Machine Learning to probe a BSM Higgs sector

Johann Brehmer

New York University

Higgs Hunting 2018

NYU

MSDSE
Moore - Sloan
Data Science Environments

LHC legacy

- What's hiding in the Higgs sector?

  ⇒ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

LHC legacy

- What's hiding in the Higgs sector?

  ⇒ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

- Challenges:

  1. Many parameters

$$
\begin{aligned}
S = \int \mathrm{d}^4x \Bigg[& \mathcal{L}_{\mathrm{SM}} + \frac{f_{\phi,2}}{\Lambda^2}\,\frac{1}{2}\,\partial_\mu(\phi^\dagger\phi)\,\partial^\mu(\phi^\dagger\phi) + \frac{f_{\phi,3}}{\Lambda^2}\,\frac{1}{3}(\phi^\dagger\phi)^3 \\
& + \frac{f_{GG}}{\Lambda^2}\,(\phi^\dagger\phi)\,G^a_{\mu\nu}\,G^{\mu\nu\,a} - \frac{f_{BB}}{\Lambda^2}\,\frac{g'^2}{4}\,(\phi^\dagger\phi)\,B_{\mu\nu}\,B^{\mu\nu} - \frac{f_{WW}}{\Lambda^2}\,\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^a_{\mu\nu}\,W^{\mu\nu\,a} \\
& + \frac{f_B}{\Lambda^2}\,\frac{\mathrm{i}g'}{2}\,(D^\mu\phi)^\dagger\,D^\nu\phi\,B_{\mu\nu} + \frac{f_W}{\Lambda^2}\,\frac{\mathrm{i}g}{2}\,(D^\mu\phi)^\dagger\,\sigma^a\,D^\nu\phi\,W^a_{\mu\nu} \\
& + \frac{f_\ell}{\Lambda^2}\,(\phi^\dagger\phi)\,\overline{L}_L\phi\,\ell_R + \frac{f_u}{\Lambda^2}\,(\phi^\dagger\phi)\,\overline{Q}_L\tilde{\phi}\,u_R + \frac{f_d}{\Lambda^2}\,(\phi^\dagger\phi)\,\overline{Q}_L\phi\,d_R \\
& + \frac{f_{G\widetilde{G}}}{\Lambda^2}\,(\phi^\dagger\phi)\,G^a_{\mu\nu}\,\widetilde{G}^{\mu\nu\,a} - \frac{f_{B\widetilde{B}}}{\Lambda^2}\,\frac{g'^2}{4}(\phi^\dagger\phi)\,B_{\mu\nu}\,\widetilde{B}^{\mu\nu} - \frac{f_{W\widetilde{W}}}{\Lambda^2}\,\frac{g^2}{4}(\phi^\dagger\phi)\,W^a_{\mu\nu}\,\widetilde{W}^{\mu\nu\,a} \Bigg]
\end{aligned}
$$

LHC legacy

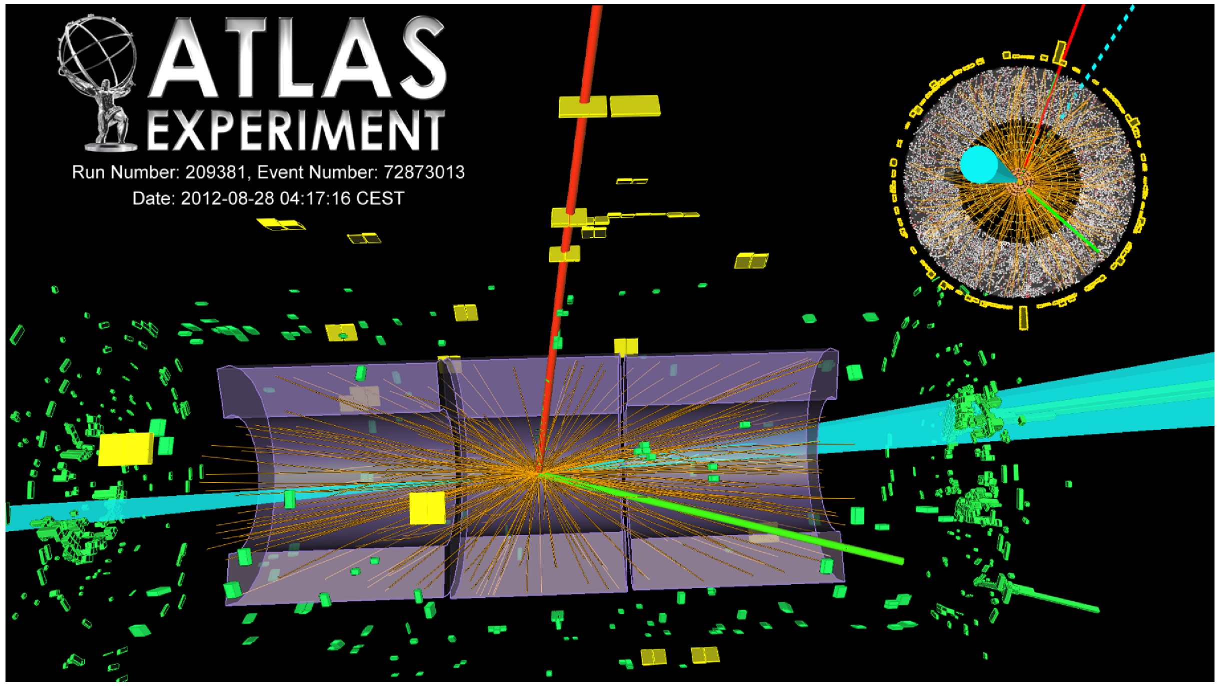
- What's hiding in the Higgs sector?

  ⇒ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

- Challenges:

  1. Many parameters

  2. Many observables



[ATLAS 1501.04943]

LHC legacy

- What's hiding in the Higgs sector?

  $\Rightarrow$ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

- Challenges:

  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects
     - No single good standard variable
     - Probability distributions overlap

LHC legacy

- What's hiding in the Higgs sector?

  ⇒ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

- Challenges:
  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects
     - No single good standard variable
     - Probability distributions overlap

[M. Yao, idea for analogy: K. Cranmer]

LHC legacy

- What's hiding in the Higgs sector?

  ⇒ constrain dimension-6 EFT operators (or Pseudo-Observables, non-linear EFT, …)

- Challenges:

  1. Many parameters
  2. Many observables
  3. Subtle kinematic effects
     - No single good standard variable
     - Probability distributions overlap
  4. Intractable likelihood…

Particle physics processes: simulation

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

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

Particle physics processes: simulation

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

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

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)

Particle physics processes: inference

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

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

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

$$\xrightarrow{\hspace{6em}}$$

Inference

Particle physics processes: inference

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

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

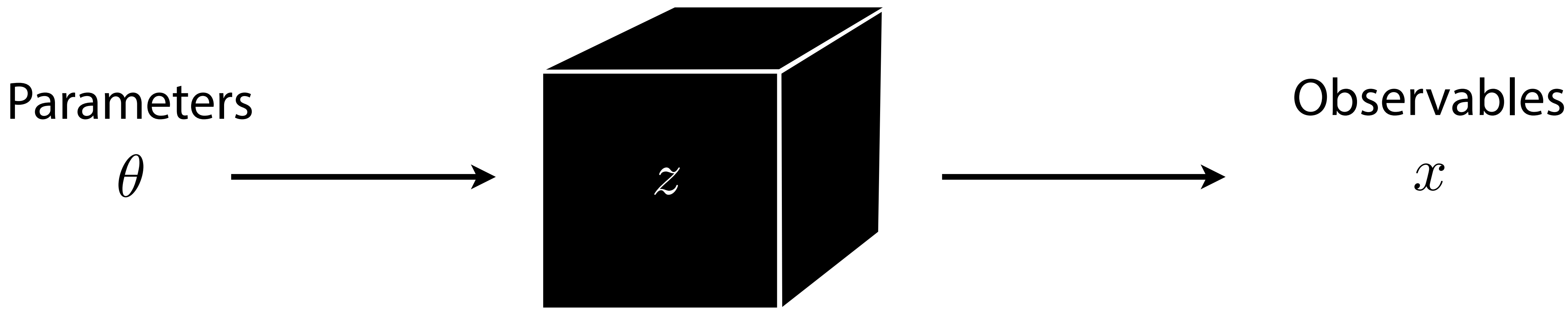
$$p(x|\theta) = \boxed{\int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p} \, p(x|z_d) \quad p(z_d|z_s) \quad p(z_s|z_p) \quad p(z_p|\theta)$$

Infeasible to calculate
the integral over this
enormous space

$\longrightarrow$

Inference

Likelihood-free inference



Prediction (simulation):

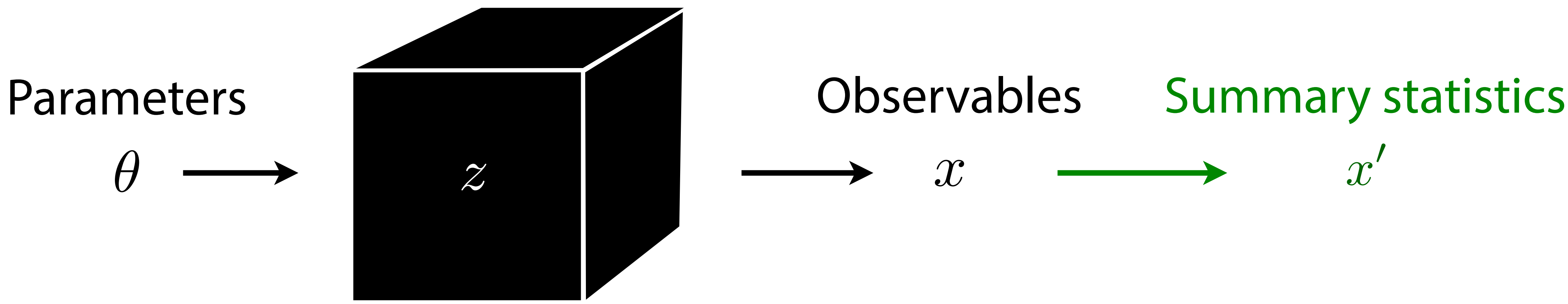
- Well-understood mechanistic model
- Simulator can generate samples

Inference:

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

Why has that not stopped us so far?

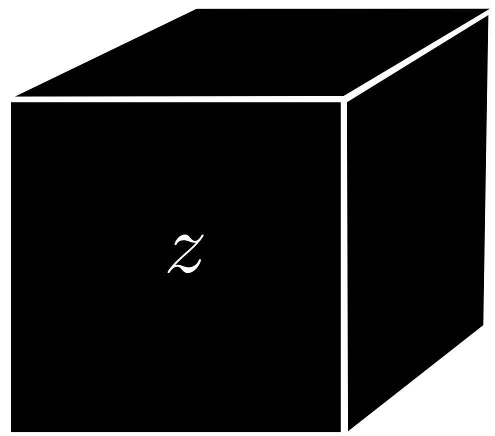
- The physicist's way:



Why has that not stopped us so far?

- The physicist's way:

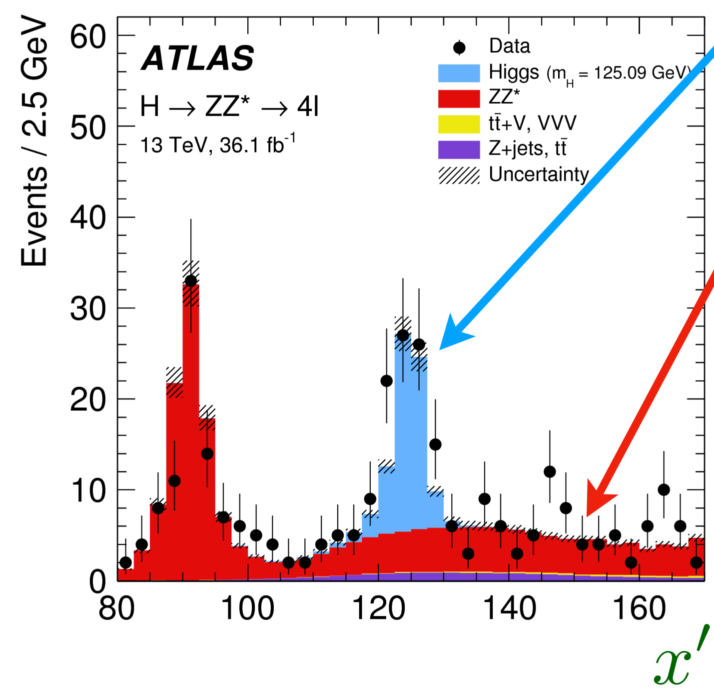
Parameters θ $\longrightarrow$



$\longrightarrow$ Observables x $\longrightarrow$ Summary statistics x'

Run simulator for different θ, fill histograms

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

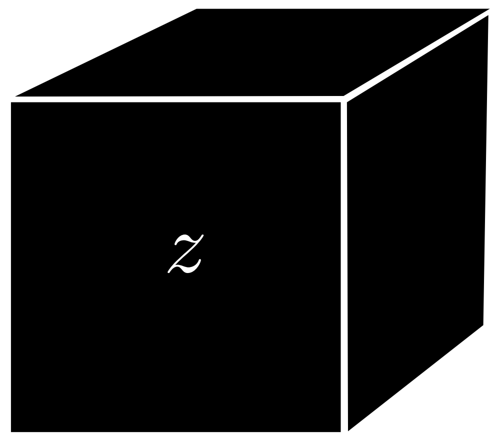


[ATLAS 1712.02304]

Why has that not stopped us so far?

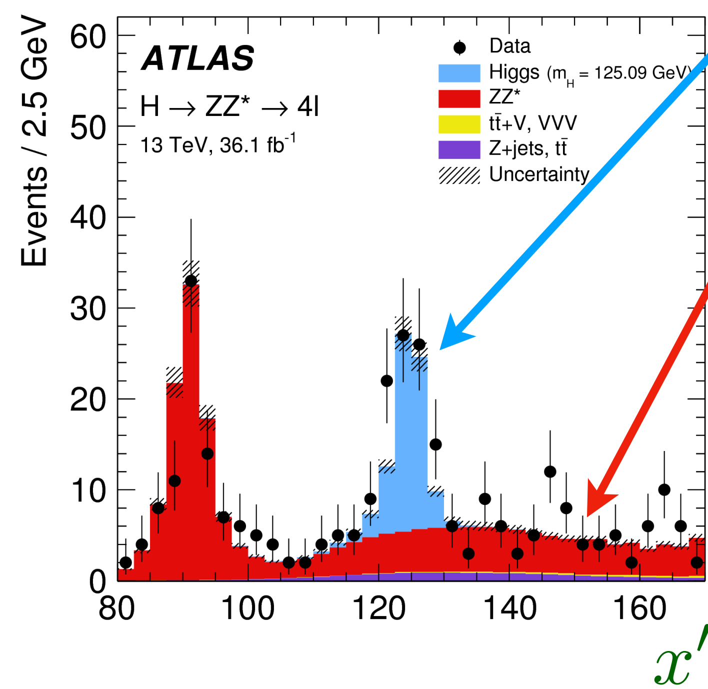
- The physicist's way:

Parameters θ $\longrightarrow$



$\longrightarrow$ Observables x $\longrightarrow$ Summary statistics x'

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



[ATLAS 1712.02304]

Run simulator for different θ, fill histograms

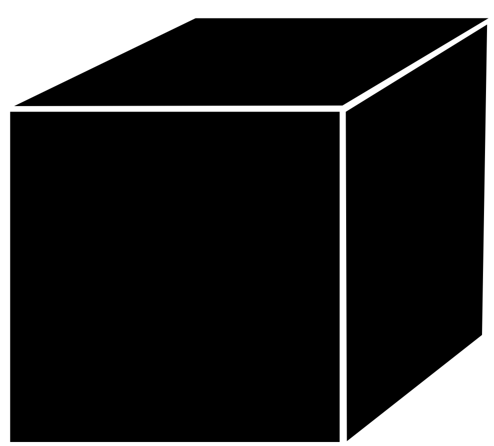
- But how to choose x'? Standard variables lose information, analysis of high-dimensional x (including correlations) often more powerful

Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables,
  Approximate Bayesian Computation
  Rely on summary statistics

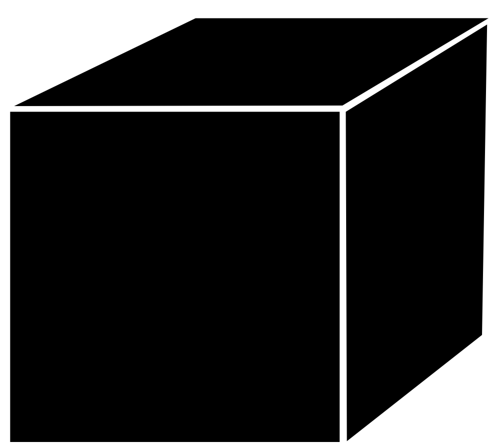
- Machine learning techniques
  Density networks, CARL, autoregressive
  models, normalizing flows, …



Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables, Approximate Bayesian Computation
  Rely on summary statistics
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …



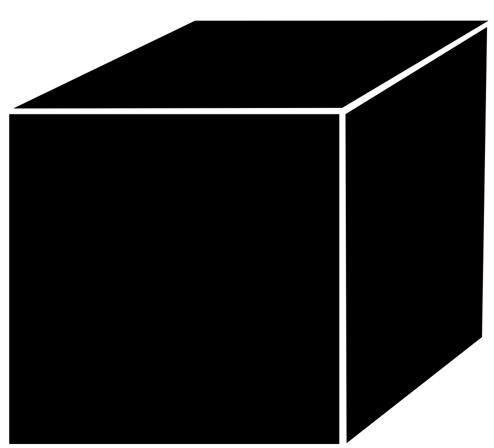
Use latent structure:

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

Likelihood-free inference methods

Treat simulator as black box:

- Histograms of observables, Approximate Bayesian Computation
  Rely on summary statistics
- Machine learning techniques
  Density networks, CARL, autoregressive models, normalizing flows, …



Use latent structure:

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

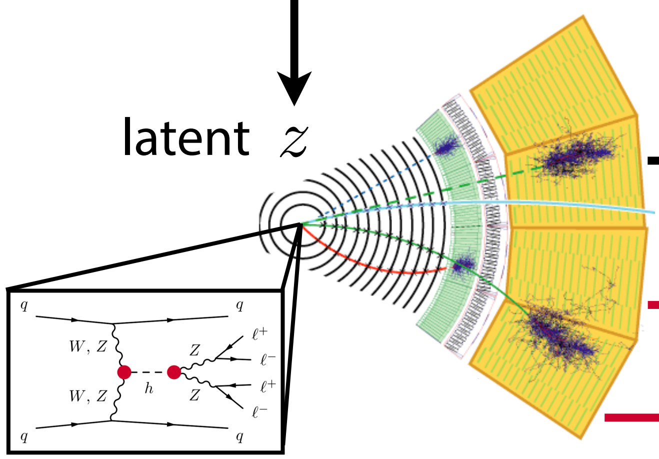
New!

- Mining gold from the simulator
  Leverage matrix-element information + machine learning
  - stronger limits
  - scales to many parameters + observables
  - no approximations on physics necessary
  - evaluation in microseconds

Mining gold to improve likelihood-free inference

parameter θ

latent z



observable x

$r(x, z|\theta)$

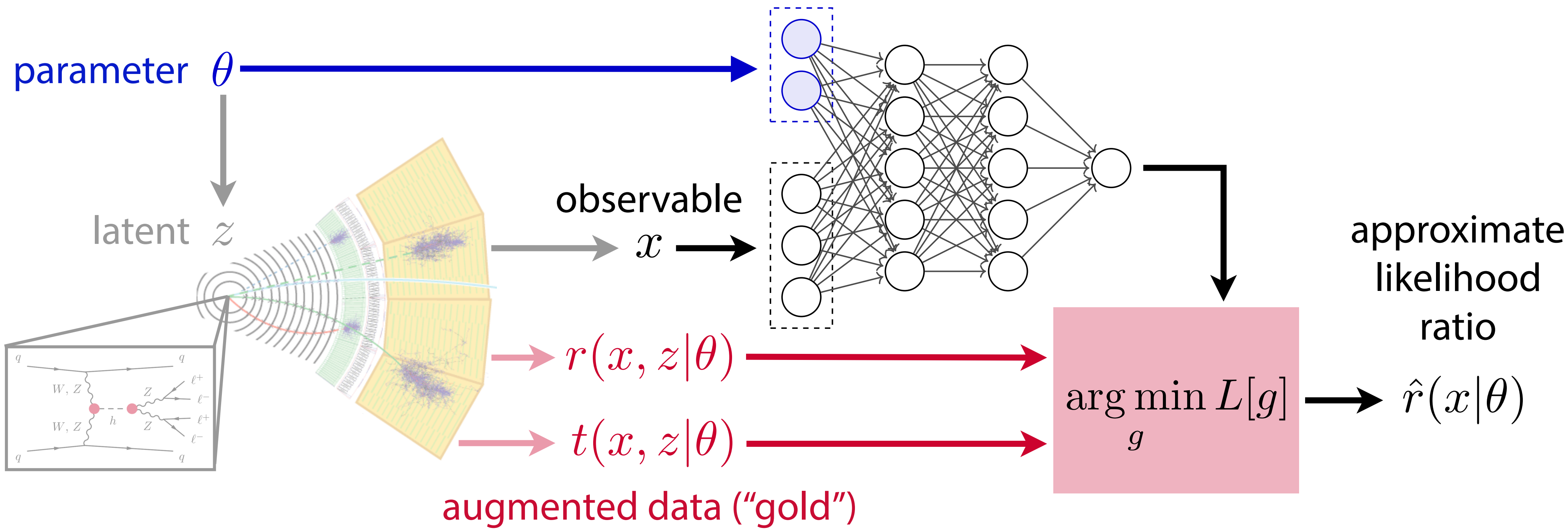
$t(x, z|\theta)$

augmented data ("gold")

1. Simulation

"Mining gold": Extract additional information from simulator

Mining gold to improve likelihood-free inference



1. Simulation

"Mining gold": Extract additional information from simulator

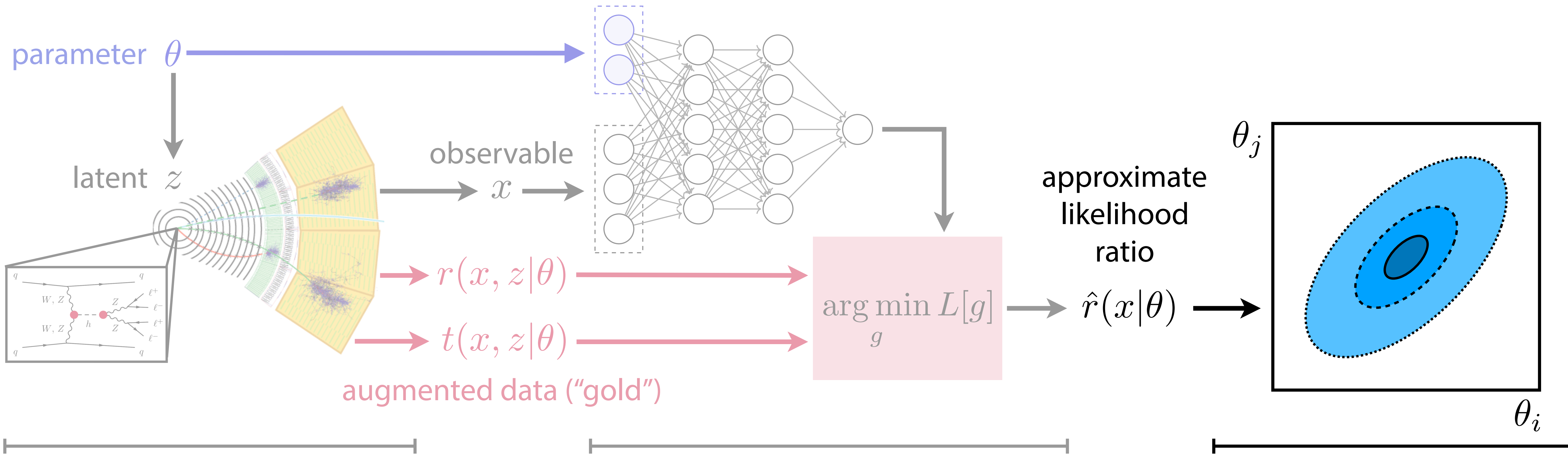
2. Machine Learning

Use this information to train precise estimator for likelihood ratio

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

observables model parameters

Mining gold to improve likelihood-free inference



1. Simulation	2. Machine Learning	3. Inference
"Mining gold": Extract additional information from simulator	Use this information to train precise estimator for likelihood ratio	Limit setting with standard hypothesis tests

Mining gold from the simulator

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

Mining gold from the simulator

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!

Mining gold from the simulator

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:

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

The value of gold

We have joint likelihood ratio

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

$$\downarrow$$

We want likelihood ratio

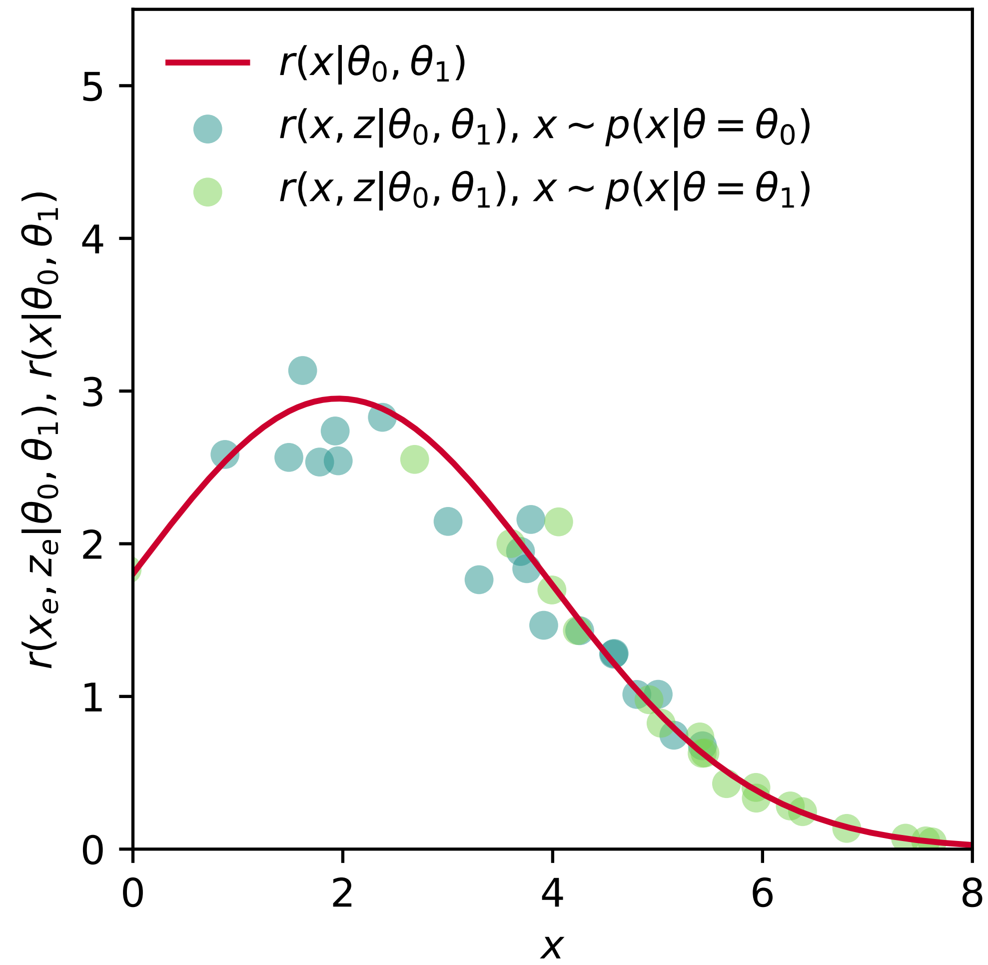
$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

The value of gold

We have joint likelihood ratio

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

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



We want likelihood ratio

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

The value of gold

We have joint likelihood ratio

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

We want likelihood ratio

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

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

$$L_r[\hat{r}(x)] = \mathbb{E}_{p(x,z|\theta_1)}\left[\Big(\hat{r}(x) - r(x, z|\theta_0, \theta_1)\Big)^2\right].$$

One can show it is minimized by

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

Machine learning

Need to minimize a functional:

Extremization

Functional

Variational family

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

Machine learning

Need to minimize a functional:

Variational family

Extremization

Functional

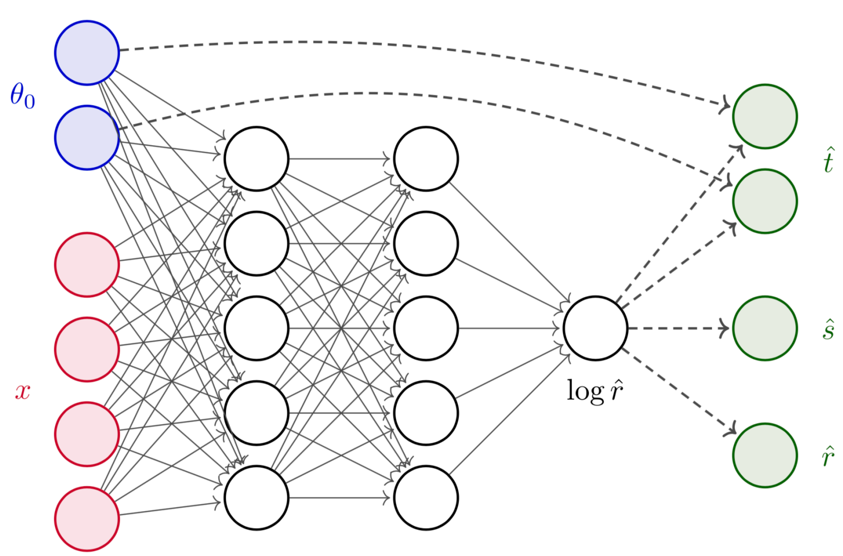
$$r(x|\theta_0, \theta_1) = \arg\min_{\hat{r}(x)} L_r[\hat{r}(x)]$$

This is exactly what machine learning does!

Loss function

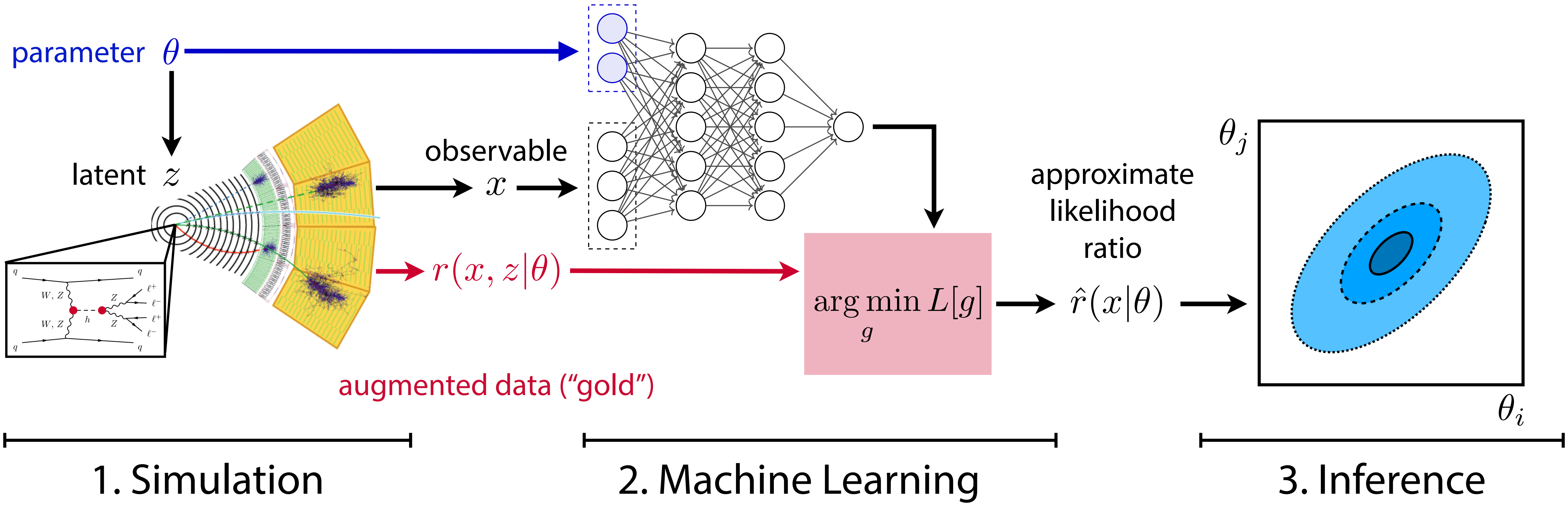
Stochastic gradient descent

Neural network

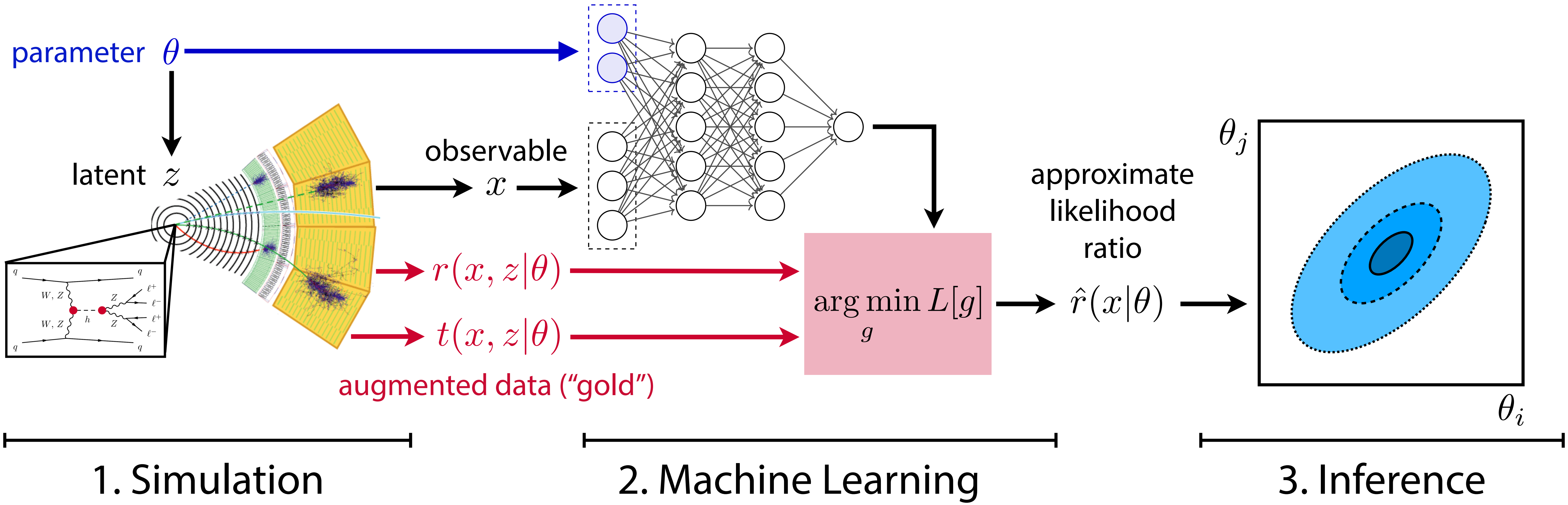


$\Rightarrow$ We implement $\hat{r}(x|\theta_0, \theta_1)$ as a neural network trained on the data available from the simulator

Mining gold to improve likelihood-free inference

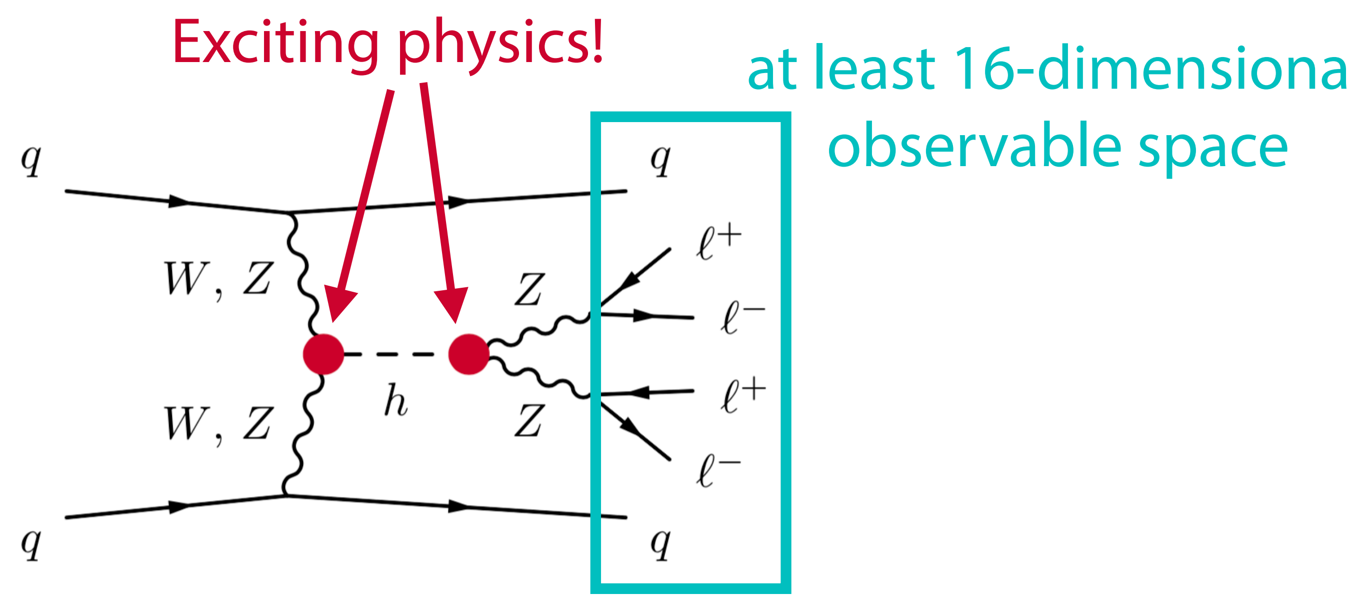


Mining gold to improve likelihood-free inference



Proof of concept

- Higgs production in weak boson fusion:

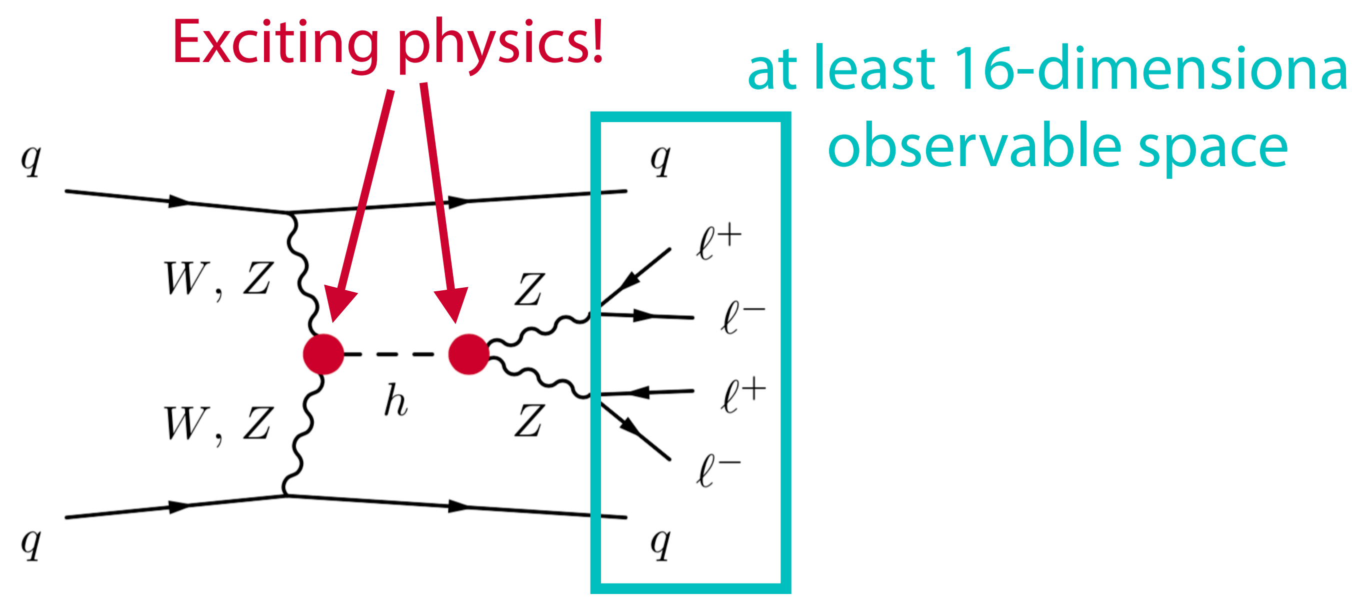


- Goal: constraints on two EFT parameters

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

Proof of concept

- Higgs production in weak boson fusion:



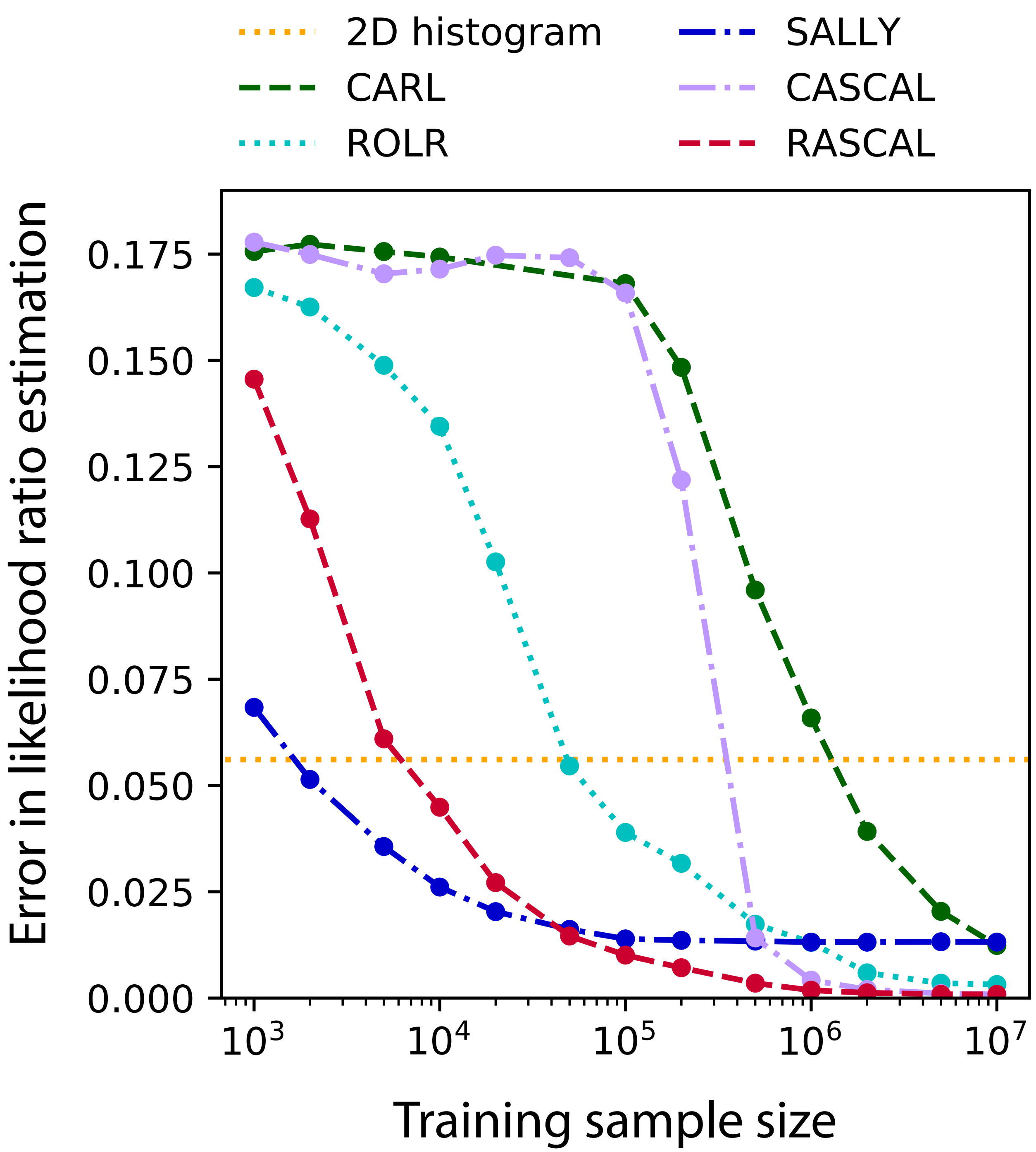
- Goal: constraints on two EFT parameters

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

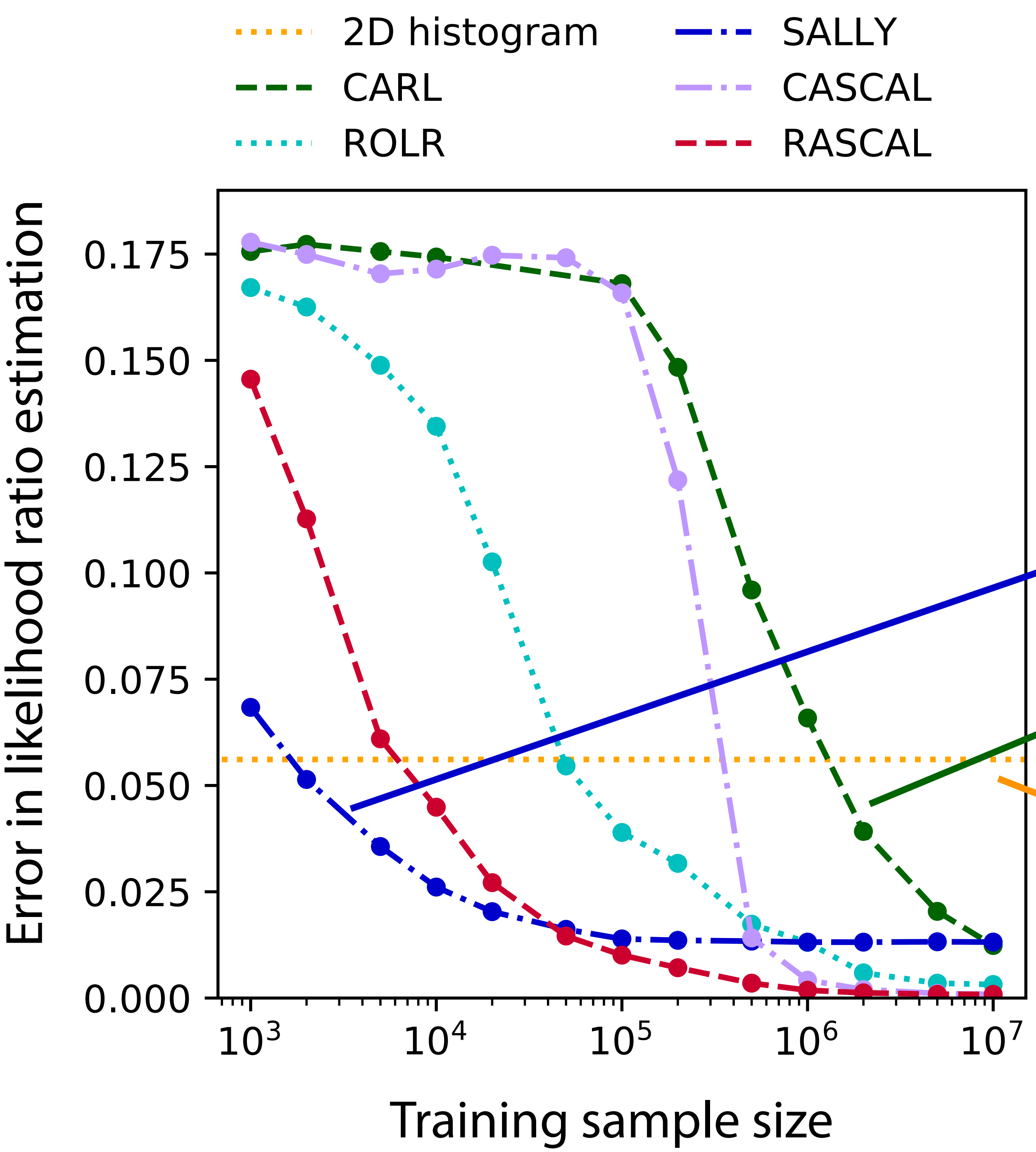
- Two setups:
  - Simplified setup in which we can compare to true likelihood
  - “Realistic” simulation with approximate detector effects

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

Less training data needed



Less training data needed



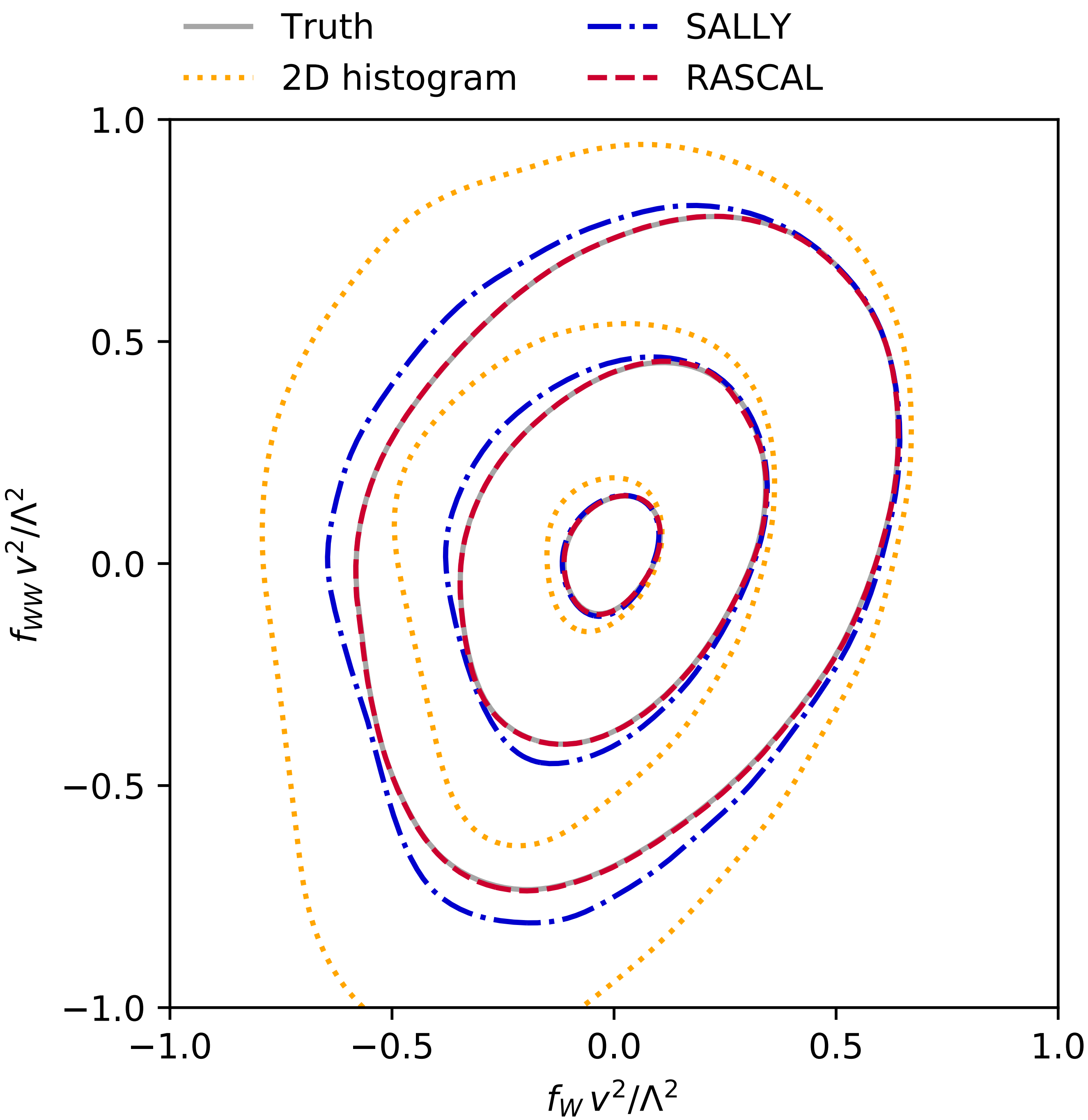
New techniques require less data than generic ML method

[CARL, see K. Cranmer, G. Louppe, J. Pavez 1506.02169]

Traditional histogram

Stronger bounds

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

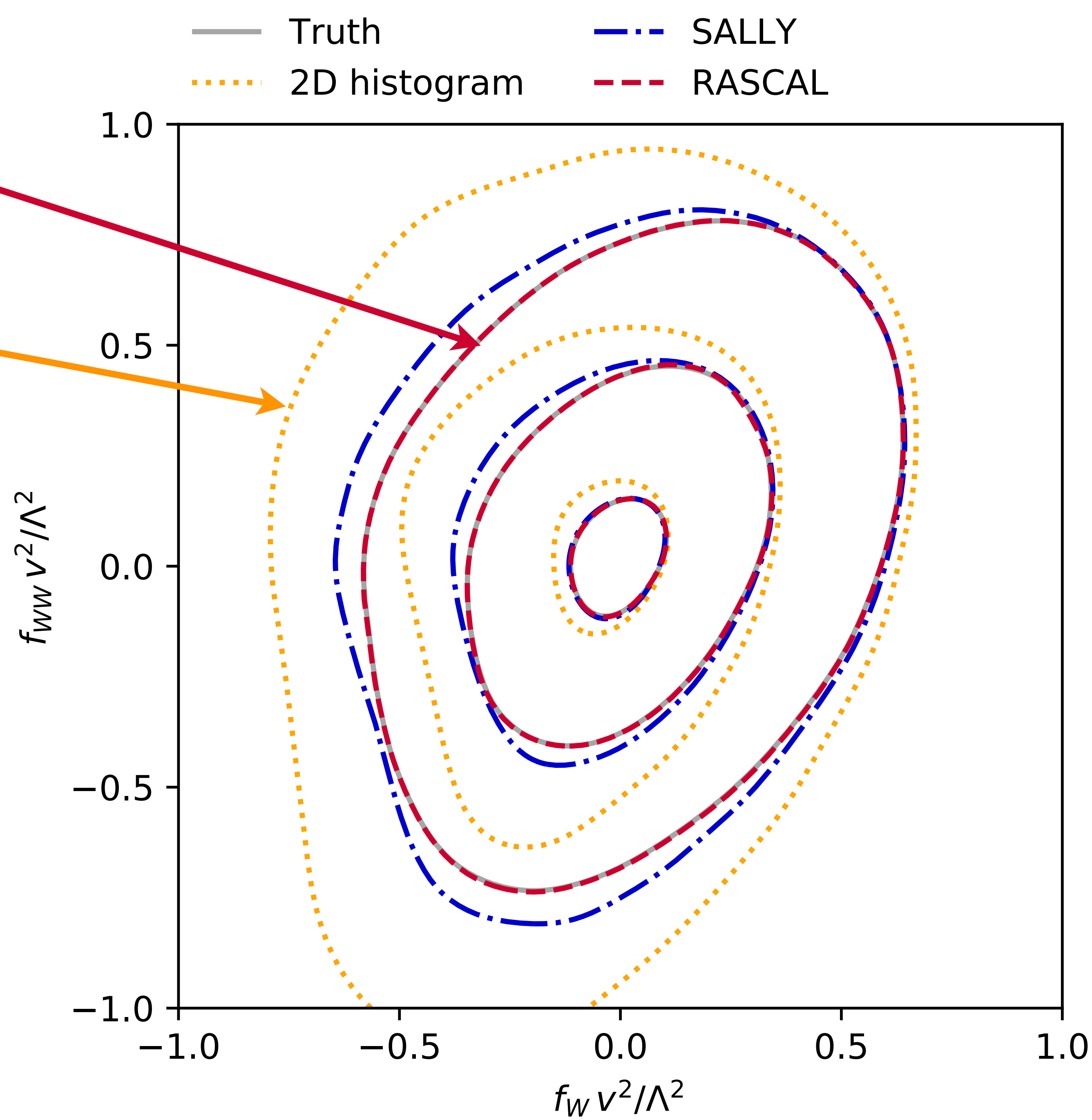


36 events, assuming SM

Stronger bounds

New technique
enables stronger
limits than
traditional histogram

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



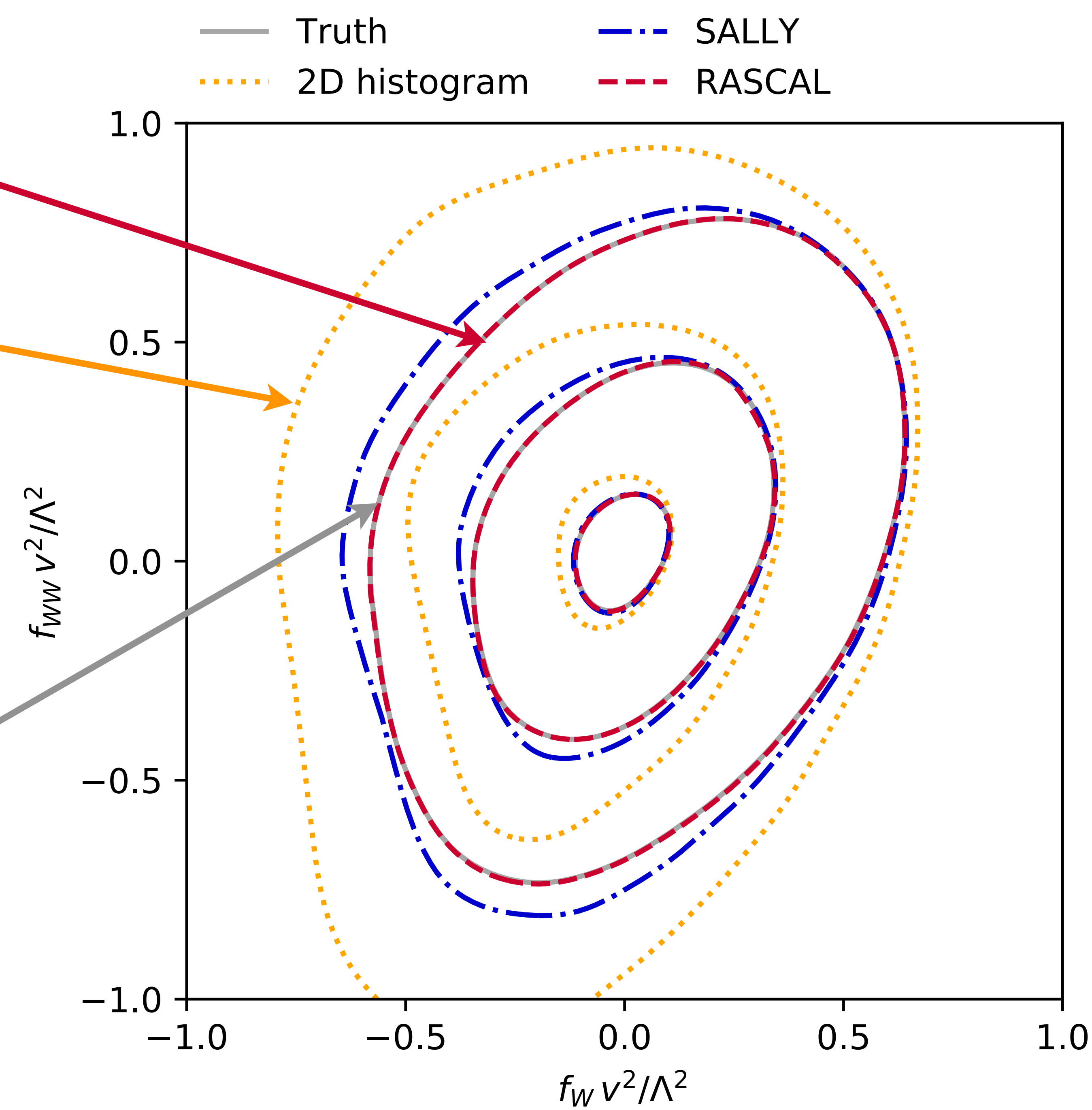
36 events, assuming SM

Stronger bounds

New technique enables stronger limits than traditional histogram

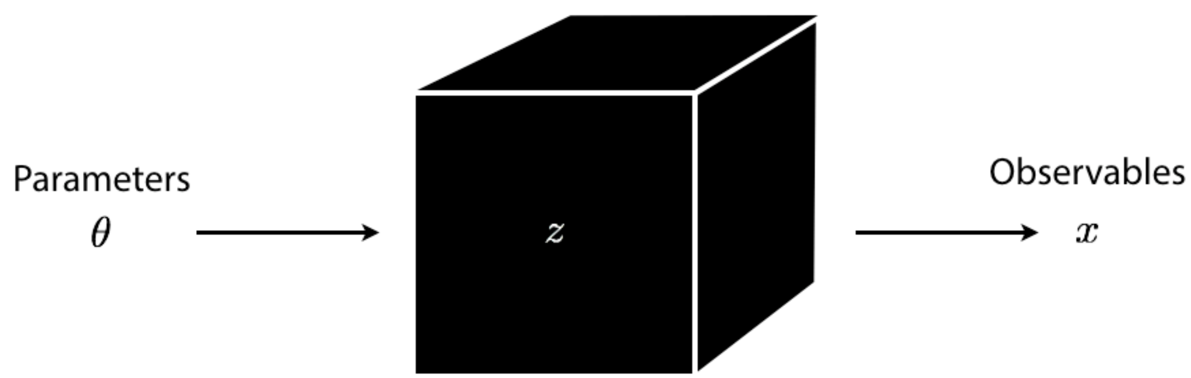
Limits from new technique virtually indistinguishable from true likelihood
(usually we don't have that)

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



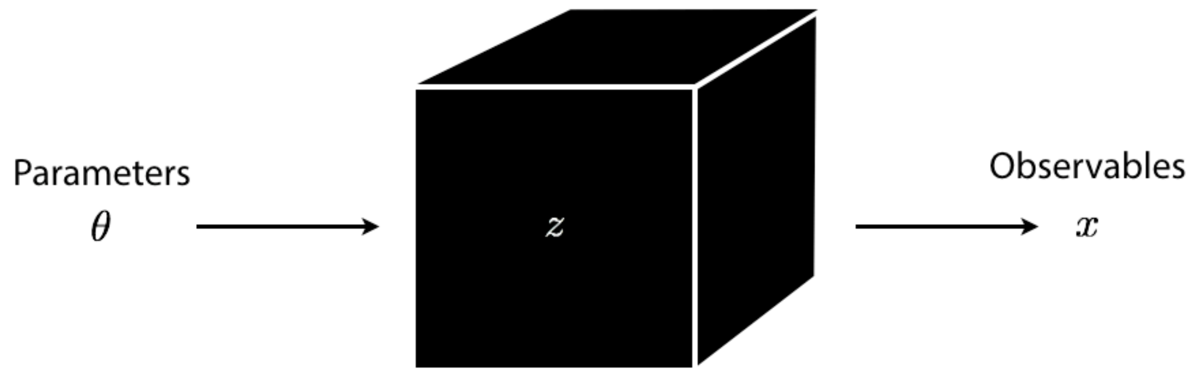
36 events, assuming SM

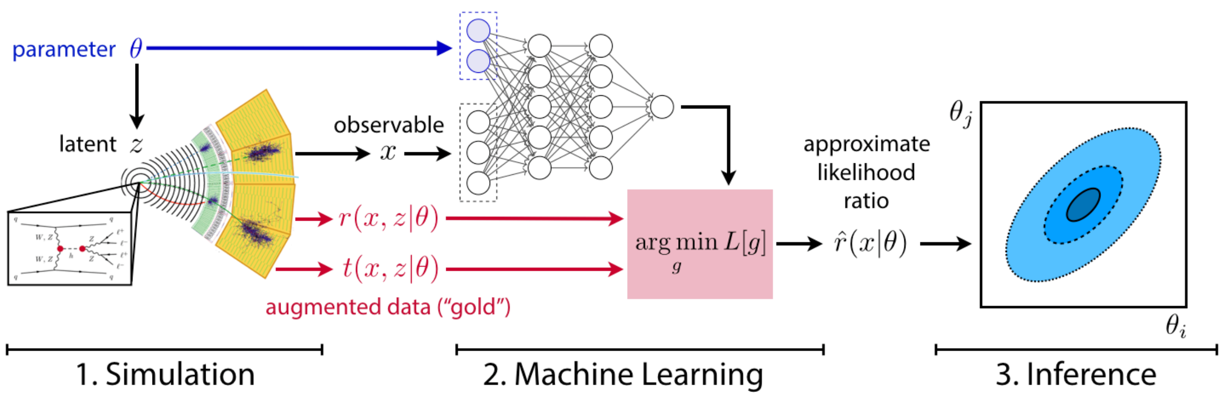
A new approach to simulator-based inference



- Many LHC analyses (and much of modern science) are based on simulations, "likelihood-free"
- Established inference methods treat simulator as black box

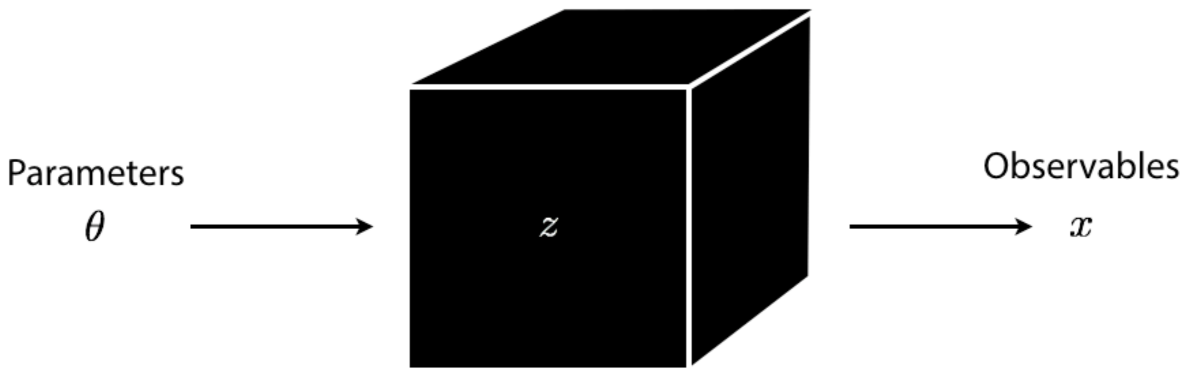
A new approach to simulator-based inference

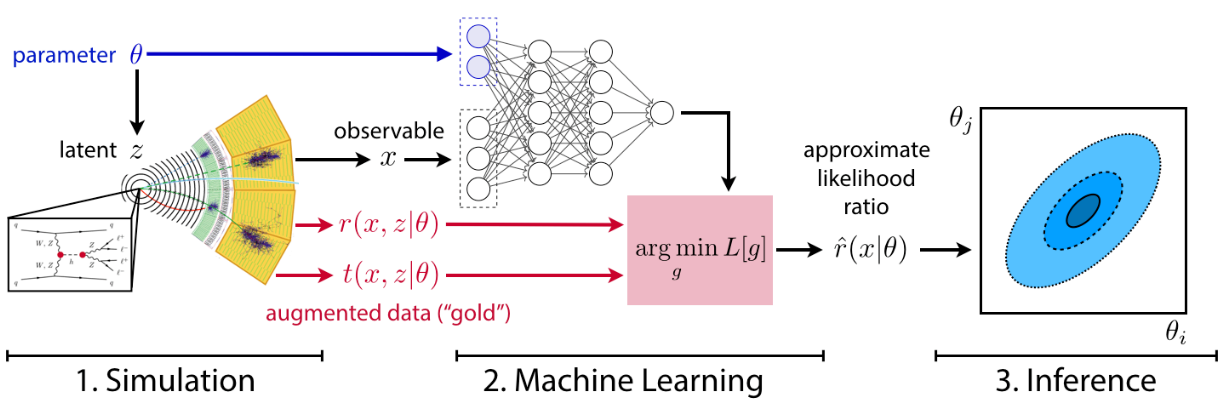


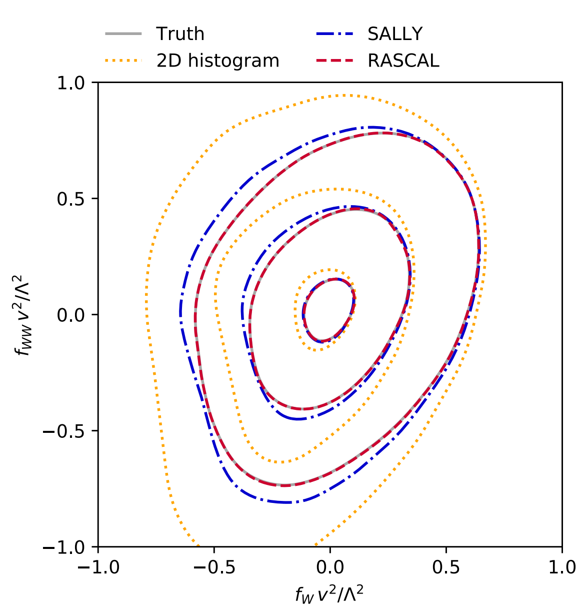


- Many LHC analyses (and much of modern science) are based on simulations, "likelihood-free"
- Established inference methods treat simulator as black box
- New inference techniques: Leverage more information from simulator + power of machine learning

A new approach to simulator-based inference







- Many LHC analyses (and much of modern science) are based on simulations, "likelihood-free"
- Established inference methods treat simulator as black box
- New inference techniques: Leverage more information from simulator + power of machine learning
- First application to Higgs physics: Stronger EFT constraints with less data

References

Kyle Cranmer
ATLAS

Gilles Louppe
ML

Juan Pavez
ML

Felix Kling
Pheno

Markus Stoye
CMS

JB, K. Cranmer, G. Louppe, and J. Pavez [1805.00013]
Constraining Effective Field Theories with Machine Learning

JB, K. Cranmer, G. Louppe, and J. Pavez [1805.00020]
A Guide to Constraining Effective Field Theories with Machine Learning

JB, G. Louppe, J. Pavez, and K. Cranmer [1805.12244]
Mining gold from implicit models to improve likelihood-free inference

M. Stoye, JB, G. Louppe, J. Pavez, and K. Cranmer In preparation
Likelihood-free inference with augmented cross-entropy estimators

JB, K. Cranmer, and F. Kling In preparation
MadMiner

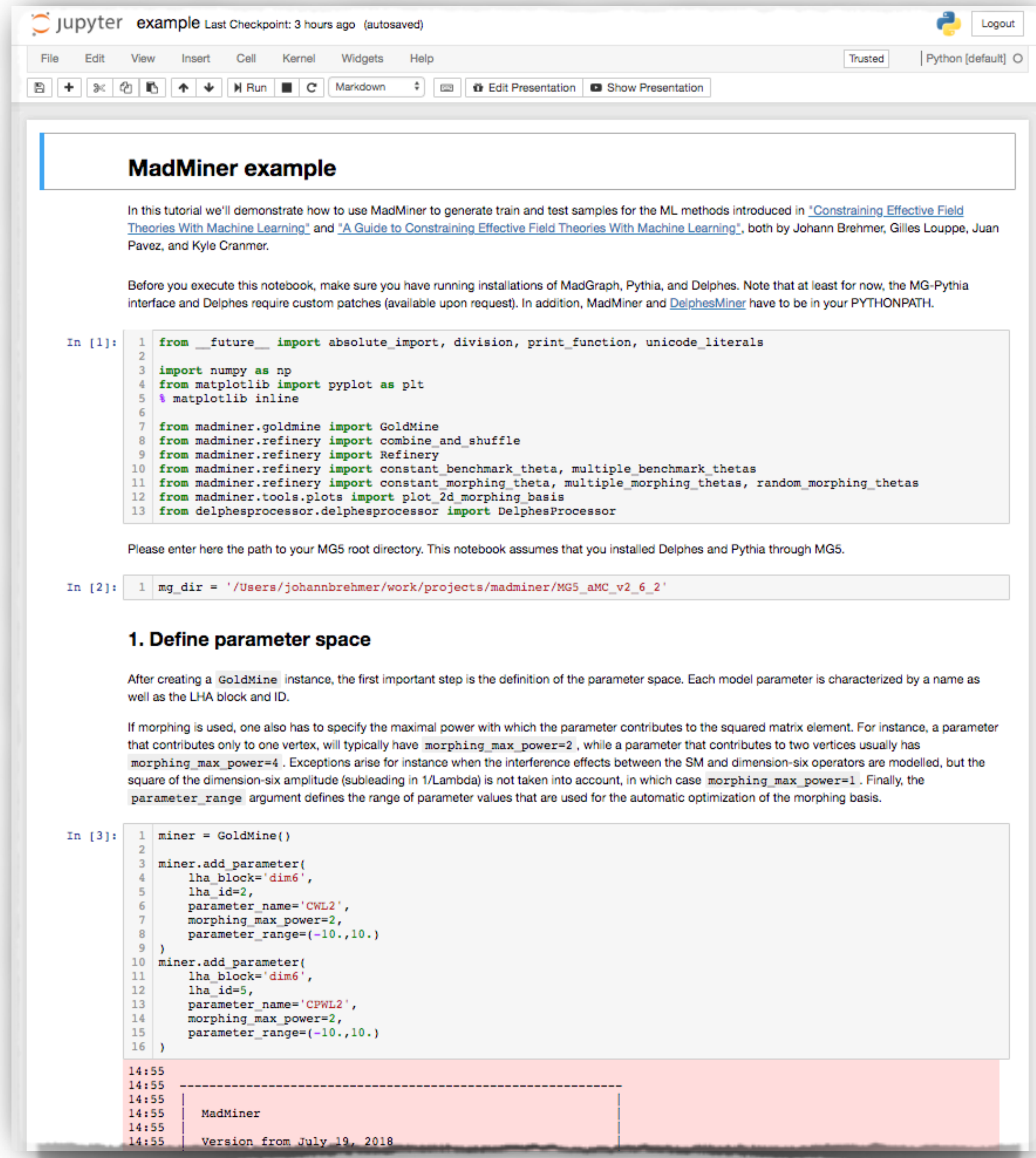
Bonus material

Can I use this?

Yes!

To make that easier, we're working on MadMiner:

- "Mining gold" from MadGraph + Pythia + detector simulation (Delphes … full sim)
- Likelihood ratio estimation with new algorithms
- User-friendly Python interface

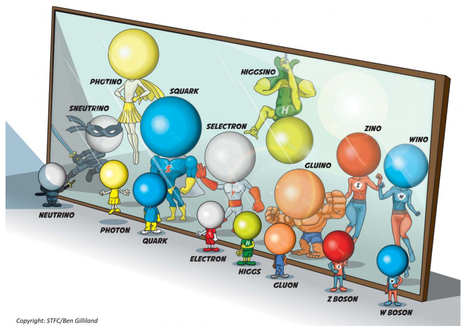


[Image sources: STFC / Ben Gililand, Sean Carroll, Friedrich Justin Bertuch 1806, symmetry]

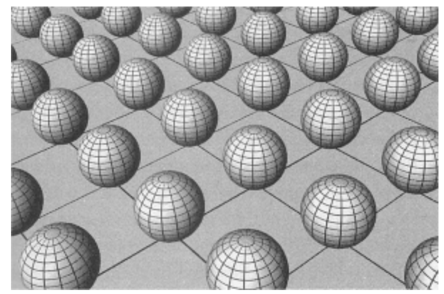
Effective field theory

Energy

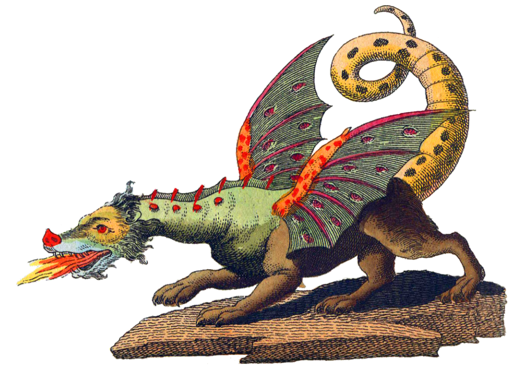
Λ



Supersymmetry?



Extra dimensions?



Dragons?

Composite Higgs?

…

Matching

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

Effective field theory

Measurements

$E \ll \Lambda$

LHC Higgs

Other LHC measurements

LEP

Flavour

…

SMEFT (Standard Model Effective Field Theory)

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

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

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

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; …]

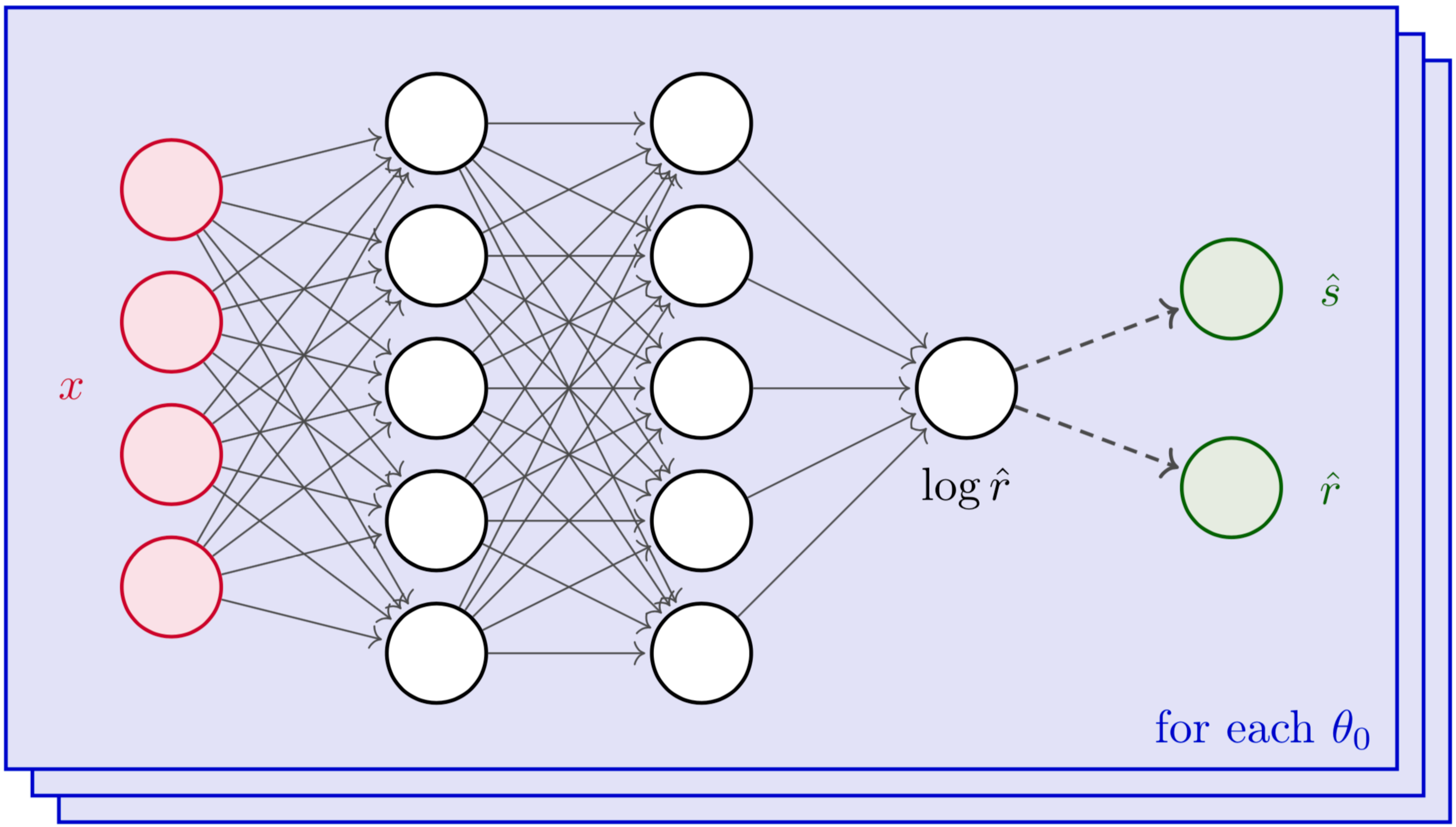
Variational calculus

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

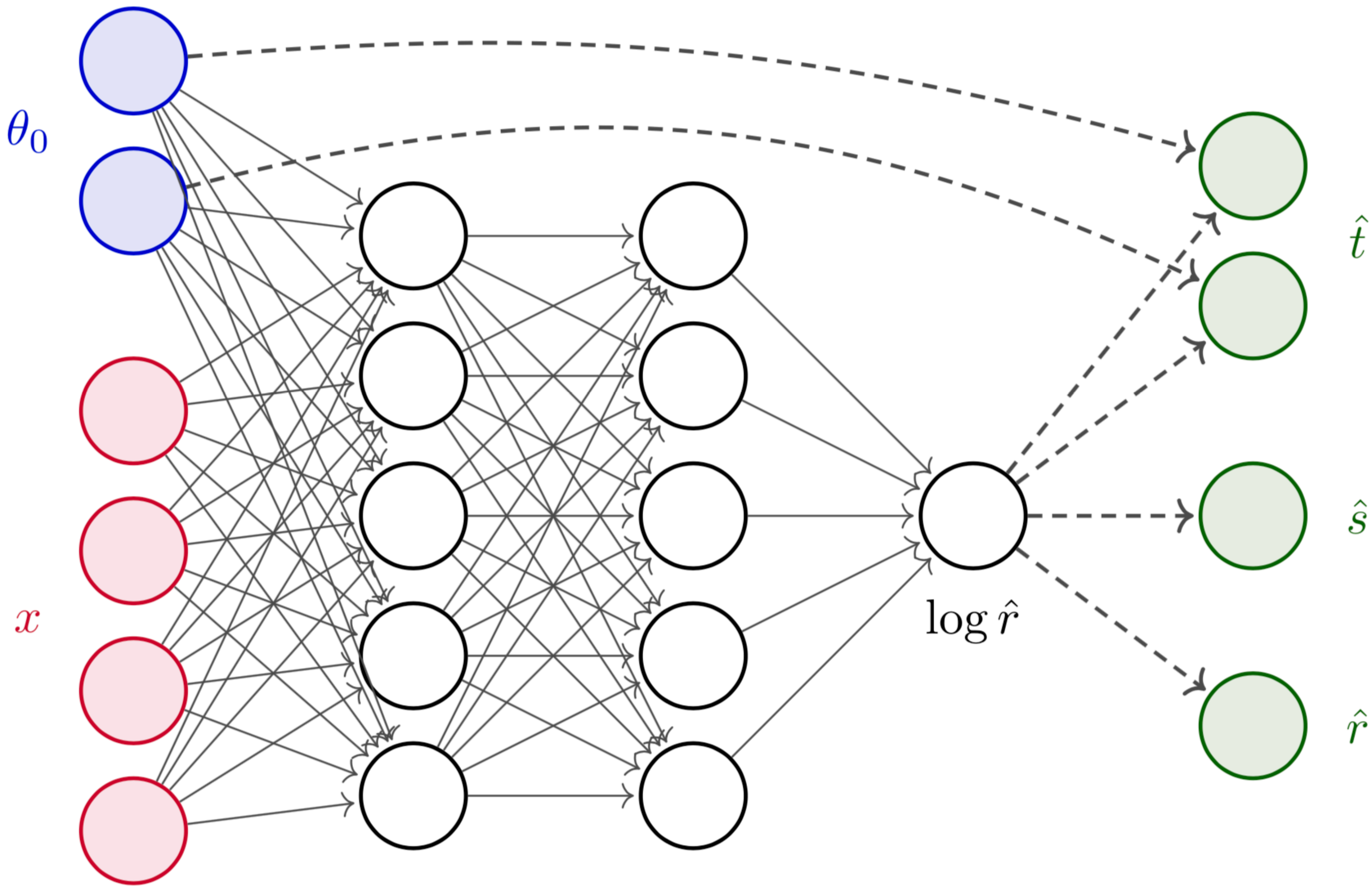
$$
0 = \left.\frac{\delta F}{\delta \hat{g}}\right|_{g^*} = 2\hat{g}\; \underbrace{\int \mathrm{d}z\; p(x, z|\theta)}_{=p(x|\theta)} - 2 \int \mathrm{d}z\; p(x, z|\theta)\, g(x, z)
$$

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

Point by point

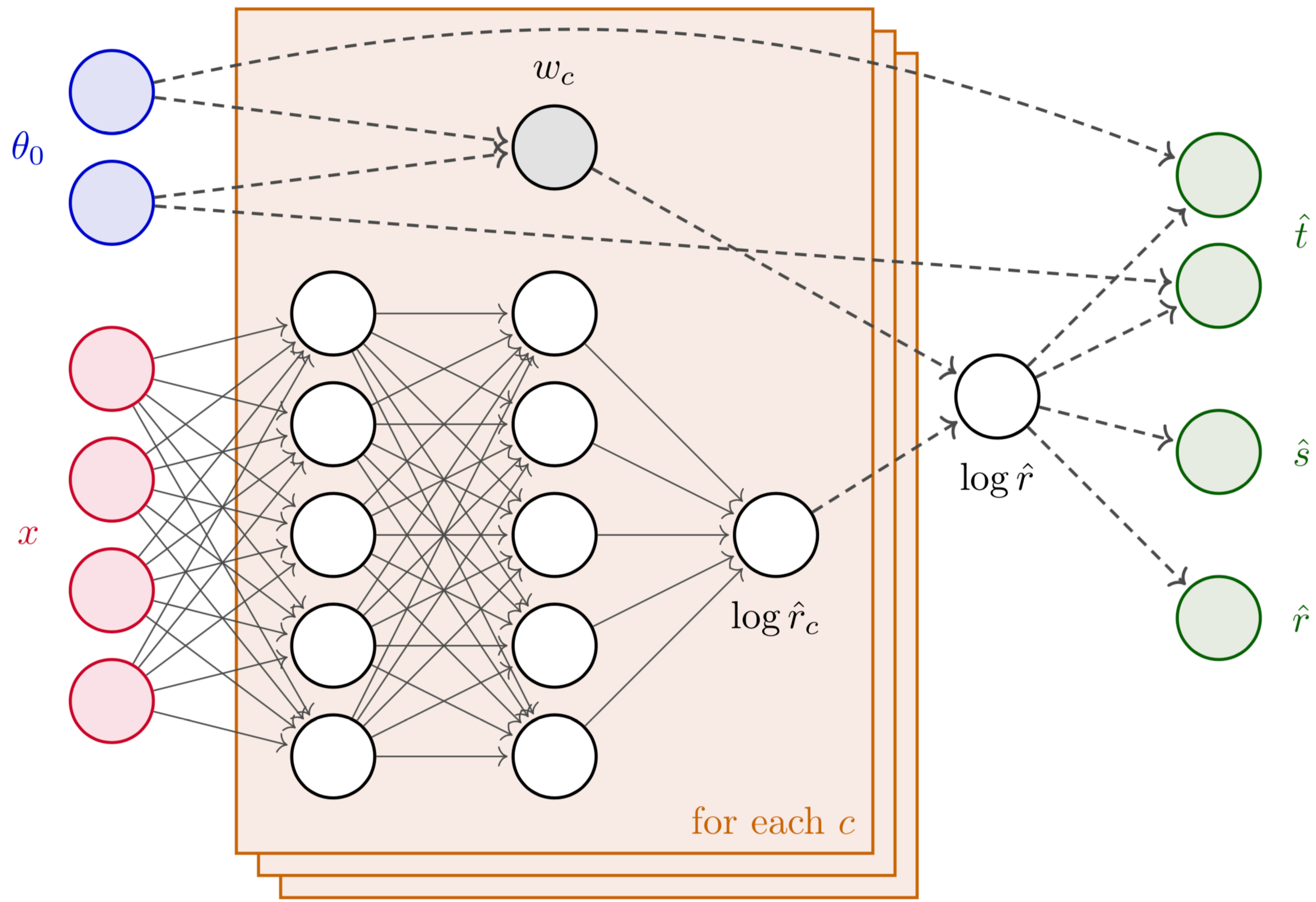


(Agnostic) parameterized estimators

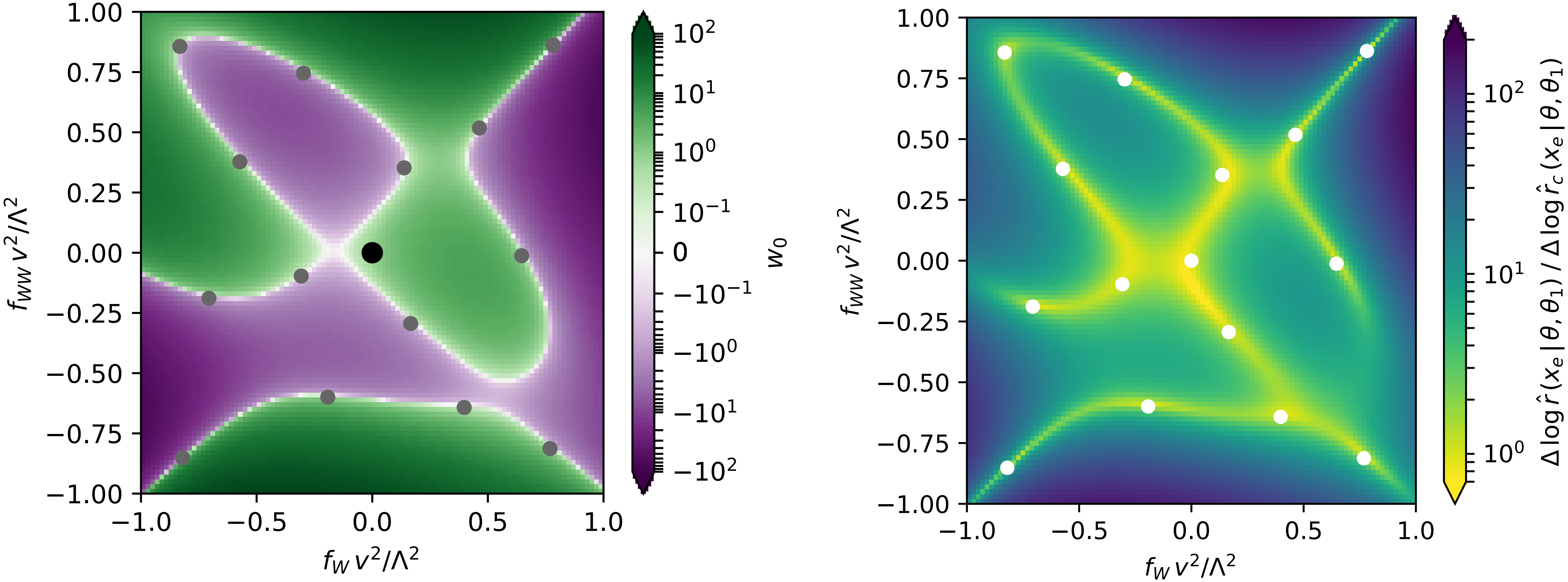


Morphing-aware parameterized estimators

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



Morphing coefficients



The score

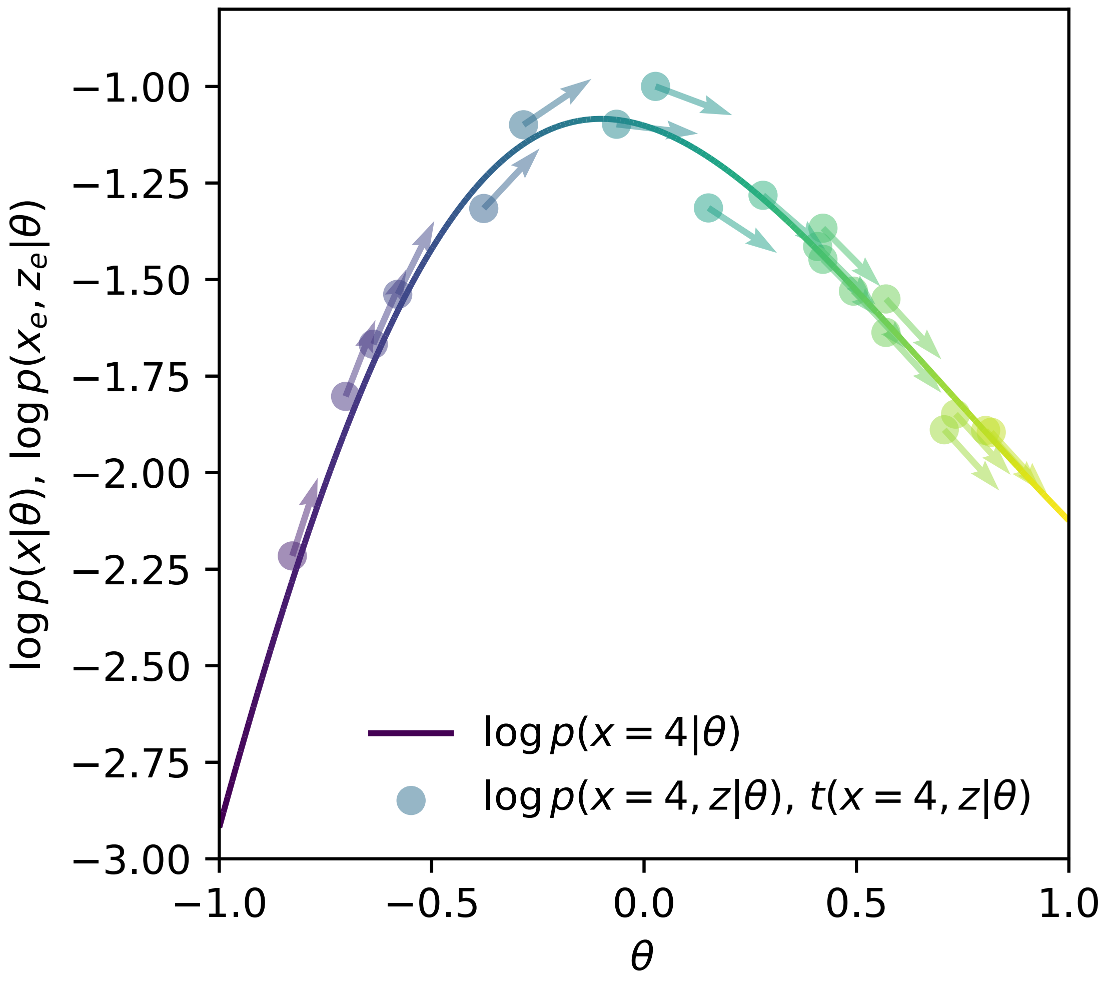
- Inference just based on the joint likelihood ratio works well, but there is another powerful piece of information

- The score

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

fully characterizes the likelihood function in the neighborhood of θ_0

- The score itself is intractable. But…



Learning the score

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

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

$$\downarrow$$

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

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

We want score

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

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

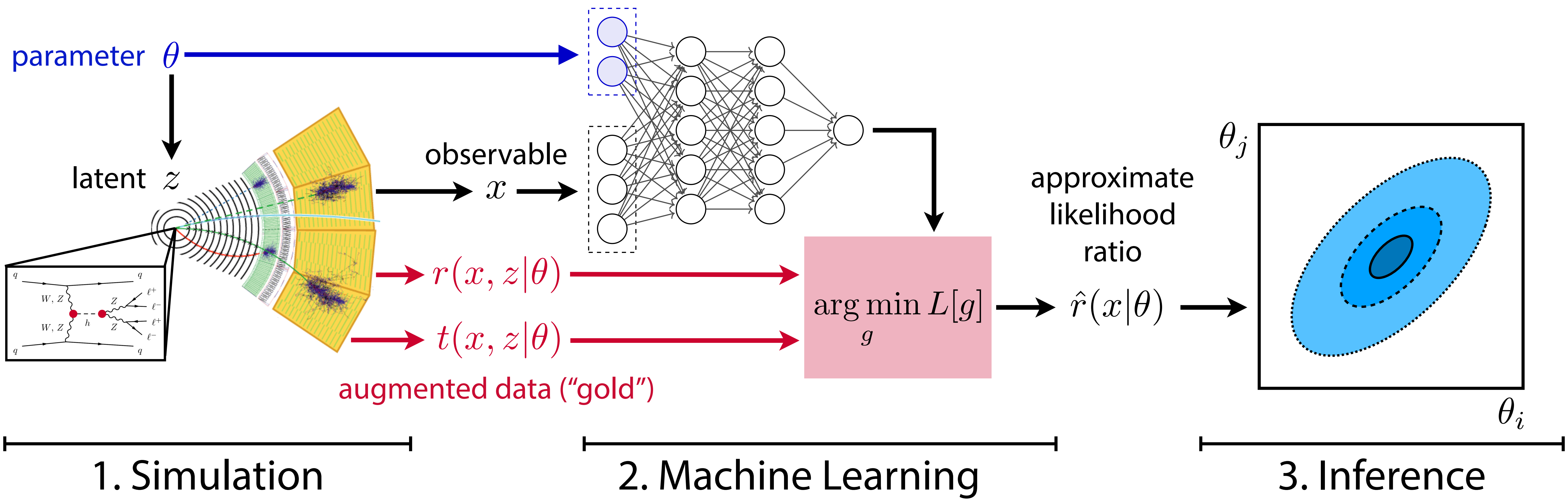
$$L_t[\hat{t}(x)] = \mathbb{E}_{p(x,z|\theta_0)}\left[\left(\hat{t}(x) - t(x, z|\theta_0)\right)^2\right].$$

One can show it is minimized by

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

Again, we implement this
with machine learning

Putting the pieces together



A family of likelihood-free inference strategies

Different strategies to combine the different pieces of information:

Method	L_{XE}	L_{MLE}	L_r	L_t	θ sampling
ABC (**A**pproximate **B**ayesian **C**omputation)					$\theta \sim \pi(\theta)$
NDE (**N**eural **d**ensity **e**stimation)		✓			$\theta \sim \pi(\theta)$
LRT / CARL (**L**ikelihood **r**atio **t**rick / **c**alibrated **a**pproximate **r**atios of **l**ikelihoods)	✓				$\theta \sim \pi(\theta)$
ROLR (**R**egression **o**n **l**ikelihood **r**atio)			✓		$\theta \sim \pi(\theta)$
SCANDAL (**Sc**ore **a**ugmented **n**eural **d**ensity **a**pproximates **l**ikelihood)		✓		✓	$\theta \sim \pi(\theta)$
CASCAL (**C**ARL **a**nd **sc**ore **a**pproximate **l**ikelihood ratio).	✓			✓	$\theta \sim \pi(\theta)$
RASCAL (**R**atio **a**nd **sc**ore **a**pproximate **l**ikelihood ratio)			✓	✓	$\theta \sim \pi(\theta)$
SALLY (**S**core **a**pproximates **l**ikelihood **l**ocall**y**)				✓	$\theta = \theta_0$
SALLINO (**S**core **a**pproximates **l**ikelihood **l**ocally **in** **o**ne dimension)				✓	$\theta = \theta_0$

A family of likelihood-free inference strategies

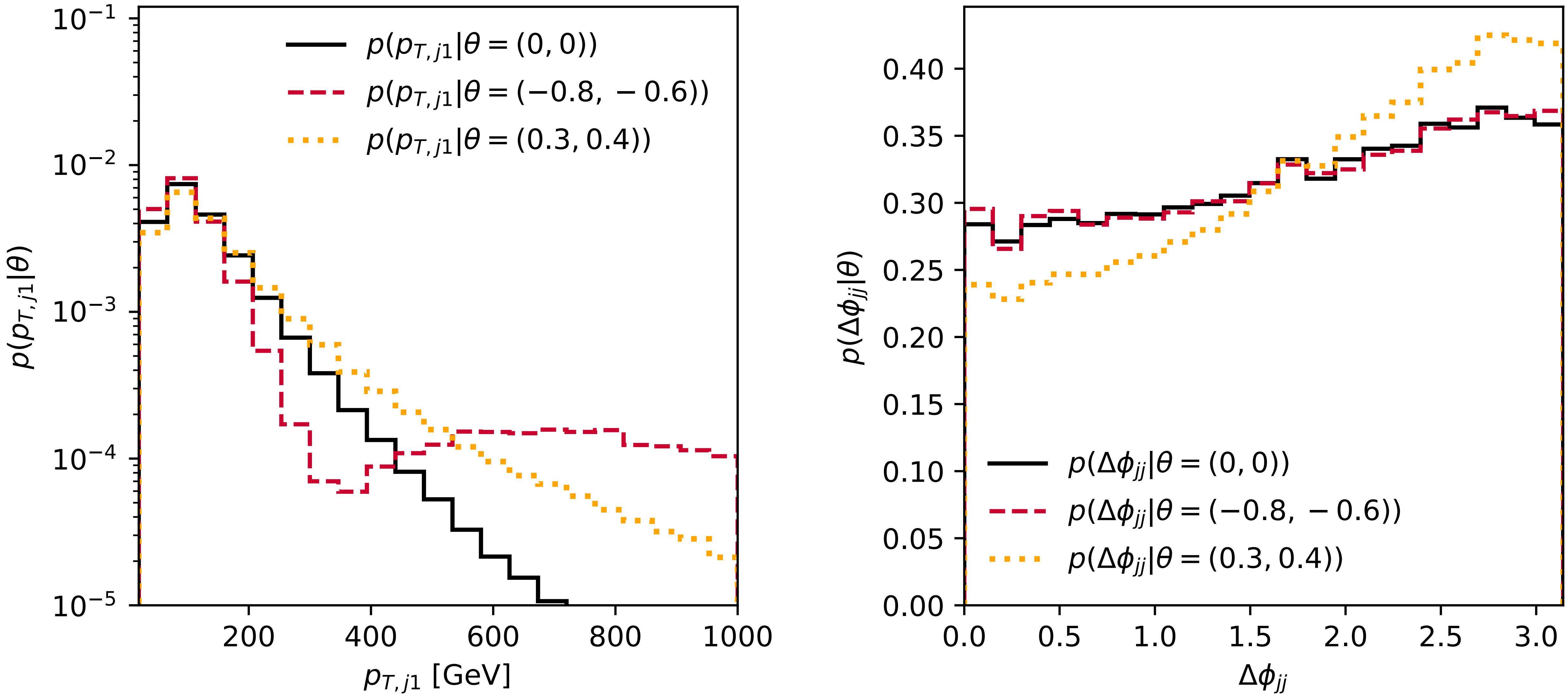
Different strategies to combine the different pieces of information:

Method	L_{XE}	L_{MLE}	L_r	L_t	θ sampling
ABC (**A**pproximate **B**ayesian **C**omputation)					$\theta \sim \pi(\theta)$
NDE (**N**eural **d**ensity **e**stimation)		✓			$\theta \sim \pi(\theta)$
LRT / CARL (**L**ikelihood **r**atio **t**rick / **c**alibrated **a**pproximate **r**atios of **l**ikelihoods)	✓				$\theta \sim \pi(\theta)$
ROLR (**R**egression **o**n **l**ikelihood **r**atio)			✓		$\theta \sim \pi(\theta)$
SCANDAL (**Sc**ore **a**ugmented **n**eural **d**ensity **a**pproximates **l**ikelihood)		✓		✓	$\theta \sim \pi(\theta)$
CASCAL (**CA**RL **a**nd **sc**ore **a**pproximate **l**ikelihood ratio).	✓			✓	$\theta \sim \pi(\theta)$
RASCAL (**R**atio **a**nd **sc**ore **a**pproximate **l**ikelihood ratio)			✓	✓	$\theta \sim \pi(\theta)$
SALLY (**S**core **a**pproximates **l**ikelihood **l**ocall**y**)				✓	$\theta = \theta_0$
SALLINO (**S**core **a**pproximates **l**ikelihood **l**ocally **in** **o**ne dimension)				✓	$\theta = \theta_0$

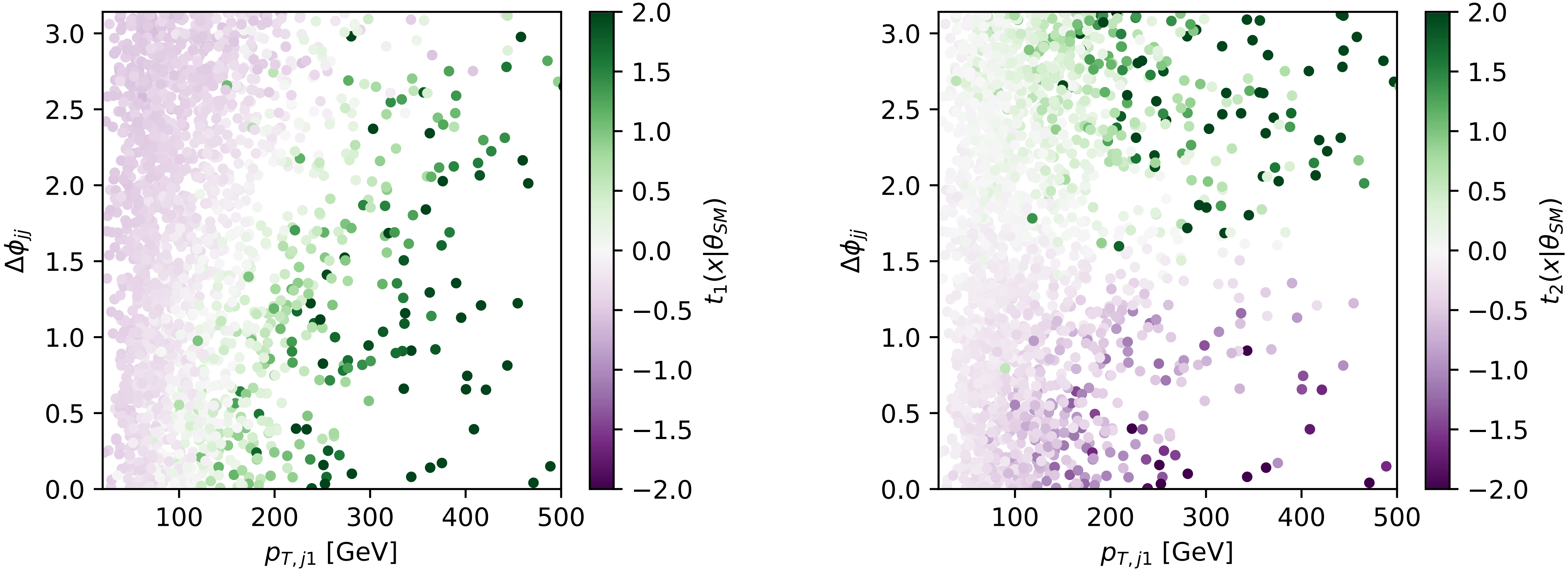
RASCAL loss function:

$$L_{\text{RASCAL}}[\hat{r}(x|\theta_0,\theta_1)] = L_r[\hat{r}(x|\theta_0,\theta_1)] + \alpha\ L_t[\nabla_{\theta_0} \log \hat{r}(x|\theta_0,\theta_1)]$$

Distributions



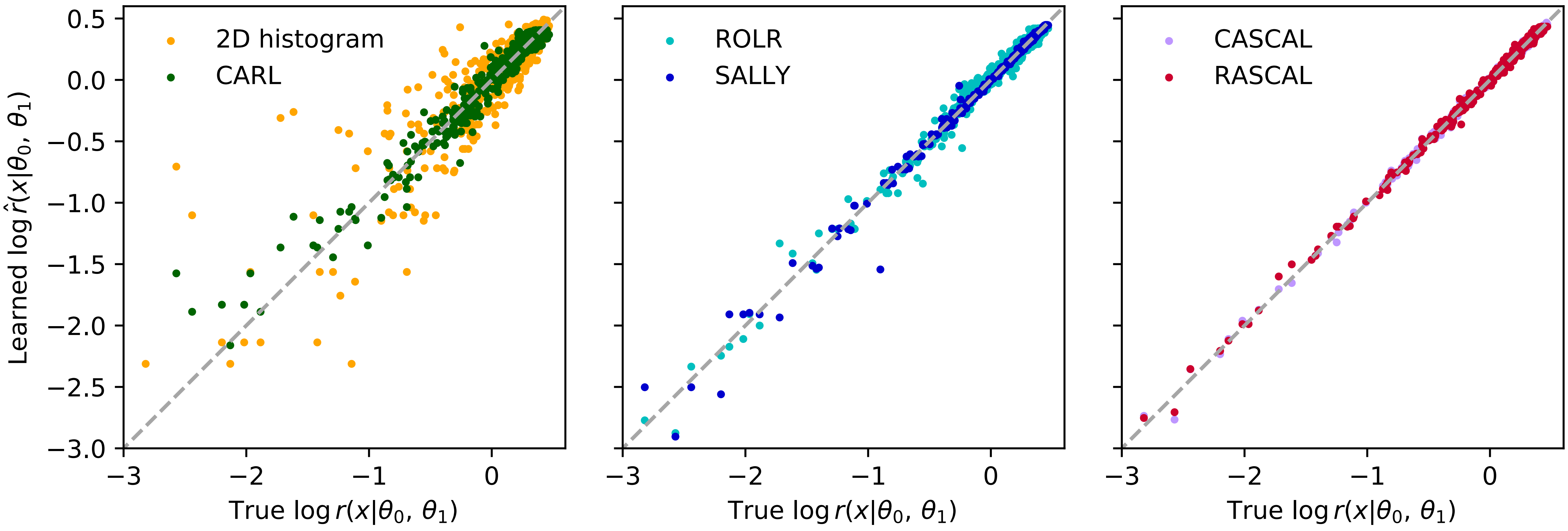
Score vs observables



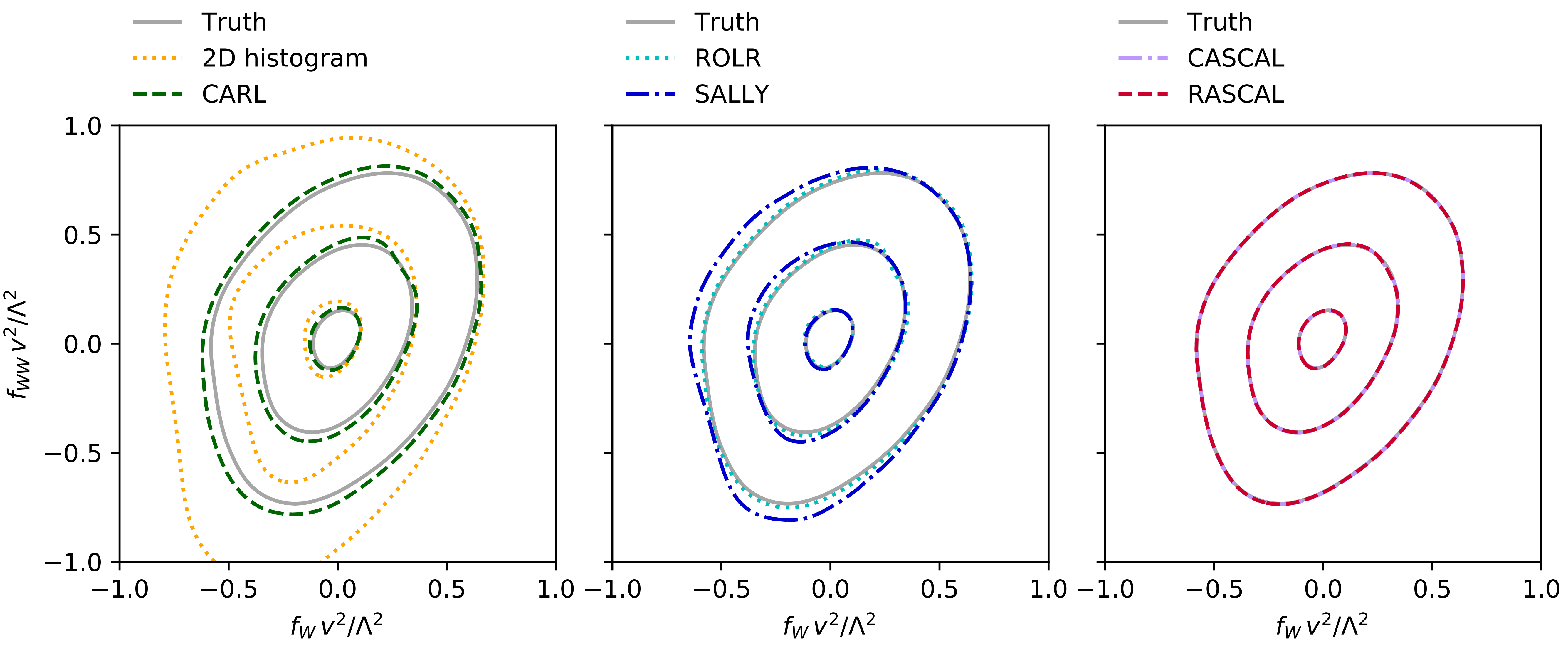
Results

Strategy	Setup	Expected MSE		Figures
		All	Trimmed	
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	

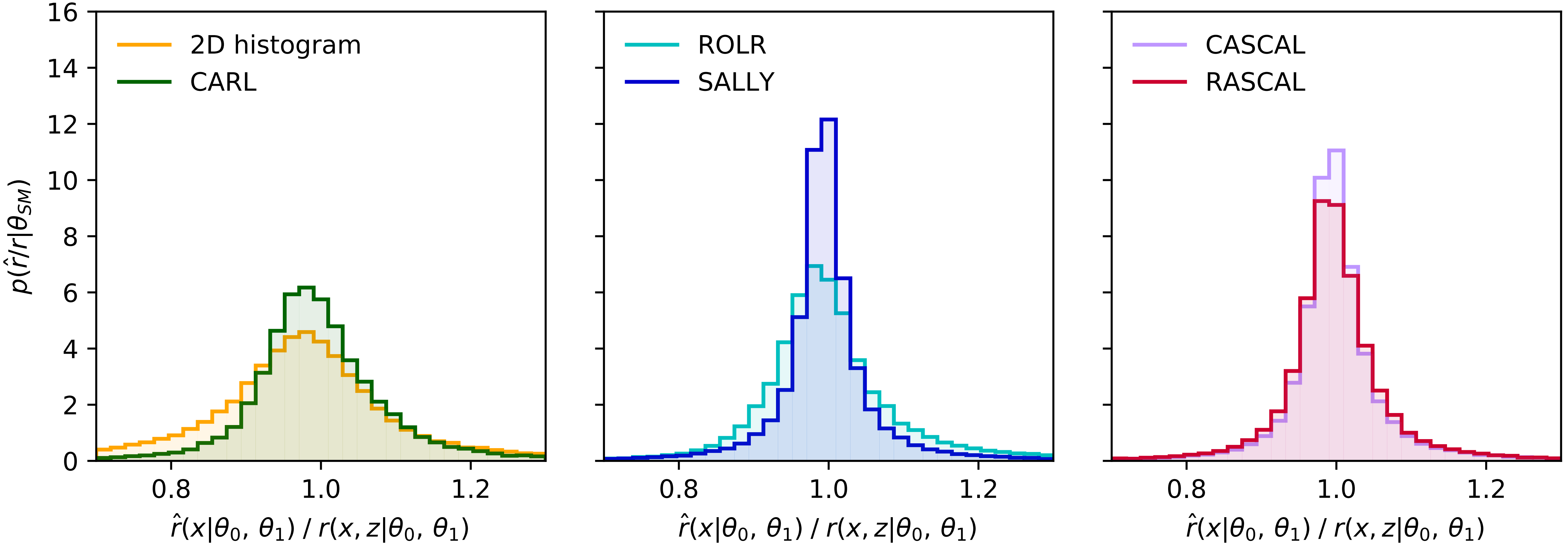
Precision of estimates



Expected limits (truth level)



Detector effects



Expected limits (with detector effects)