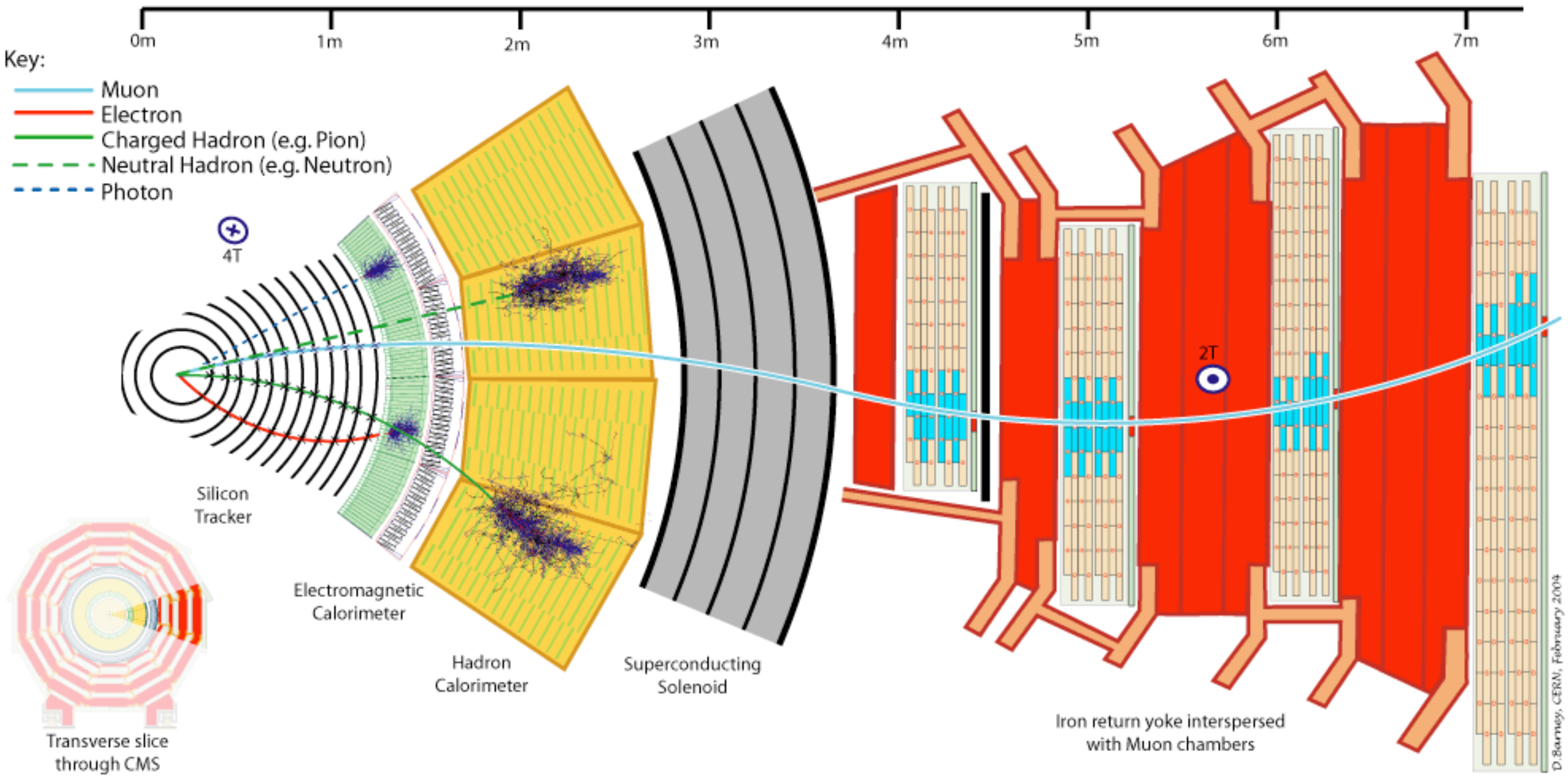


[Image source: F. Krauss]

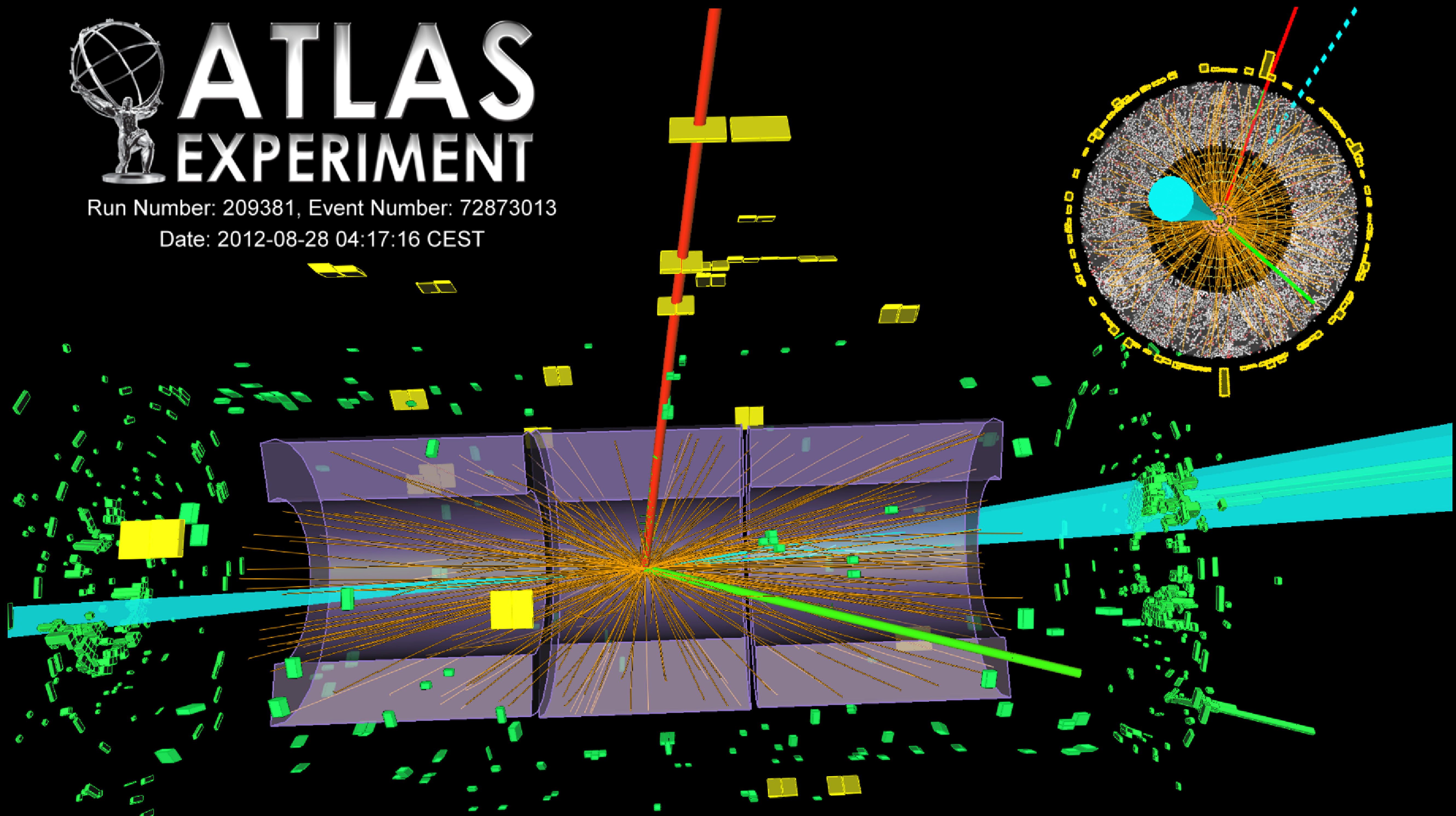


[Image source: ATLAS]

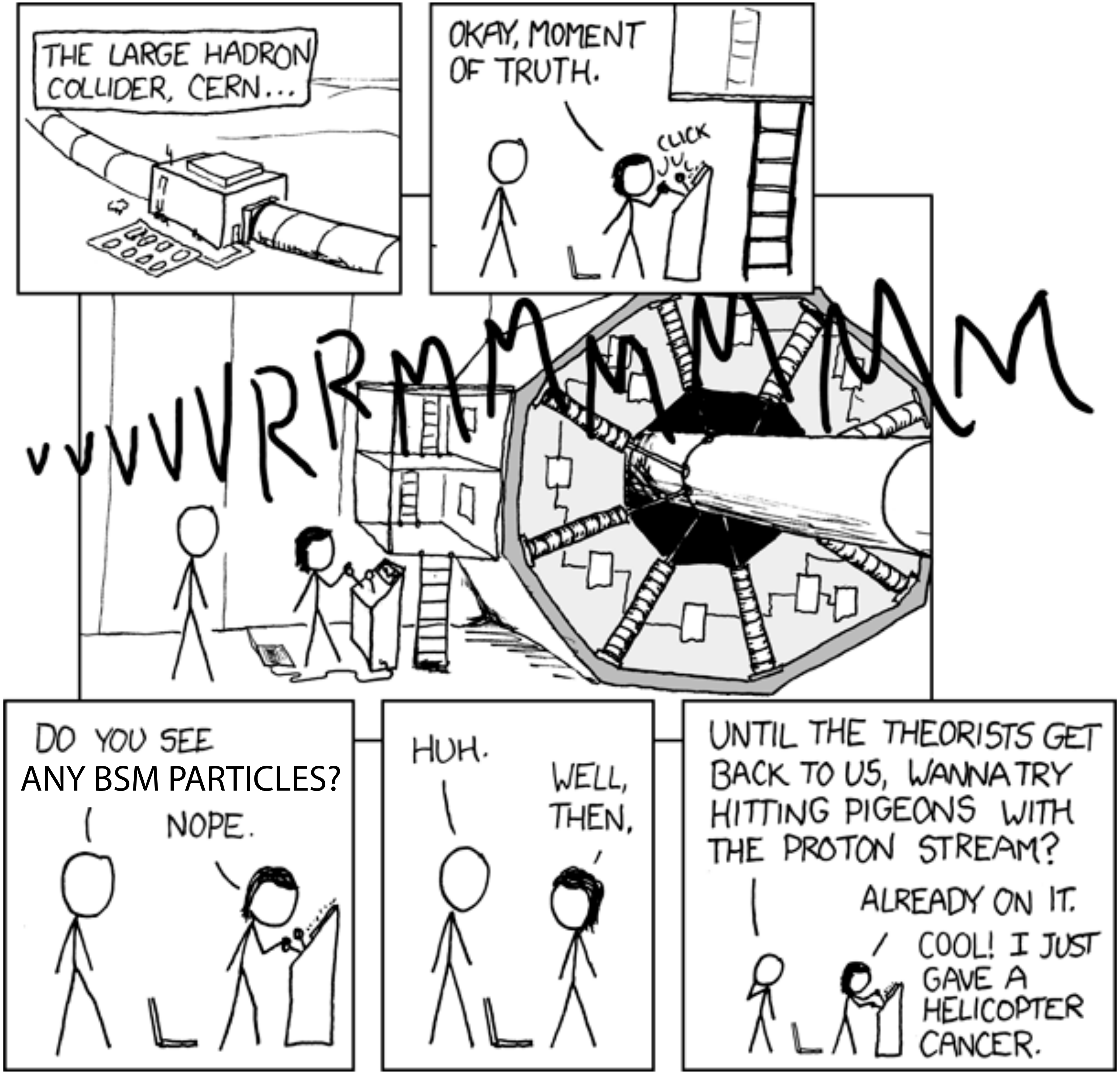
Particle, meet detector



[Image source: CMS]



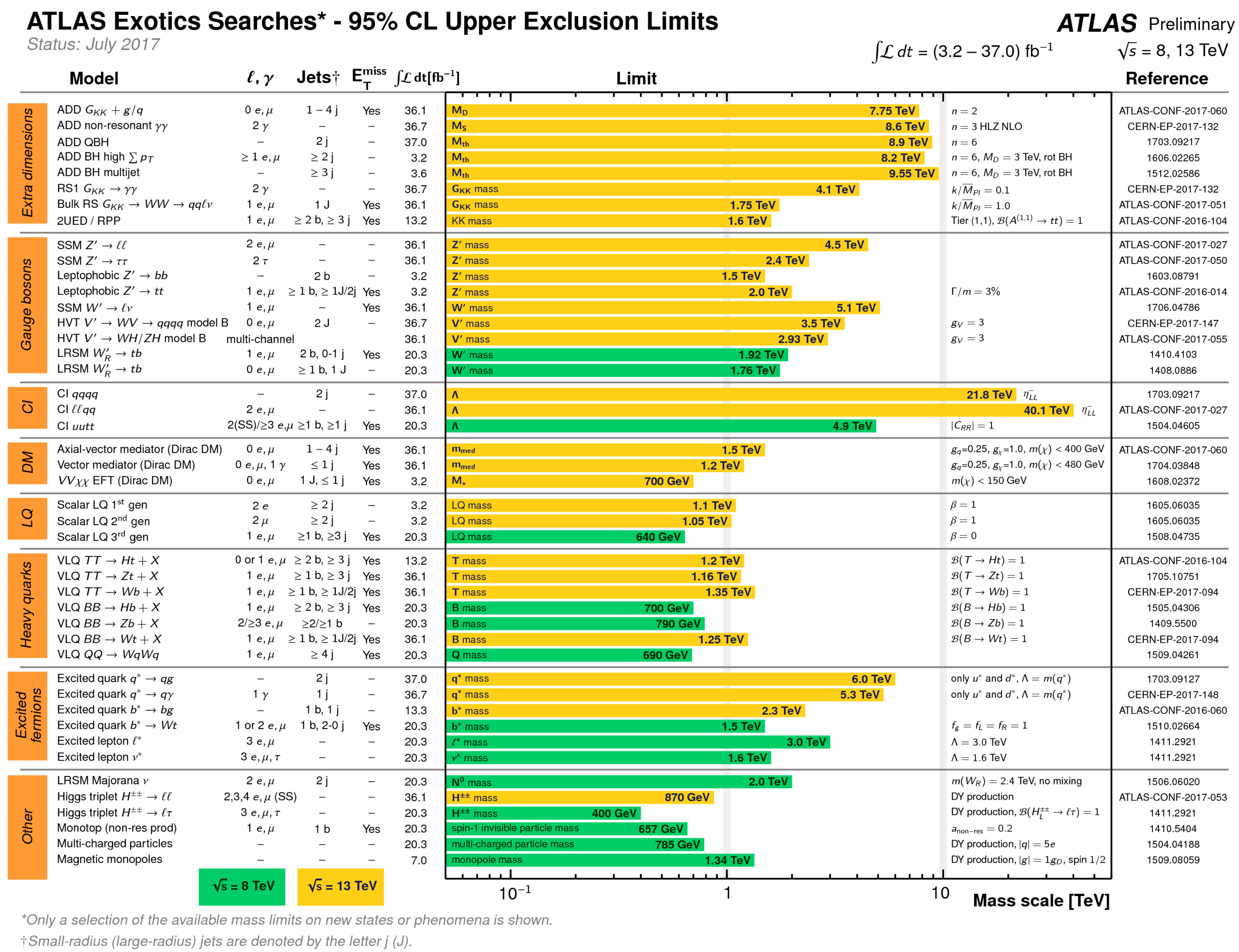
[Image source: ATLAS]


THE LARGE HADRON COLLIDER, CERN...
OKAY, MOMENT OF TRUTH.
CLICK
VVVVVRRMMMMMMM
DO YOU SEE ANY BSM PARTICLES?
NOPE.
HUH.
WELL, THEN,
UNTIL THE THEORISTS GET BACK TO US, WANNA TRY HITTING PIGEONS WITH THE PROTON STREAM?
ALREADY ON IT.
COOL! I JUST GAVE A HELICOPTER CANCER.

[Source: xkcd]

Where is everyone?

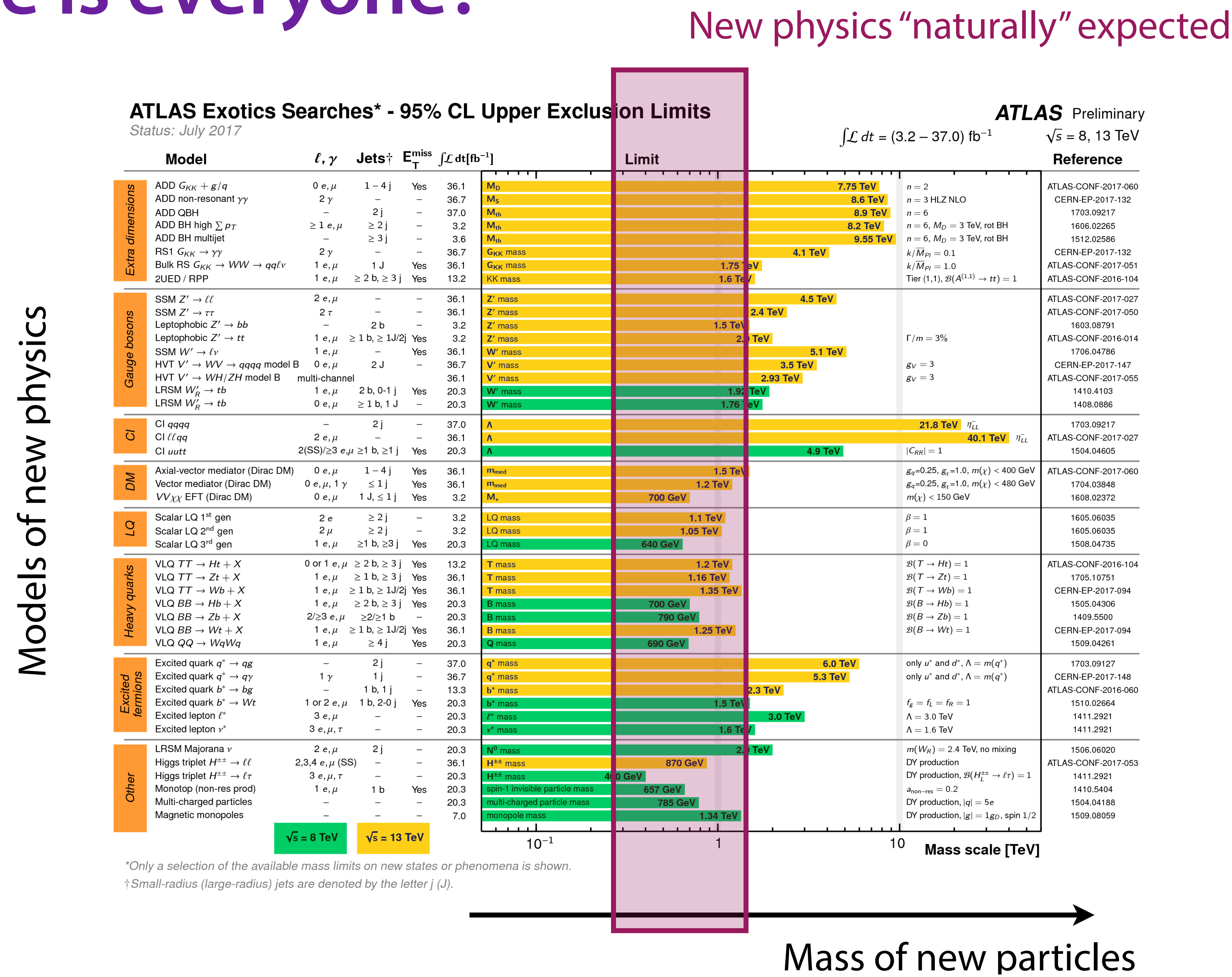
Models of new physics



Mass of new particles

[Source: ATLAS]

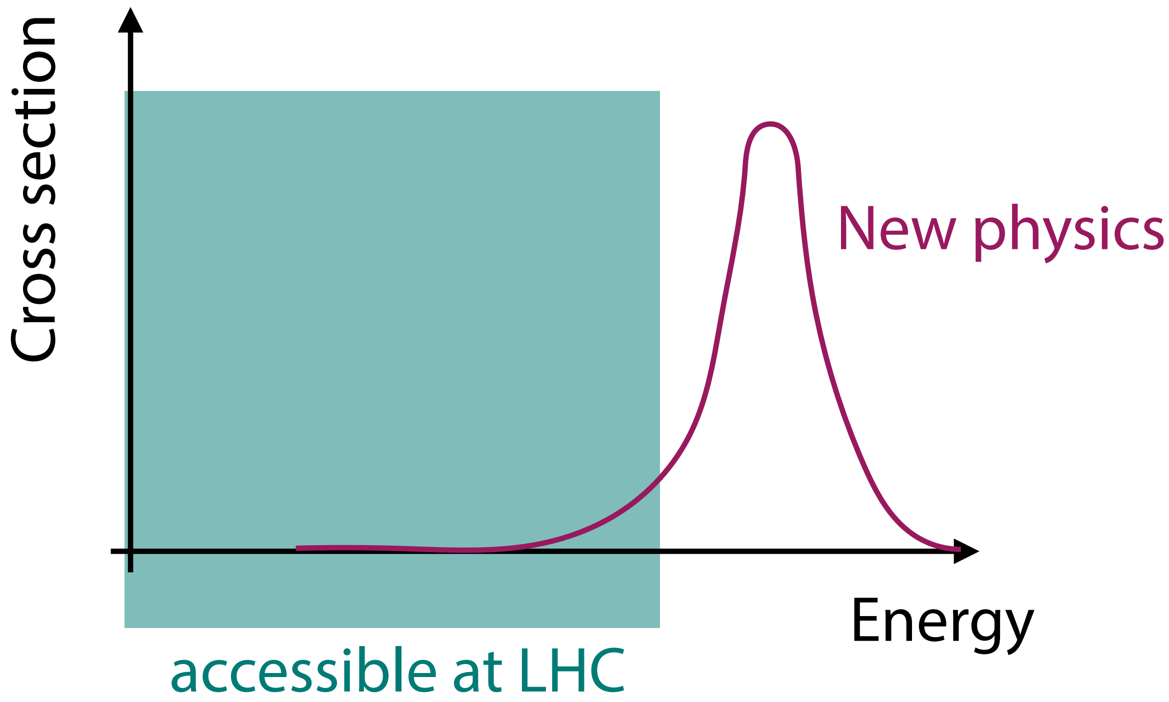
Where is everyone?

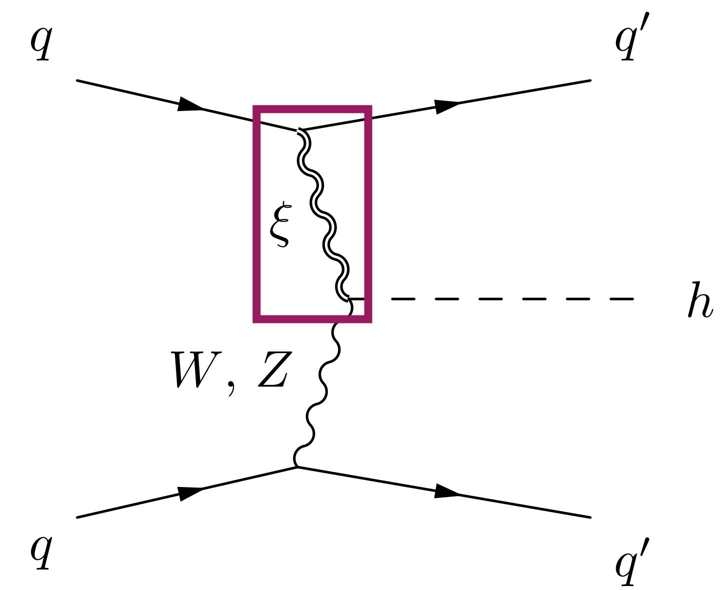


[Source: ATLAS]

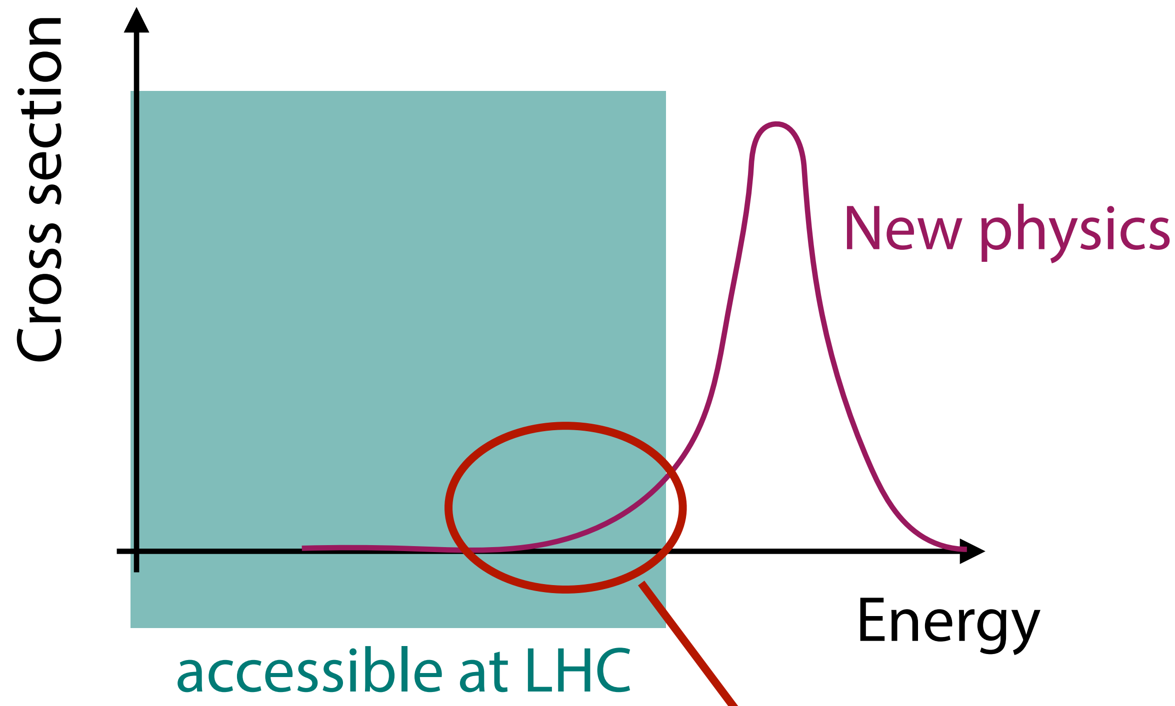
⇒ New physics is heavier than hoped for

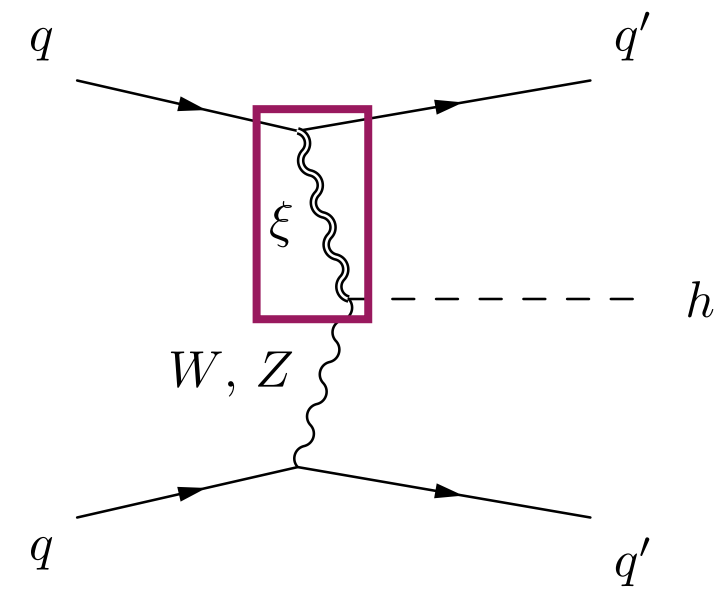
Indirect effects of heavy new physics





Indirect effects of heavy new physics



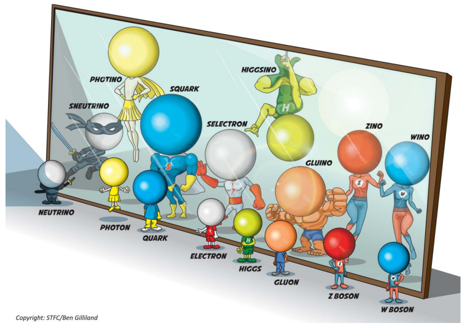


Can we find such effects in the interactions of SM particles?

[Image sources: STFC / Ben Gililand, Sean Carroll, Friedrich Justin Bertuch 1806, symmetry]

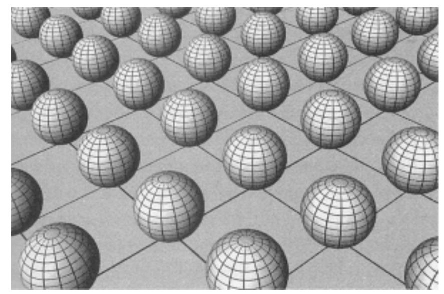
Effective field theory

Energy

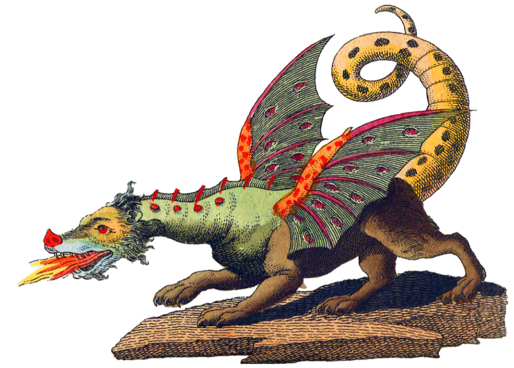


Λ

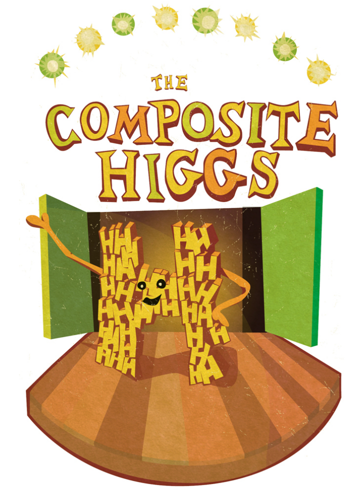
Supersymmetry?



Extra dimensions?



Dragons?



Composite Higgs?

…

$E \ll \Lambda$

LHC Higgs

Other LHC measurements

LEP

Flavour

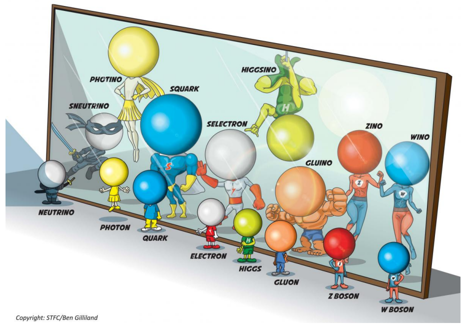
…

[Image sources: STFC / Ben Gililand, Sean Carroll, Friedrich Justin Bertuch 1806, symmetry]

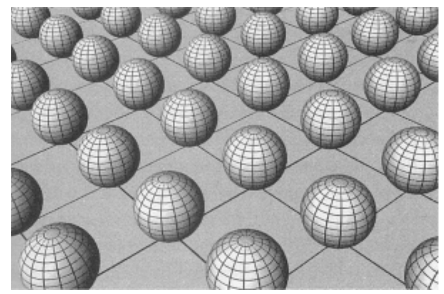
Effective field theory

Energy

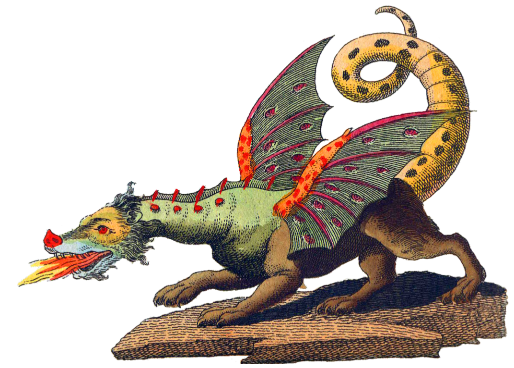
Λ



Supersymmetry?



Extra dimensions?



Dragons?

Composite Higgs?

…

Matching

$$\mathcal{L}_{\rm EFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{f_i}{\Lambda^2}\,\mathcal{O}_i + \dots$$

Effective field theory

Measurements

$E \ll \Lambda$

LHC Higgs

Other LHC measurements

LEP

Flavour

…

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

SMEFT (Standard Model Effective Field Theory)

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2}\, \mathcal{O}_i + \dots$$

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

SMEFT (Standard Model Effective Field Theory)

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{f_i}{\Lambda^2}\, \mathcal{O}_i + \ldots$$

Operators

- all possible interactions between SM particles mediated by new physics
- fixed by SM particles + SM symmetries + expansion in $1/\Lambda$, independent of high-energy physics
- affect rates + kinematics

$$\begin{aligned}
\mathcal{O}_{\phi,1} &= (D_\mu\phi)^\dagger \phi\, \phi^\dagger D^\mu \phi & \mathcal{O}_{GG} &= (\phi^\dagger\phi)\, G^a_{\mu\nu}\, G^{\mu\nu\, a} \\
\mathcal{O}_{\phi,2} &= \frac{1}{2}\, \partial_\mu(\phi^\dagger\phi)\, \partial^\mu(\phi^\dagger\phi) & \mathcal{O}_{BB} &= -\frac{g'^2}{4}\, (\phi^\dagger\phi)\, B_{\mu\nu}\, B^{\mu\nu} \\
\mathcal{O}_{\phi,3} &= \frac{1}{3}(\phi^\dagger\phi)^3 & \mathcal{O}_{WW} &= -\frac{g^2}{4}\, (\phi^\dagger\phi)\, W^a_{\mu\nu}\, W^{\mu\nu\, a} \\
\mathcal{O}_{\phi,4} &= (\phi^\dagger\phi)\, (D_\mu\phi)^\dagger\, D^\mu\phi & \mathcal{O}_{BW} &= -\frac{g\, g'}{4}\, (\phi^\dagger\sigma^a\phi)\, B_{\mu\nu}\, W^{\mu\nu\, a} \\
& & \mathcal{O}_{B} &= \frac{\mathrm{i}g'}{2}\, (D^\mu\phi)^\dagger\, D^\nu\phi\, B_{\mu\nu} \\
& & \mathcal{O}_{W} &= \frac{\mathrm{i}g}{2}\, (D^\mu\phi)^\dagger\, \sigma^a\, D^\nu\phi\, W^a_{\mu\nu}
\end{aligned}$$

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

SMEFT (Standard Model Effective Field Theory)

Operators

- all possible interactions between SM particles mediated by new physics
- fixed by SM particles + SM symmetries + expansion in $1/\Lambda$, independent of high-energy physics
- affect rates + kinematics

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2}\,\mathcal{O}_i + \ldots$$

Wilson coefficients

- precise measurement of these parameters is one of the most important goals of the LHC
- can be translated to high-energy physics parameters

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

SMEFT (Standard Model Effective Field Theory)

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i + \dots$$

Operators

- all possible interactions between SM particles mediated by new physics
- fixed by SM particles + SM symmetries + expansion in $1/\Lambda$, independent of high-energy physics
- affect rates + kinematics

Higher-order terms

- suppressed by additional factors of E^2/Λ^2

Wilson coefficients

- precise measurement of these parameters is one of the most important goals of the LHC
- can be translated to high-energy physics parameters

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

SMEFT (Standard Model Effective Field Theory)

Operators

- all possible interactions between SM particles mediated by new physics
- fixed by SM particles + SM symmetries + expansion in $1/\Lambda$, independent of high-energy physics
- affect rates + kinematics

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2}\,\mathcal{O}_i + \ldots$$

Higher-order terms

- suppressed by additional factors of E^2/Λ^2

Wilson coefficients

- precise measurement of these parameters is one of the most important goals of the LHC
- can be translated to high-energy physics parameters

What makes the EFT measurements so difficult?

LHC processes

Parameters of interest

Theory parameters

θ

Evolution

LHC processes

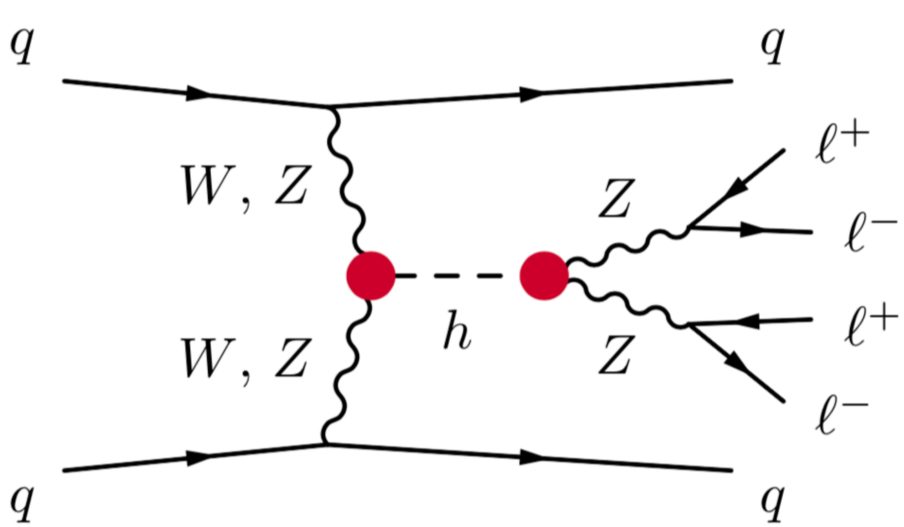
Latent variables

Parameters of interest

Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$



Evolution

LHC processes

Latent variables

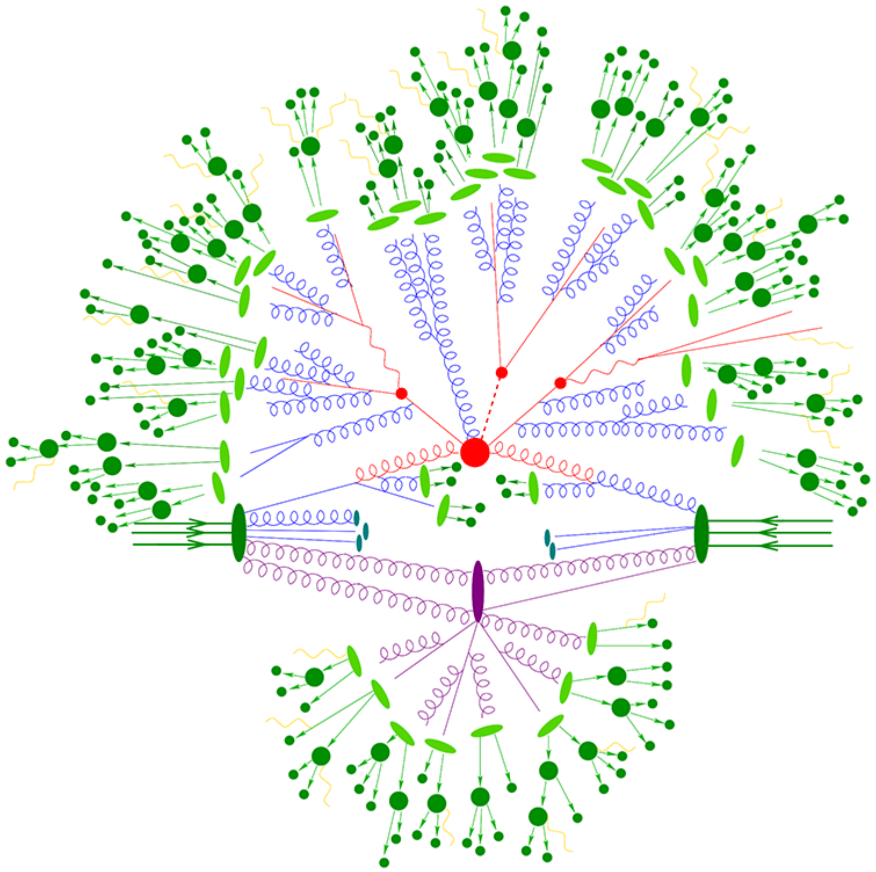
Parameters of interest

Shower splittings

Parton-level momenta

Theory parameters

$$z_s \longleftarrow z_p \longleftarrow \theta$$



Evolution

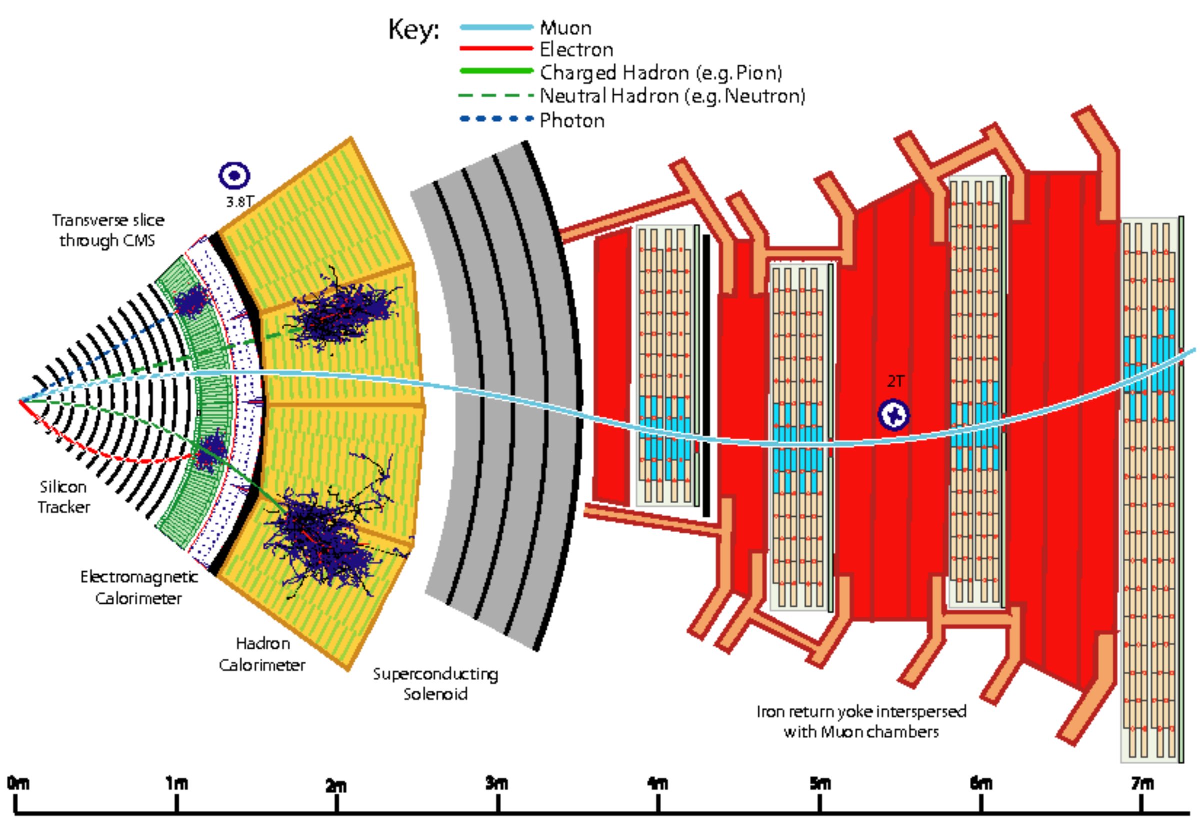
LHC processes

Latent variables

Parameters of interest

Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

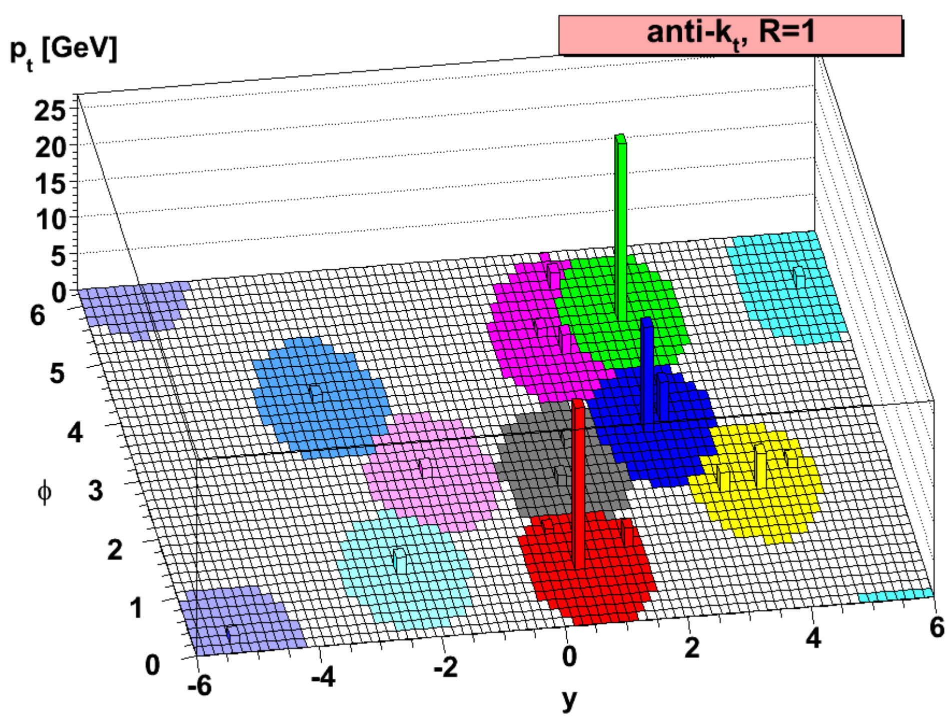


Evolution

LHC processes

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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



[Image source: M. Cacciari, G. Salam, G. Soyez 0802.1189]

Evolution

LHC processes

Features | Latent variables | Parameters of interest

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

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

Sample from $p(x|z_d)$ $p(z_d|z_s)$ $p(z_s|z_p)$ $p(z_p|\theta)$

GEANT4 A SIMULATION TOOLKIT

MADGRAPH5_aMC@NLO

Prediction (simulation)

LHC processes

Features | Latent variables | Parameters of interest

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

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

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

Inference

LHC processes

Features | Latent variables | Parameters of interest

Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

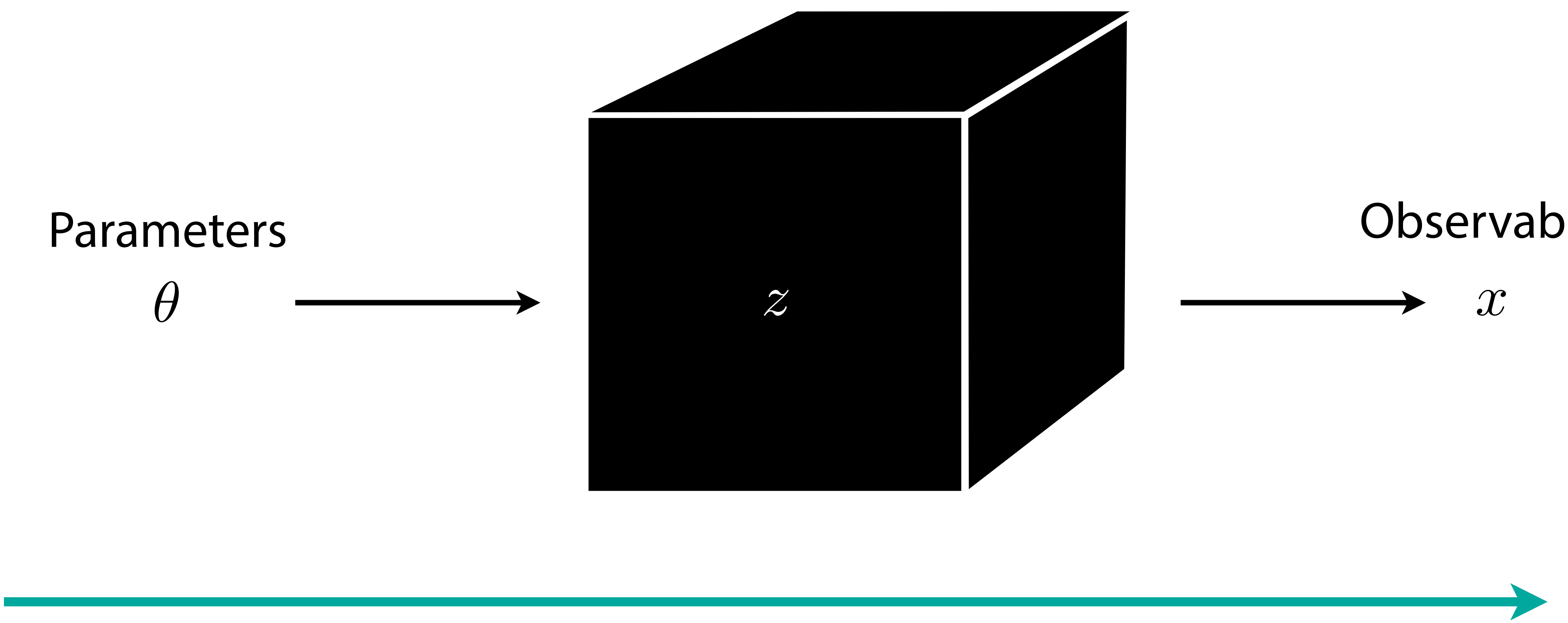
$$x \longleftarrow z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$

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

Infeasible to calculate the integral over this enormous space

$\longrightarrow$ Inference

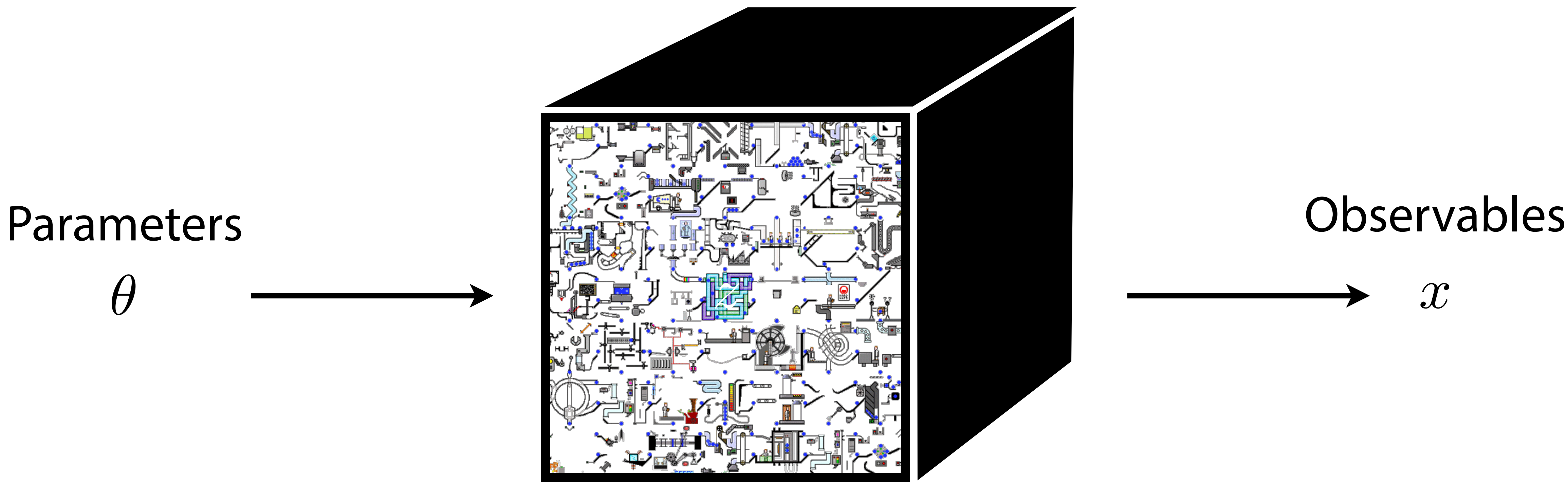
Likelihood-free inference



Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

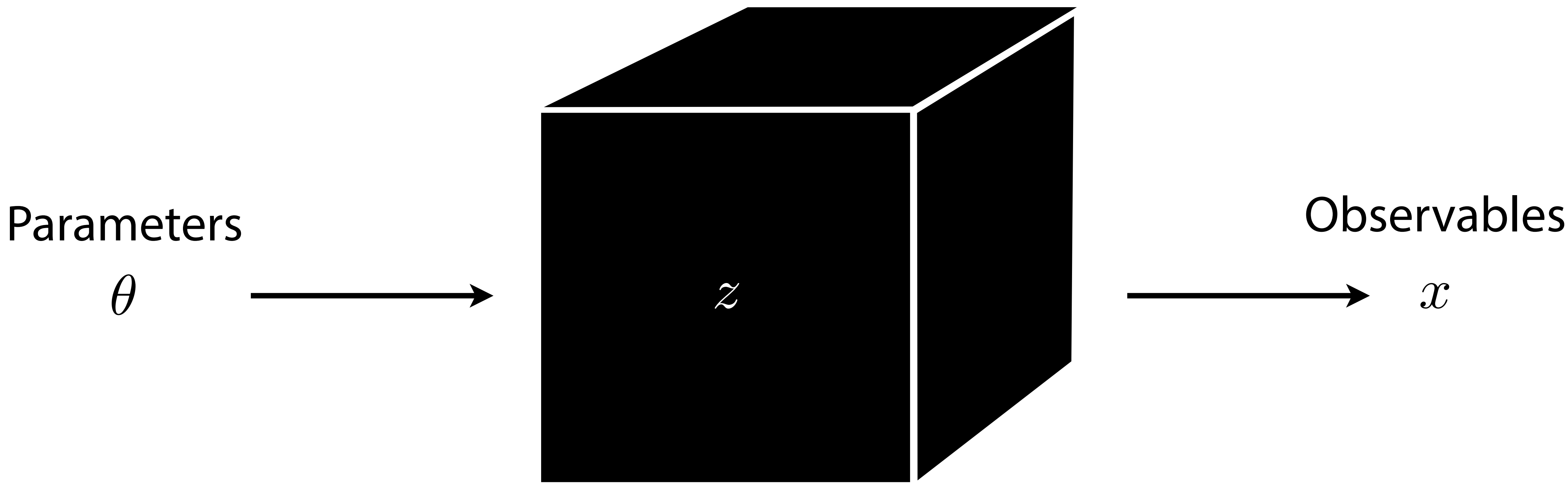
Likelihood-free inference



Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

Likelihood-free inference



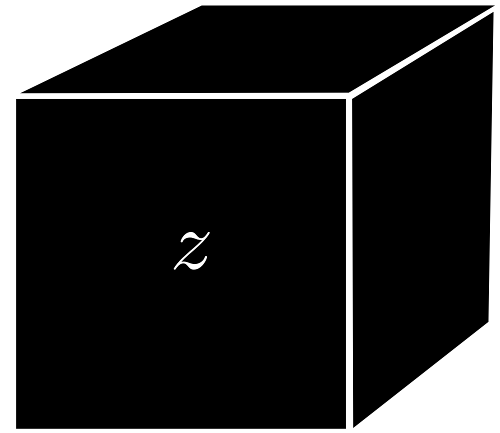
Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

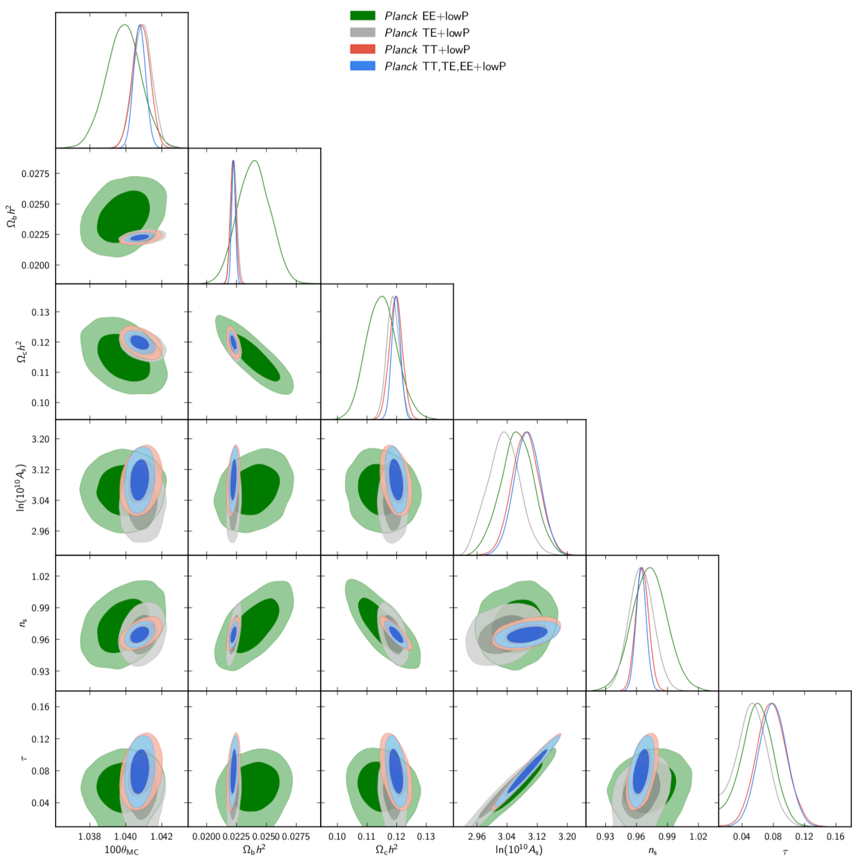
Inference:

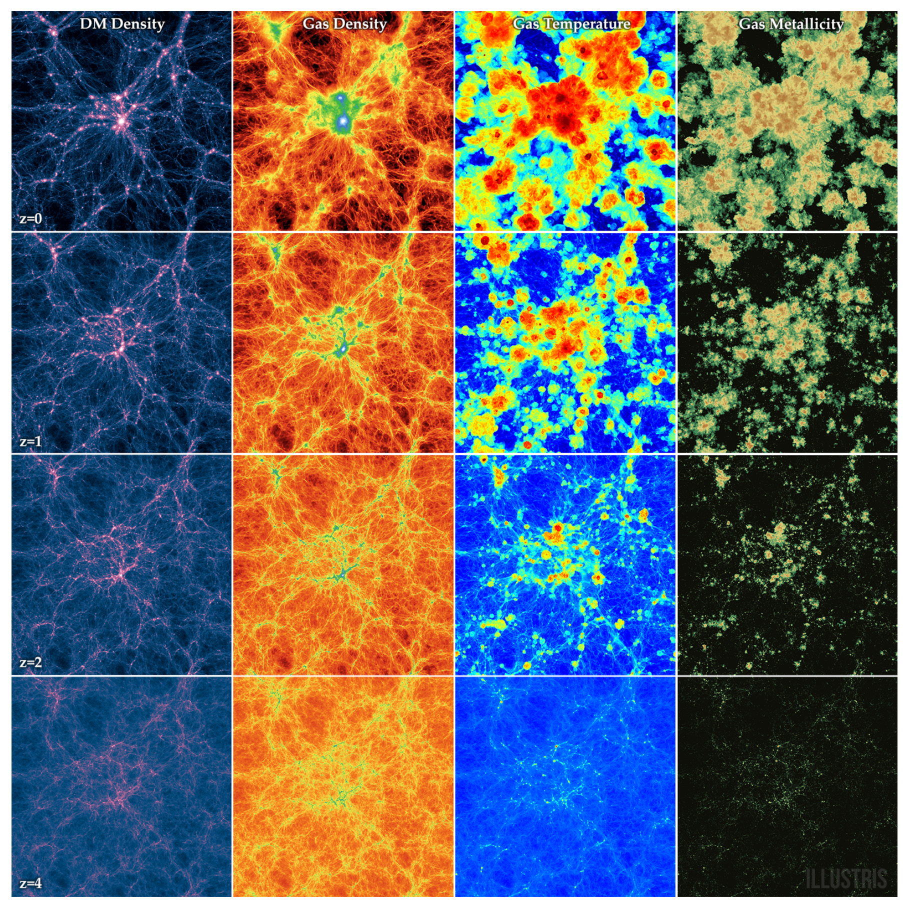
- Likelihood function $p(x|\theta)$ is intractable
- Goal: estimator $\hat{p}(x|\theta)$

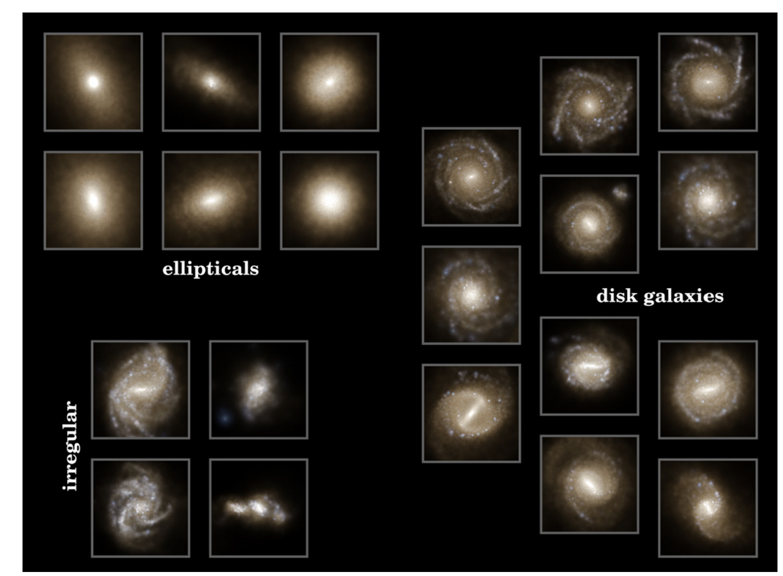
Cosmological N-body simulations

Parameters θ $\longrightarrow$ 

$\longrightarrow$ Observables x







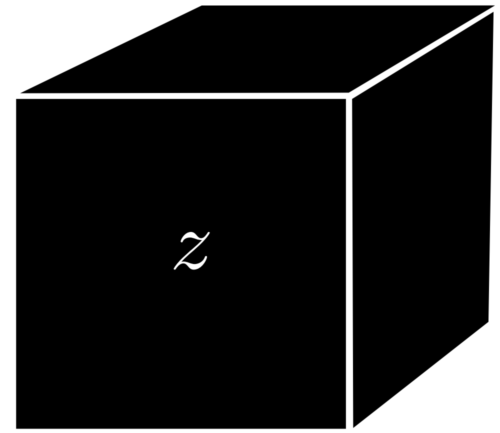
[Source: Planck 1502.01589]

[Source: Illustris 1405.2921]

Computational topography

Parameters

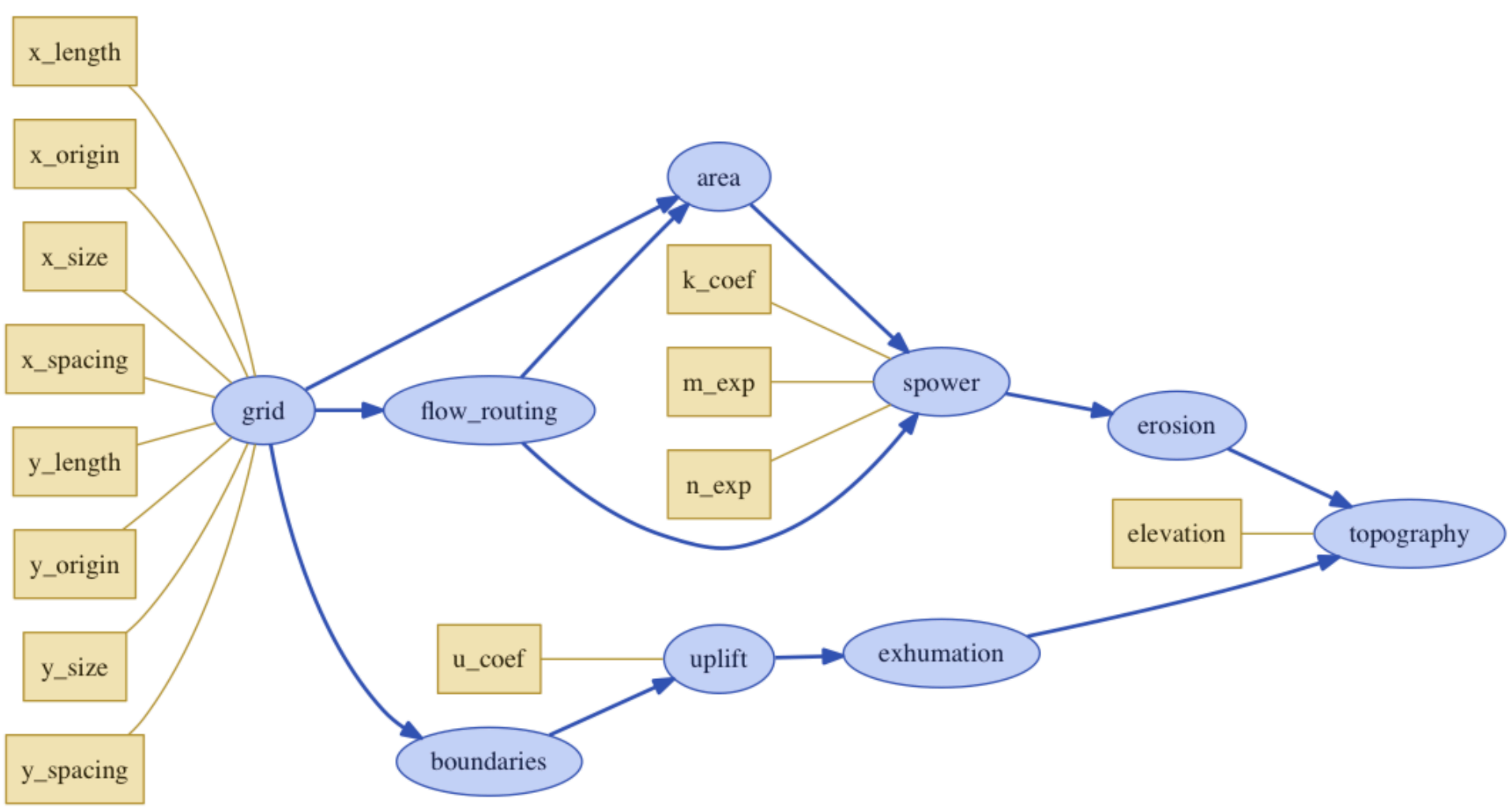
θ $\longrightarrow$



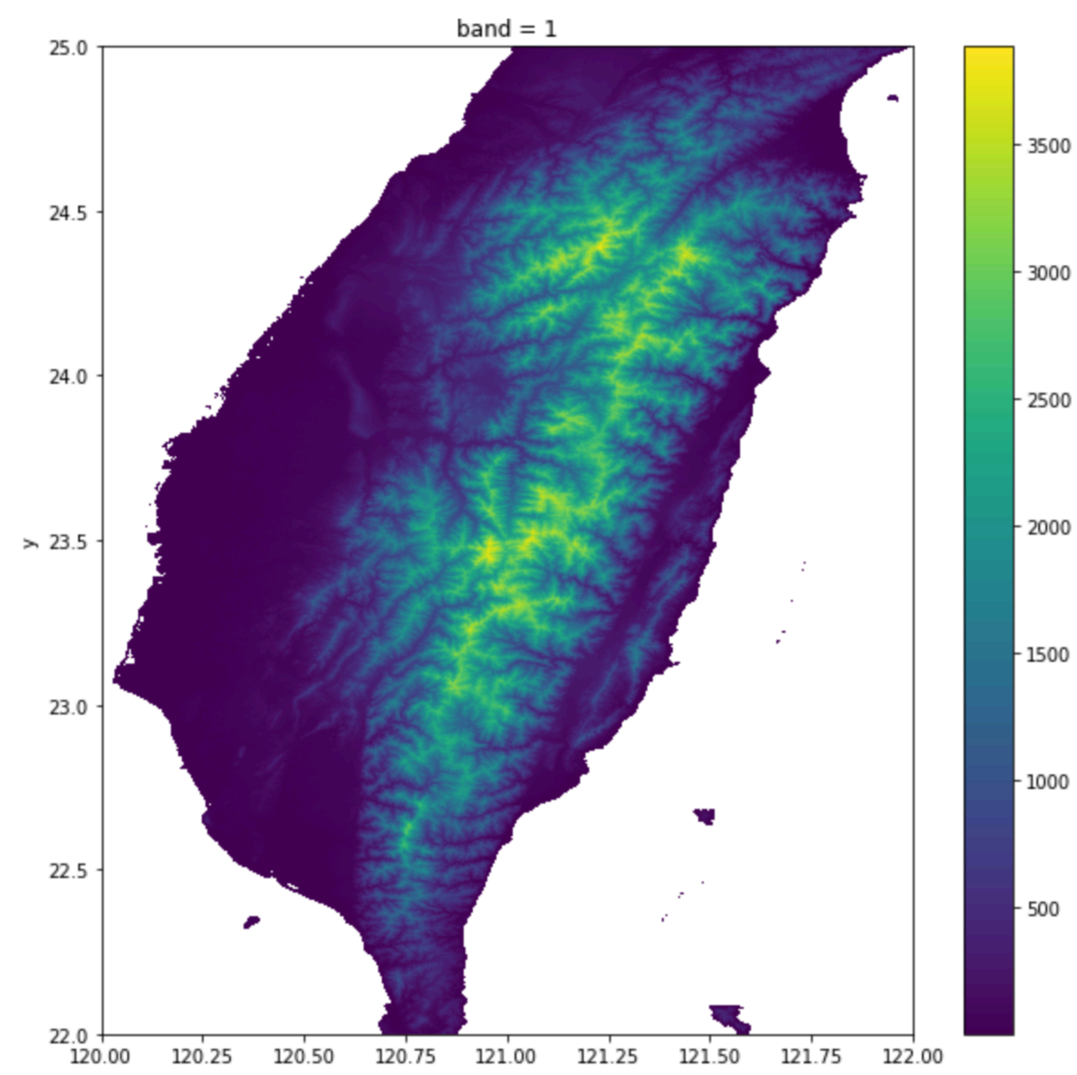
$\longrightarrow$

Observables

x

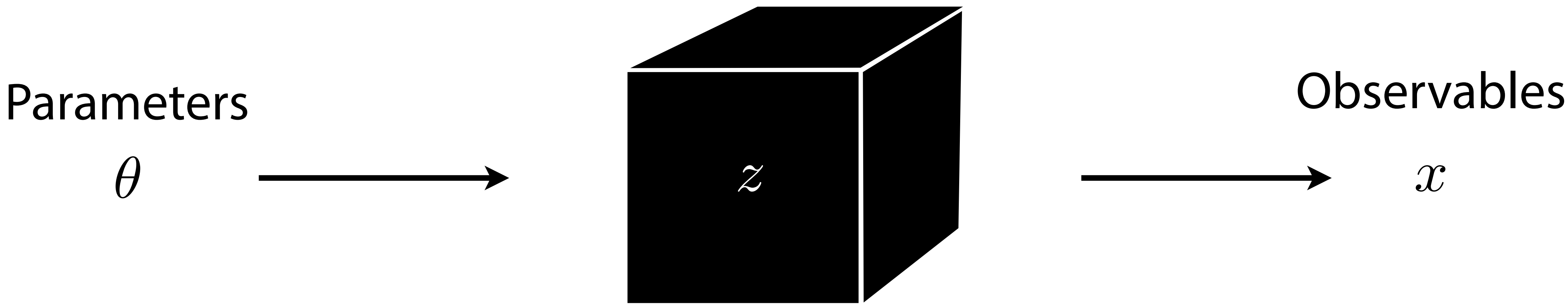


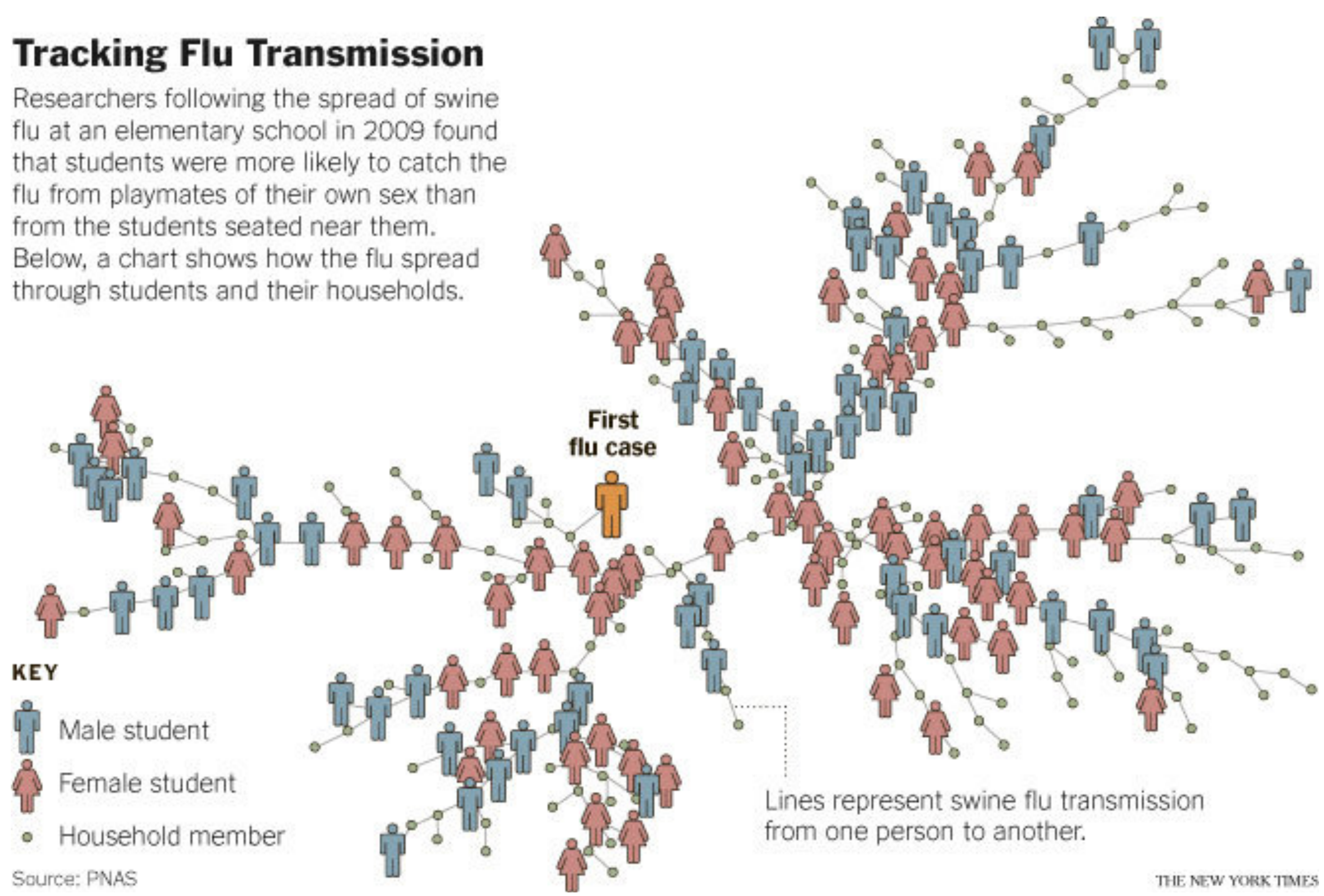
We create a simulation setup for this model, run it, and then plot the final topography (after 1 million years of simulation).



[Source: Benoit Bovy]

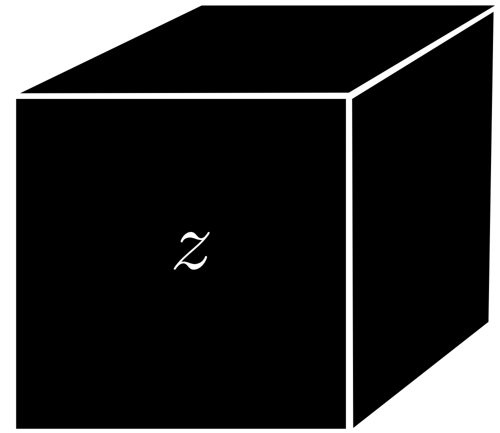
Epidemiology

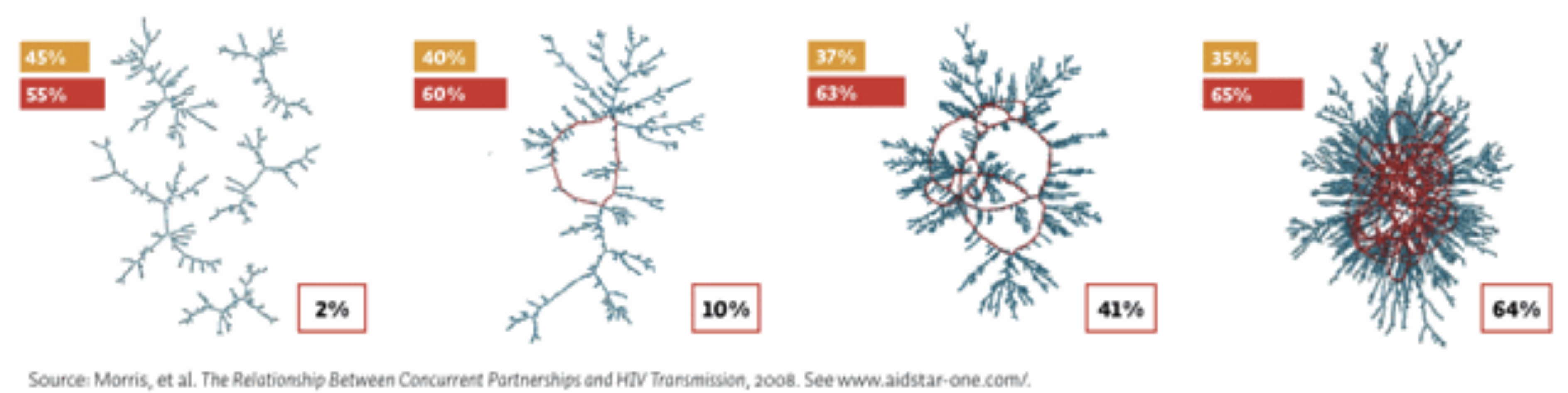




[Source: New York Times]

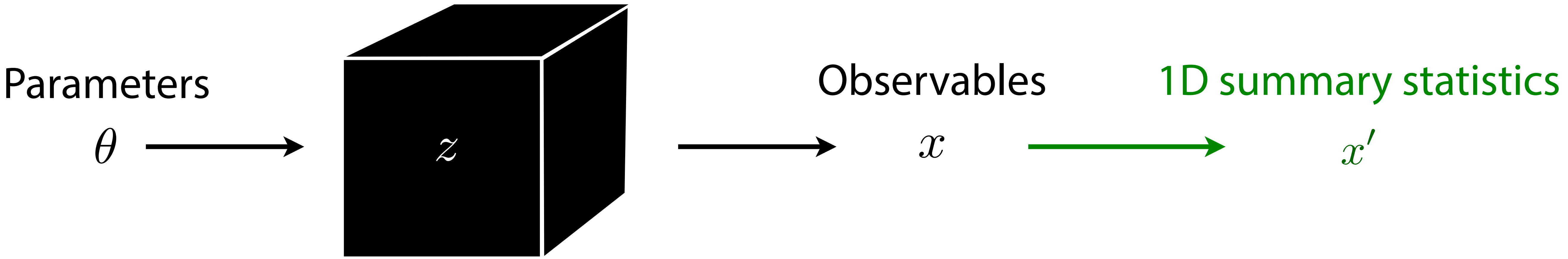
Epidemiology

Parameters θ →  → Observables x



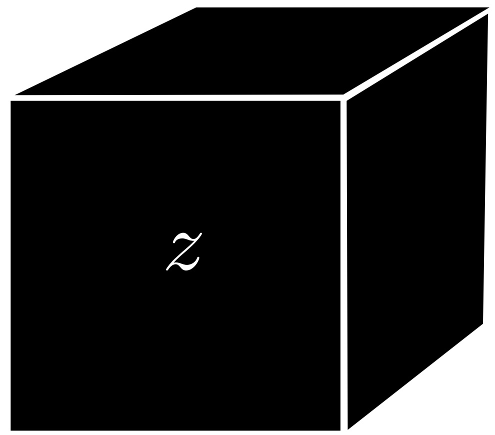
[Source: Morris et al 2008]

Why has that not stopped us so far?



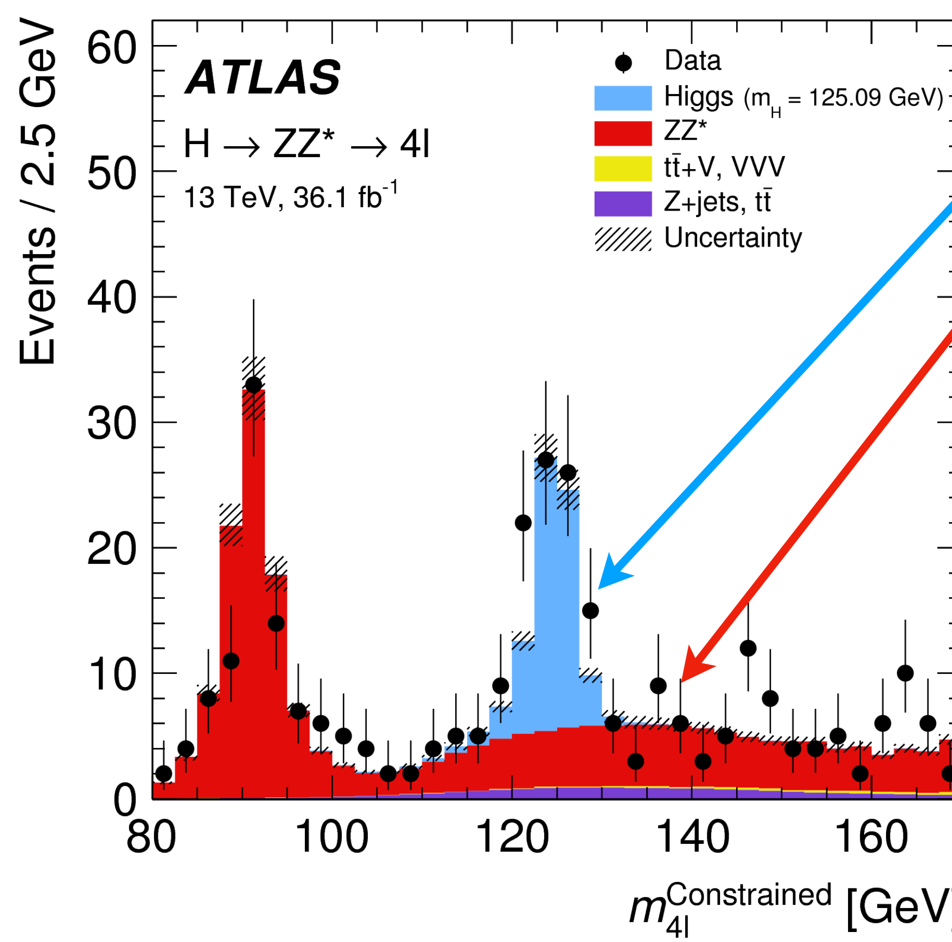
Why has that not stopped us so far?

Parameters θ



Observables x

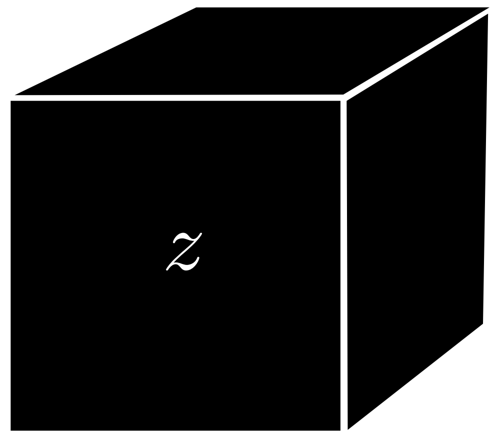
1D summary statistics x'



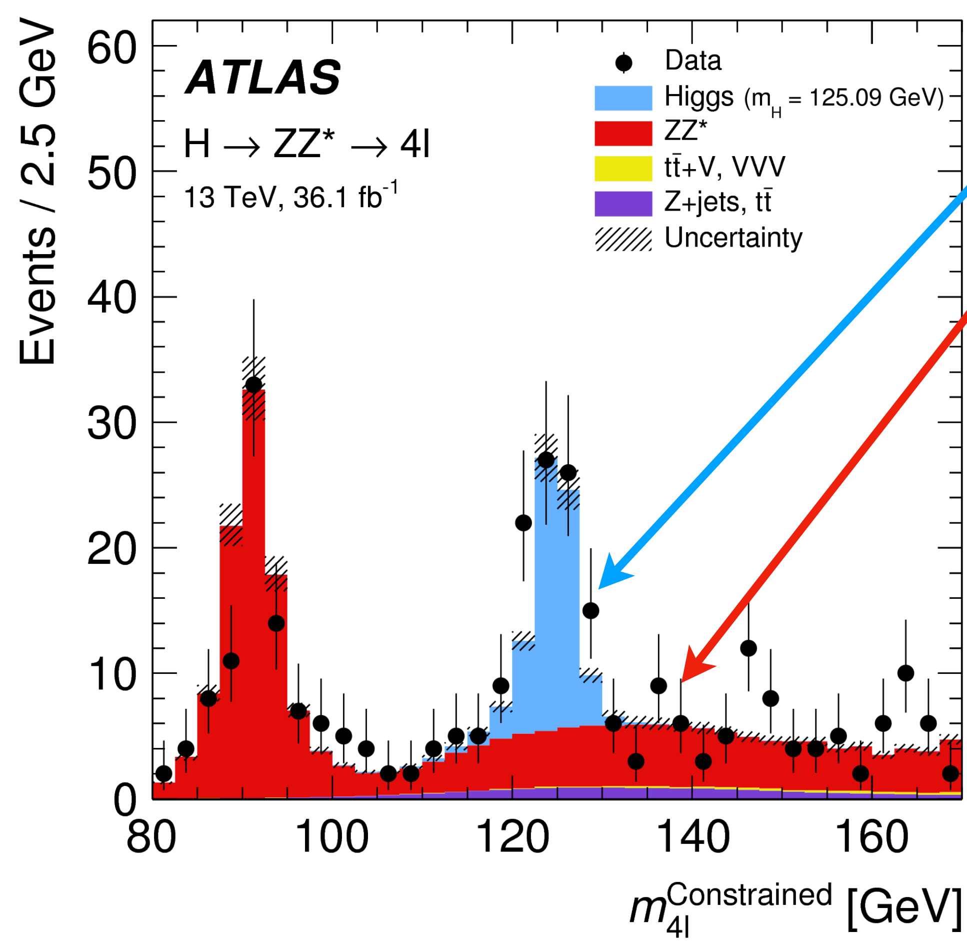
Run simulator for different θ, fill histograms

x'

Why has that not stopped us so far?

Parameters θ →  → Observables x → 1D summary statistics x'

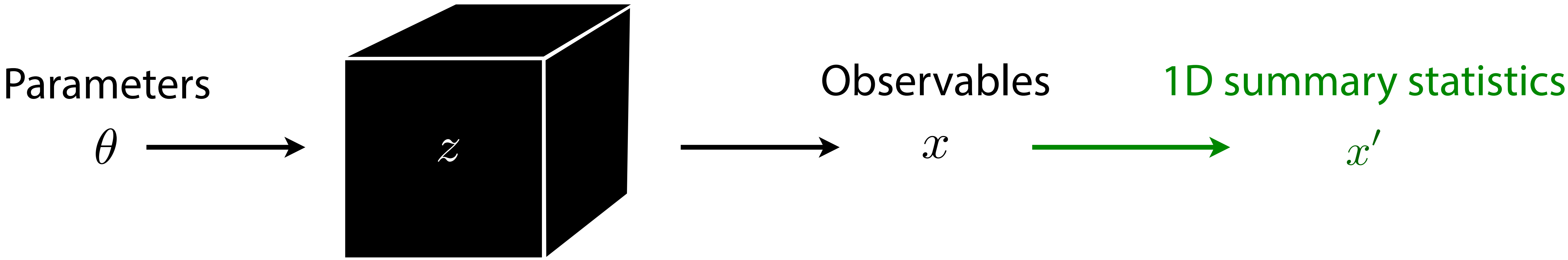
$$\hat{p}(x|\theta) = p(x'|\theta) =$$



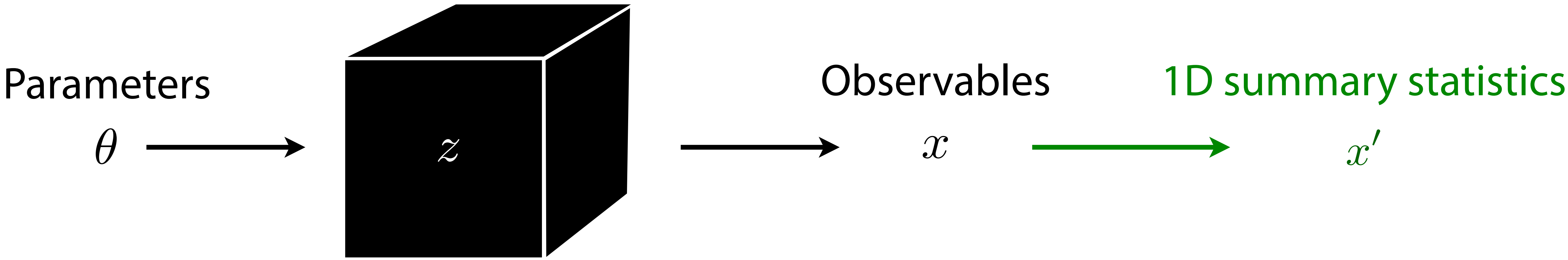
x'

Run simulator for different θ, fill histograms

Does the histogram method scale?

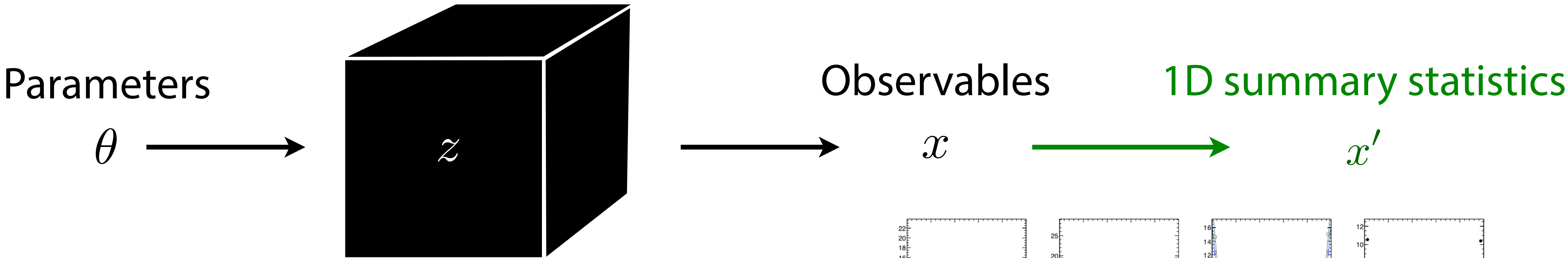


Does the histogram method scale?

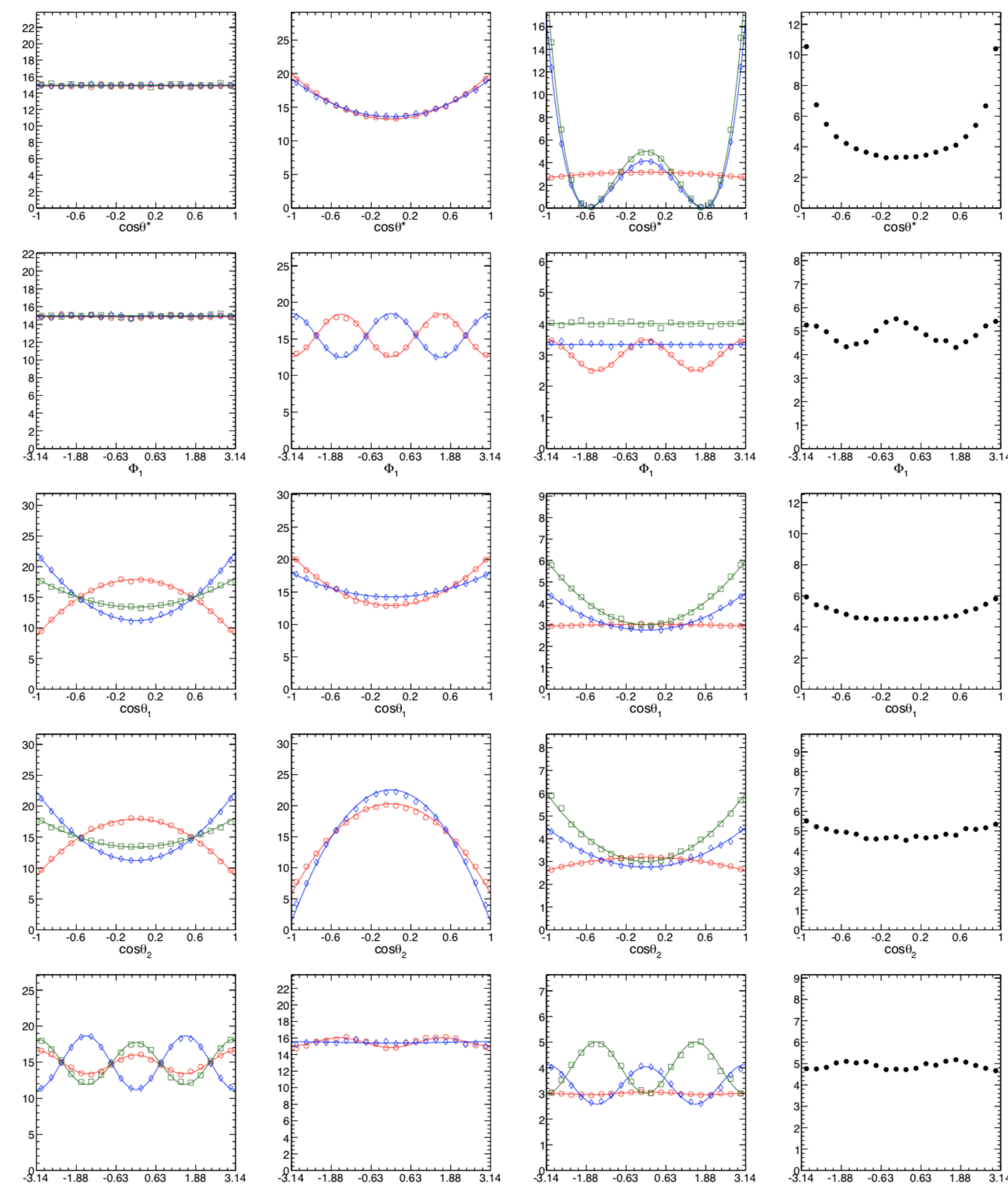


- Choosing x' is difficult and problem-dependent:
  Reconstructed masses, momenta, angles, …
- EFT measurements:
  often there is no single good variable — compressing to any x' loses information!
  [JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
  JB, F. Kling, T. Plehn , T. Tait 1712.02350]
- Ideally: analyse high-dimensional x including all correlations

Does the histogram method scale?

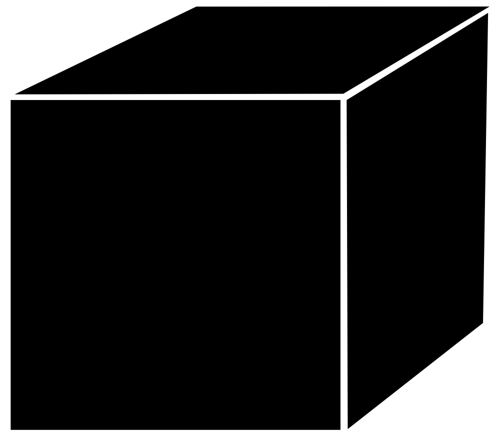


- Choosing x' is difficult and problem-dependent: Reconstructed masses, momenta, angles, …
- EFT measurements: often there is no single good variable — compressing to any x' loses information!
  [JB, K. Cranmer, F. Kling, T. Plehn 1612.05261; JB, F. Kling, T. Plehn , T. Tait 1712.02350]
- Ideally: analyse high-dimensional x including all correlations



[Source: Bolognesi et al. 1208.4018

An incomplete list of established methods



- Histograms of individual kinematic variables
  Summary statistics
- Approximate Bayesian Computation
  Summary statistics
- Machine Learning techniques
  CARL, autoregressive models, normalizing flows, …

- Matrix Element Method / Optimal Observables
  Neglect or approximate shower + detector,
  explicitly calculate integral

$$\hat{p}(x|\theta) = \int \mathrm{d}z_p \, \tilde{p}(x|z_p) \, p(z_p|\theta)$$

How can we get more information out of the data?

Learning Effective Theories, Effectively

- New analysis techniques
    - Based on high-dimensional x
      (no summary statistics necessary!)
    - Leverages structure of particle physics processes + power of machine learning

- Appropriate for large-scale EFT measurements at LHC
    - Scales well to many observables and parameters
    - Supports full shower & detector simulation

Learning Effective Theories, Effectively

- New analysis techniques
  - Based on high-dimensional x
    (no summary statistics necessary!)
  - Leverages structure of particle physics processes + power of machine learning

- Appropriate for large-scale EFT measurements at LHC
  - Scales well to many observables and parameters
  - Supports full shower & detector simulation

- Goal: estimator $\hat{r}(x|\theta_0, \theta_1)$ for likelihood ratio

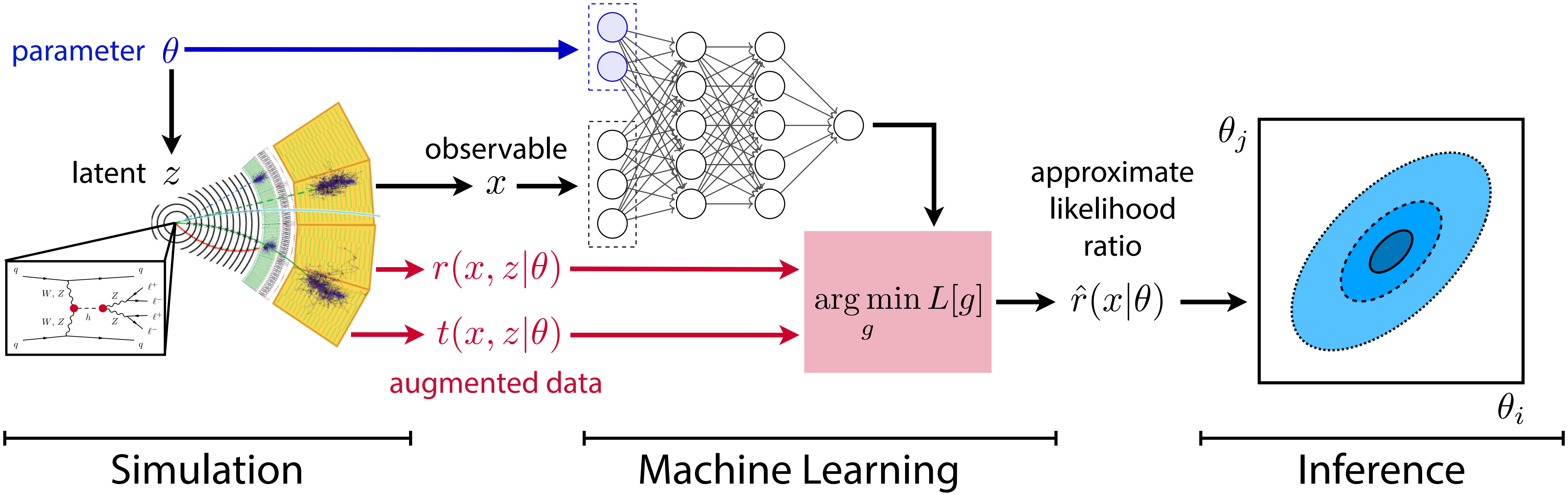
observables

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

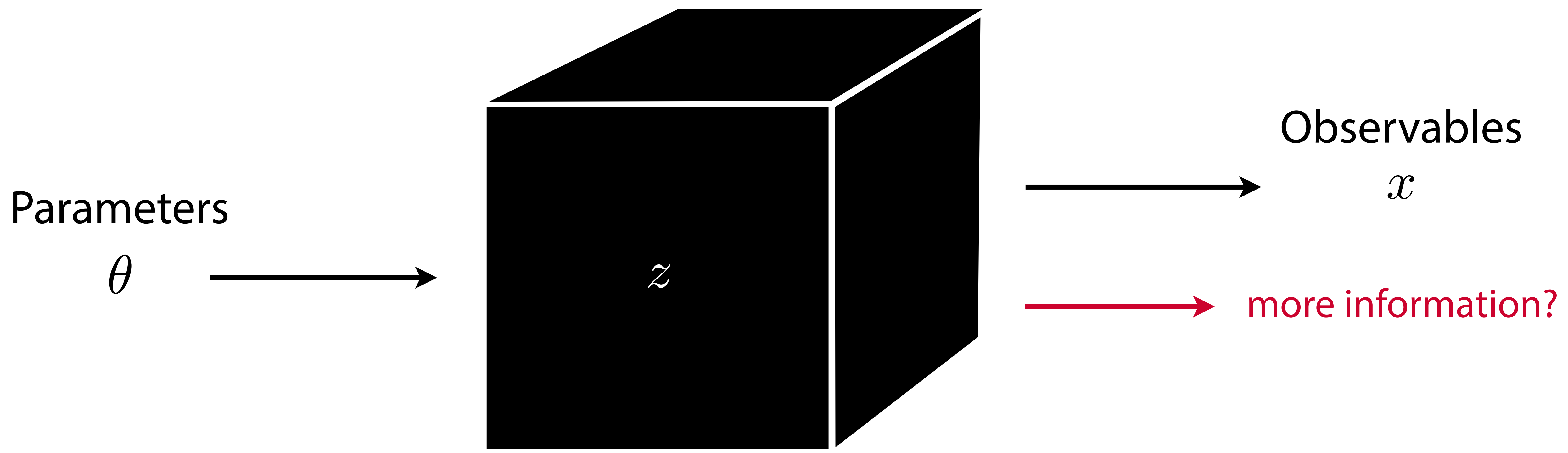
theory parameters

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

Sneak Preview



Opening the black box



Opening the black box

Features | Latent variables | Parameters of interest

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

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

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

Intractable integrals

Opening the black box

Features

Latent variables

Parameters of interest

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

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

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

This parton-level likelihood can be evaluated!

Opening the black box

Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

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

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

This parton-level likelihood can be evaluated!

⇒ We can calculate the "joint" likelihood ratio conditional on a specific evolution:

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

("How much more likely is this specific evolution assuming θ_0 compared to θ_1?")

From joint likelihood ratio to likelihood ratio

We have joint likelihood ratio

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

We want likelihood ratio

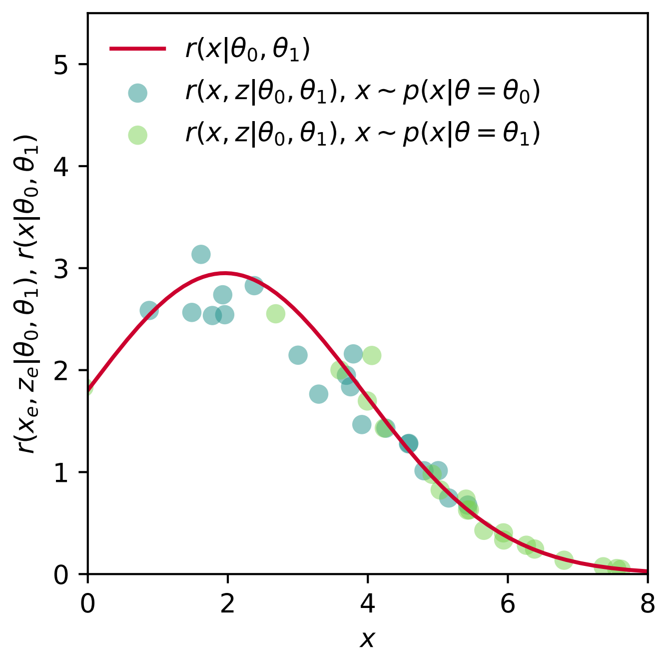
$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

From joint likelihood ratio to likelihood ratio

We have joint likelihood ratio

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

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



We want likelihood ratio

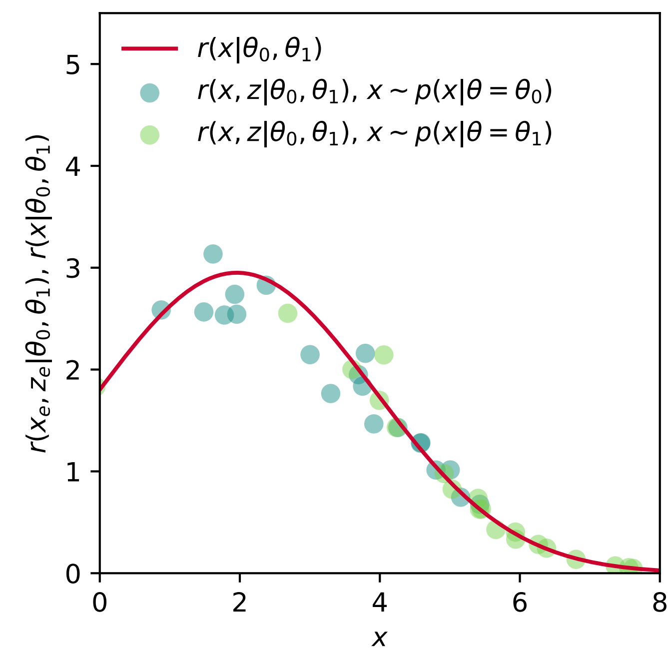
$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

From joint likelihood ratio to likelihood ratio

We have joint likelihood ratio

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

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



With $r(x, z|\theta_0, \theta_1)$,
we can define a functional $L[g]$ that is
minimized by $r(x|\theta_0, \theta_1)$:

$$\hat{r}(x|\theta_0, \theta_1) = \arg\min_g L[g] \rightarrow r(x|\theta_0, \theta_1)$$

We want likelihood ratio

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

Machine Learning = applied calculus of variations

Need to minimize a functional:

Extremization

Functional

Variational family

$$\hat{r}(x|\theta_0, \theta_1) = \arg\min_g L[g]$$

Machine Learning = applied calculus of variations

Need to minimize a functional:

Extremization

Functional

Variational family

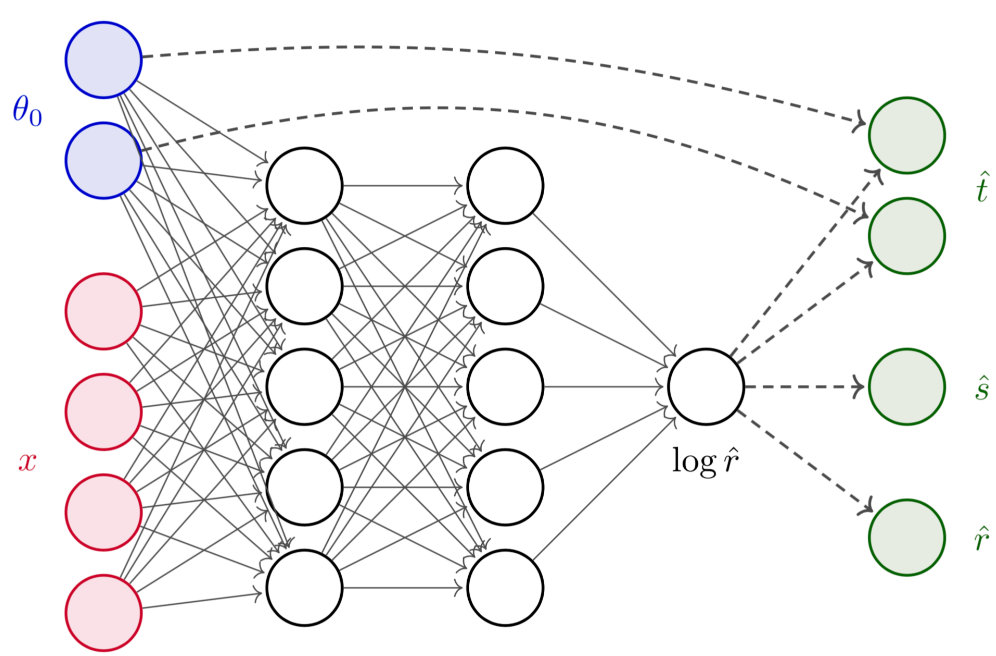
$$\hat{r}(x|\theta_0, \theta_1) = \arg\min_{g} L[g]$$

This is exactly what
Machine Learning does!

Loss function

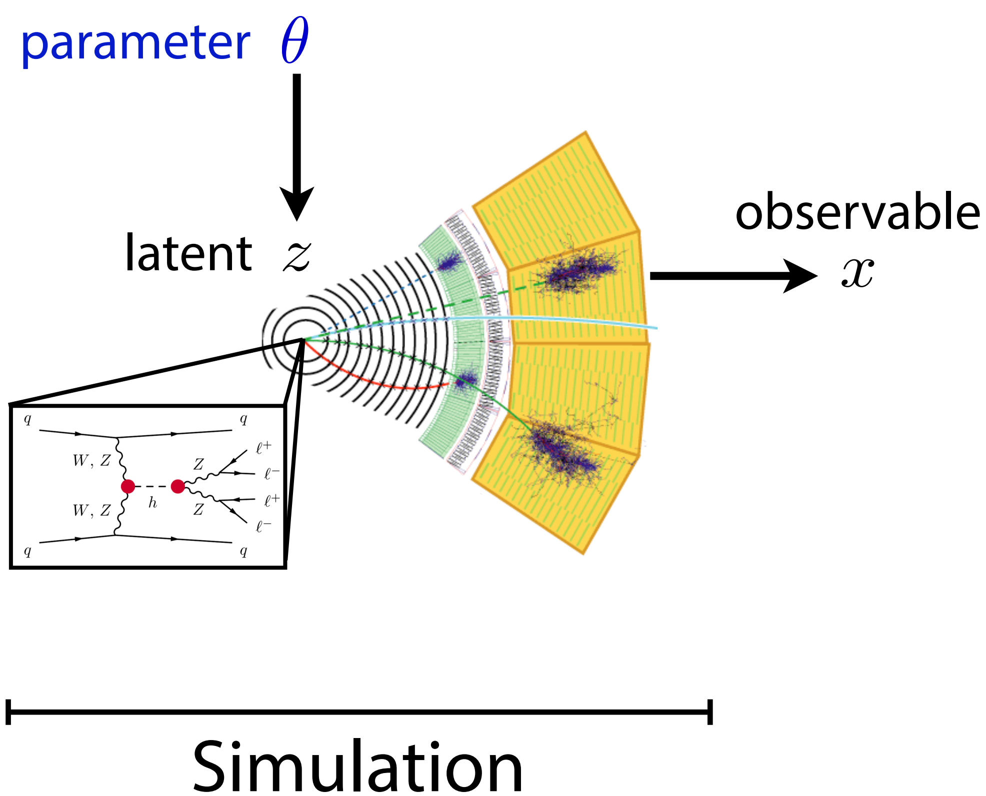
Stochastic Gradient Descent

Neural network

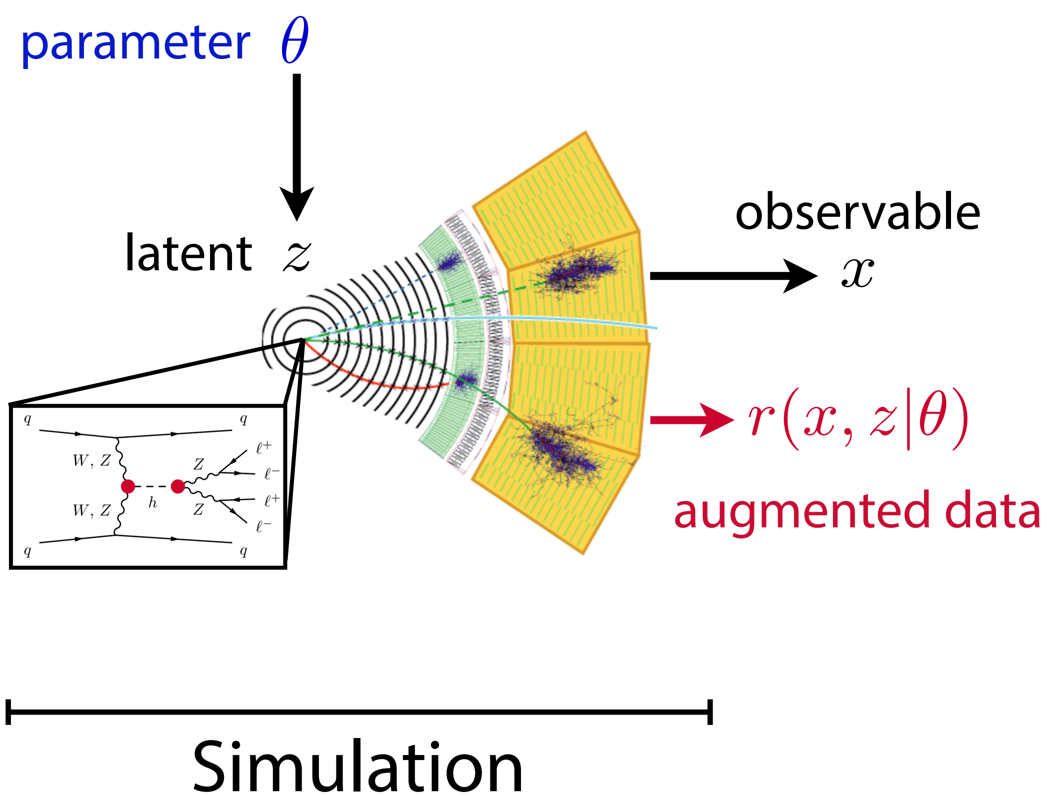


$\Rightarrow$ We implement $\hat{r}(x|\theta_0, \theta_1)$
as a Neural Network trained on the
data available from the simulator

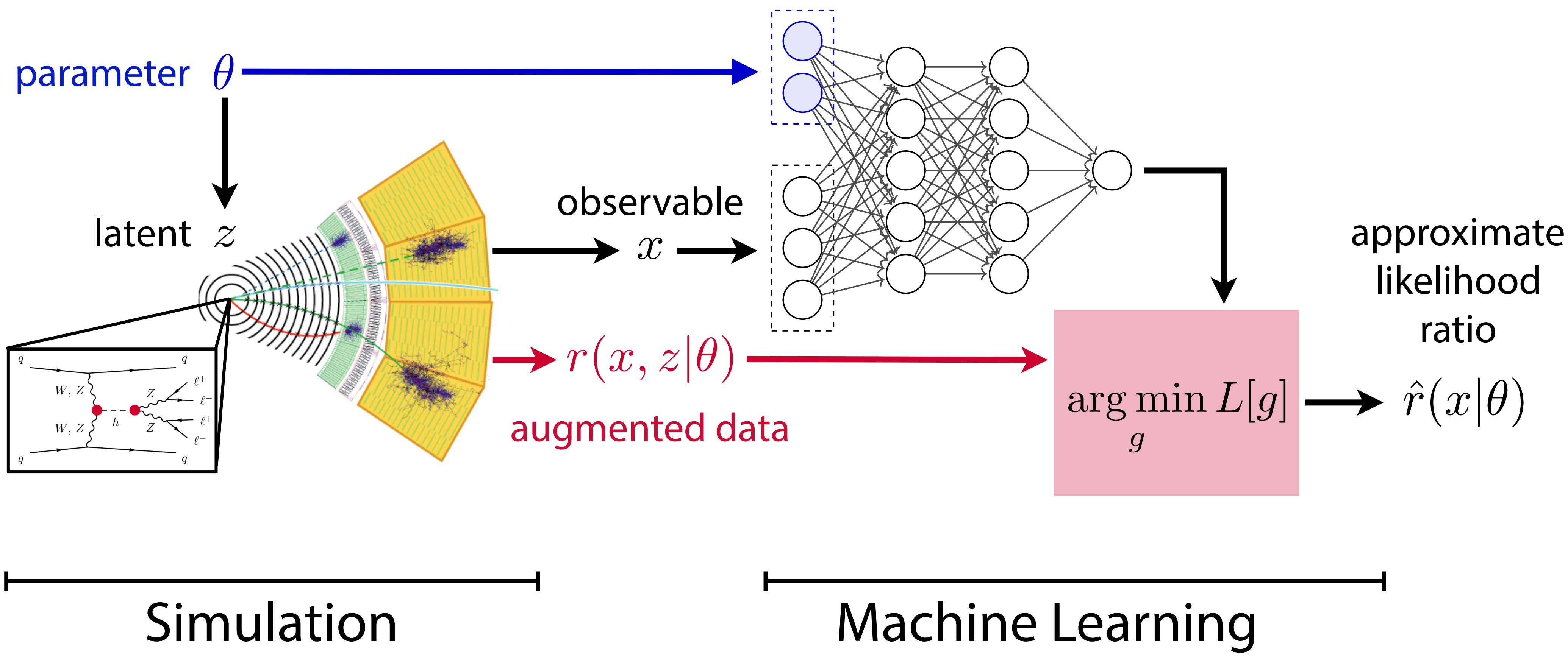
What we have so far



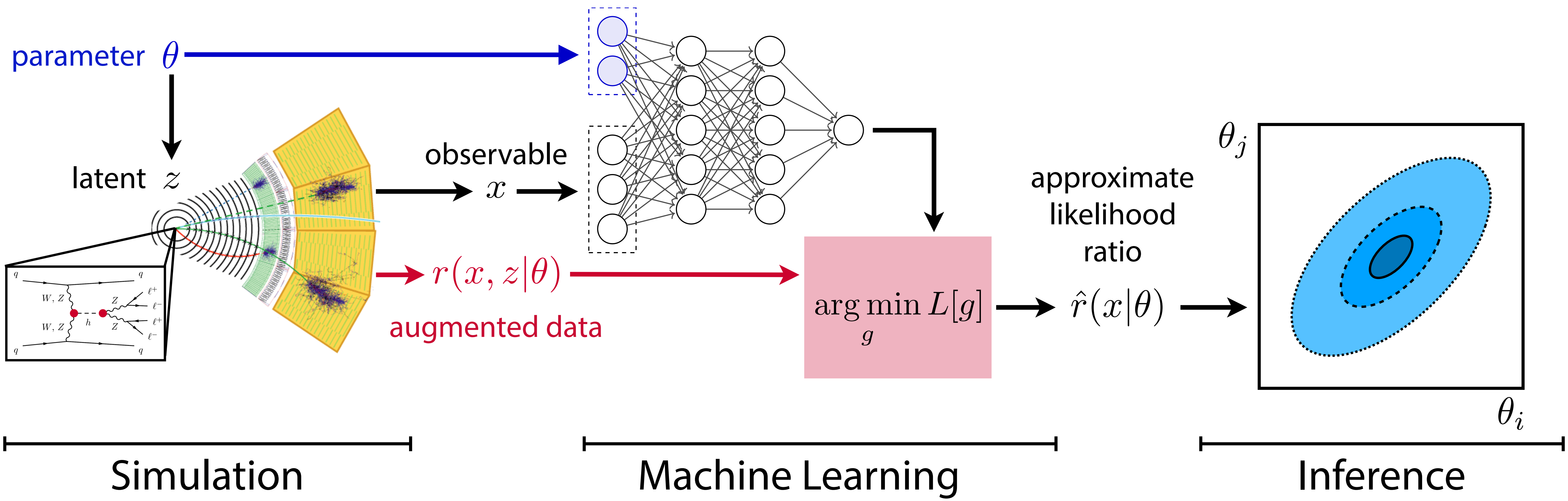
What we have so far



What we have so far



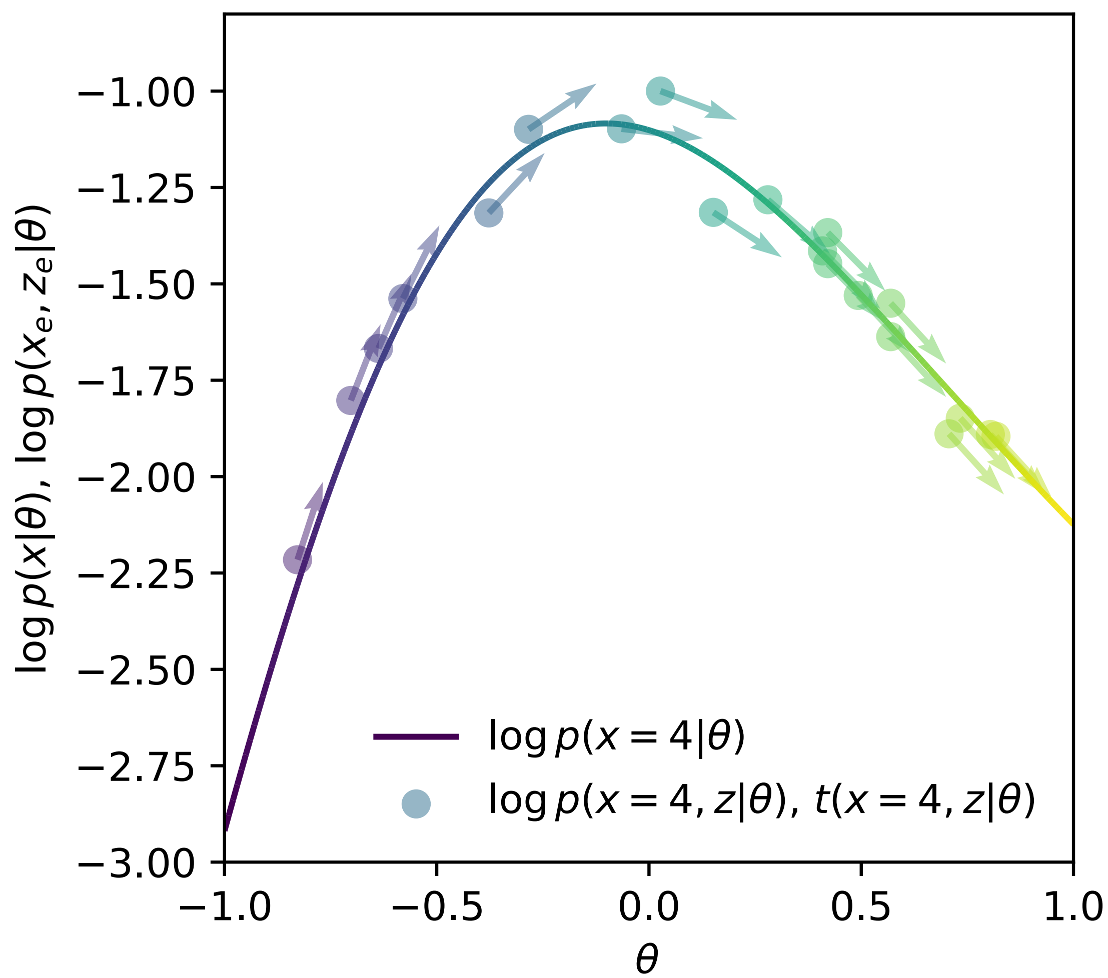
What we have so far



One more piece: the score

Score $t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$: tangent vectors of likelihood function

("How much more likely does the observation x get when I move in θ space?")



Learning the score

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

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

We want score

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

Learning the score

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

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

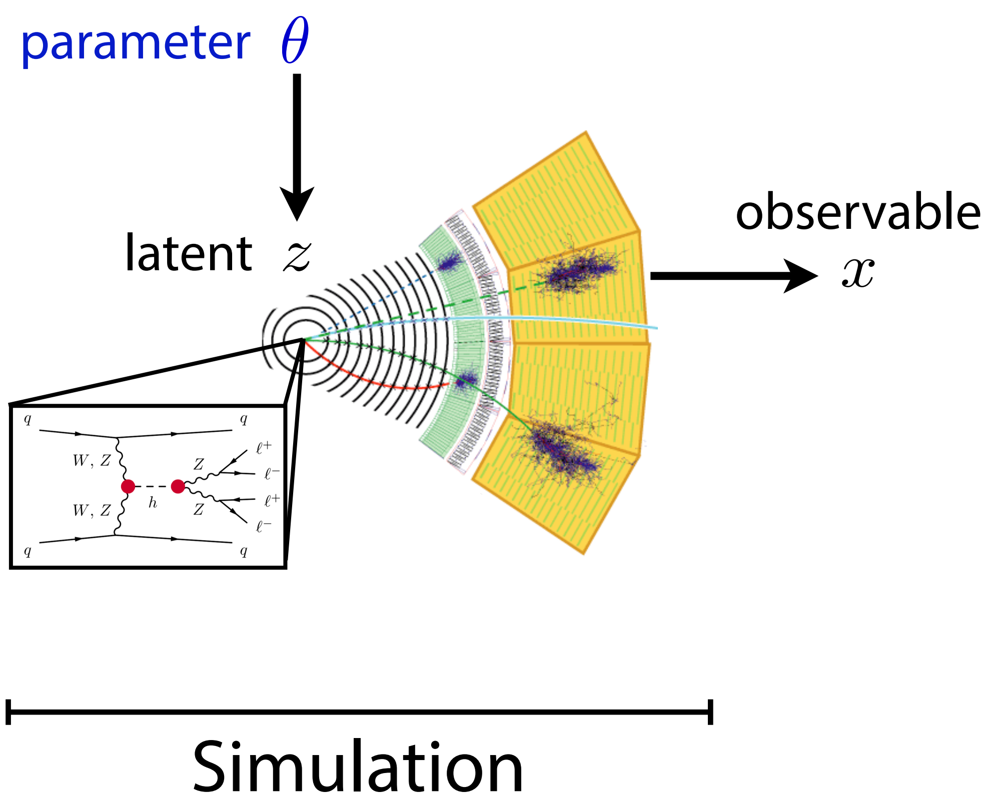
From $t(x, z|\theta_0)$,
we can learn $\hat{t}(x|\theta_0)$:

$$\hat{t}(x|\theta_0) = \arg\min_g L[g] \rightarrow t(x|\theta_0)$$

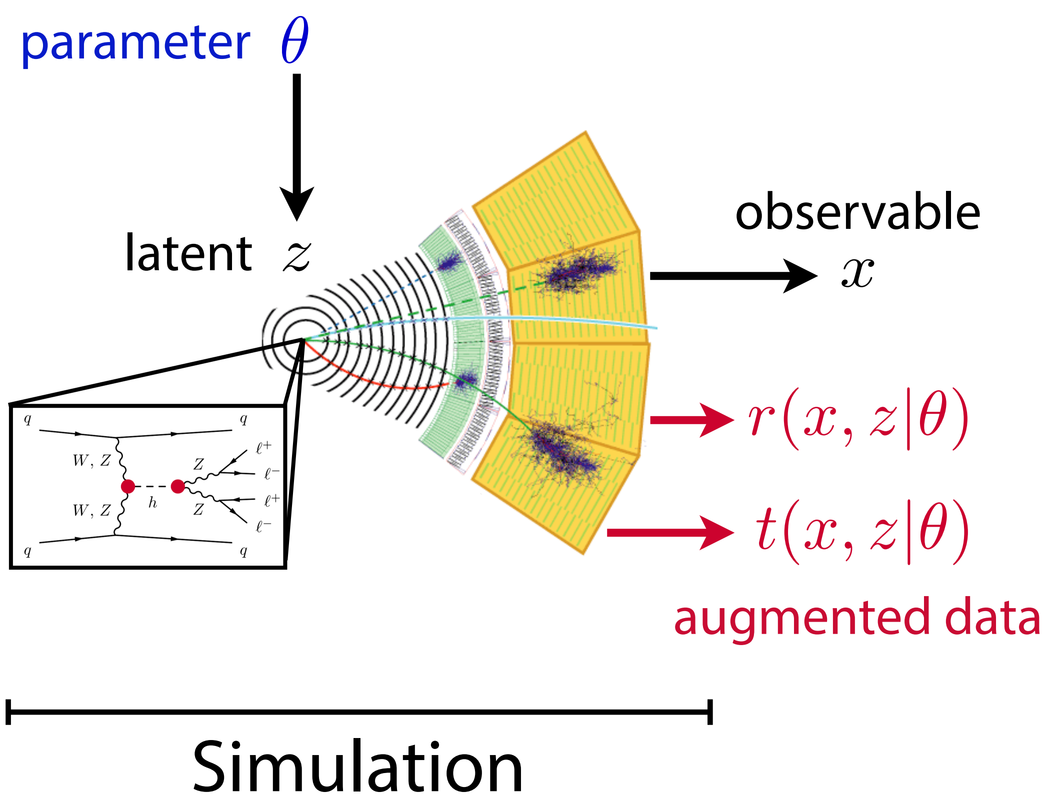
We want score

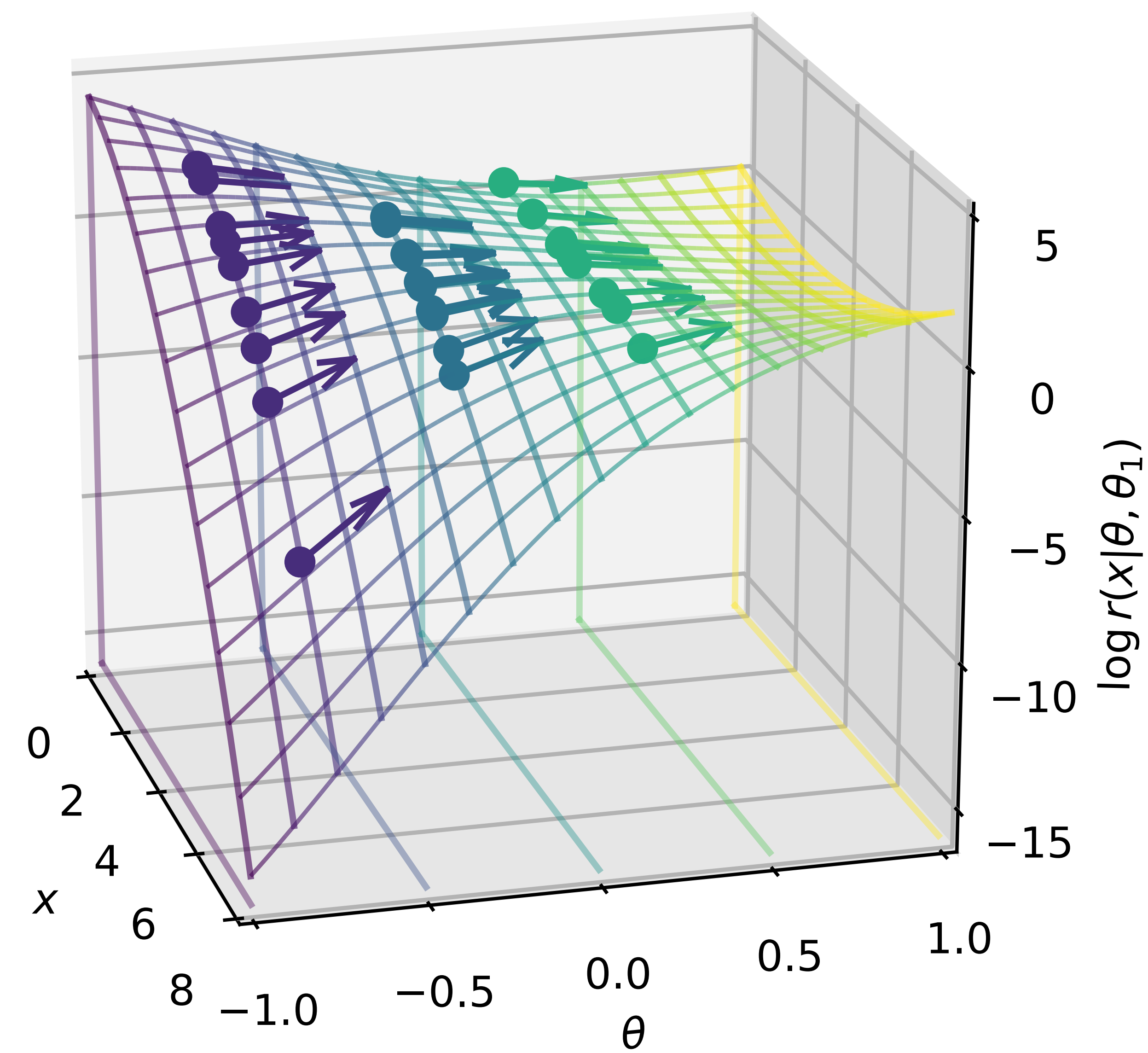
$$t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta)\Big|_{\theta_0}$$

RASCAL (Ratio And SCore Approximate Likelihood ratio)

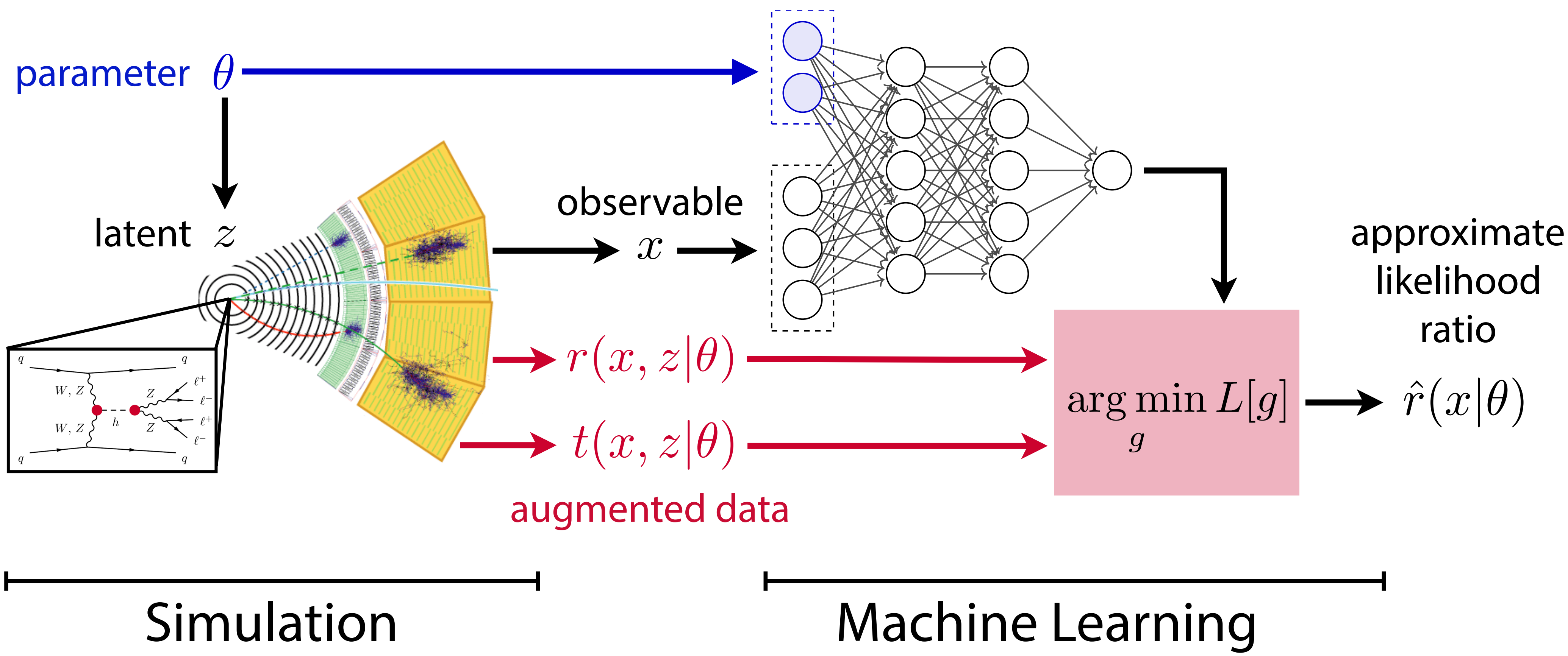


RASCAL (Ratio And SCore Approximate Likelihood ratio)

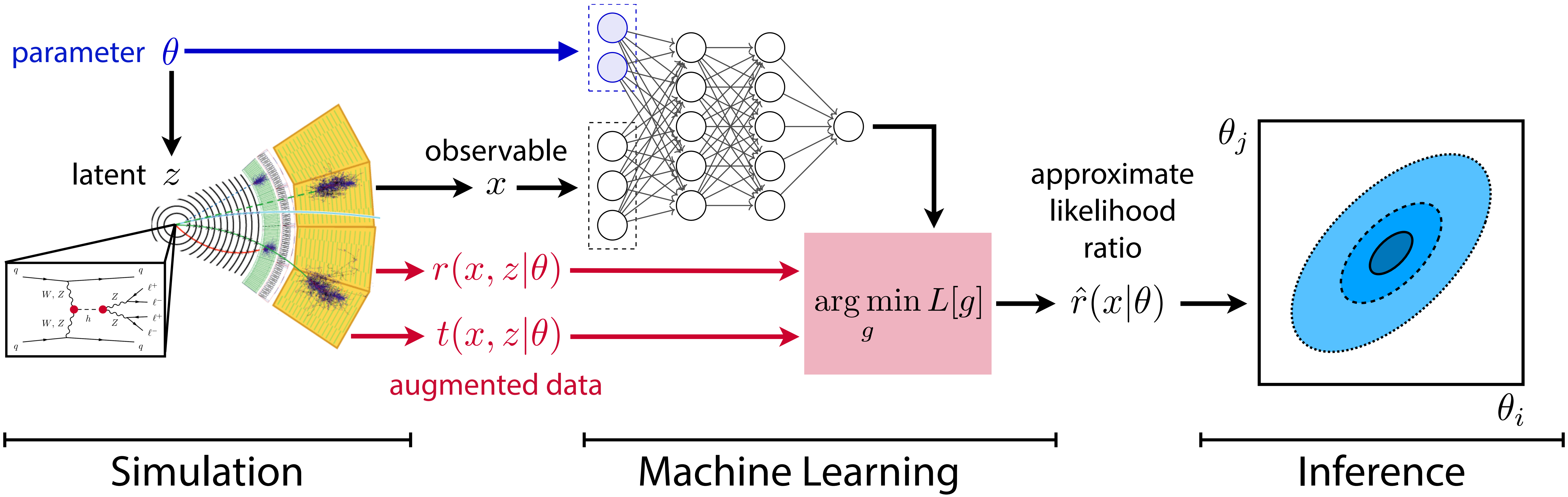




RASCAL (Ratio And SCore Approximate Likelihood ratio)



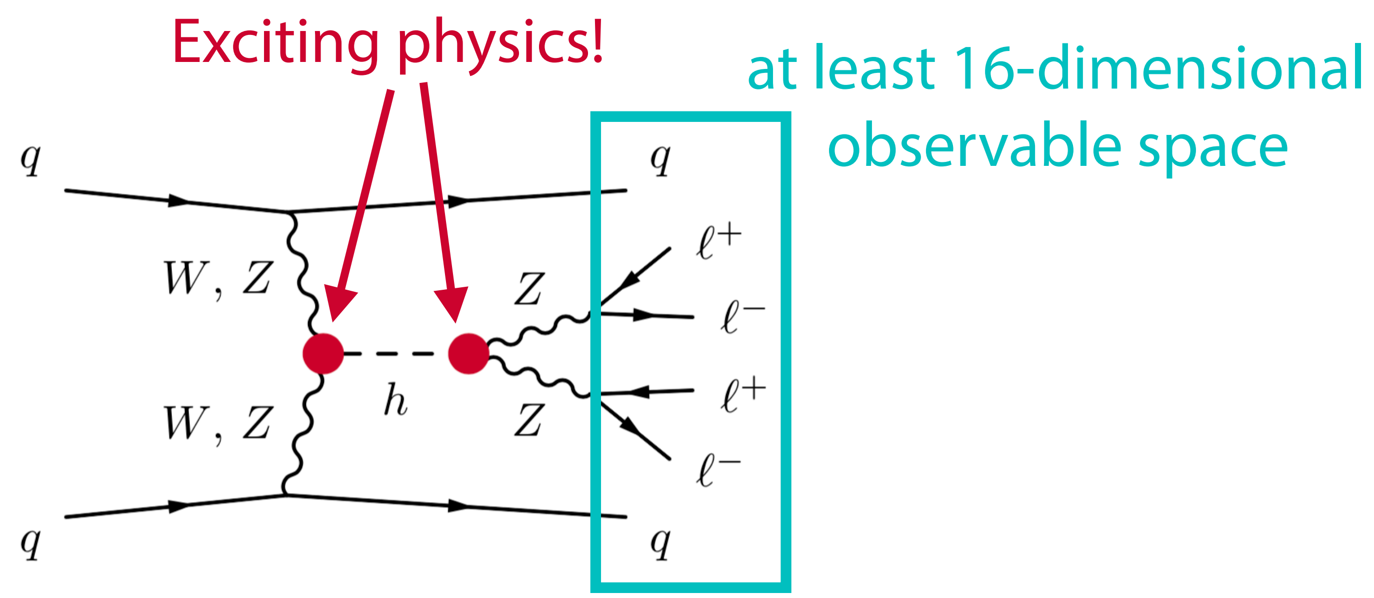
RASCAL (Ratio And SCore Approximate Likelihood ratio)



How well does it work?

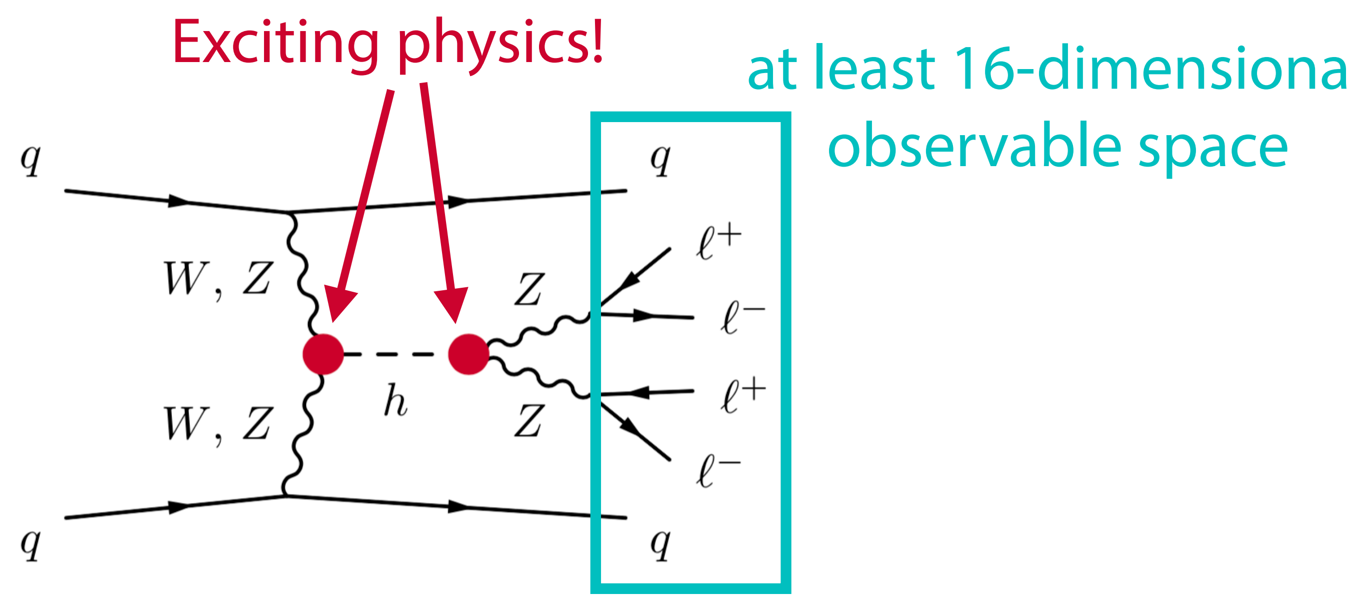
Proof of concept

- Higgs production in weak boson fusion:



Proof of concept

- Higgs production in weak boson fusion:

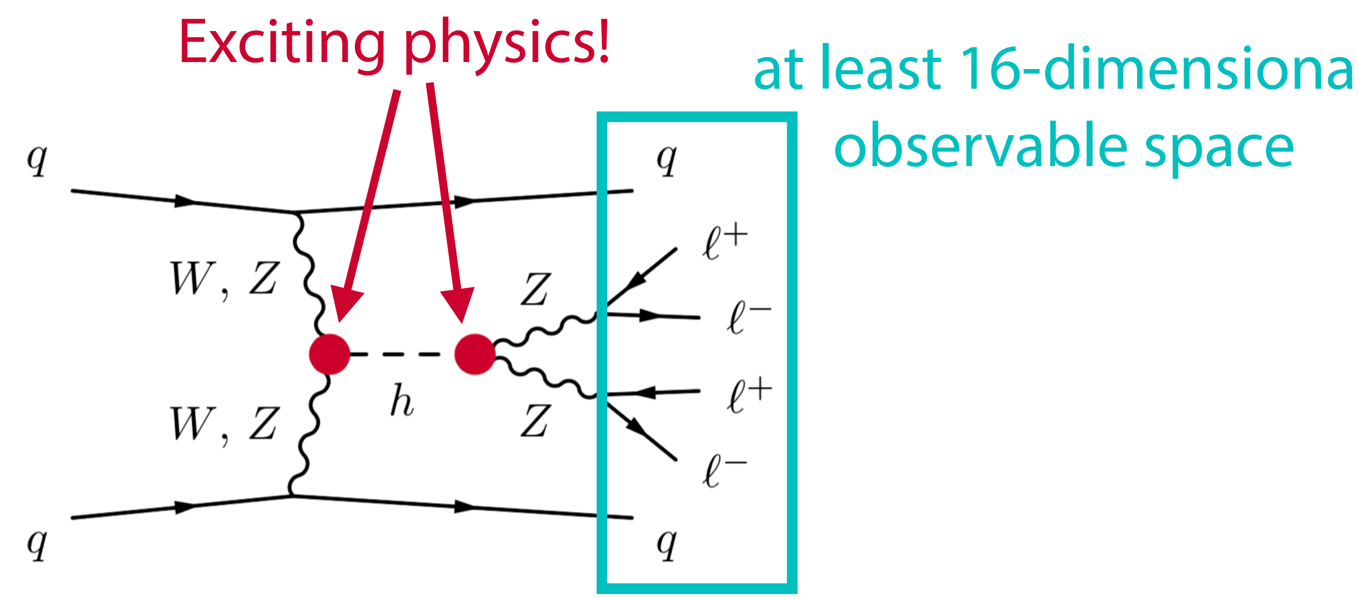


- Goal: constraints on two theory parameters

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

Proof of concept

- Higgs production in weak boson fusion:



- Goal: constraints on two theory parameters

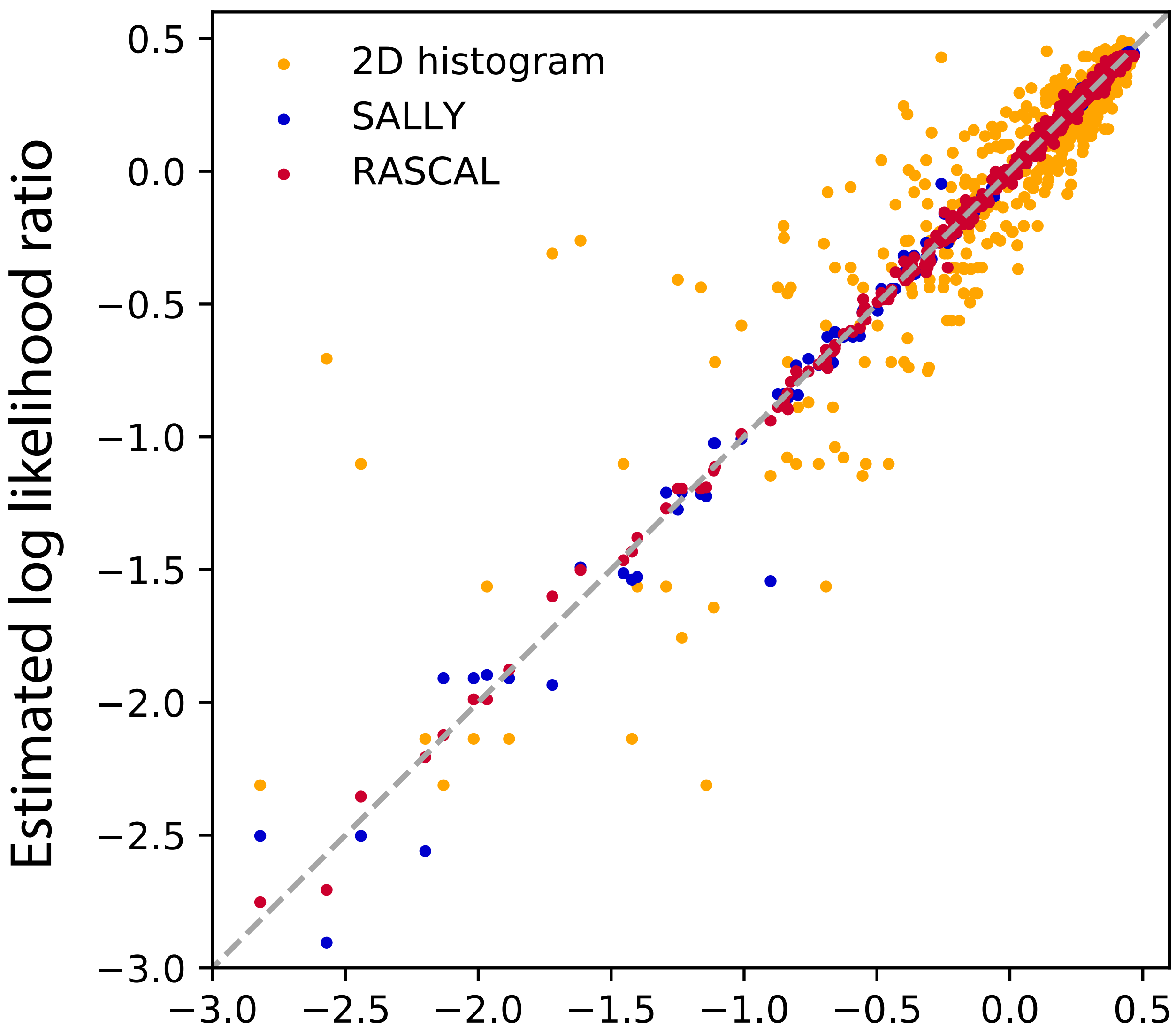
$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \boxed{\frac{f_W}{\Lambda^2}} \underbrace{\frac{ig}{2} \, (D^\mu \phi)^\dagger \, \sigma^a \, D^\nu \phi \, W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}} \underbrace{\frac{g^2}{4} \, (\phi^\dagger \phi) \, W^a_{\mu\nu} \, W^{\mu\nu\, a}}_{\mathcal{O}_{WW}}$$

- Two setups:
    - Simplified setup in which we can compare to true likelihood
    - “Realistic” simulation with approximate detector effects

- Simulation: MadGraph [J. Alwall et al. 1405.0301] + MadMax [K. Cranmer, T. Plehn hep-ph/0605268; T. Plehn, P. Schichtel, D. Wiegand 1311.2591]

Precise likelihood ratio estimates

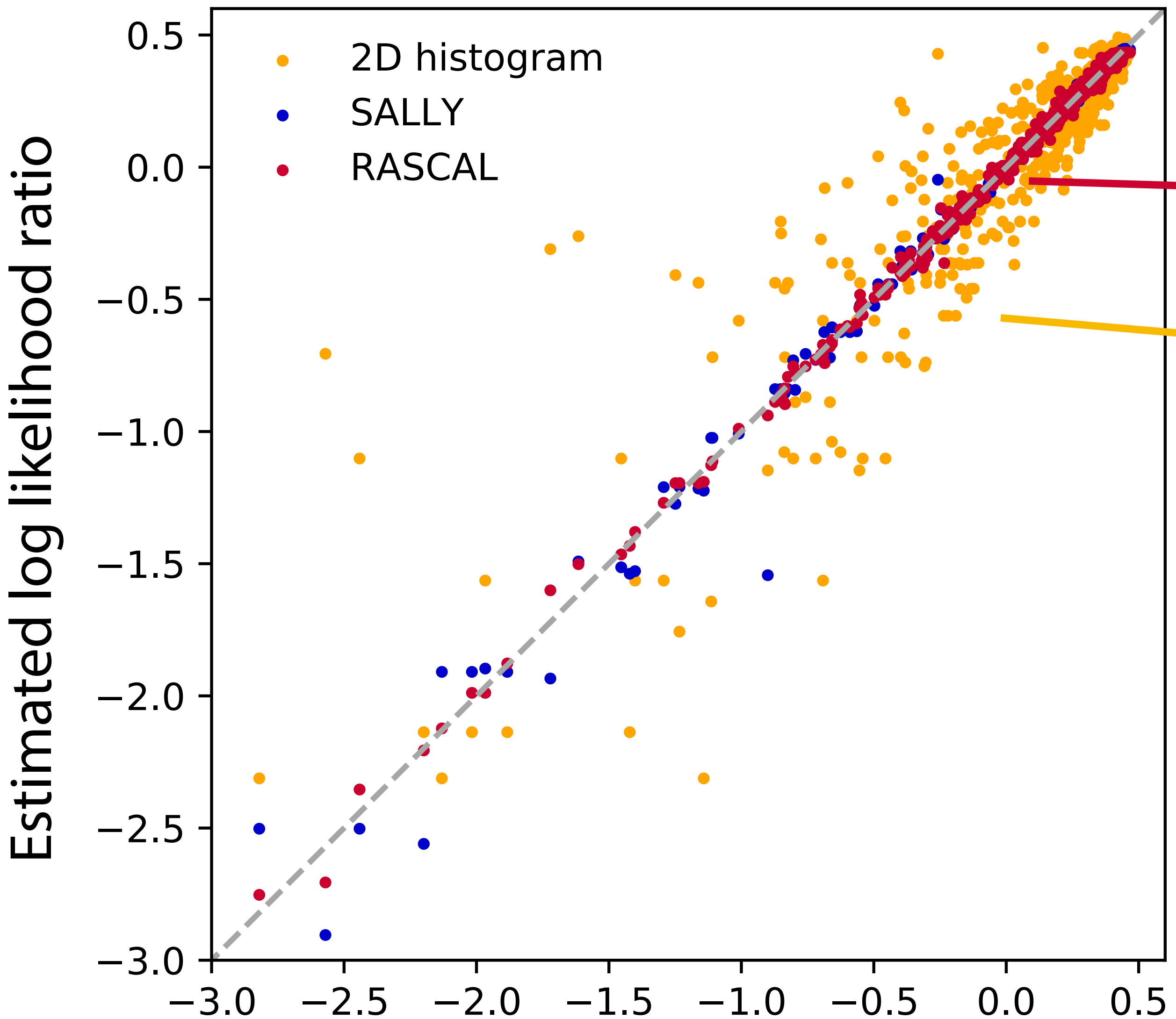
Each point = 1 event



True log likelihood ratio
(usually don't have this)

Precise likelihood ratio estimates

Each point = 1 event



True log likelihood ratio

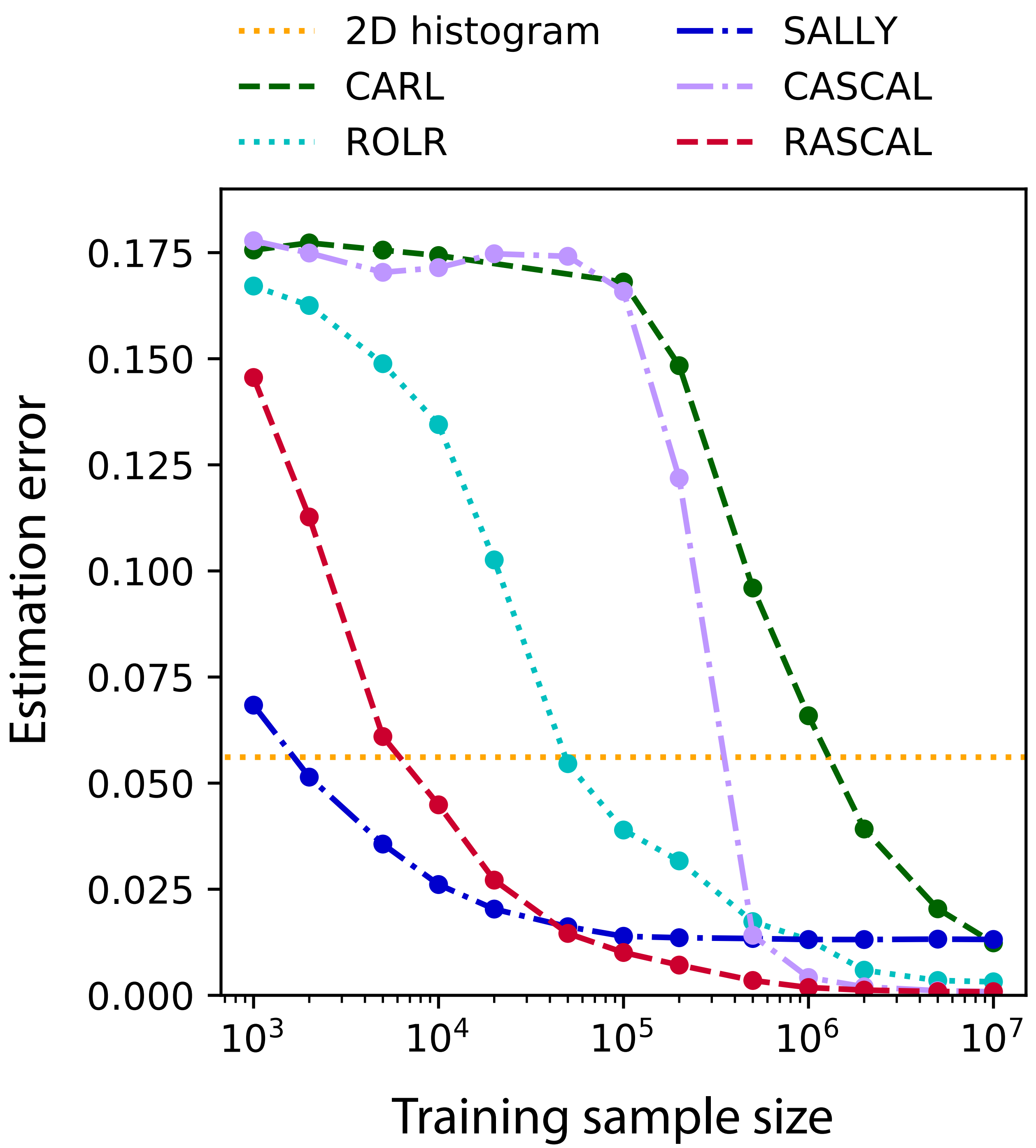
(usually don't have this)

RASCAL

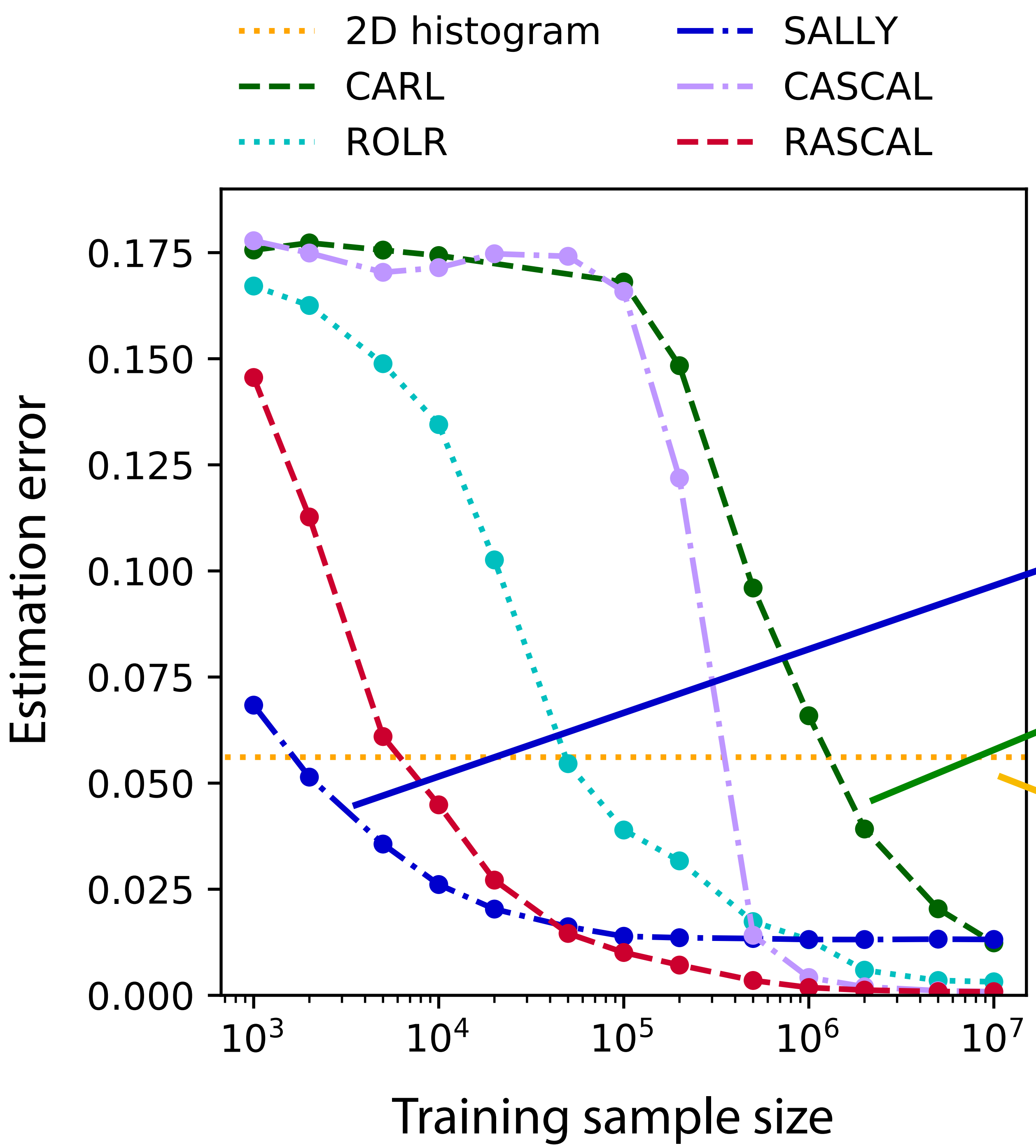
much more precise than

traditional histogram

Less training data needed



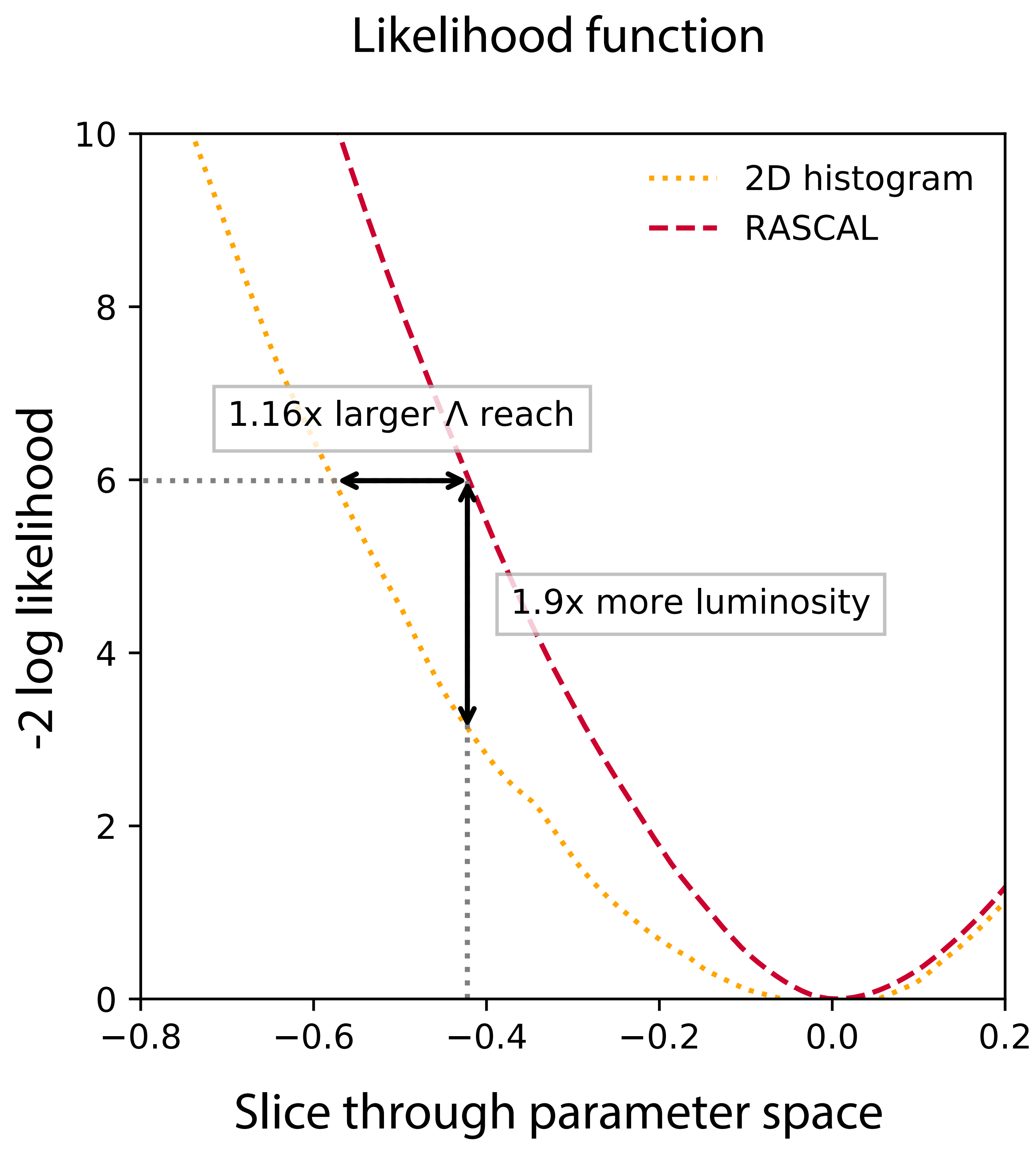
Less training data needed



New techniques
require less data than
generic ML method
[CARL, see K. Cranmer, G. Louppe, J. Pavez 1506.02169]

Traditional Histogram

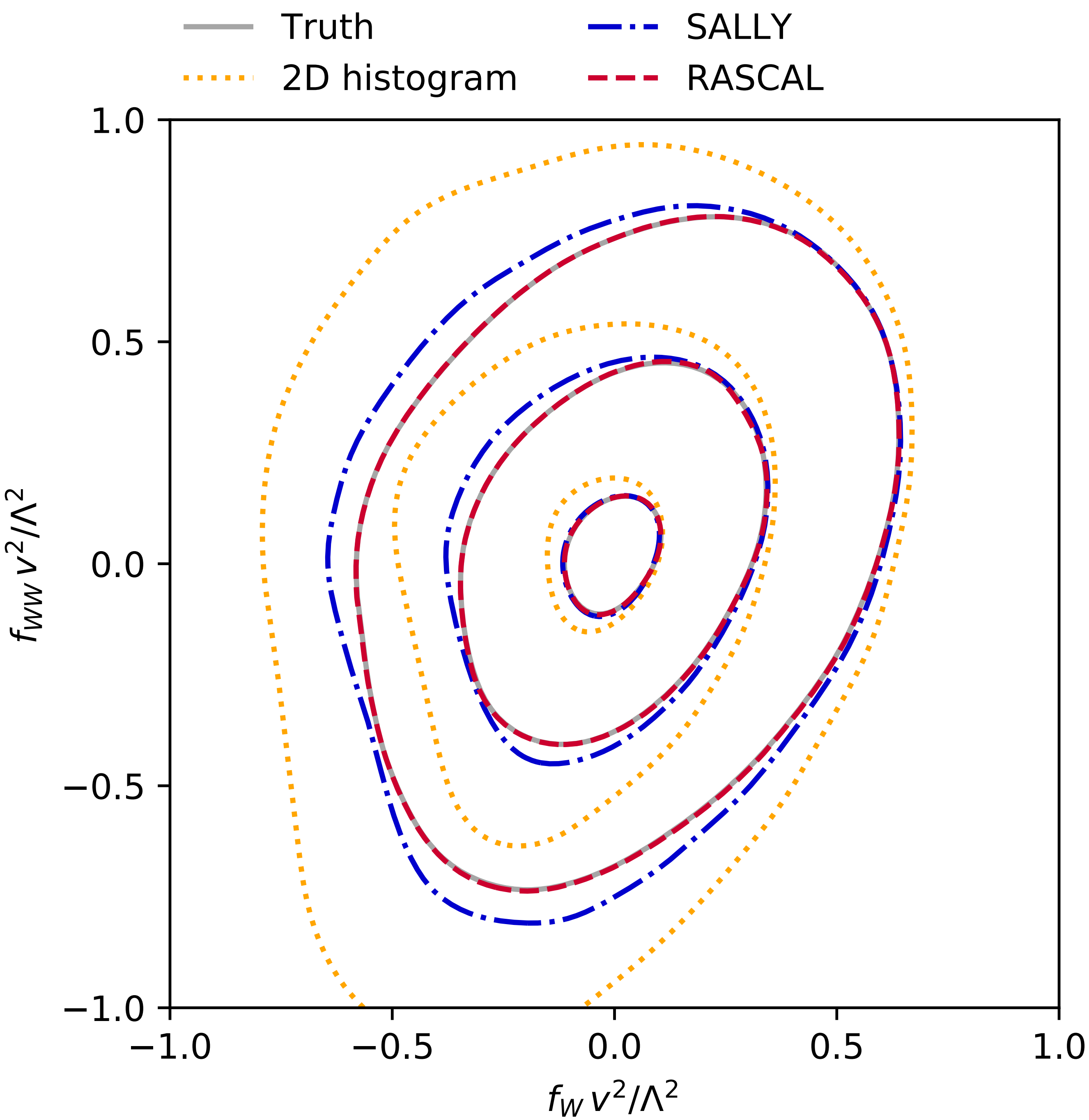
Better sensitivity



36 events, assuming SM

Stronger bounds

Expected exclusion limits at 68%, 95%, 99.7% CL

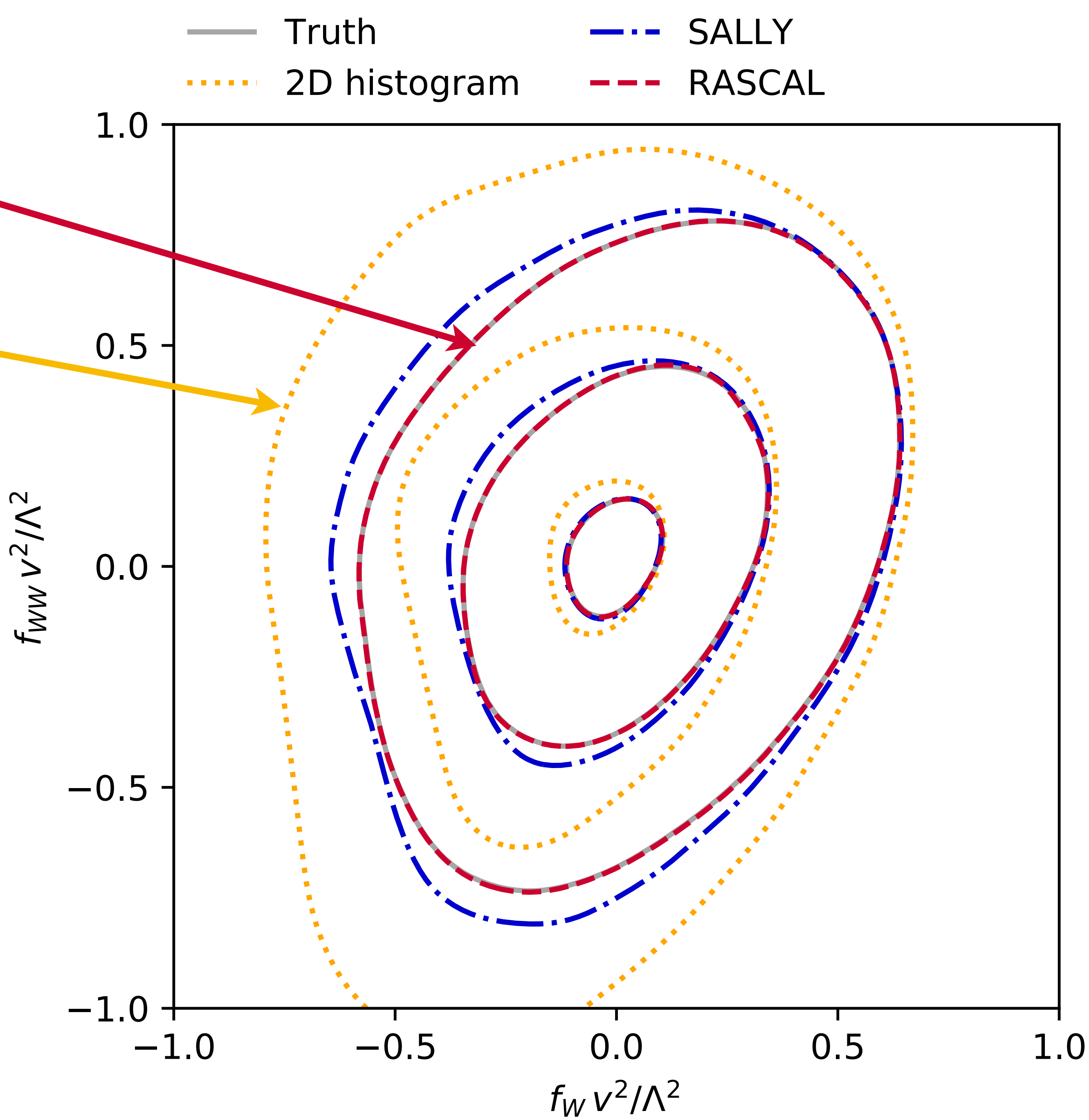


36 events, assuming SM

Stronger bounds

RASCAL enables stronger limits than traditional histogram

Expected exclusion limits at 68%, 95%, 99.7% CL



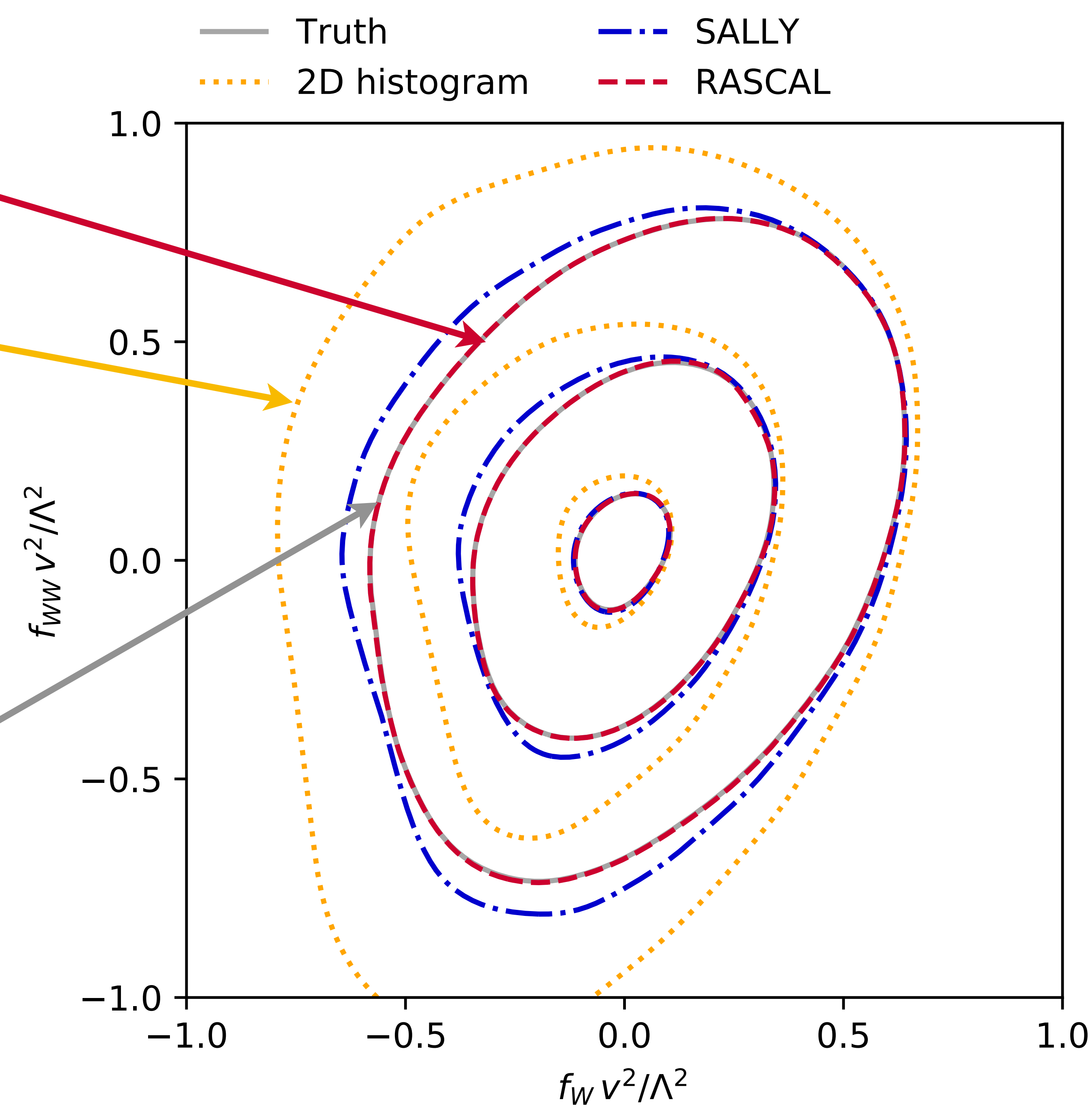
36 events, assuming SM

Stronger bounds

RASCAL enables stronger limits than traditional histogram

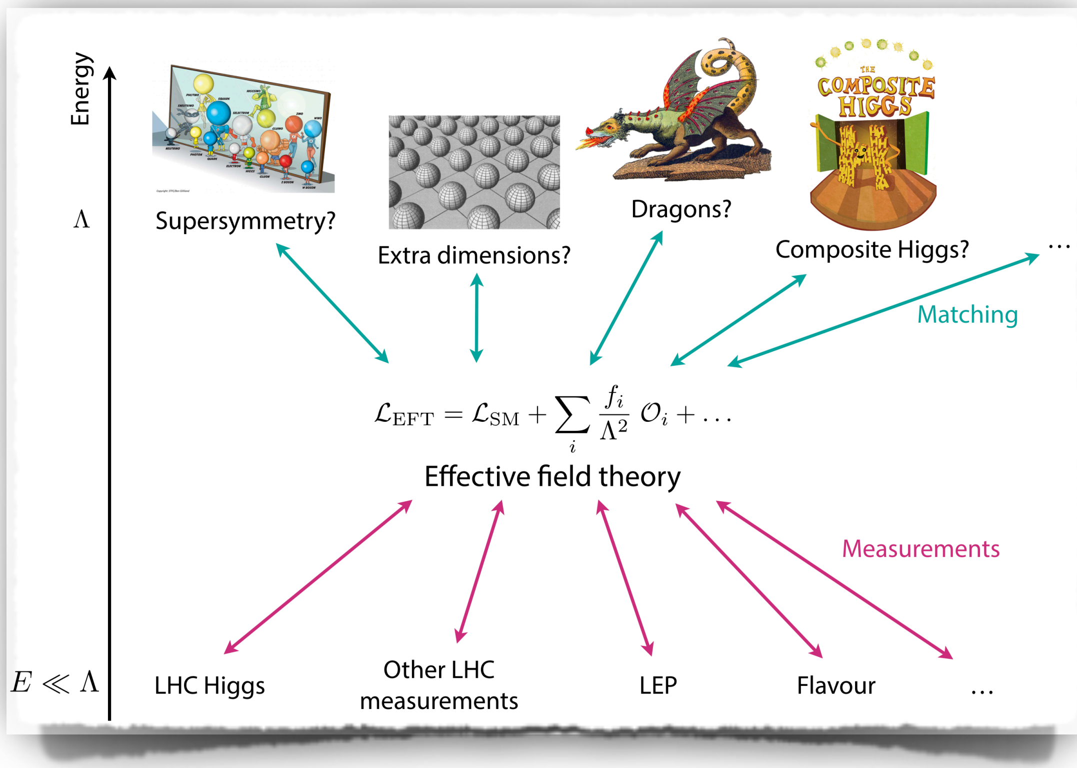
Limits from RASCAL virtually indistinguishable from true likelihood
(usually we don't have that)

Expected exclusion limits at 68%, 95%, 99.7% CL



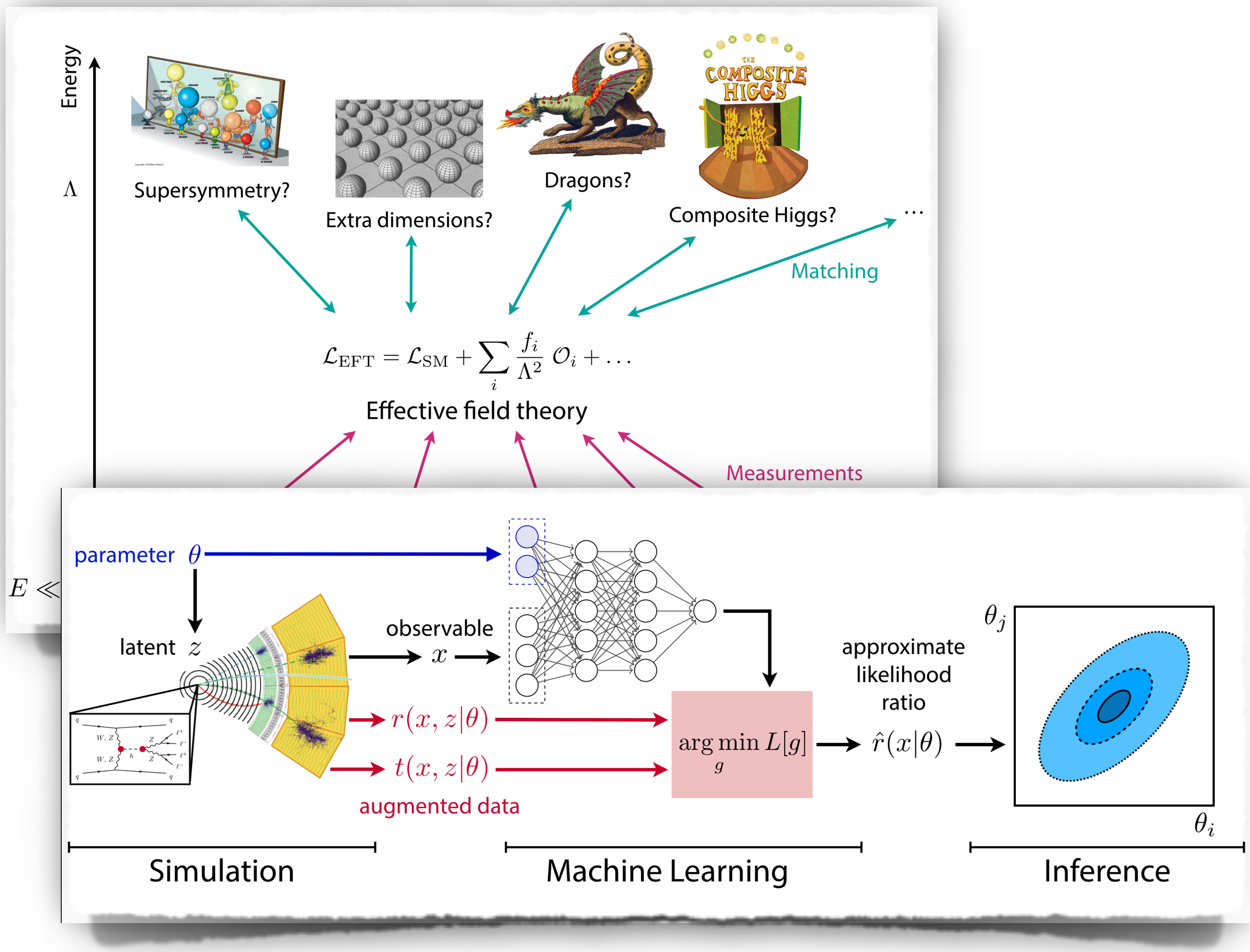
36 events, assuming SM

Conclusions



- EFT measurements at LHC might point us to exciting new physics

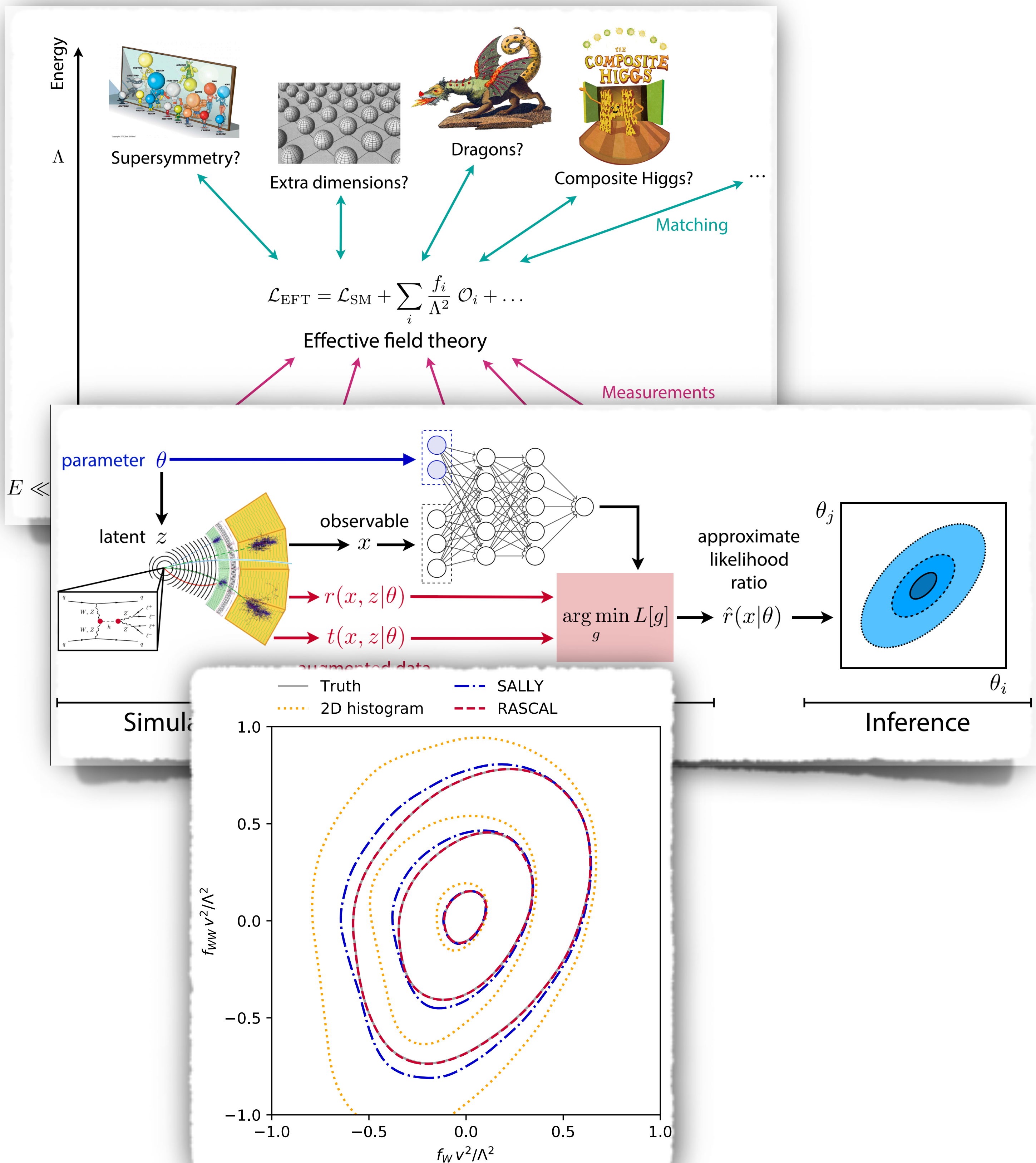
Conclusions



- EFT measurements at LHC might point us to exciting new physics

- New analysis techniques for difficult measurements: Knowledge of physics structure + power of Machine Learning

Conclusions



- EFT measurements at LHC might point us to exciting new physics

- New analysis techniques for difficult measurements: Knowledge of physics structure + power of Machine Learning

- First tests:
  stronger bounds with less data

Collaborators

Kyle Cranmer

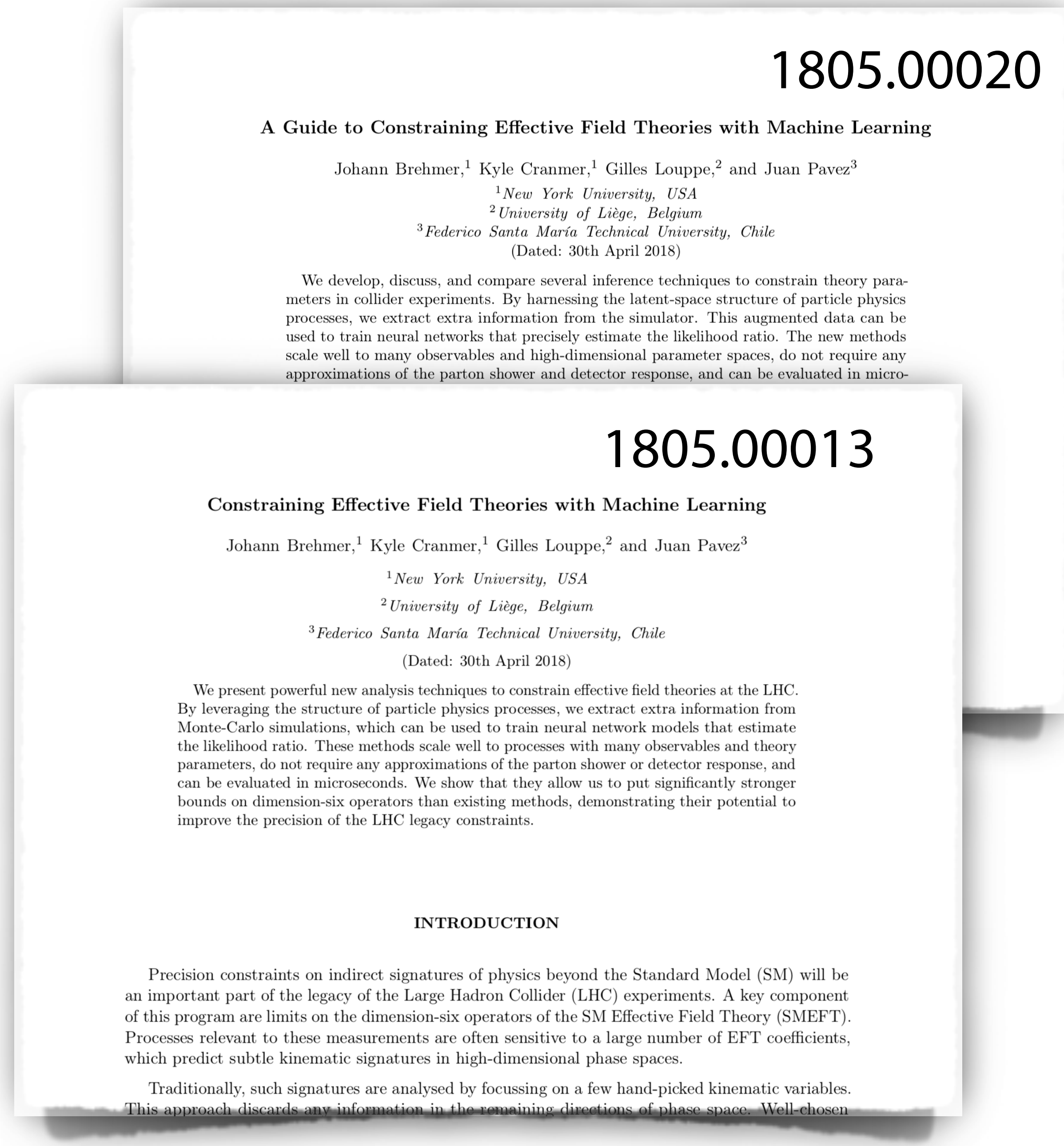
Gilles Louppe

Juan Pavez

Tilman Plehn

Felix Kling

Tim Tait



1805.00020

A Guide to Constraining Effective Field Theories with Machine Learning

Johann Brehmer,[1] Kyle Cranmer,[1] Gilles Louppe,[2] and Juan Pavez[3]

[1]*New York University, USA*
[2]*University of Liège, Belgium*
[3]*Federico Santa María Technical University, Chile*
(Dated: 30th April 2018)

We develop, discuss, and compare several inference techniques to constrain theory parameters in collider experiments. By harnessing the latent-space structure of particle physics processes, we extract extra information from the simulator. This augmented data can be used to train neural networks that precisely estimate the likelihood ratio. The new methods scale well to many observables and high-dimensional parameter spaces, do not require any approximations of the parton shower and detector response, and can be evaluated in micro-

1805.00013

Constraining Effective Field Theories with Machine Learning

Johann Brehmer,[1] Kyle Cranmer,[1] Gilles Louppe,[2] and Juan Pavez[3]

[1]*New York University, USA*
[2]*University of Liège, Belgium*
[3]*Federico Santa María Technical University, Chile*
(Dated: 30th April 2018)

We present powerful new analysis techniques to constrain effective field theories at the LHC. By leveraging the structure of particle physics processes, we extract extra information from Monte-Carlo simulations, which can be used to train neural network models that estimate the likelihood ratio. These methods scale well to processes with many observables and theory parameters, do not require any approximations of the parton shower or detector response, and can be evaluated in microseconds. We show that they allow us to put significantly stronger bounds on dimension-six operators than existing methods, demonstrating their potential to improve the precision of the LHC legacy constraints.

INTRODUCTION

Precision constraints on indirect signatures of physics beyond the Standard Model (SM) will be an important part of the legacy of the Large Hadron Collider (LHC) experiments. A key component of this program are limits on the dimension-six operators of the SM Effective Field Theory (SMEFT). Processes relevant to these measurements are often sensitive to a large number of EFT coefficients, which predict subtle kinematic signatures in high-dimensional phase spaces.

Traditionally, such signatures are analysed by focussing on a few hand-picked kinematic variables. This approach discards any information in the remaining directions of phase space. Well-chosen

Thanks to Kyle Cranmer for ~~letting me steal~~ inspiring many slides!

Unless stated otherwise, images are our own work, or from Wikimedia Commons and in the public domain.

Deleted scenes

Dictionary

Symbol	Physics meaning	Machine learning abstraction
x	Set of all observables	Features
v	One or two kinematic variables	Low-dimensional summary statistics / engineered feature
$z \equiv z_{\text{parton}}$	Parton-level four-momenta	Latent variables
z_{shower}	Parton shower trajectories	Latent variables
z_{detector}	Detector interactions	Latent variables
$z_{\text{all}} = (z_{\text{parton}}, z_{\text{shower}}, z_{\text{detector}})$	Full simulation history of event	All latent variables
θ	Theory parameters (Wilson coefficients)	Parameters of interest
$\hat{\theta}$	Best fit for theory parameters	Estimator for parameters of interest
$p(x\|\theta)$	Distributions of observables given theory parameters	Intractable likelihood
$p(z\|\theta)$	Parton-level distributions from matrix element	Tractable likelihood of latent variables
$p(x\|z)$	Effect of shower, detector, reconstruction	Intractable density defined through stochastic generative process
$r(x\|\theta_0, \theta_1)$	Likelihood ratio between hypotheses θ_0, θ_1, see Eq. (11).	
$\hat{r}(x\|\theta_0, \theta_1)$	Estimator for likelihood ratio	
$t(x\|\theta)$	Score, see Eq. (14).	
$\hat{t}(x\|\theta)$	Estimator for score	
x_e, z_e	Event	Data point
θ_o	Wilson coefficient for one operator	Individual parameter of interest
θ_c, $w_c(z)$, $p_c(x)$	Morphing basis points, coefficients, densities, see Eq. (6).	

For the Bayesians among us

$$\hat{r}(x|\theta) \approx \frac{p(x|\theta)}{\int \mathrm{d}\theta' p(x|\theta')\, p(\theta')}$$

$$p(\theta|x) = \frac{p(x|\theta)\, p(\theta)}{\int \mathrm{d}\theta' p(x|\theta')\, p(\theta')}$$

$$\approx \hat{r}(x|\theta)\, p(\theta)$$

Available information

Quantity		General likelihood-free	Particle physics
Samples	$\{x_e\}$	✓	✓
Likelihood	$p(x_e \vert \theta)$		
Likelihood ratio	$r(x_e \vert \theta_0, \theta_1)$		$*$
Score	$t(x_e \vert \theta)$		$*$
Latent state	$\{x_e, z_e\}$		✓
Joint likelihood	$p(x_e, z_{\text{all},e} \vert \theta)$		
Joint likelihood ratio	$r(x_e, z_{\text{all},e} \vert \theta_0, \theta_1)$		✓
Joint score	$t(x_e, z_{\text{all},e} \vert \theta)$		✓

Variational calculus

$$
\begin{aligned}
L[\hat{g}(x)] &= \int \mathrm{d}x \, \mathrm{d}z \; p(x,z|\theta) \; |g(x,z) - \hat{g}(x)|^2 \\
&= \int \mathrm{d}x \; \underbrace{\left[\hat{g}^2(x) \int \mathrm{d}z \; p(x,z|\theta) - 2\hat{g}(x) \int \mathrm{d}z \; p(x,z|\theta) \, g(x,z) + \int \mathrm{d}z \; p(x,z|\theta) \, g^2(x,z) \right]}_{F(x)}
\end{aligned}
$$

$$
0 = \left. \frac{\delta F}{\delta \hat{g}} \right|_{g^*} = 2\hat{g} \underbrace{\int \mathrm{d}z \; p(x,z|\theta)}_{=p(x|\theta)} - 2 \int \mathrm{d}z \; p(x,z|\theta) \, g(x,z)
$$

$$
g^*(x) = \frac{1}{p(x|\theta)} \int \mathrm{d}z \; p(x,z|\theta) \, g(x,z)
$$

Loss functions

- Ratio estimator:

$$L[\hat{r}(x|\theta_0,\theta_1)] = \frac{1}{N} \sum_{(x_e,z_e)\sim p(x,z|\theta_1)} |r(x_e, z_{\text{all},e}|\theta_0,\theta_1) - \hat{r}(x|\theta_0,\theta_1)|^2$$

- Score estimator:

$$L[\hat{t}(x|\theta)] = \frac{1}{N} \sum_{(x_e,z_e)\sim p(x,z|\theta)} |t(x_e, z_{\text{all},e}|\theta) - \hat{t}(x|\theta)|^2$$

- Combined (RASCAL):

$$L[\hat{r}(x|\theta_0,\theta_1)] = \frac{1}{N} \sum_{(x_e,z_e,y_e)} \Bigg[y_e\, |r(x_e, z_e|\theta_0,\theta_1) - \hat{r}(x|\theta_0,\theta_1)|^2 + (1-y_e) \left| \frac{1}{r(x_e, z_e|\theta_0,\theta_1)} - \frac{1}{\hat{r}(x|\theta_0,\theta_1)} \right|^2 + \alpha\, y_e \, \left| t(x_e, z_e|\theta) - \hat{t}(x|\theta_0) \right|^2 \Bigg]$$

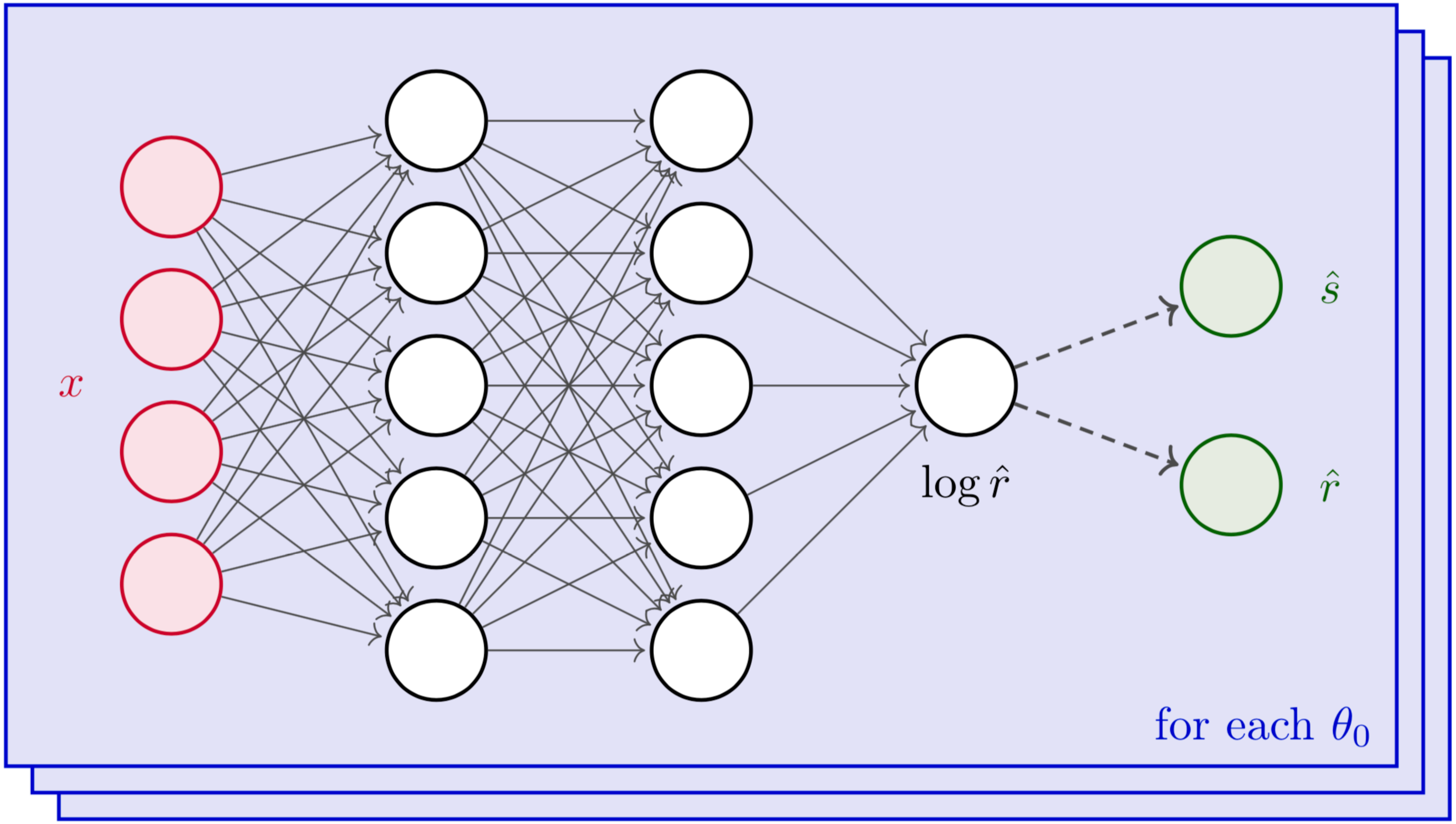
Three classes of estimators

- Point by point
- (Agnostic) parameterized
- Morphing-aware:

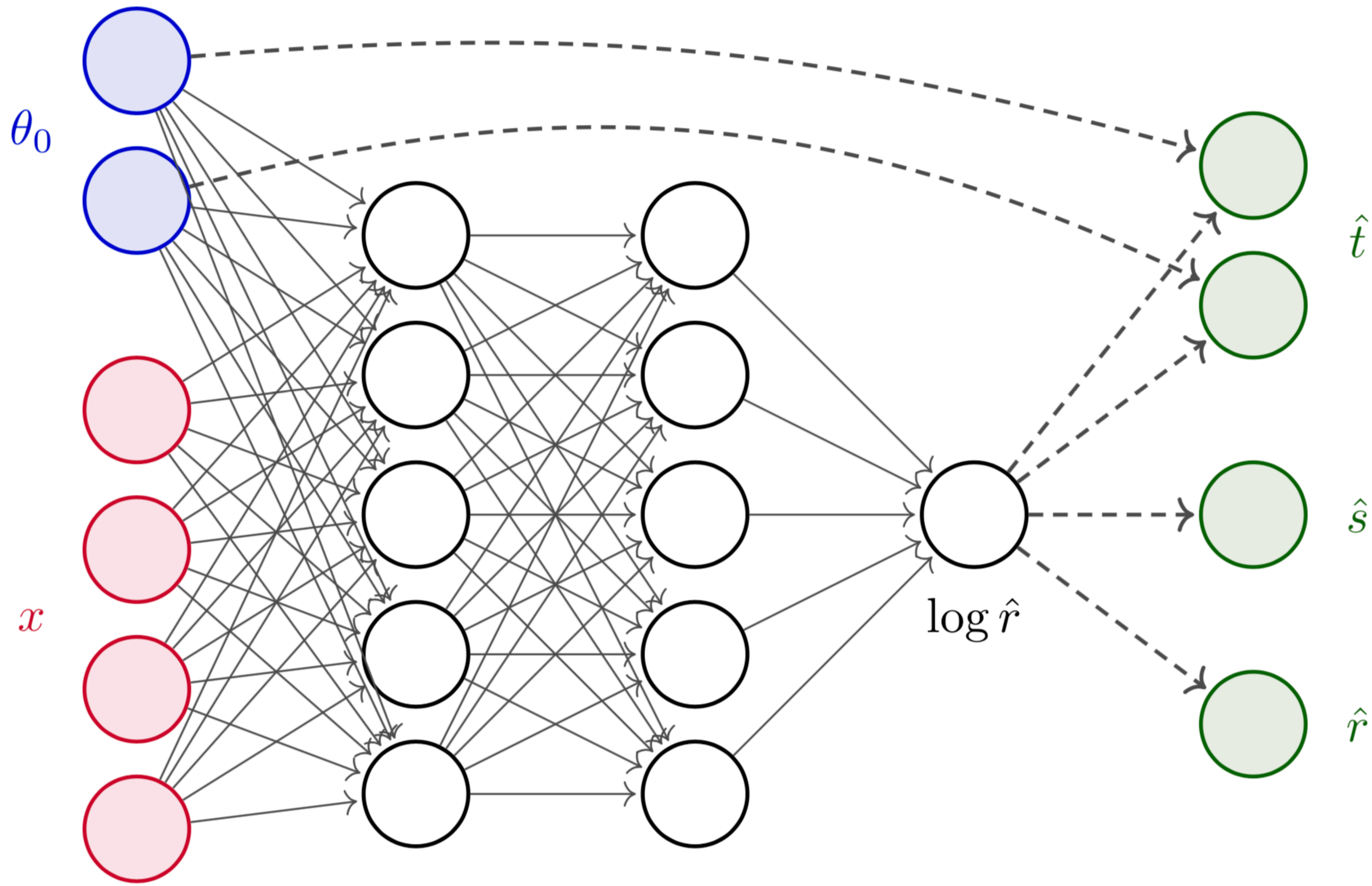
$$p(x|\theta) = \sum_c w_c(\theta)\, p_c(x)$$

$$\hat{r}(x|\theta_0, \theta_1) = \sum_c w_c(\theta_0)\, \hat{r}_c(x)$$

Point by point

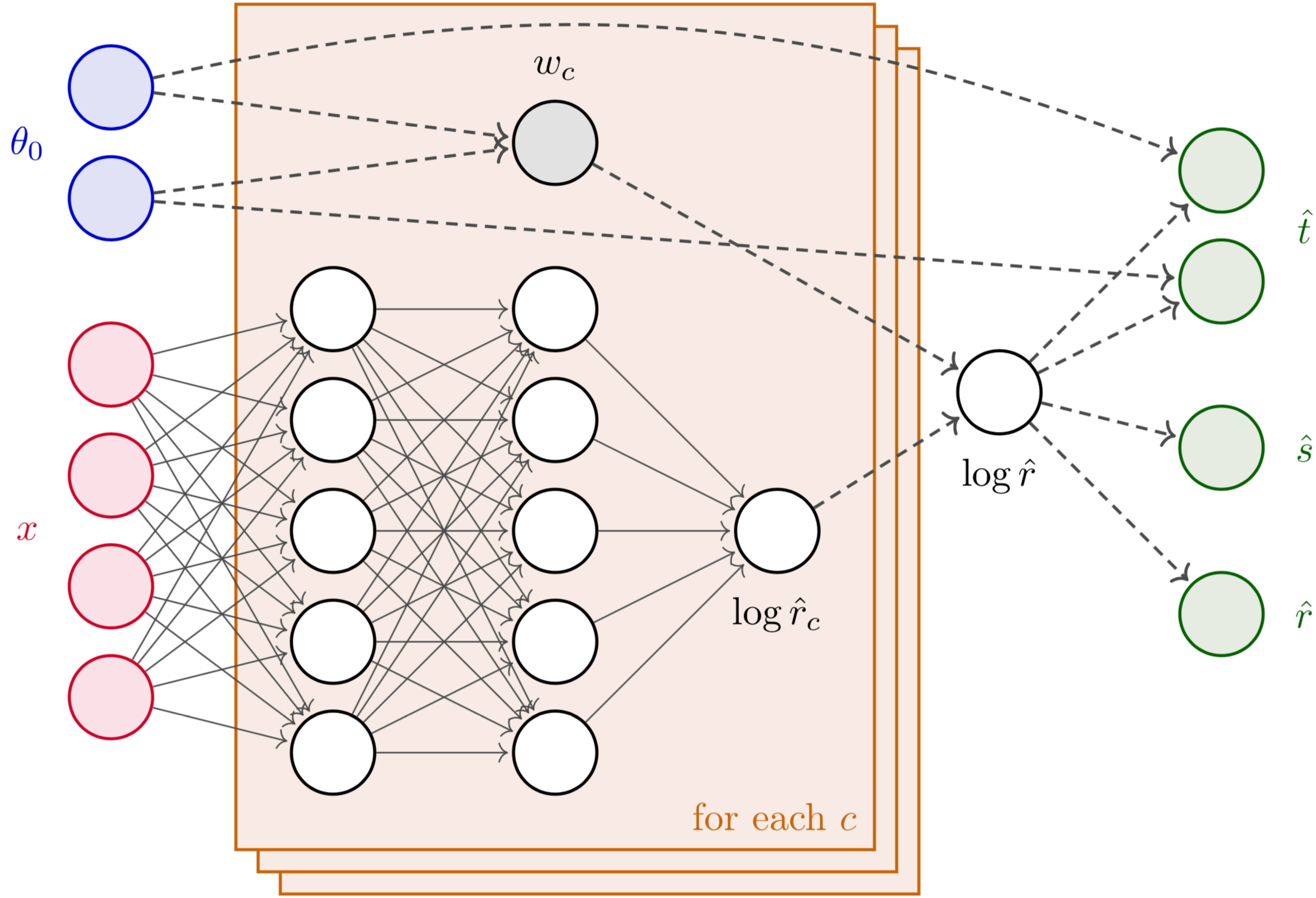


(Agnostic) parameterized

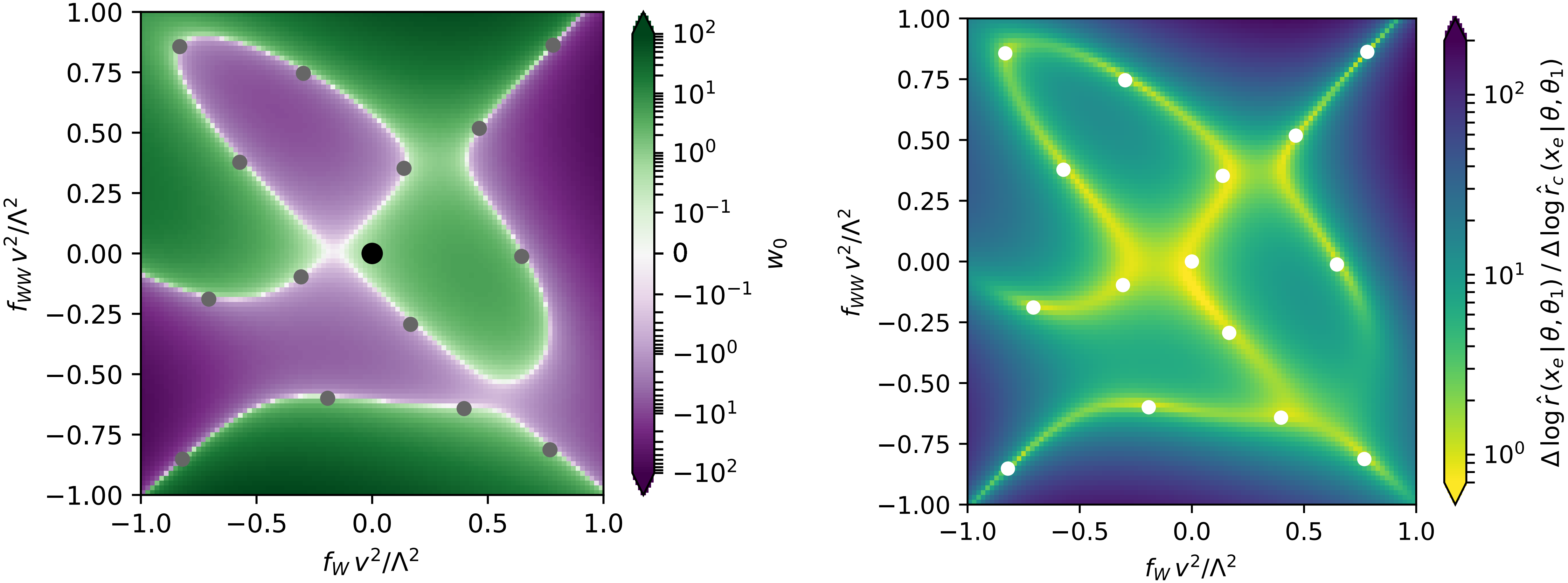


Morphing-aware

$$\hat{r}(x|\theta_0, \theta_1) = \sum_c w_c(\theta_0)\, \hat{r}_c(x)$$



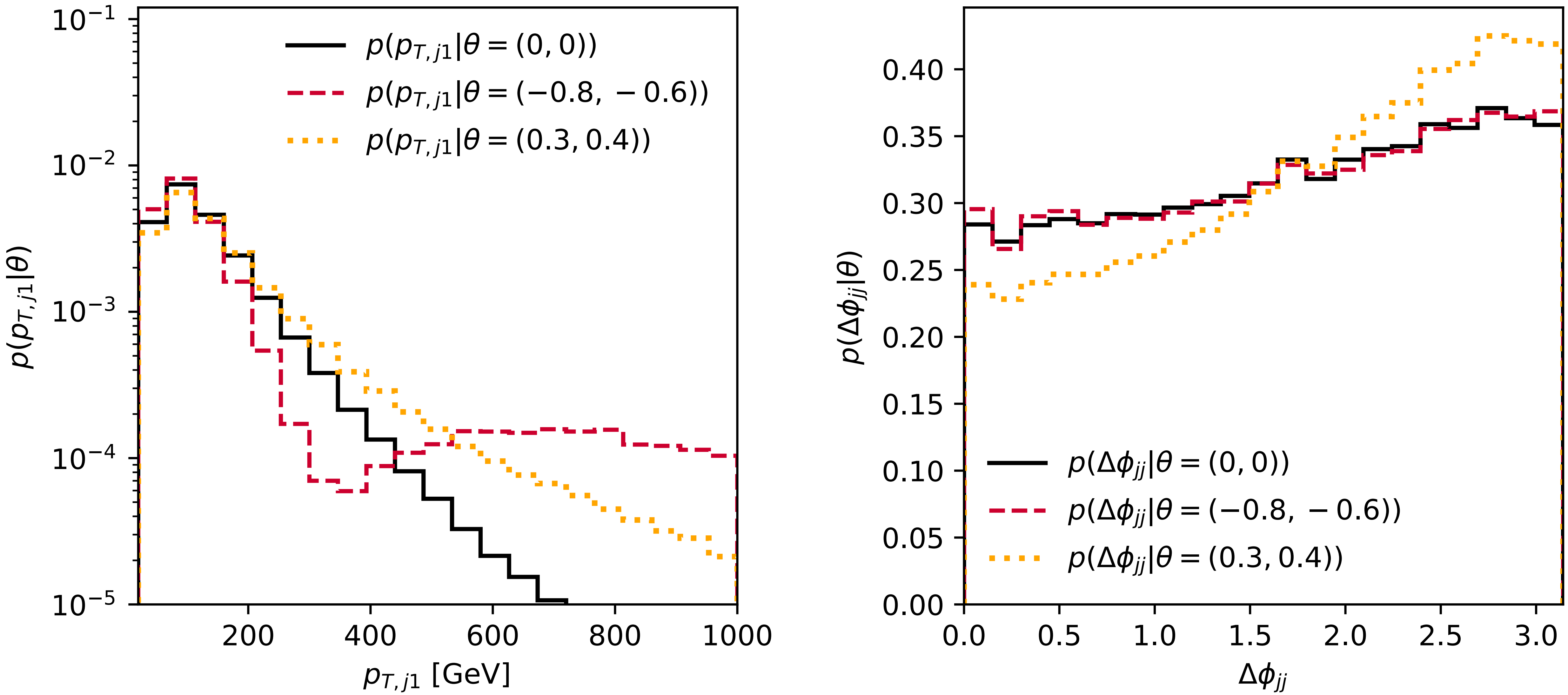
Morphing coefficients



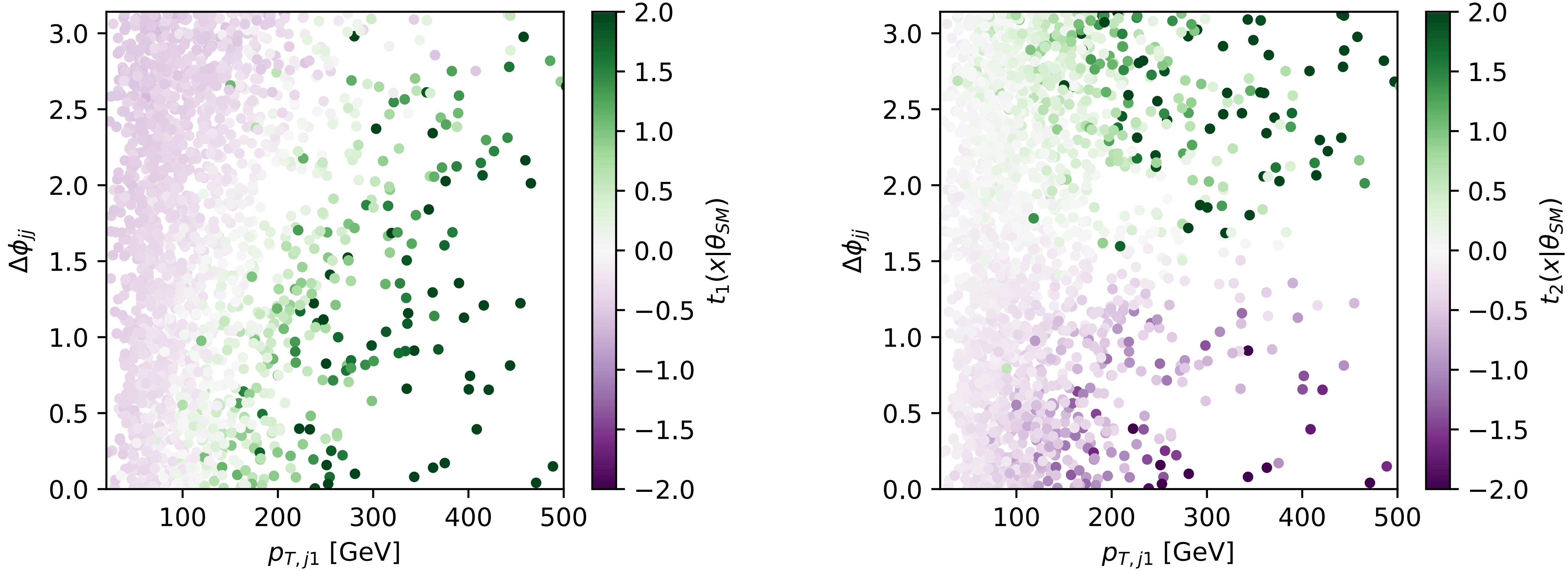
Analysis techniques

Strategy	Estimator versions			Loss function			Asymptotically exact
	PbP	Param	Aware	CE	Ratio	Score	
Histograms	✓		(✓)				
AFC	✓		(✓)				
CARL	✓	✓	✓	✓			✓
ROLR	✓	✓	✓		✓		✓
SALLY	✓		(✓)			✓	
SALLINO	✓		(✓)			✓	
CASCAL		✓	✓	✓		✓	✓
RASCAL		✓	✓		✓	✓	✓

Distributions



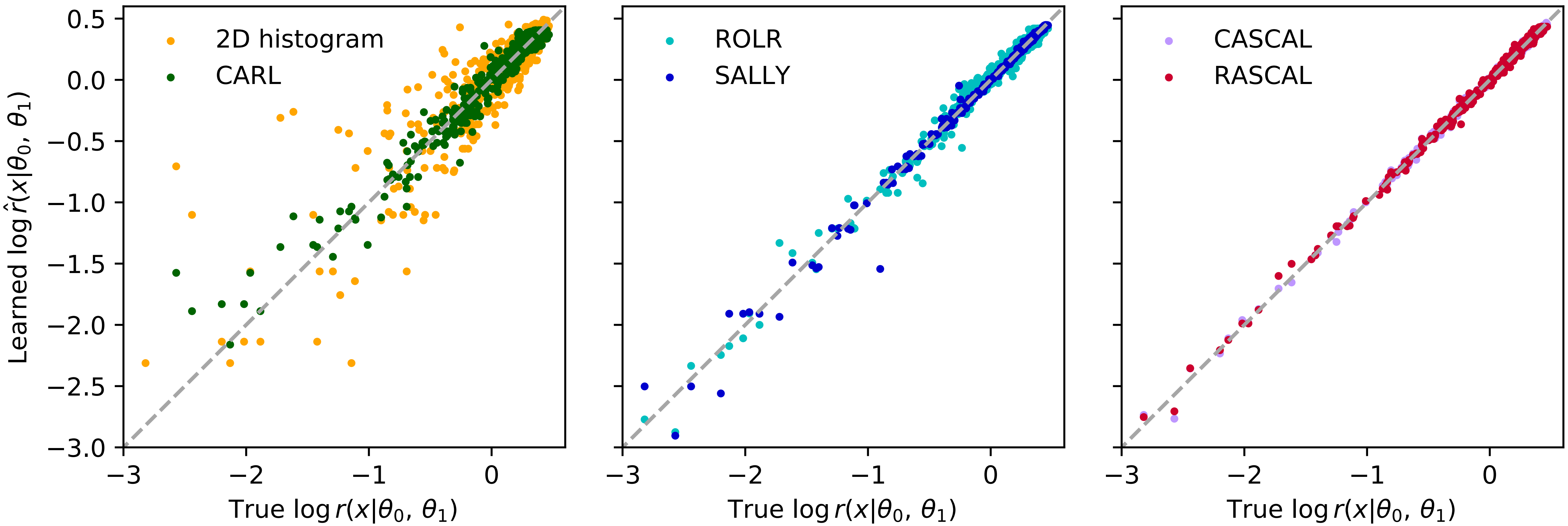
Score vs observables



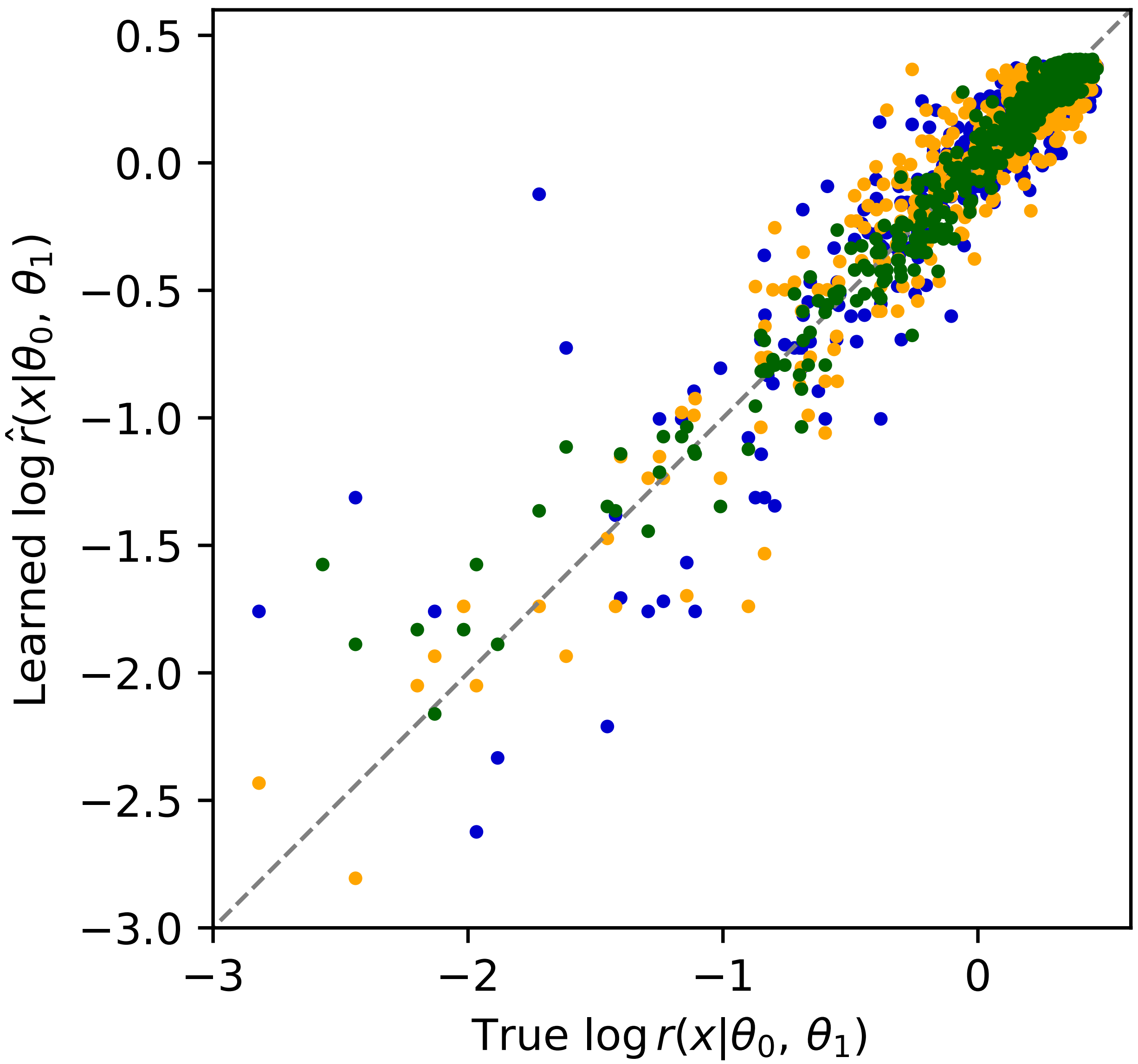
Results

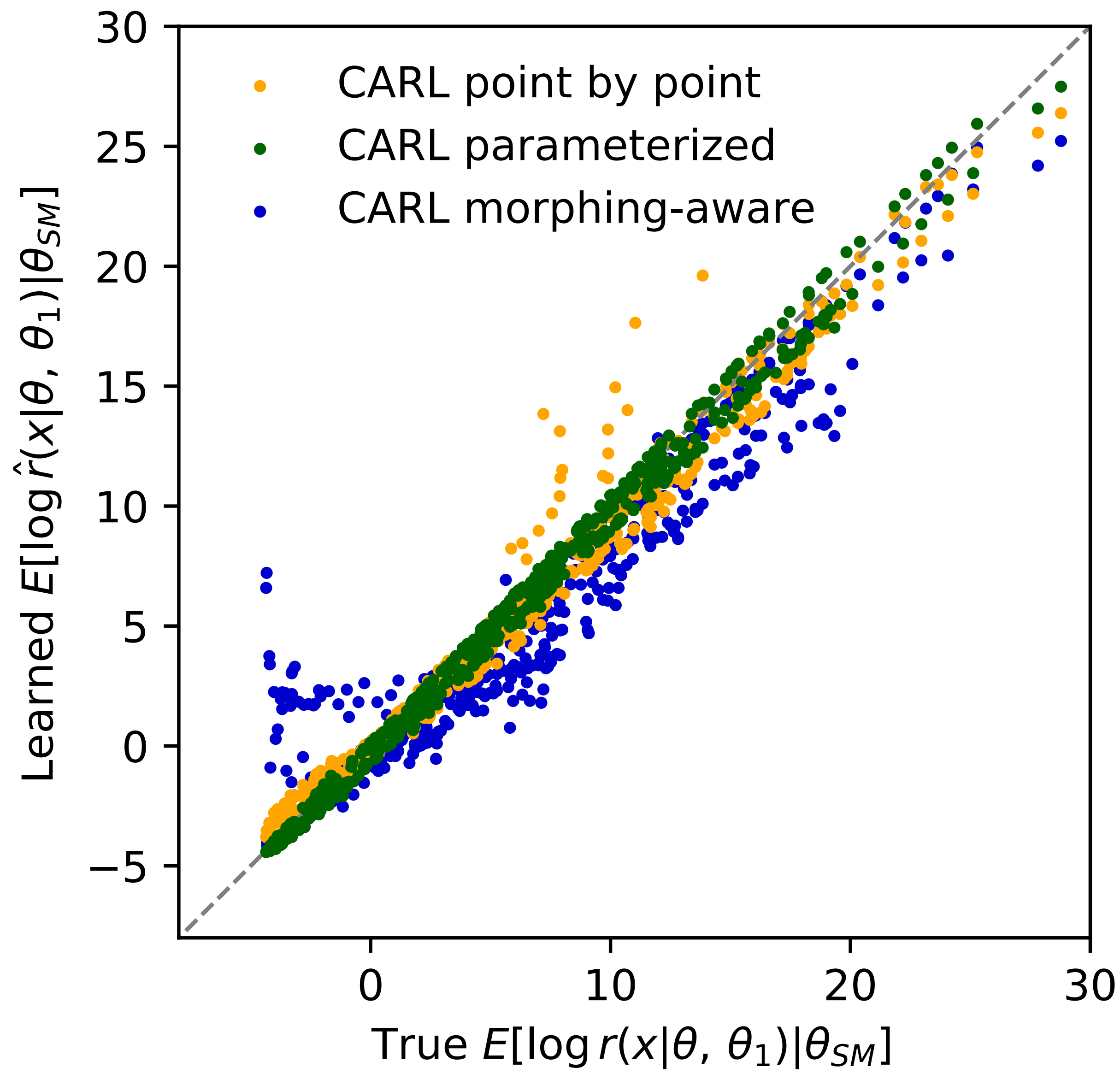
Strategy	Setup	Expected MSE All	Expected MSE Trimmed	Figures
Histogram	$p_{T,j1}, \Delta\phi_{jj}$	**0.056**	0.0106	✓
	$p_{T,j1}$	0.088	0.0230	
	$\Delta\phi_{jj}$	0.160	0.0433	
AFC	$p_{T,j1}, \Delta\phi_{jj}$	0.059	**0.0091**	
	$p_{T,j1}, m_{Z2}, m_{jj}, \Delta\eta_{jj}, \Delta\phi_{jj}$	0.078	0.0101	
CARL (PbP)	PbP	0.030	0.0111	Fig. 12
CARL (parameterized)	Baseline	0.012	**0.0026**	✓
	Random θ	**0.012**	0.0028	
CARL (morphing-aware)	Baseline	0.076	0.0200	Fig. 12
	Random θ	0.086	0.0226	
	Morphing basis	0.156	0.0618	
ROLR (PbP)	PbP	0.005	0.0022	
ROLR (parameterized)	Baseline	0.003	0.0017	✓
	Random θ	**0.003**	**0.0014**	
ROLR (morphing-aware)	Baseline	0.024	0.0063	
	Random θ	0.022	0.0052	
	Morphing basis	0.130	0.0485	
SALLY		**0.013**	**0.0002**	✓
SALLINO		0.021	0.0006	
CASCAL (parameterized)	Baseline	**0.001**	**0.0002**	✓
	Random θ	0.001	0.0002	
CASCAL (morphing-aware)	Baseline	0.136	0.0427	
	Random θ	0.092	0.0268	
	Morphing basis	0.040	0.0081	
RASCAL (parameterized)	Baseline	0.001	0.0004	✓
	Random θ	**0.001**	**0.0004**	
RASCAL (morphing-aware)	Baseline	0.125	0.0514	
	Random θ	0.132	0.0539	
	Morphing basis	0.031	0.0072	

Precision of estimates

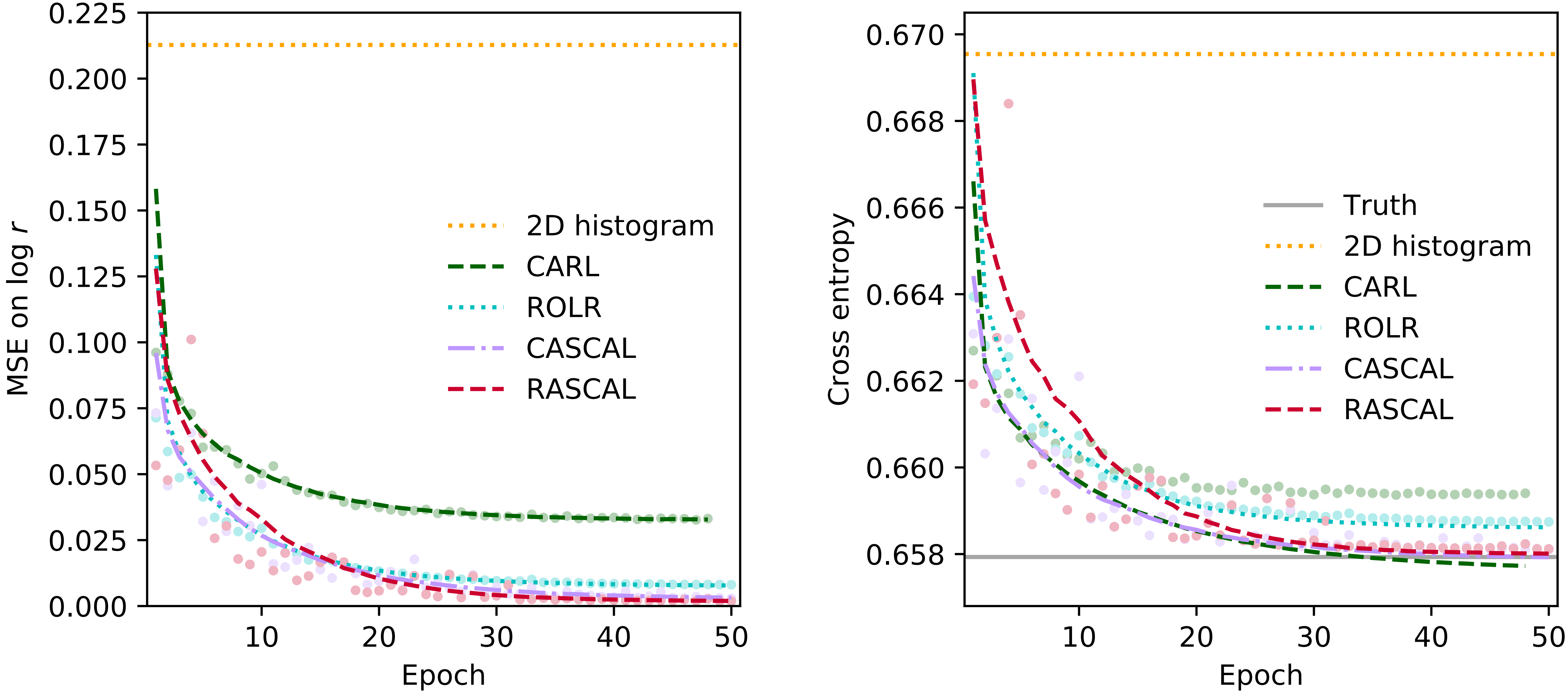


PbP vs parameterized vs morphing-aware

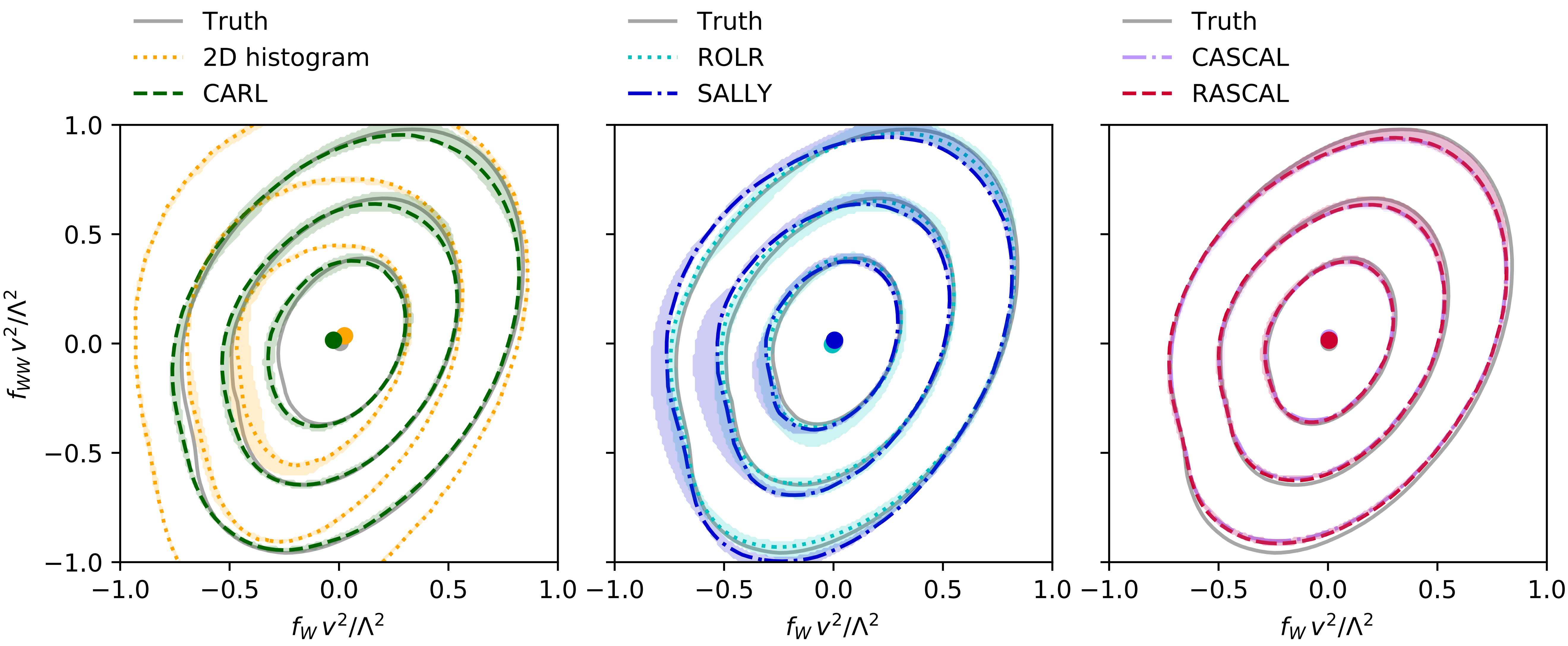




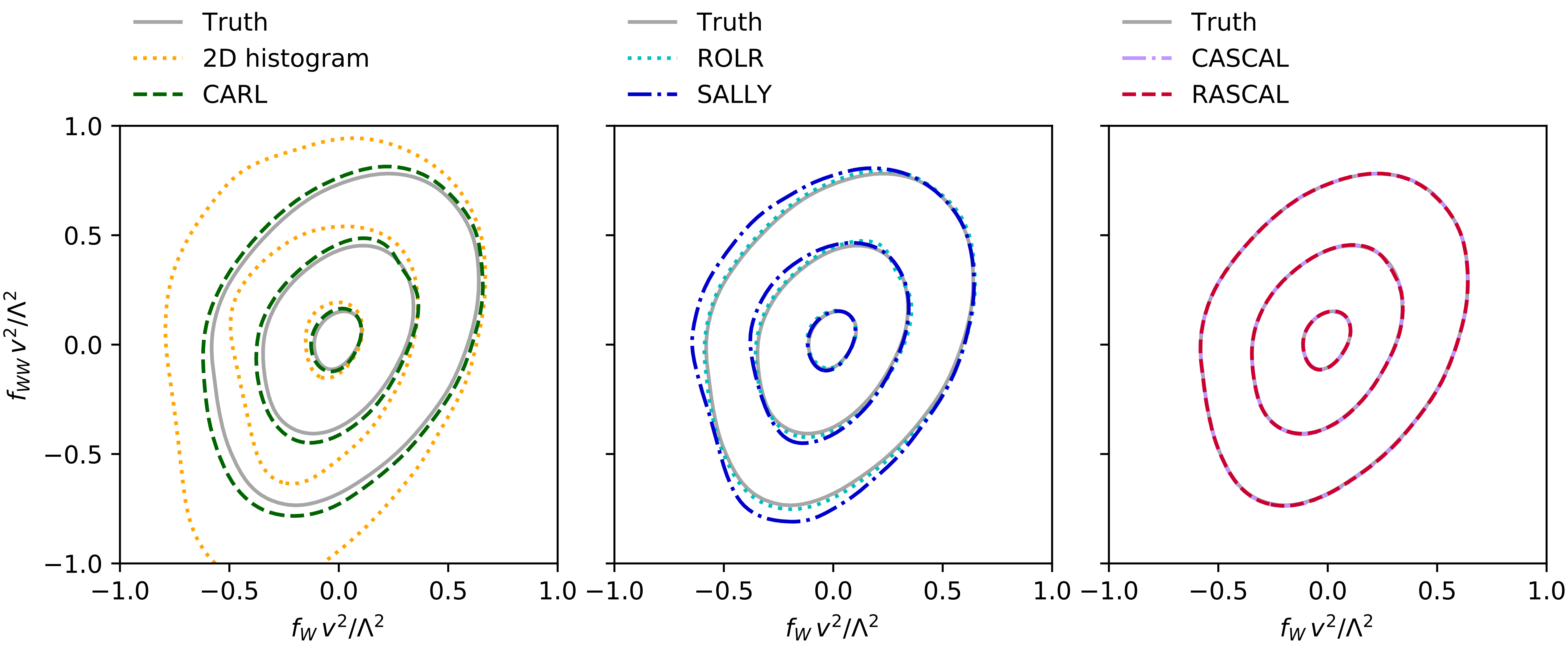
Learning curves

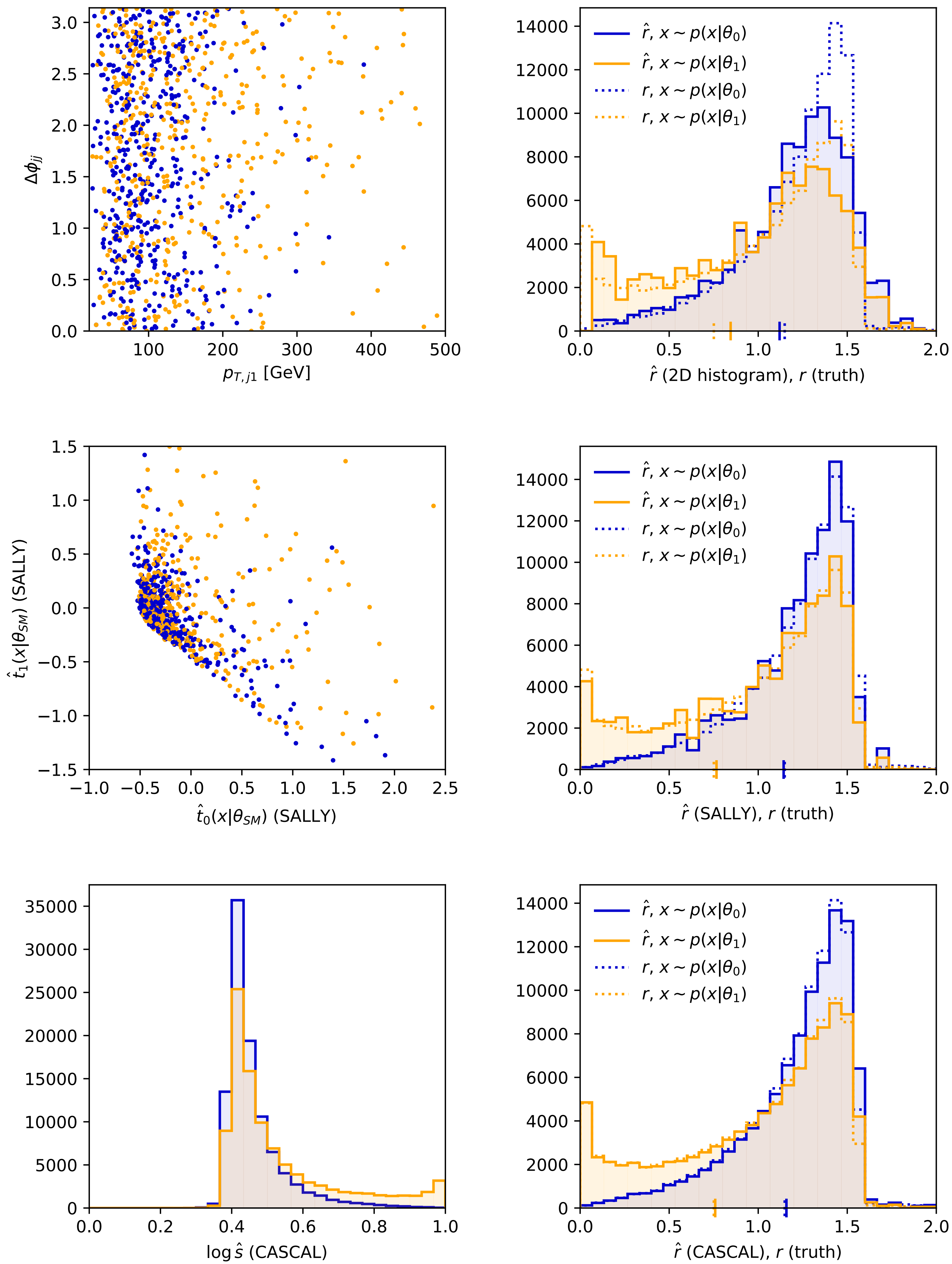


Expected limits (asymptotics)

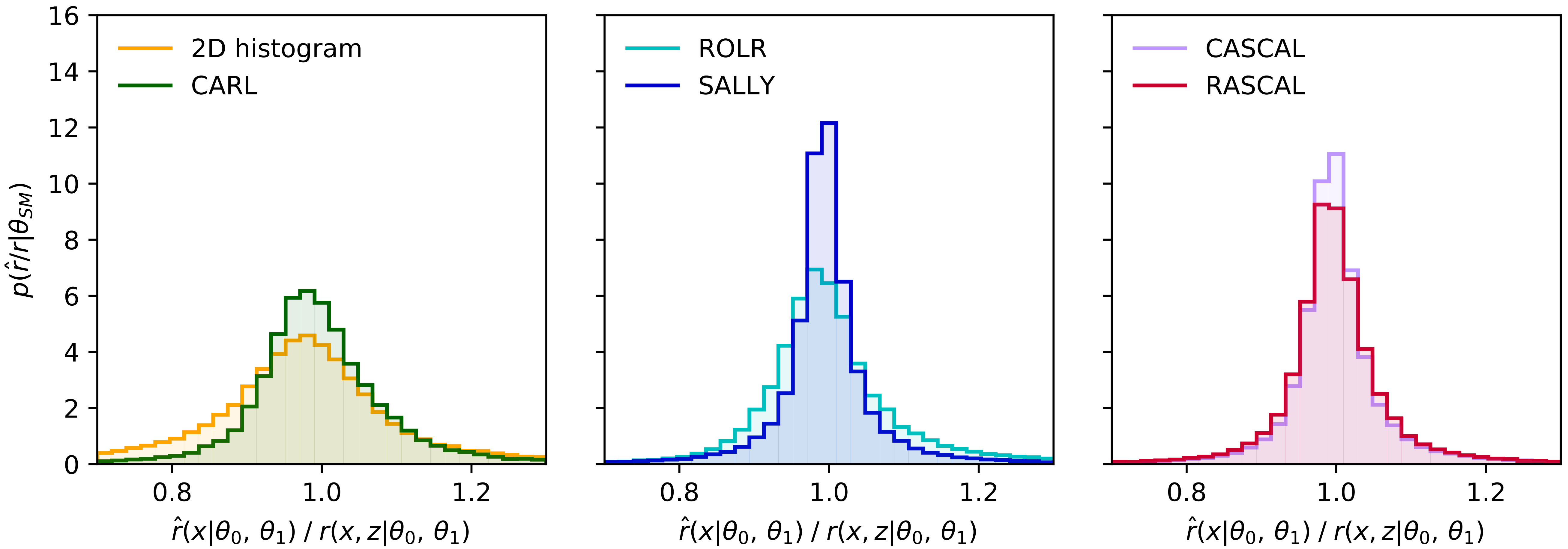


Expected limits (Neyman constr.)

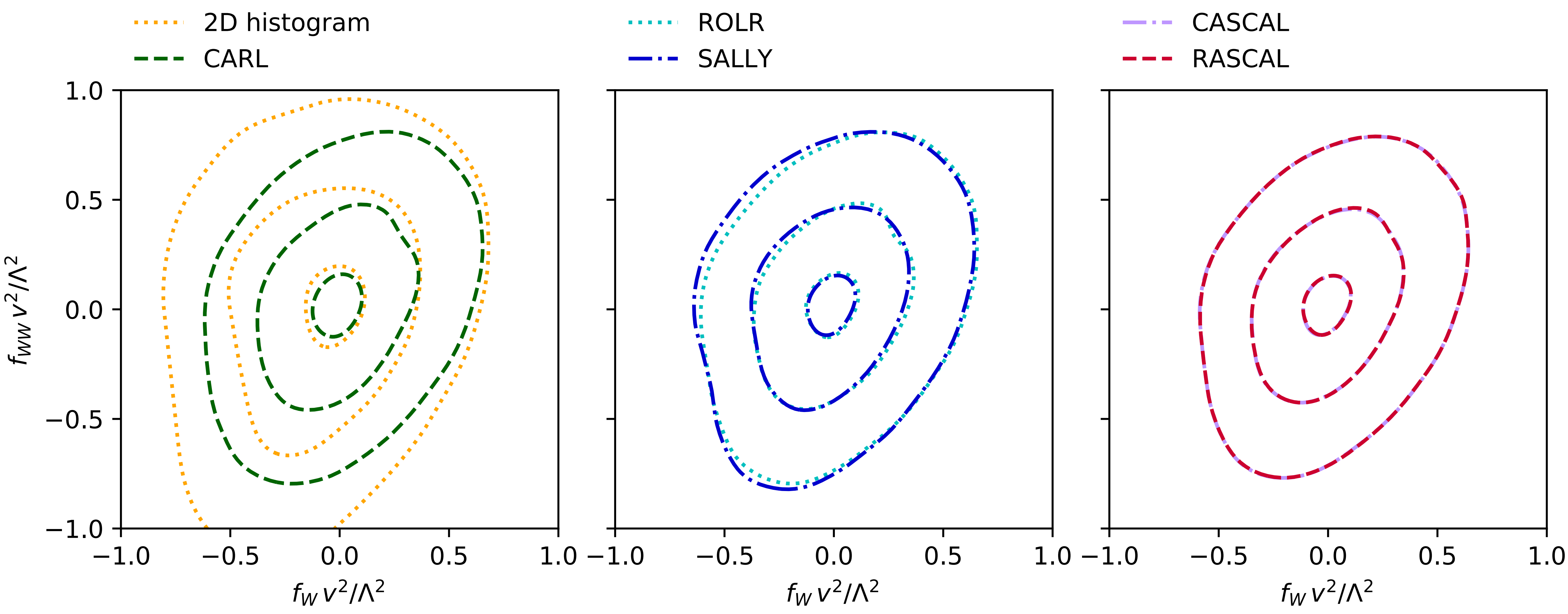



$\Delta\phi_{jj}$
$p_{T,j1}$ [GeV]
$\hat{r}$, $x \sim p(x|\theta_0)$
$\hat{r}$, $x \sim p(x|\theta_1)$
r, $x \sim p(x|\theta_0)$
r, $x \sim p(x|\theta_1)$
$\hat{r}$ (2D histogram), r (truth)
$\hat{t}_1(x|\theta_{SM})$ (SALLY)
$\hat{t}_0(x|\theta_{SM})$ (SALLY)
$\hat{r}$ (SALLY), r (truth)
$\log \hat{s}$ (CASCAL)
$\hat{r}$ (CASCAL), r (truth)

Detector effects



Constraints with detector effects



The End