# Learning to constrain new physics

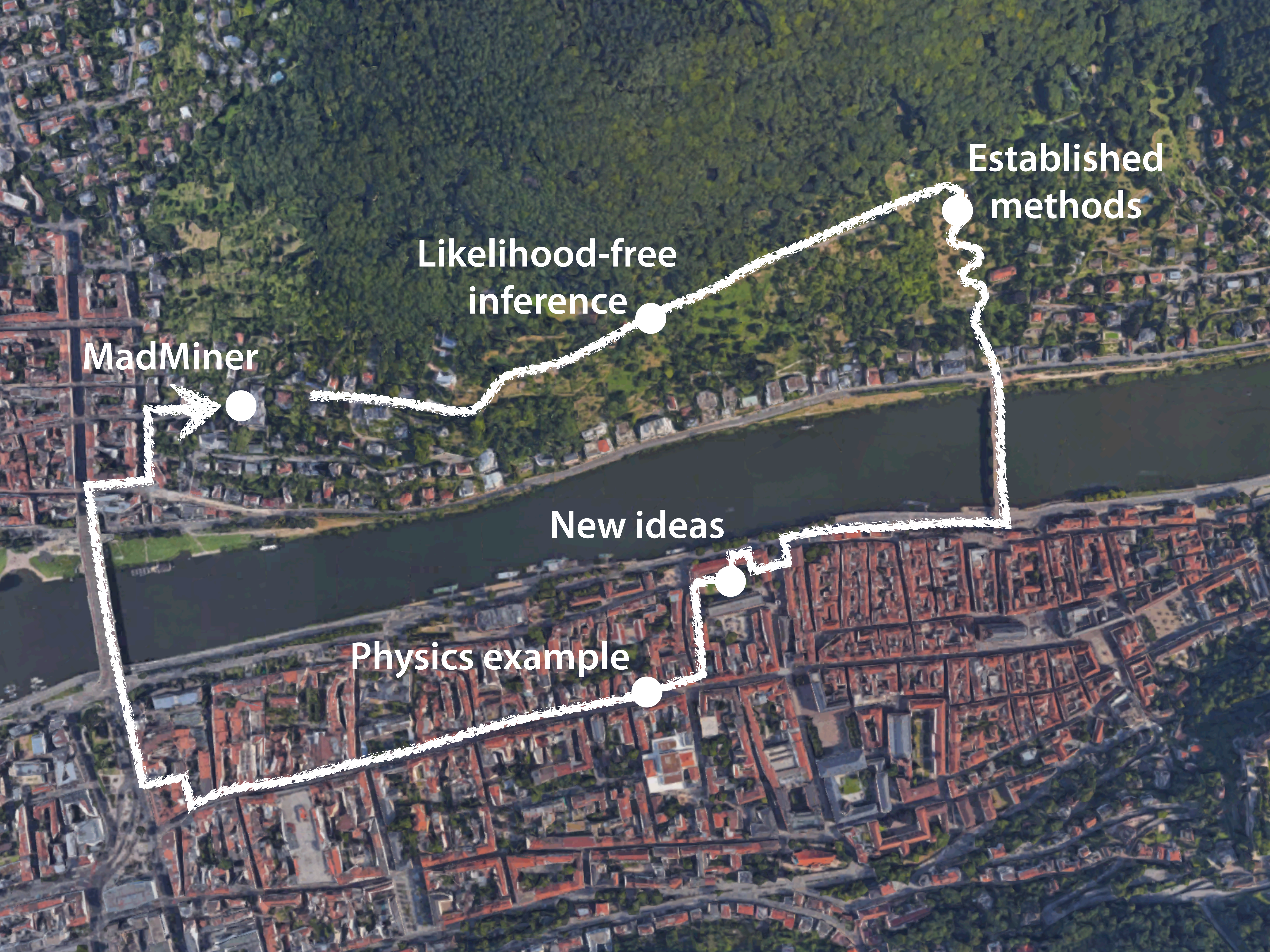
Johann Brehmer

New York University

Heidelberg
August 21, 2018

NYU

MSDSE
Moore - Sloan
Data Science Environments


Established
methods
Likelihood-free
inference
MadMiner
New ideas
Physics example

# Likelihood-free inference

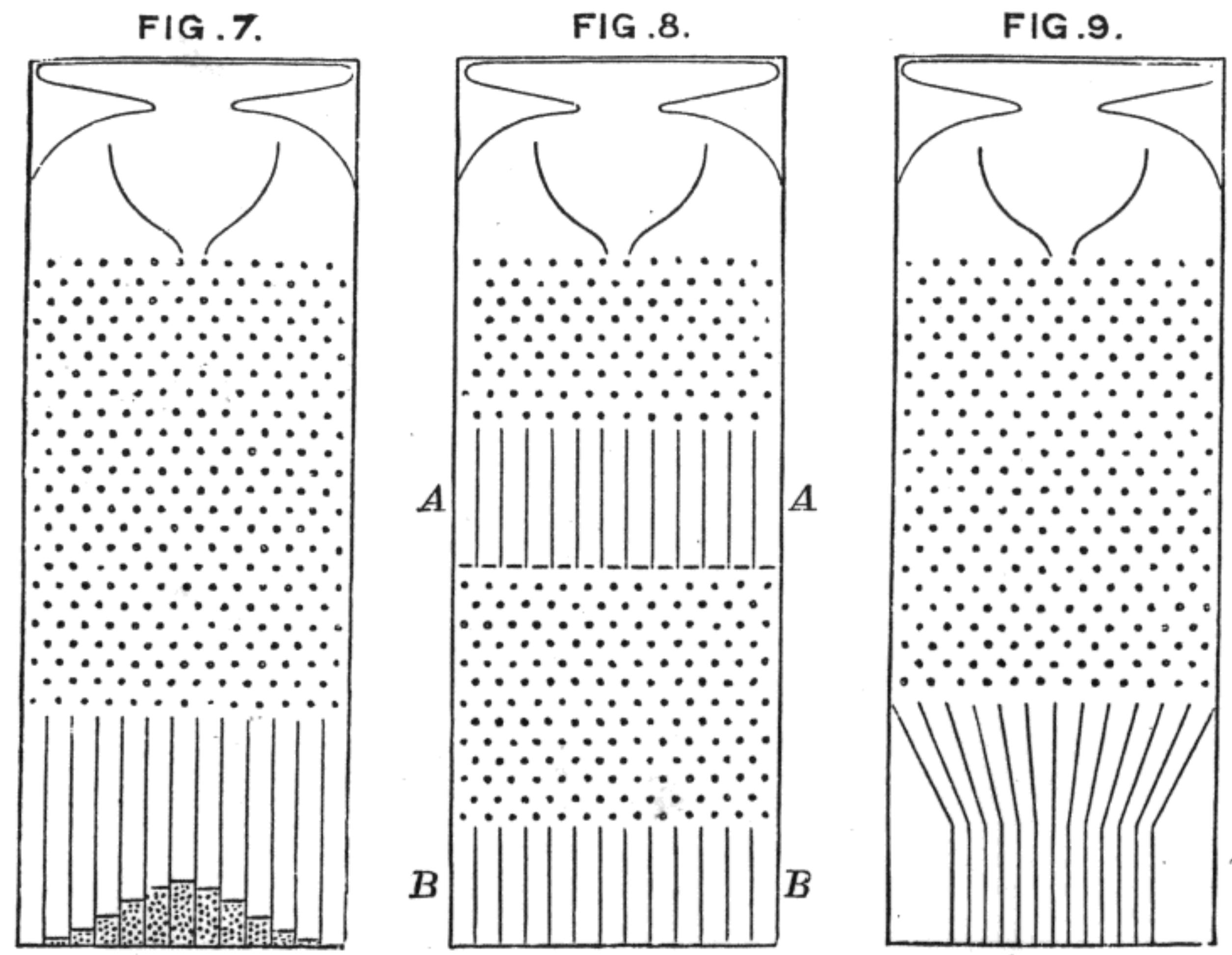
# The Galton board


[Source: galtonboard.com]

# The Galton board

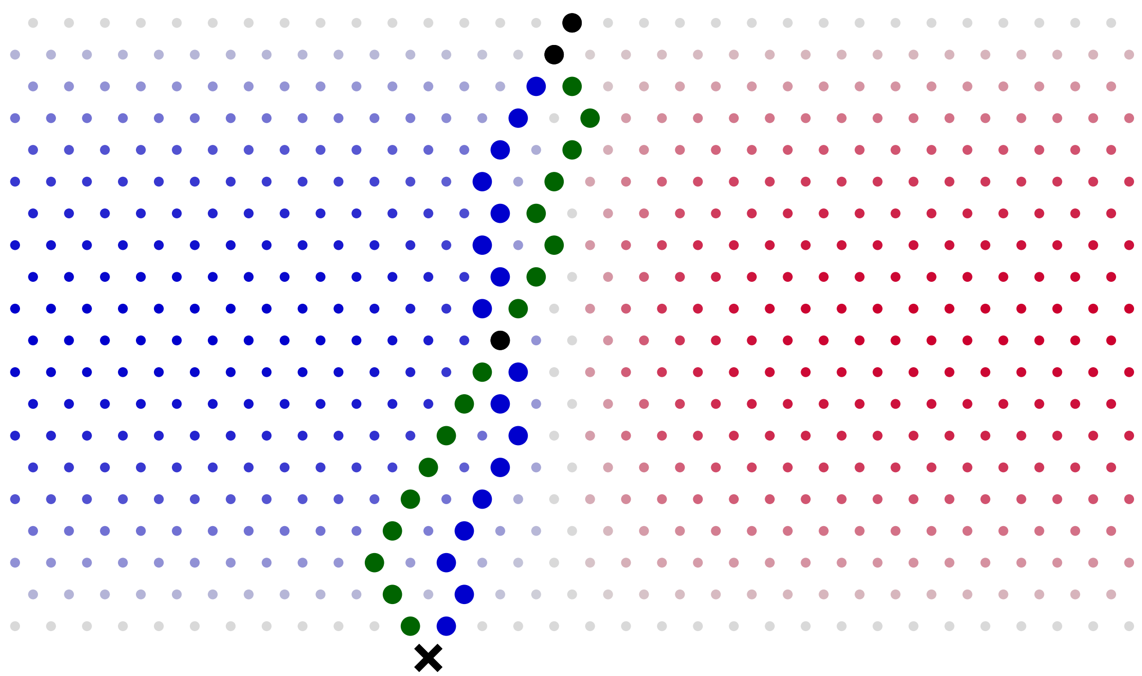

[Source: galtonboard.com]

# The Galton board




[Source: F. Galton 1889]

# Probabilities



Probability of ending in bin $x$:

$$p(x) = \int \mathrm{d}z \; p(x, z)$$

Sum over
all trajectories
("latent variables")

Probability of
each path $z$
from start to $x$

# The generalized Galton board

- What if probability to go left at a nail is not always 0.5, but some (known) function of some parameters $\theta$ ?
- Prediction: given $\theta$, generate observations $\{x_i\}\ldots$ just drop balls!
- Inference: given observations $\{x_i\}$, what are the most likely values for $\theta$ ?

# The generalized Galton board

- What if probability to go left at a nail is not always 0.5, but some (known) function of some parameters $\theta$ ?

- Prediction: given $\theta$, generate observations $\{x_i\}\ldots$ just drop balls!

- Inference: given observations $\{x_i\}$, what are the most likely values for $\theta$ ?

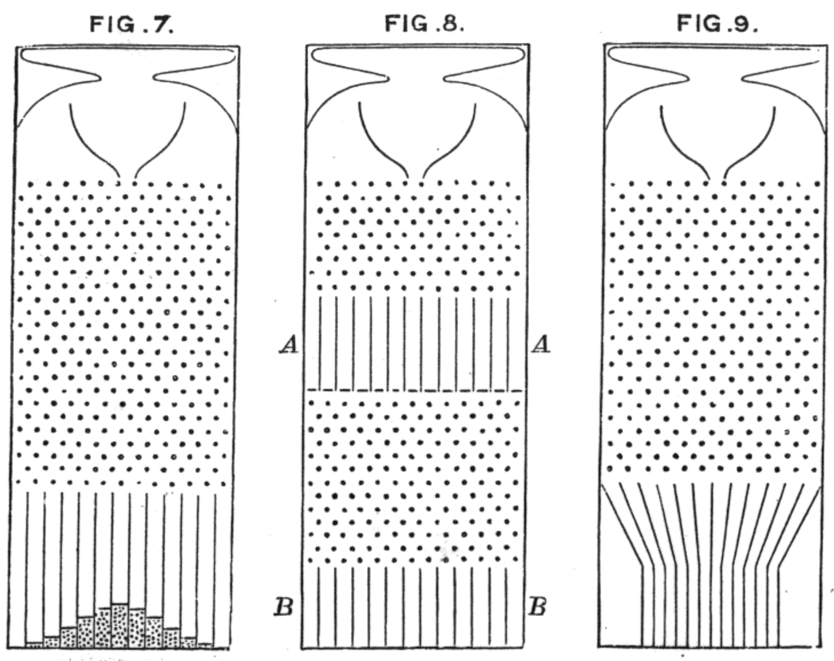
"Easy" problem if we can evaluate likelihood $p(x_i|\theta)$. But

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

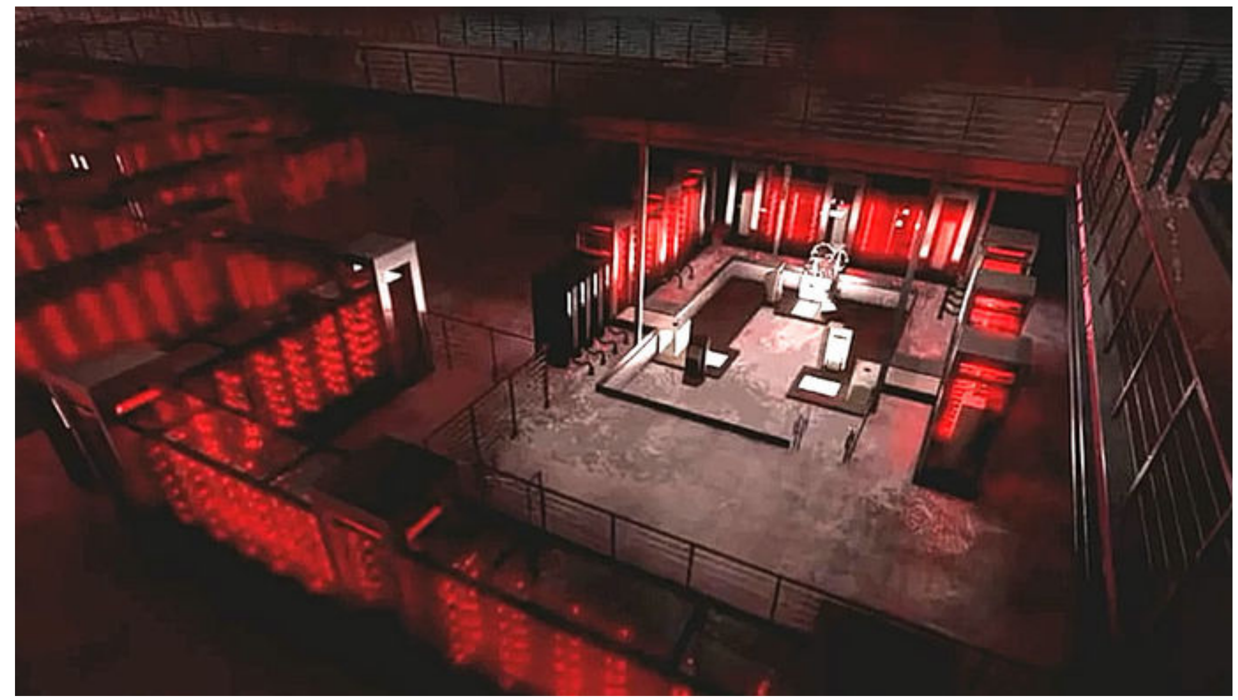
The number of possible paths $z$ can be huge, and it becomes impossible to calculate the integral

⇒ Likelihood not tractable, only implicitly defined through "simulator"

# Galton board: metaphor for simulator-based science



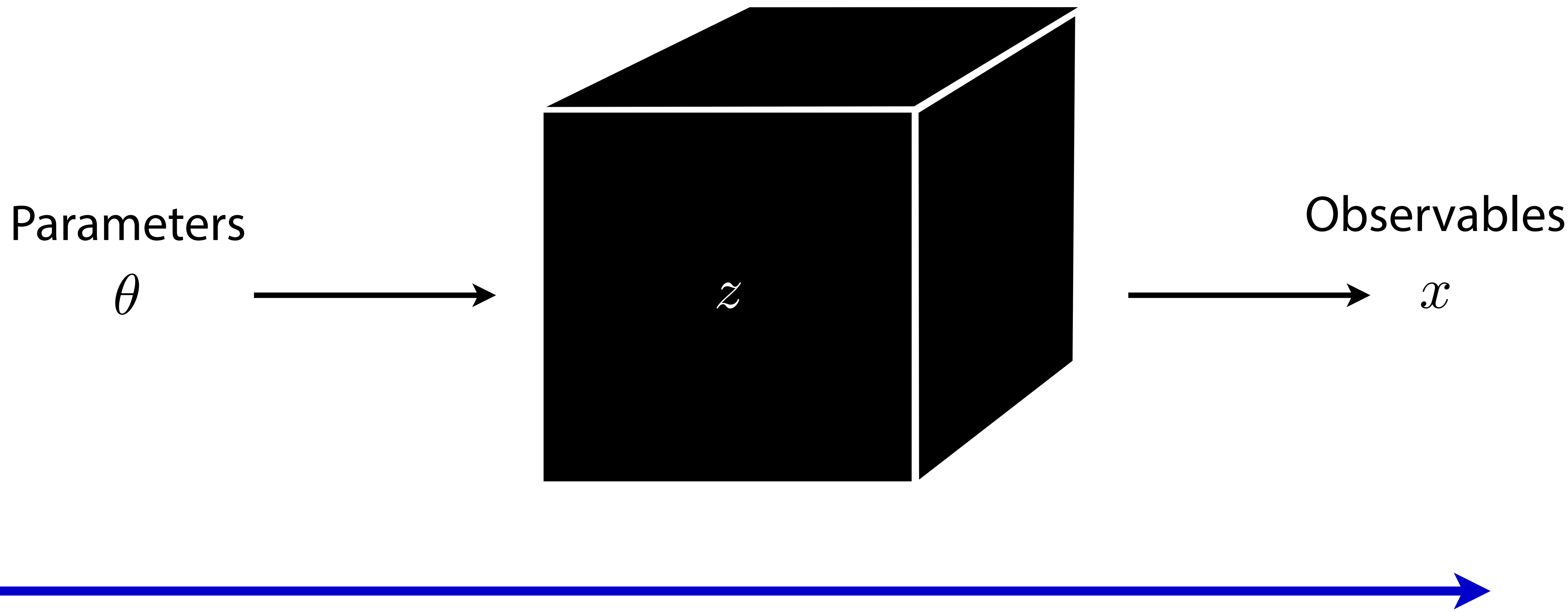

[Source: F. Galton 1889]



[Source: HBO 2018]

| | | |
|---|---|---|
| Galton board device | $\longrightarrow$ | Computer simulation |
| Parameters $\theta$ | $\longrightarrow$ | Parameters of interest $\theta$ |
| Bins $x$ | $\longrightarrow$ | Observables $x$ |
| Path $z$ | $\longrightarrow$ | Latent variables $z$ (stochastic execution trace through simulator) |

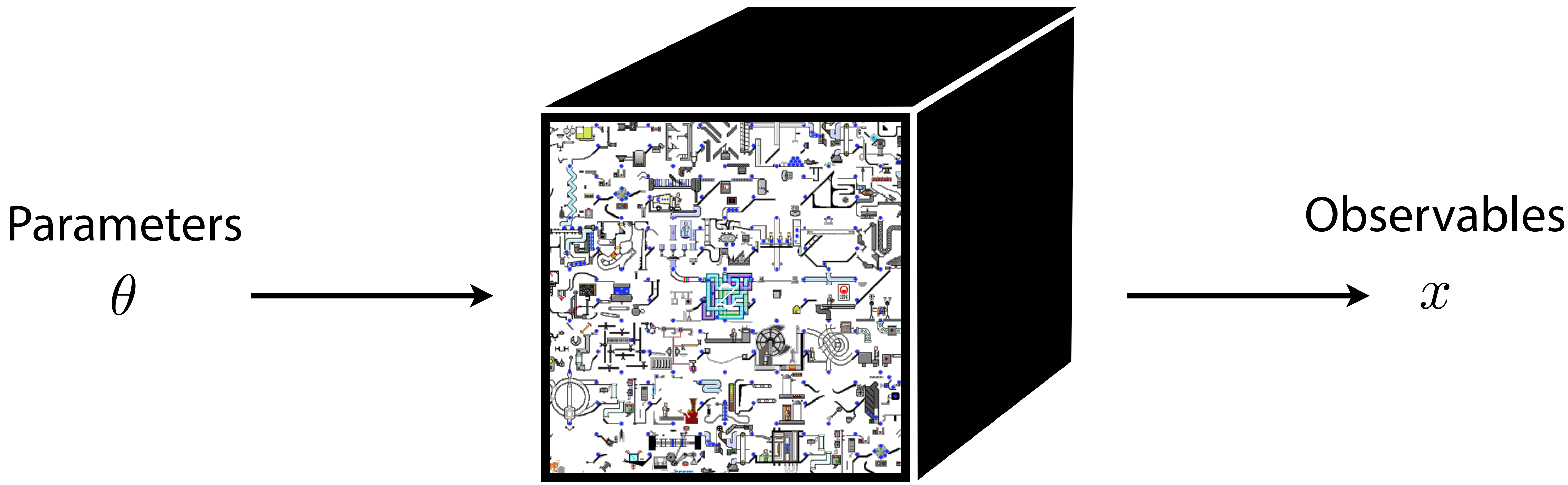
# Likelihood-free inference




Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

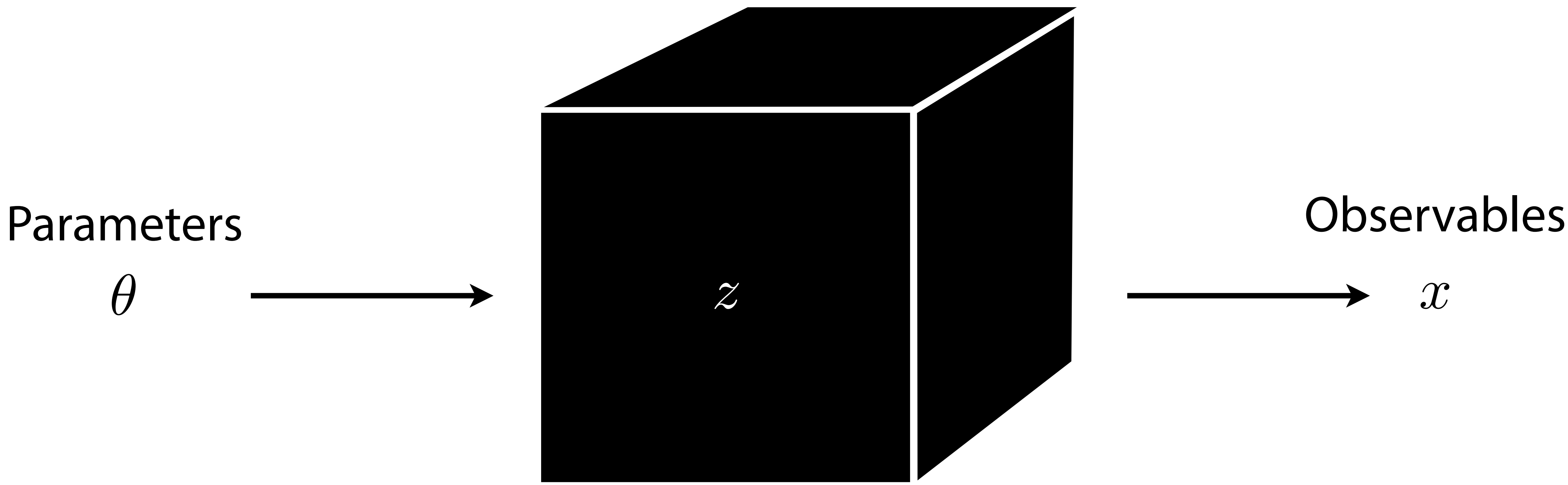
# Likelihood-free inference




Prediction (simulation):

- Well-understood mechanistic model
- Simulator can generate samples

# Likelihood-free inference




Prediction (simulation):

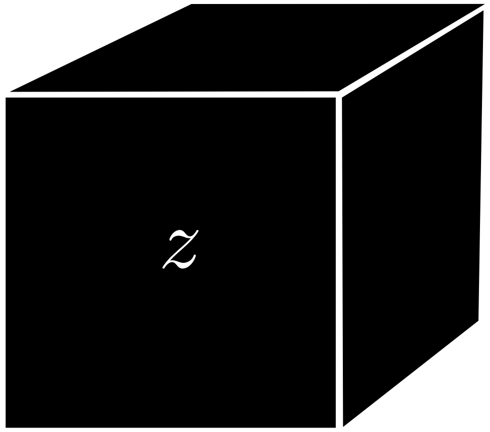
- Well-understood mechanistic model
- Simulator can generate samples

Inference:

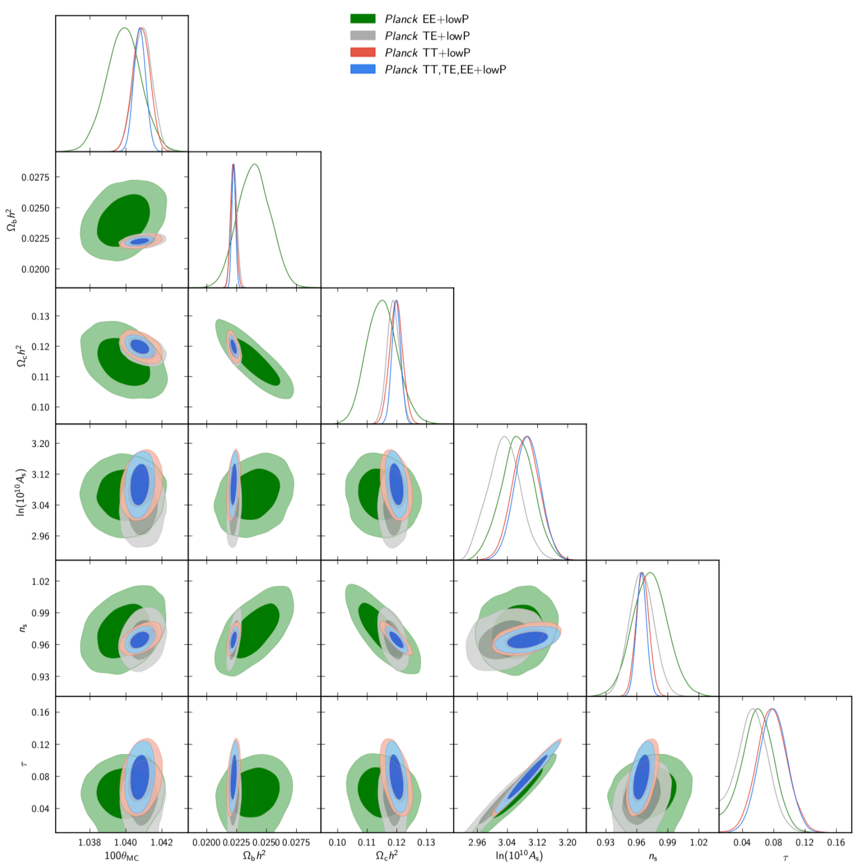
- Likelihood function $p(x|\theta)$ is intractable
- Goal: estimator $\hat{p}(x|\theta)$

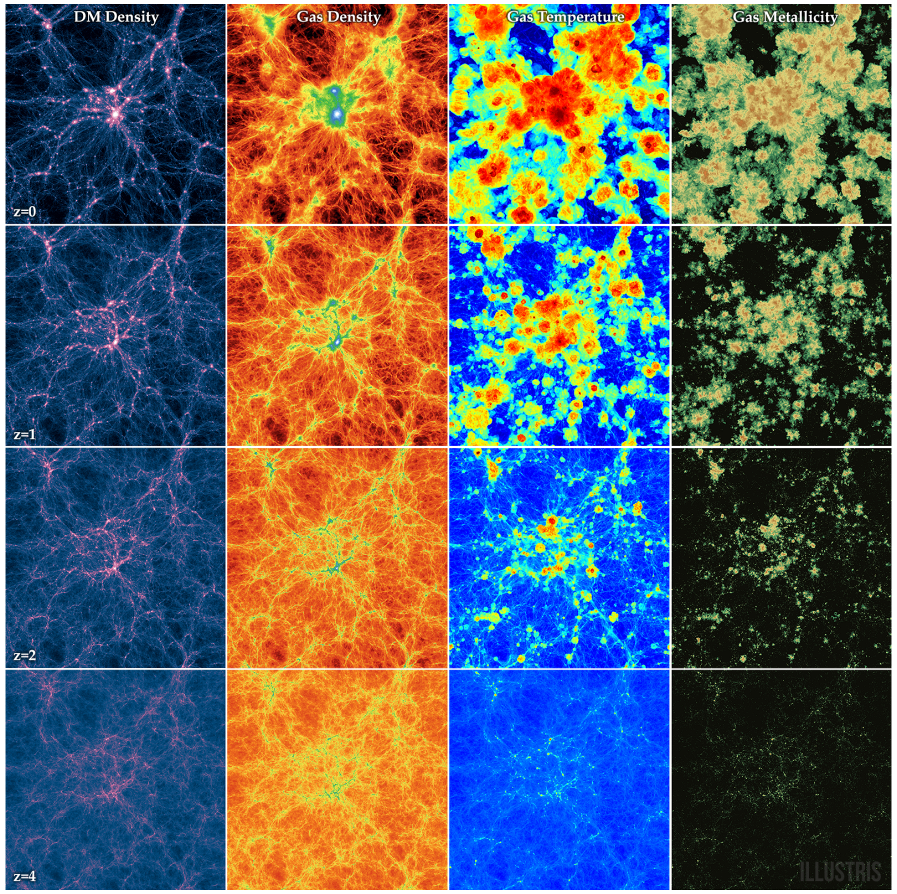
# Cosmological N-body simulations

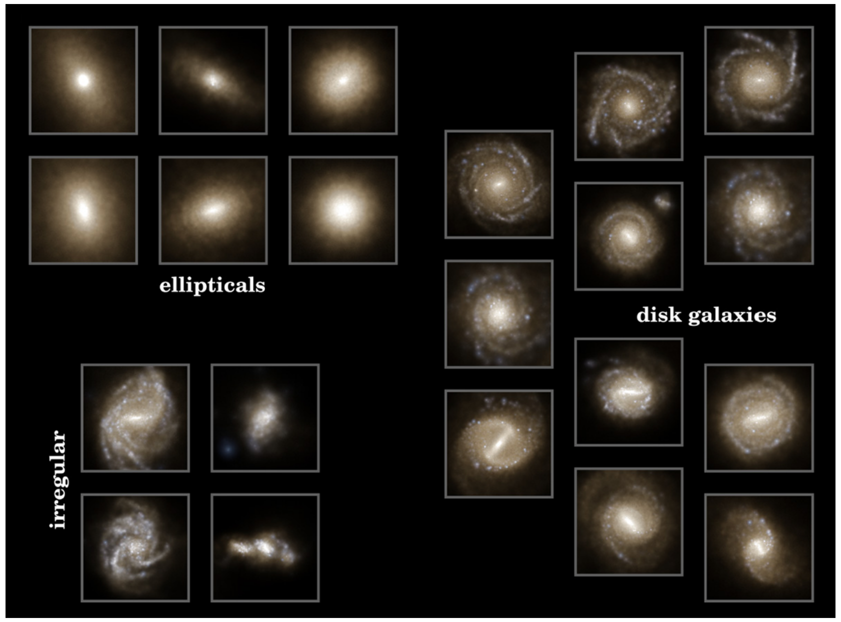
Parameters $\theta$ $\longrightarrow$




$\longrightarrow$ Observables $x$



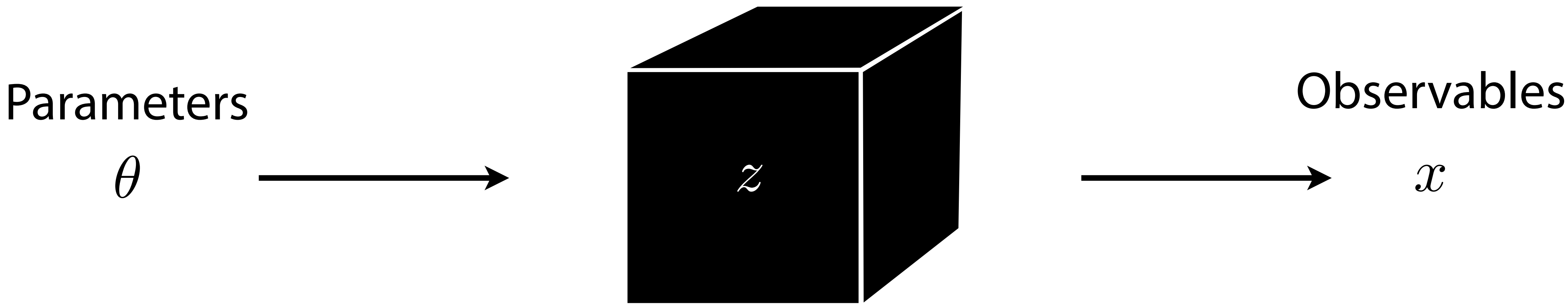



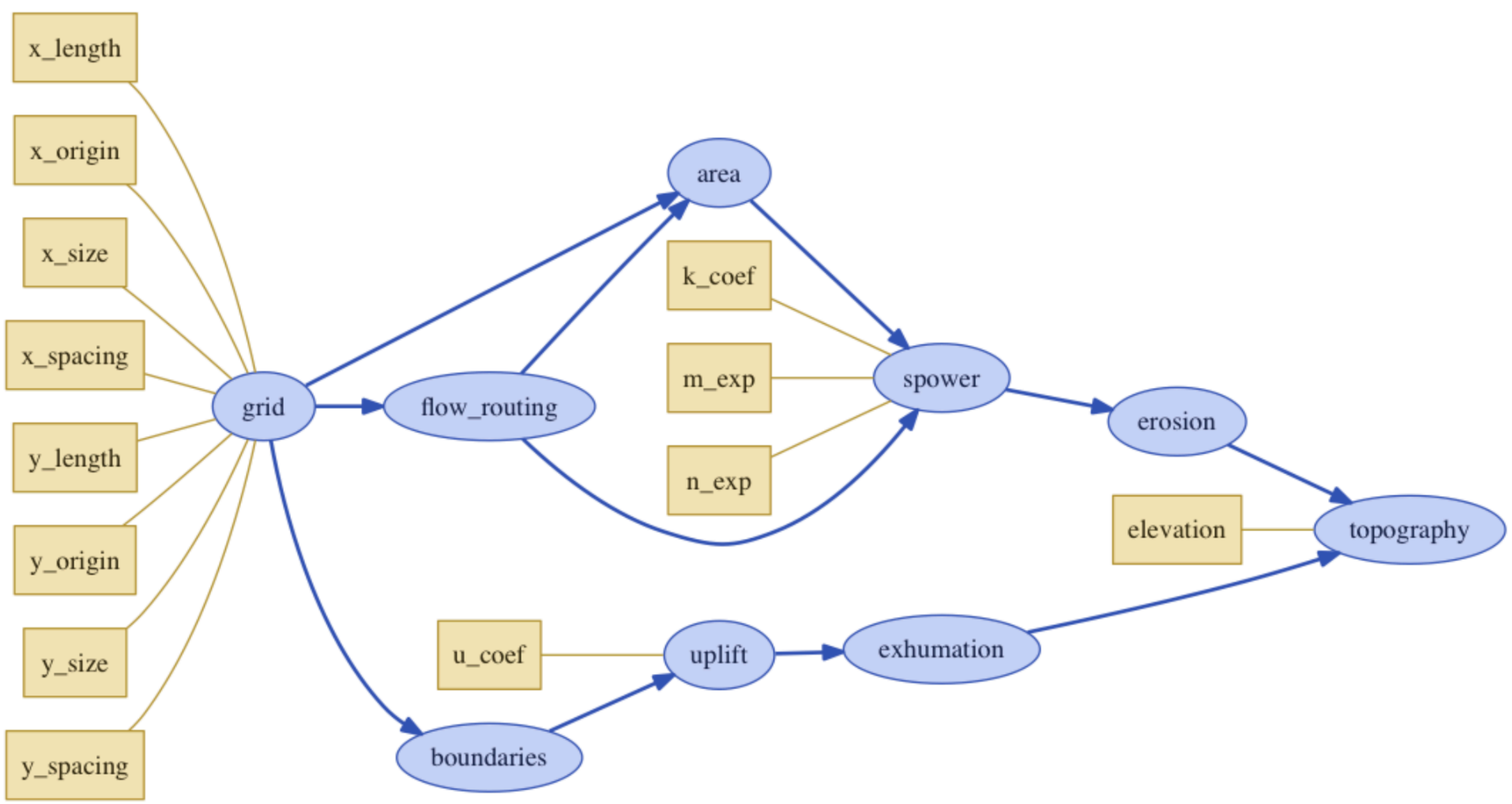



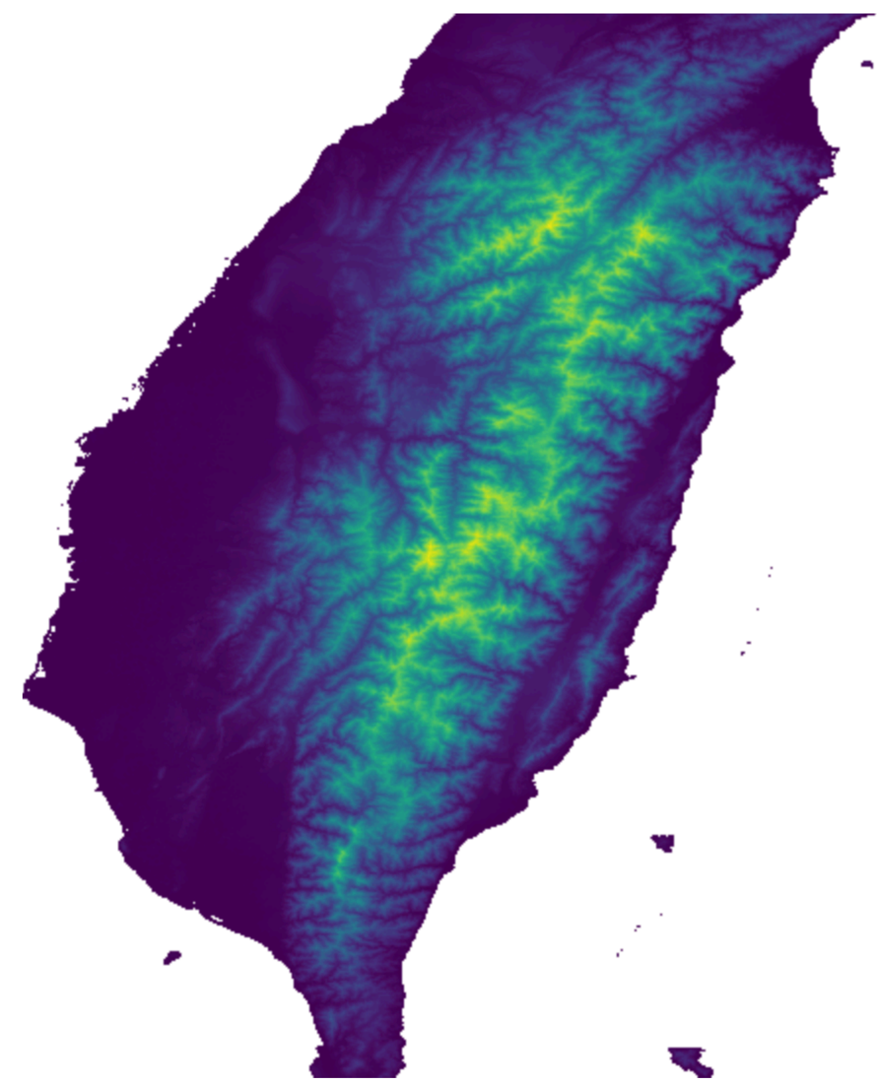

[Source: Planck 1502.01589]

[Source: Illustris 1405.2921]

# Computational topography

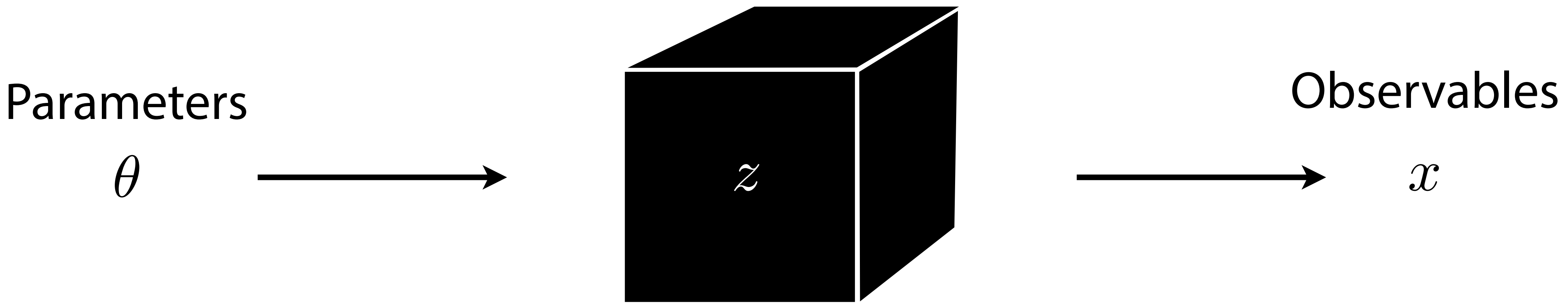


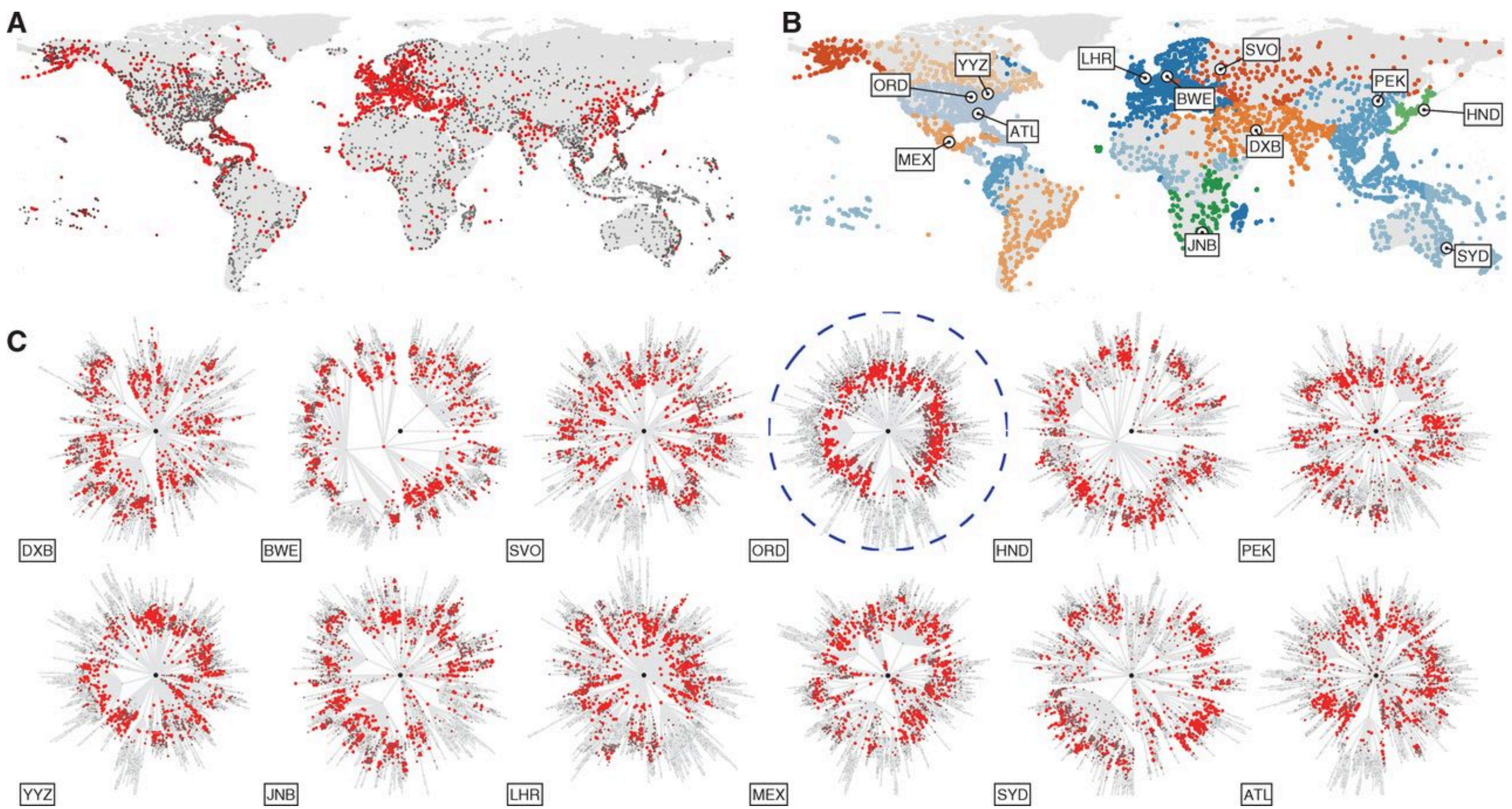



[Source: Benoit Bovy]

# Epidemiology




[Source: D. Brockmann, D. Helbing 2013]

# Particle physics processes: simulation

| Features | Latent variables | | | Parameters of interest |
|---|---|---|---|---|
| Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters |
| $x$ | $z_d$ | $z_s$ | $z_p$ | $\theta$ |

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

# 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 $p(x|z_d)$ $p(z_d|z_s)$ $p(z_s|z_p)$ $p(z_p|\theta)$

Prediction (simulation)

# Particle physics processes: inference

Features | Latent variables | Parameters of interest

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

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

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

$\longrightarrow$ Inference

# Particle physics processes: inference

Features | Latent variables | Parameters of interest

| Observables | Detector interactions | Shower splittings | Parton-level momenta | Theory parameters |
|---|---|---|---|---|
| $x$ | $z_d$ | $z_s$ | $z_p$ | $\theta$ |

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

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

Infeasible to calculate the integral over this enormous space

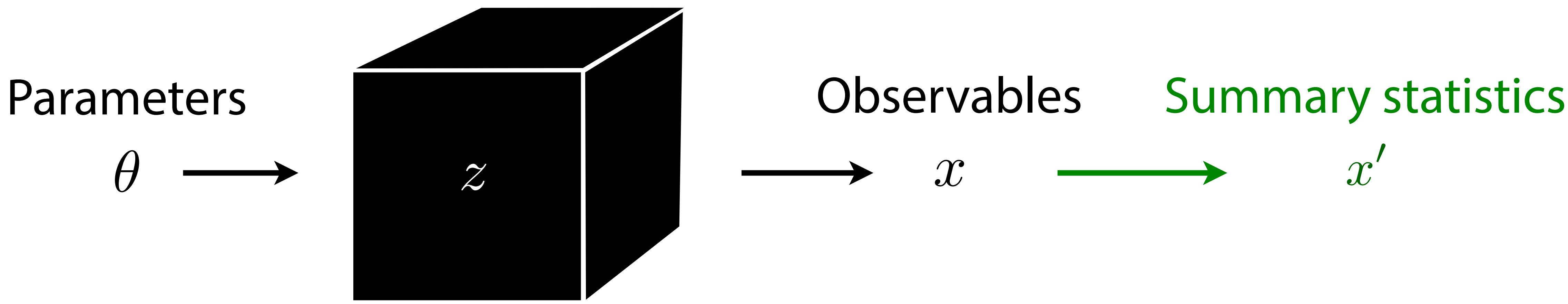
$\longrightarrow$

Inference

Why has that not stopped us so far?

# Why has that not stopped us so far?

- The physicist's way:

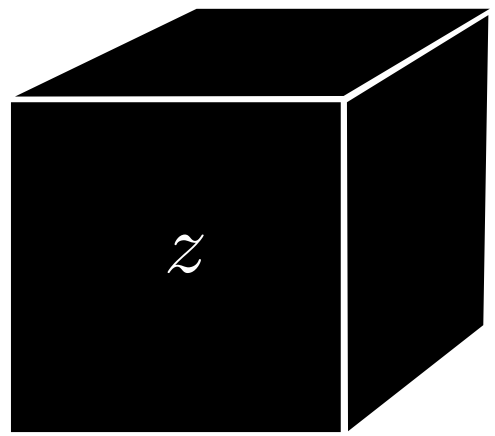


# Why has that not stopped us so far?

- The physicist's way:

Parameters

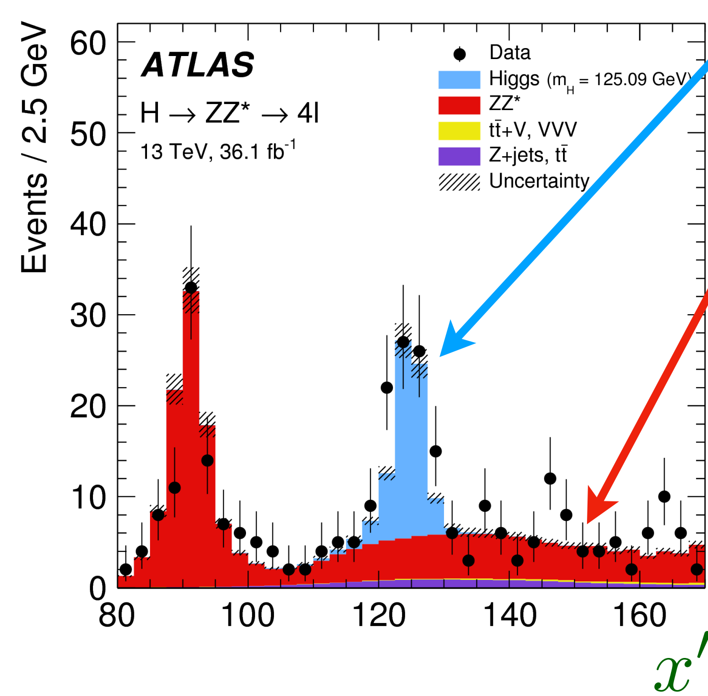
$\theta \longrightarrow$




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

Run simulator for different $\theta$, fill histograms

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

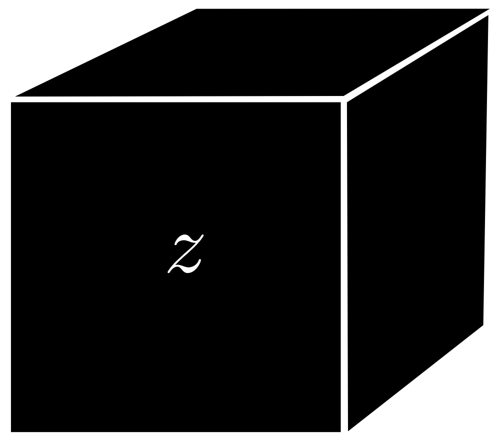



[ATLAS 1712.02304]

# Why has that not stopped us so far?

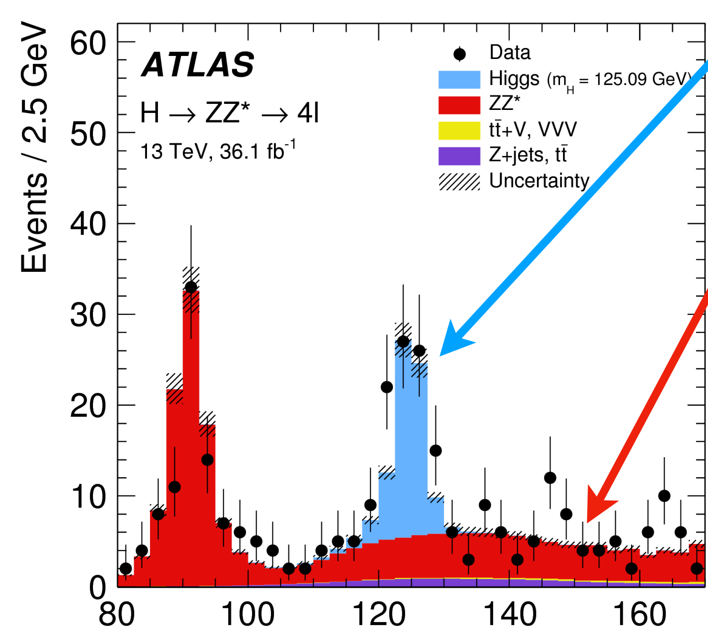
- The physicist's way:

Parameters $\theta$ $\longrightarrow$




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

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




$x'$

[ATLAS 1712.02304]

Run simulator for different $\theta$, fill histograms

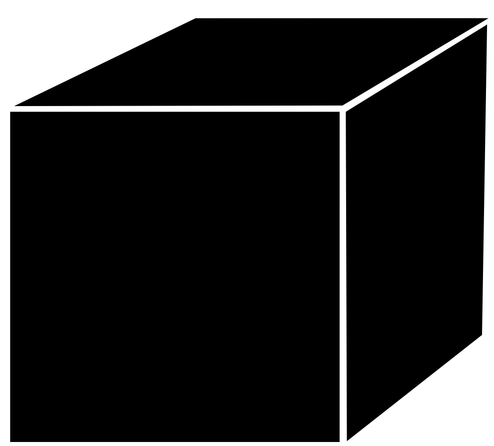
- But how to choose $x'$? Standard variables lose information, analysis of high-dimensional $x$ (including correlations) often more powerful

# Likelihood-free inference methods

Treat simulator as black box:

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

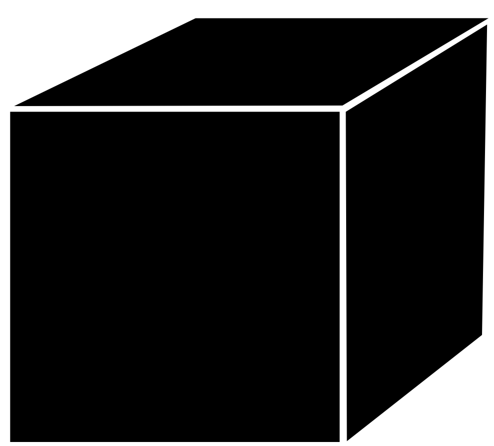
- Machine learning techniques
  Density networks, CARL, autoregressive
  models, normalizing flows, …



# 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, …



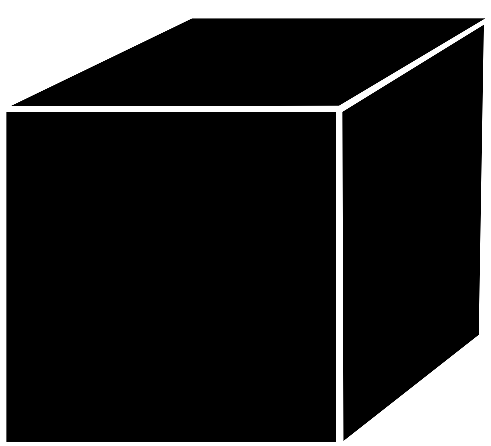
Use latent structure:

- 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:

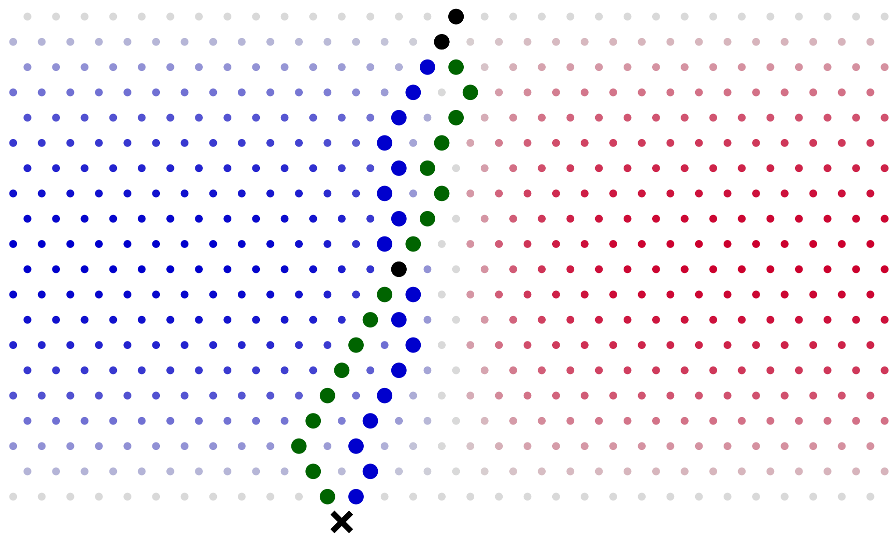
- Matrix Element Method / Optimal Observables
  Neglect or approximate shower + detector, explicitly calculate $z$ integral

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 from the simulator

# Back to the Galton board



- Likelihood

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

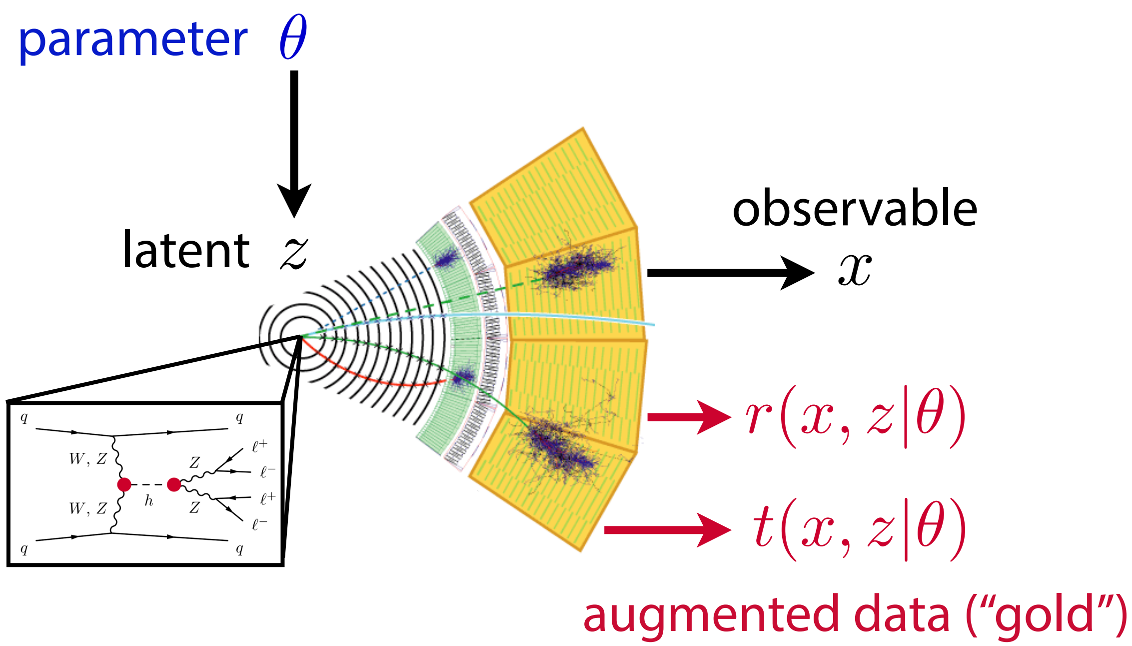
is usually intractable because of the integral over all possible paths $z$

- But: we can calculate the probability of each individual path

$$p(x, z|\theta) = \prod_{\text{nails } i} p_i(x, z|\theta)$$

Turns out that that can be useful…

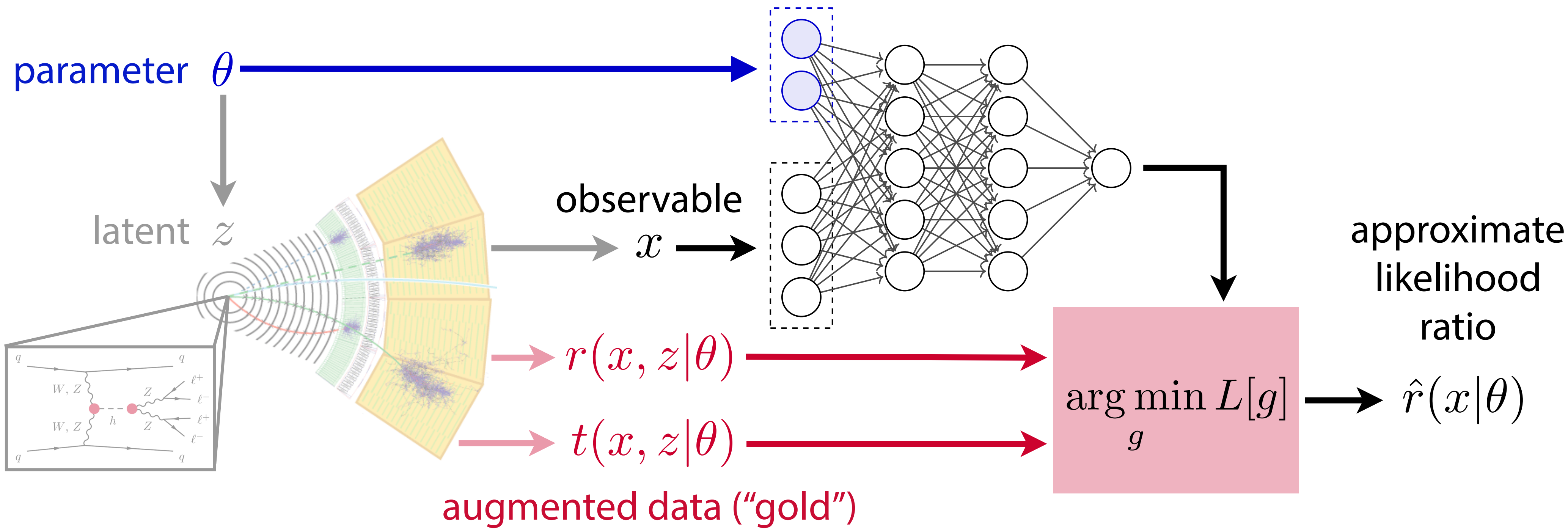
# Bird's-eye view




1. Simulation

"Mining gold": Extract additional information from simulator

# Bird's-eye view




1. Simulation

"Mining gold": Extract additional information from simulator

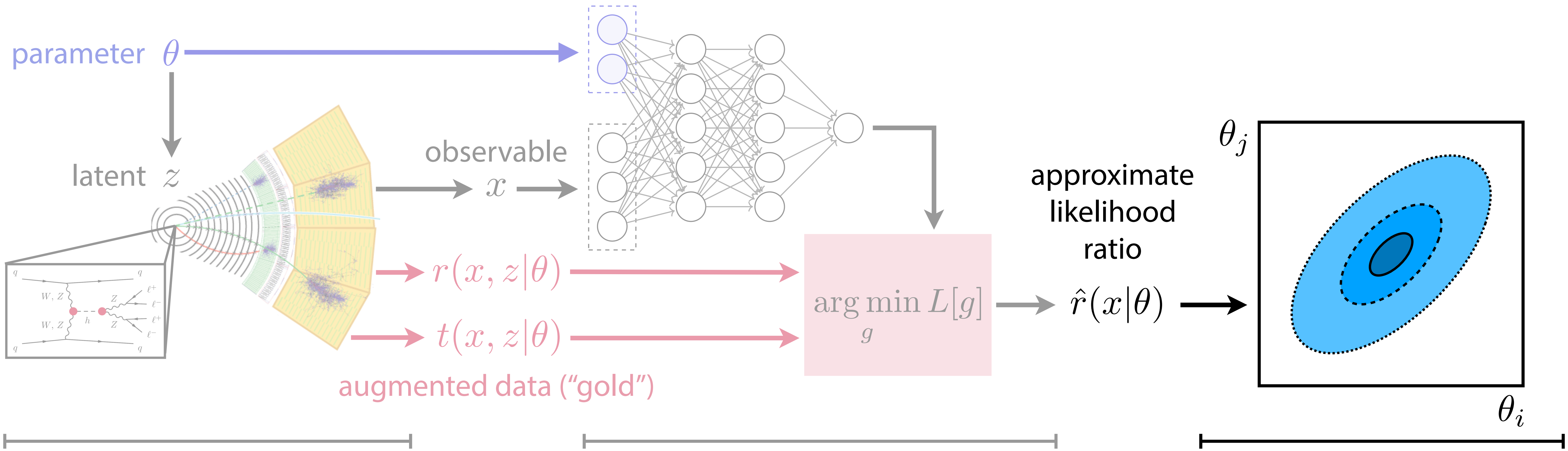
2. Machine Learning

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

# Bird's-eye view




## 1. Simulation

"Mining gold": Extract additional information from simulator

## 2. Machine Learning

Use this information to train precise estimator for likelihood ratio

## 3. Inference

Limit setting with standard hypothesis tests


DEGUSSA
FEINGOLD

# 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 |
|---|---|---|---|---|
| $x$ | $z_d$ | $z_s$ | $z_p$ | $\theta$ |

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

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

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

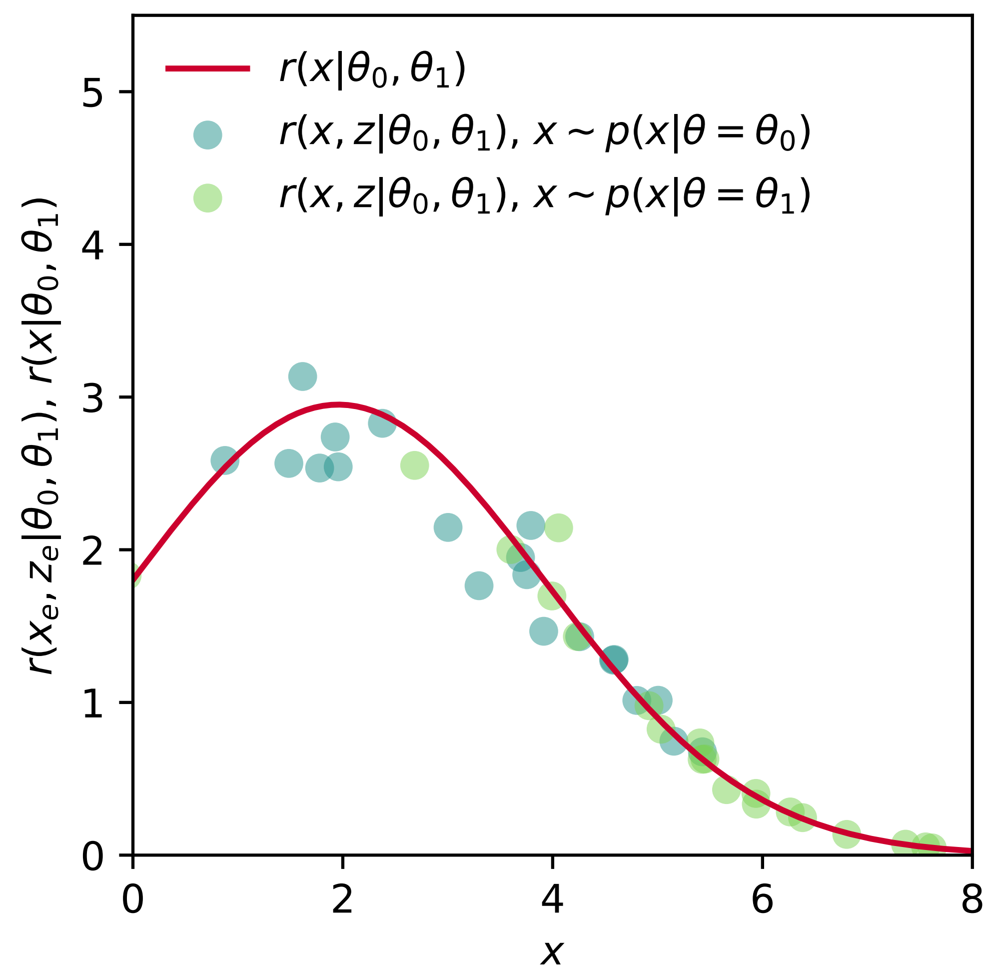
$$r(x|\theta_0, \theta_1) \equiv \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

# 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

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

One can show it is minimized by

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

# 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:

Extremization

Functional

Variational family

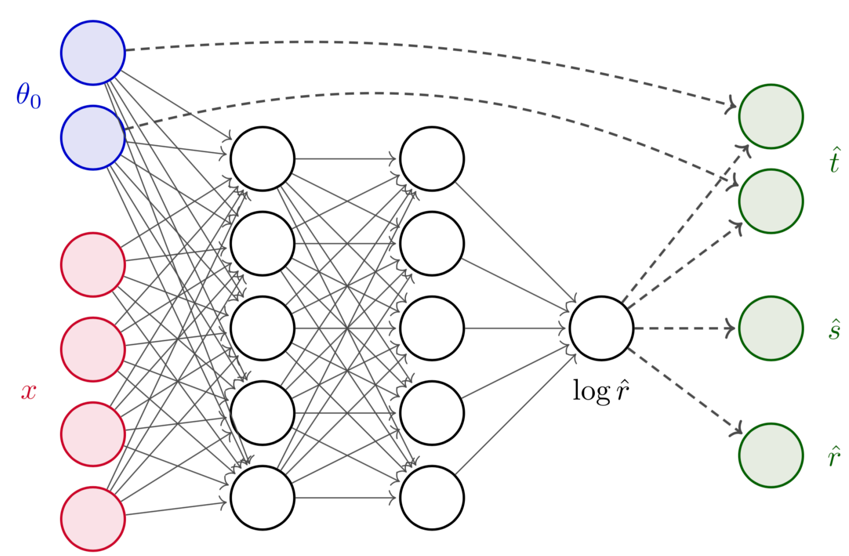
$$r(x|\theta_0, \theta_1) = \underset{\hat{r}(x)}{\arg\min}\, L_r[\hat{r}(x)]$$

This is exactly what machine learning does!

Loss function

Stochastic gradient descent

Neural network

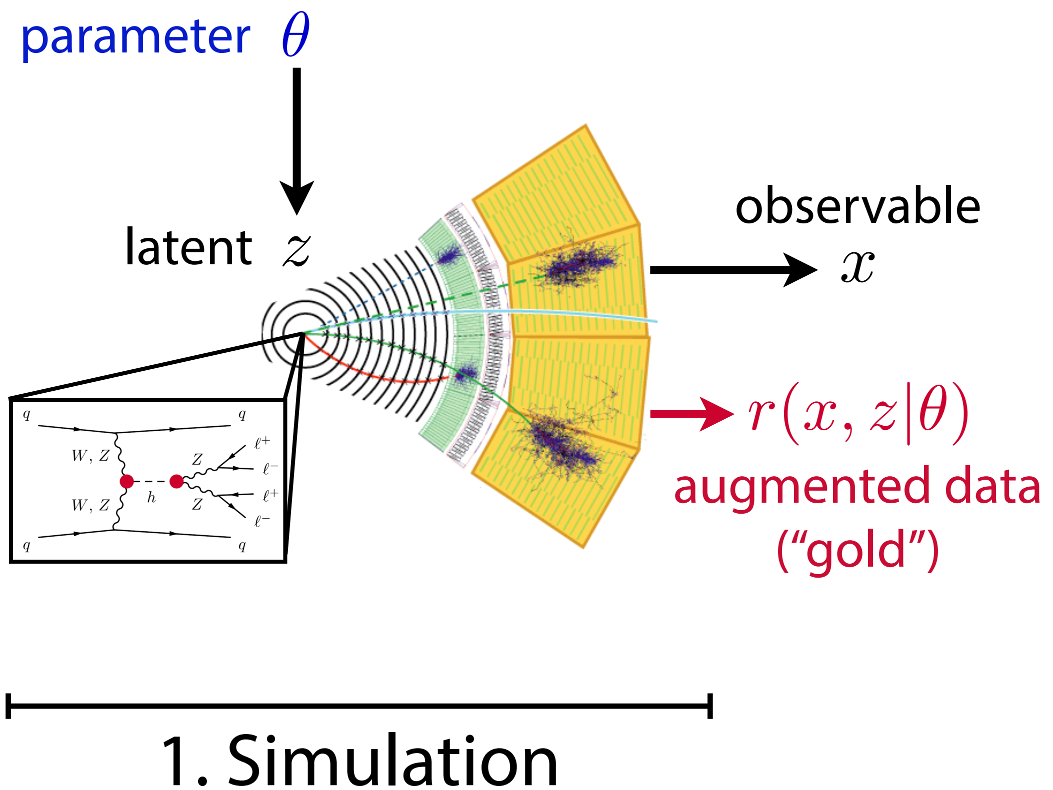



$\Rightarrow$ We implement $\hat{r}(x|\theta_0, \theta_1)$ as a neural network trained on the data available from the simulator

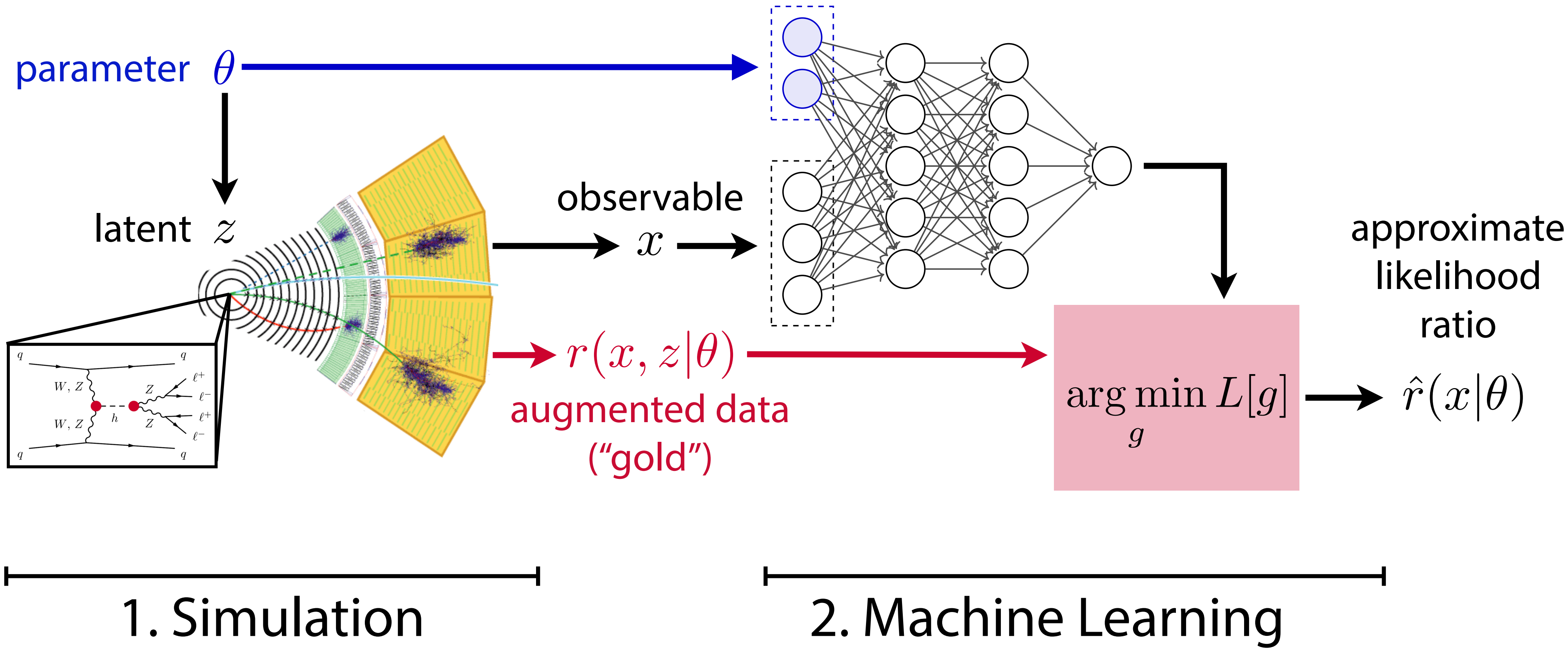
# Parameterized ratio estimators

- Missing: explain parameterized likelihood ratio estimator in detail
- Briefly explain how to get to limits

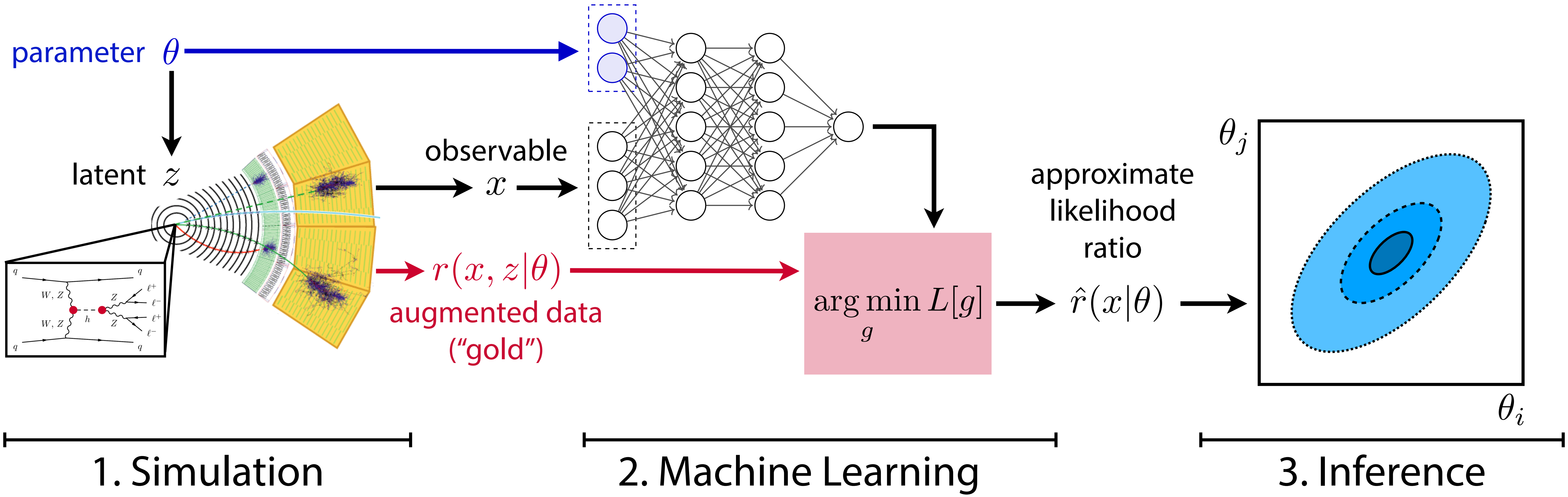
# What we have so far



# What we have so far



# What we have so far



# The score

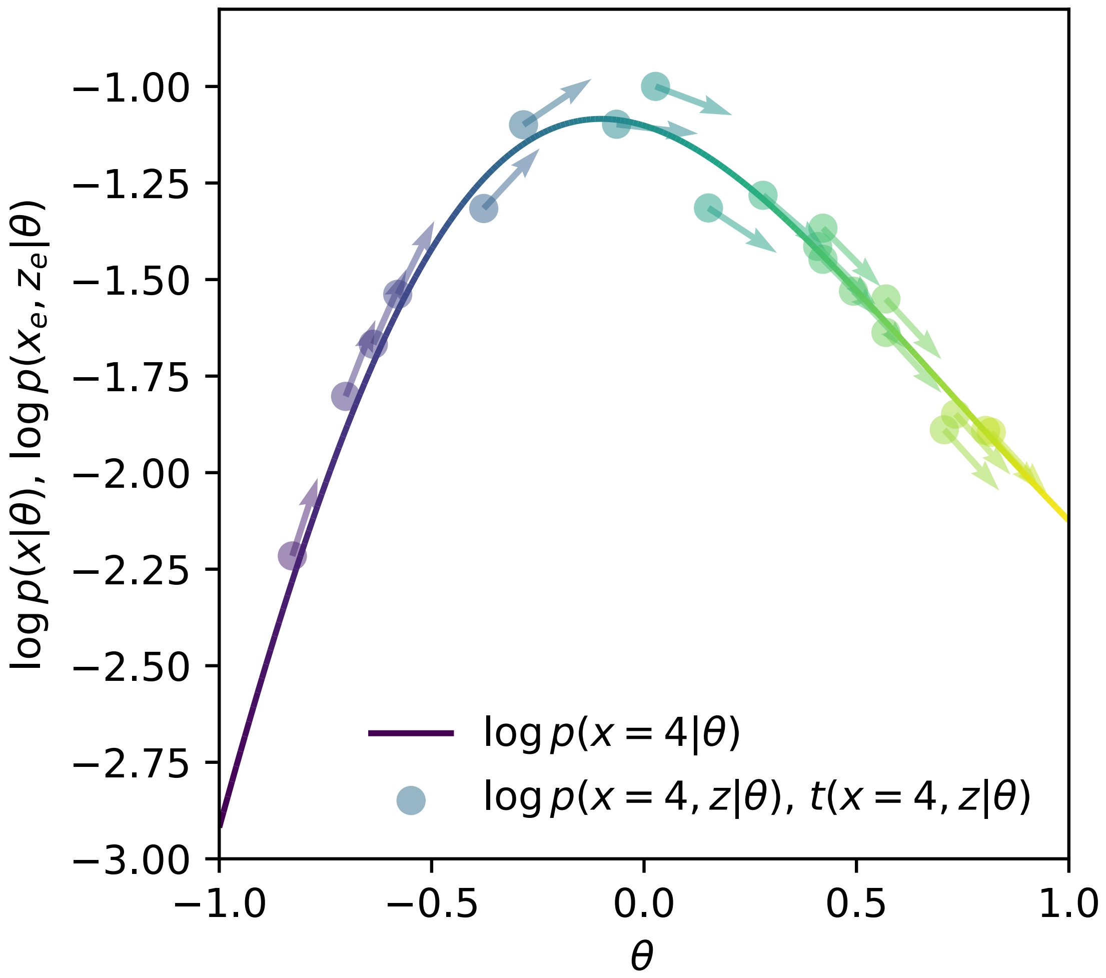
- Inference just based on the joint likelihood ratio works well, but there is another powerful piece of information

- 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 $\theta_0$

- The score itself is intractable. But…



# Learning the score

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

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

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

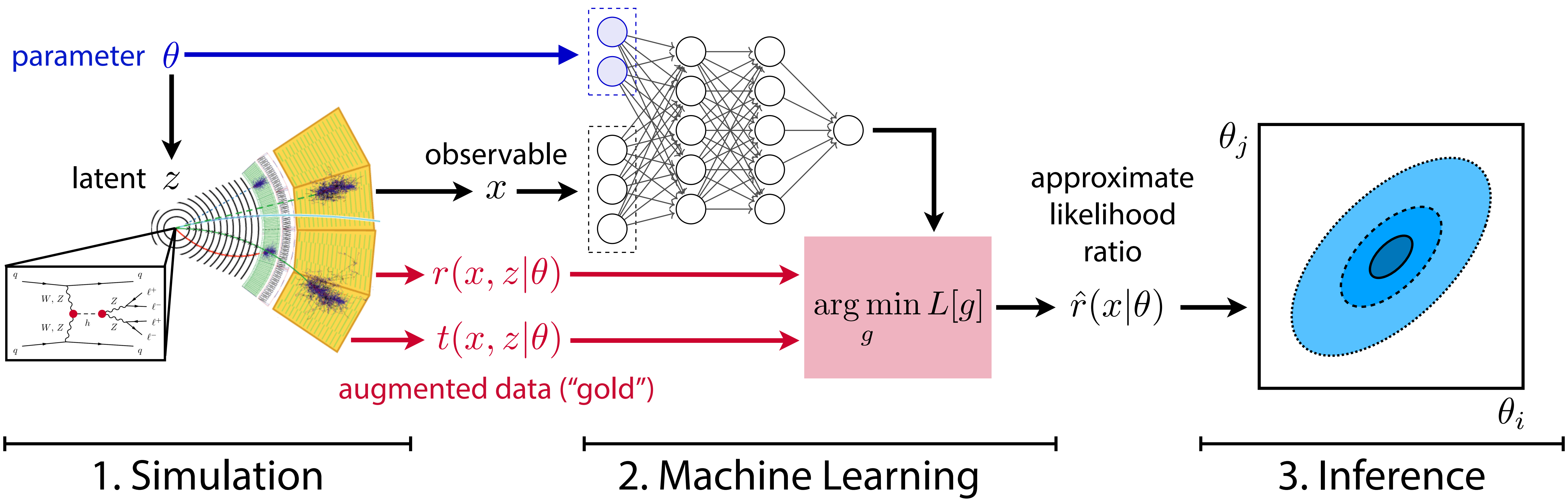
$$L_t[\hat{t}(x)] = \mathbb{E}_{p(x,z|\theta_0)}\left[\left(\hat{t}(x) - t(x,z|\theta_0)\right)^2\right].$$

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 | $L_{\mathrm{XE}}$ | $L_{\mathrm{MLE}}$ | $L_r$ | $L_t$ | $\theta$ 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 and **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

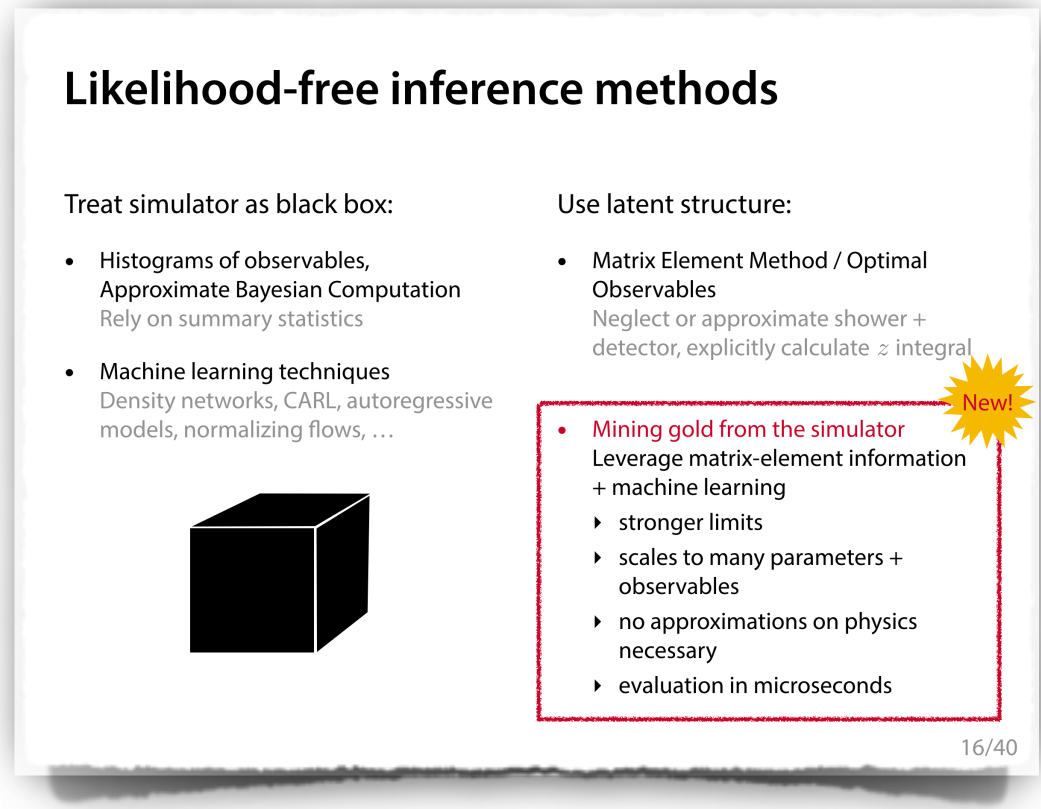
Different strategies to combine the different pieces of information:

| Method | $L_{\mathrm{XE}}$ | $L_{\mathrm{MLE}}$ | $L_r$ | $L_t$ | $\theta$ 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:

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

# Comparison to established methods

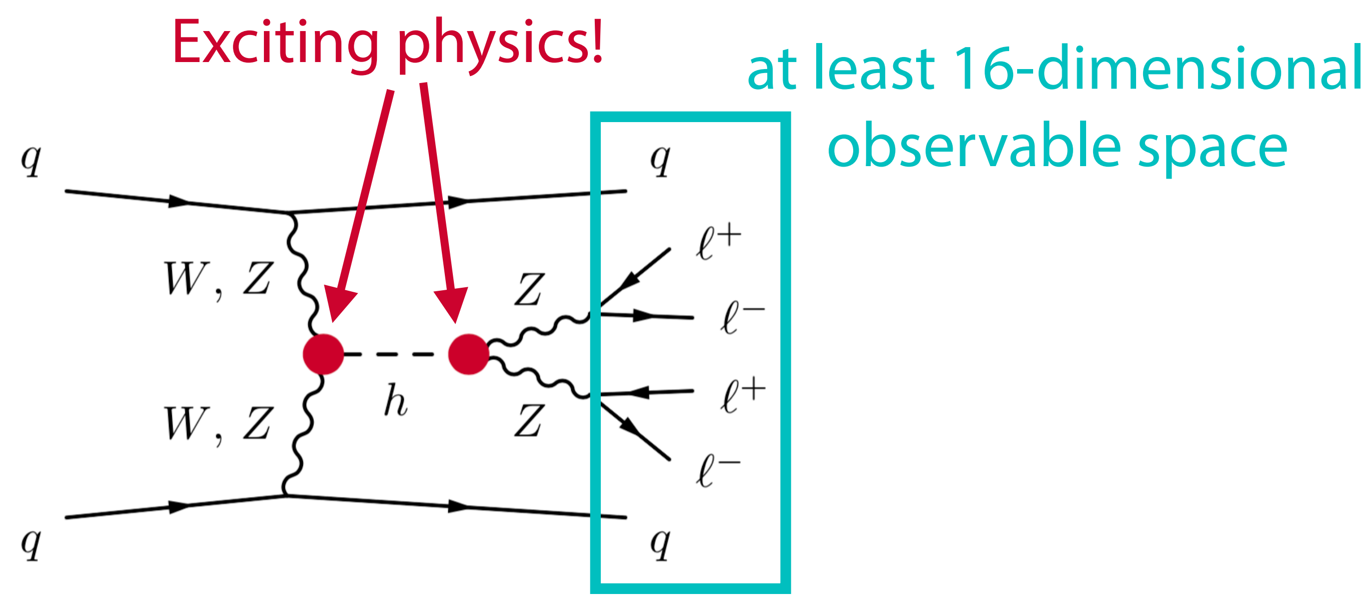



| | Histograms | NDE | MEM / OO | RASCAL etc |
|---|---|---|---|---|
| High-dimensional observables | | ✓ | ✓ | ✓ |
| Realistic shower, detector sim. | ✓ | ✓ | | ✓ |
| Asymptotically exact | | ✓ | | ✓ |
| Uses matrix element information | | | ✓ | ✓ |
| Evaluation | fast | fast | expensive | fast |

# Proof of concept

# Proof of concept

- Higgs production in weak boson fusion:

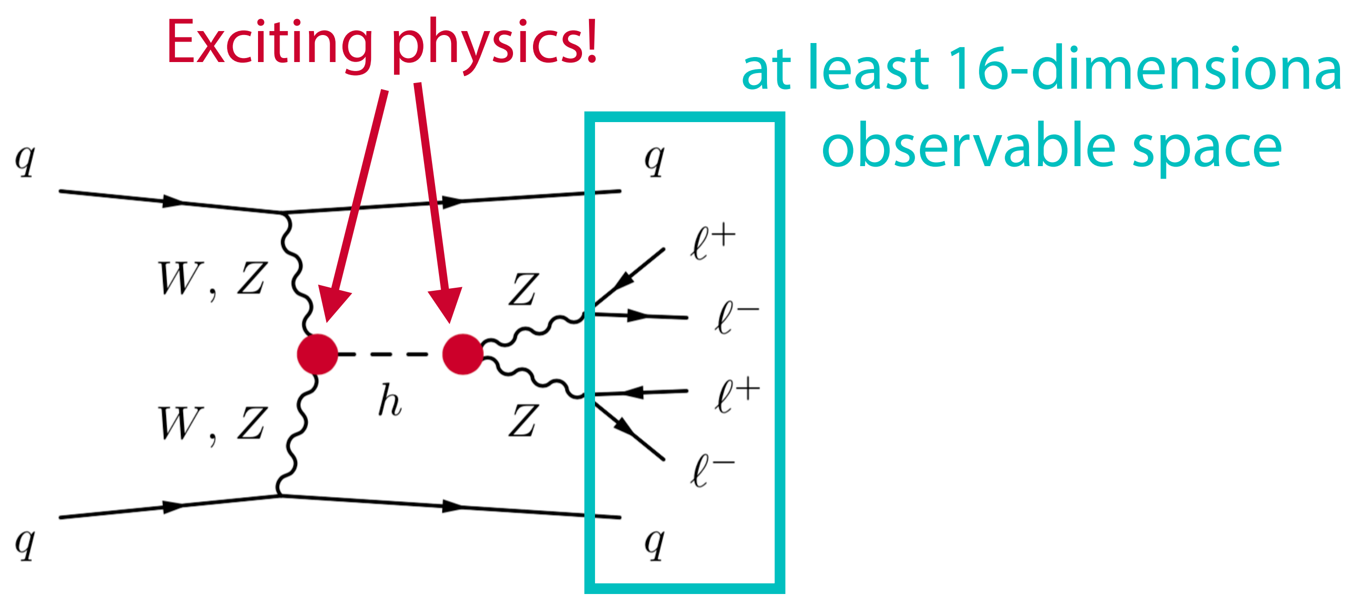



- Goal: constraints on two EFT parameters

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

# Proof of concept

- Higgs production in weak boson fusion:




- Goal: constraints on two EFT parameters

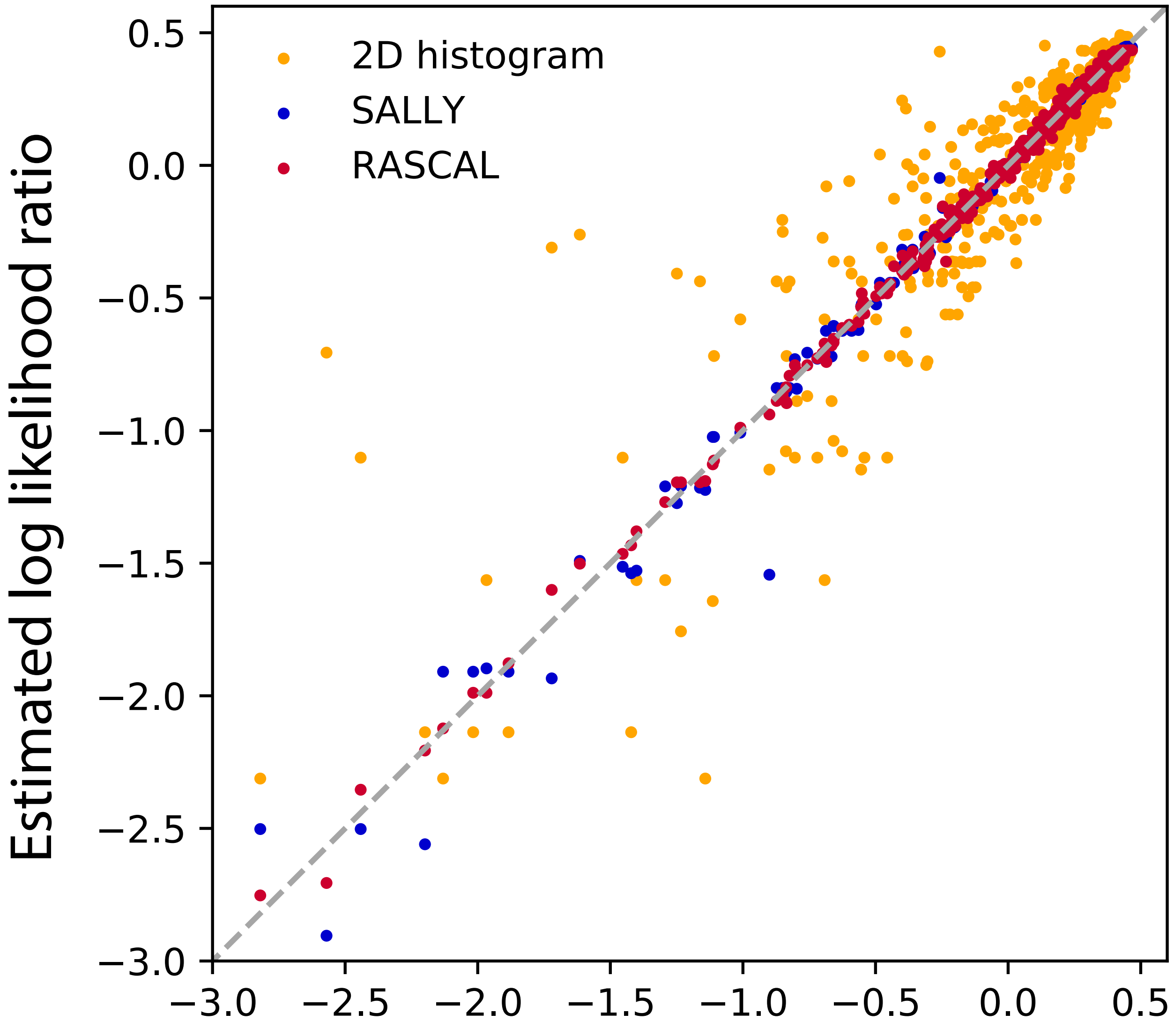
$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \boxed{\frac{f_W}{\Lambda^2}} \underbrace{\frac{ig}{2}\,(D^\mu\phi)^\dagger\,\sigma^a\,D^\nu\phi\,W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}} \underbrace{\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^a_{\mu\nu}\,W^{\mu\nu\,a}}_{\mathcal{O}_{WW}}$$

- 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]

# Precise likelihood ratio estimates

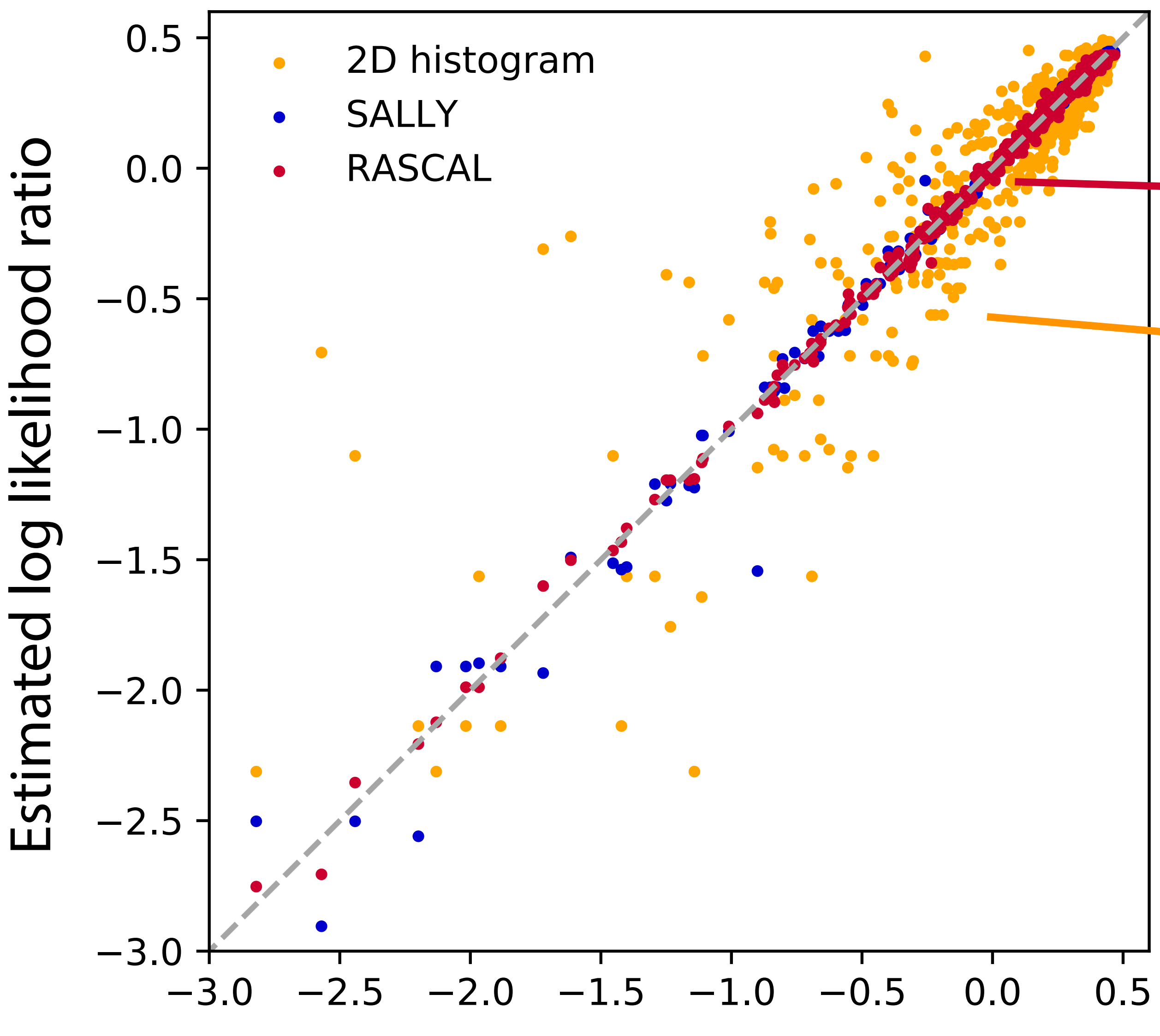
Each point = 1 event




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

# Precise likelihood ratio estimates

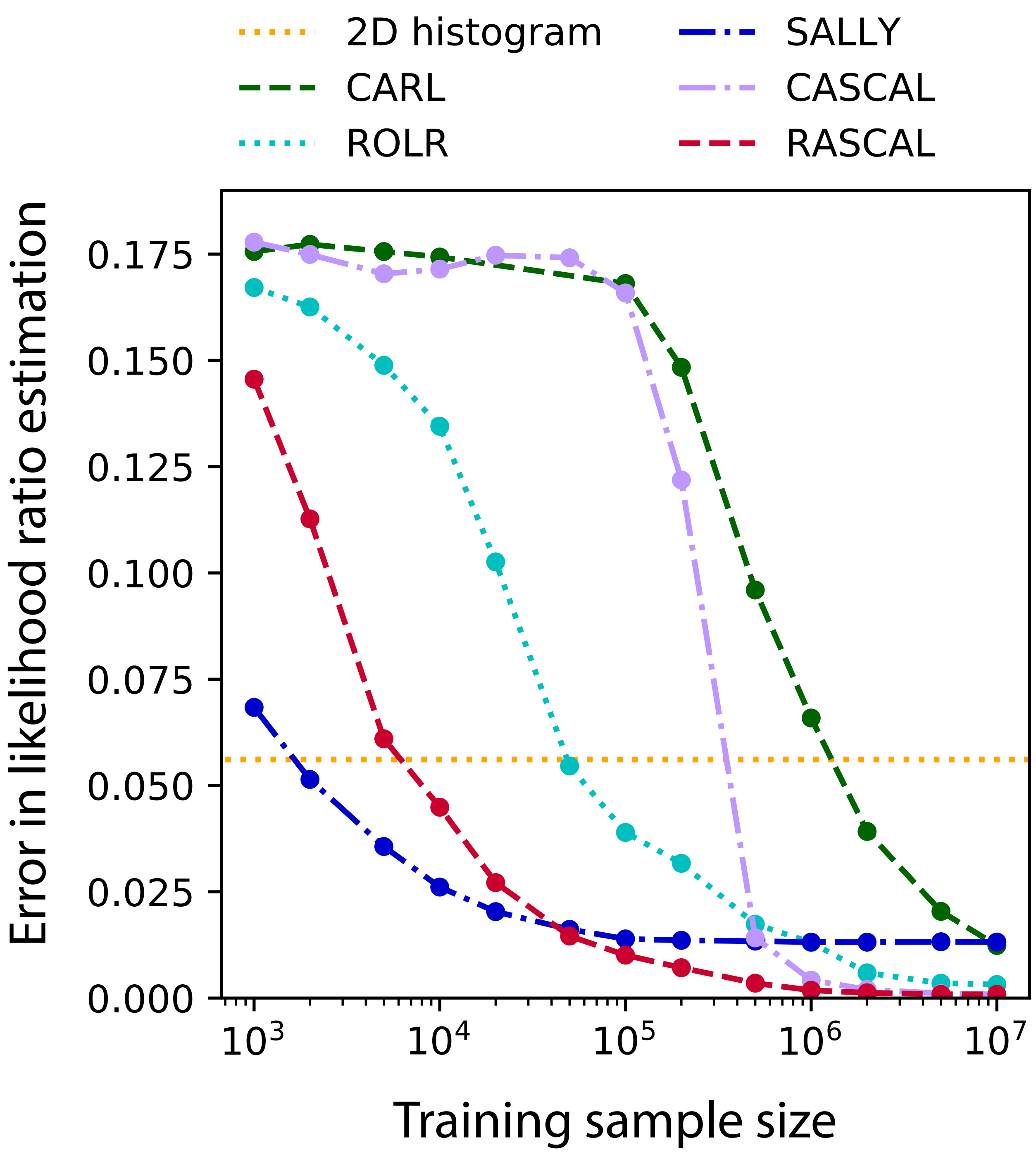
Each point = 1 event



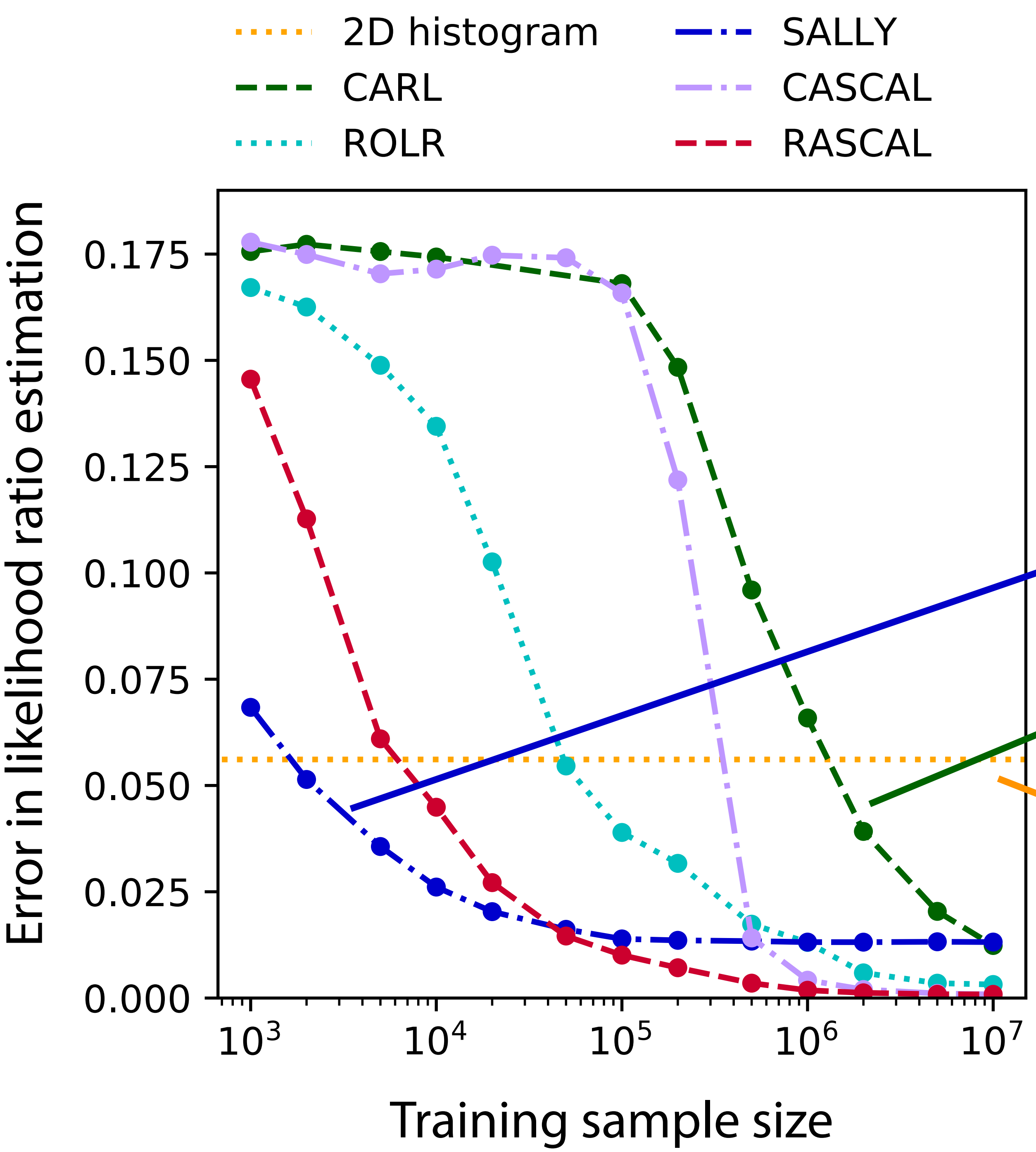

True log likelihood ratio
(usually don't have this)

RASCAL
much more precise than
traditional histogram

# Less training data needed



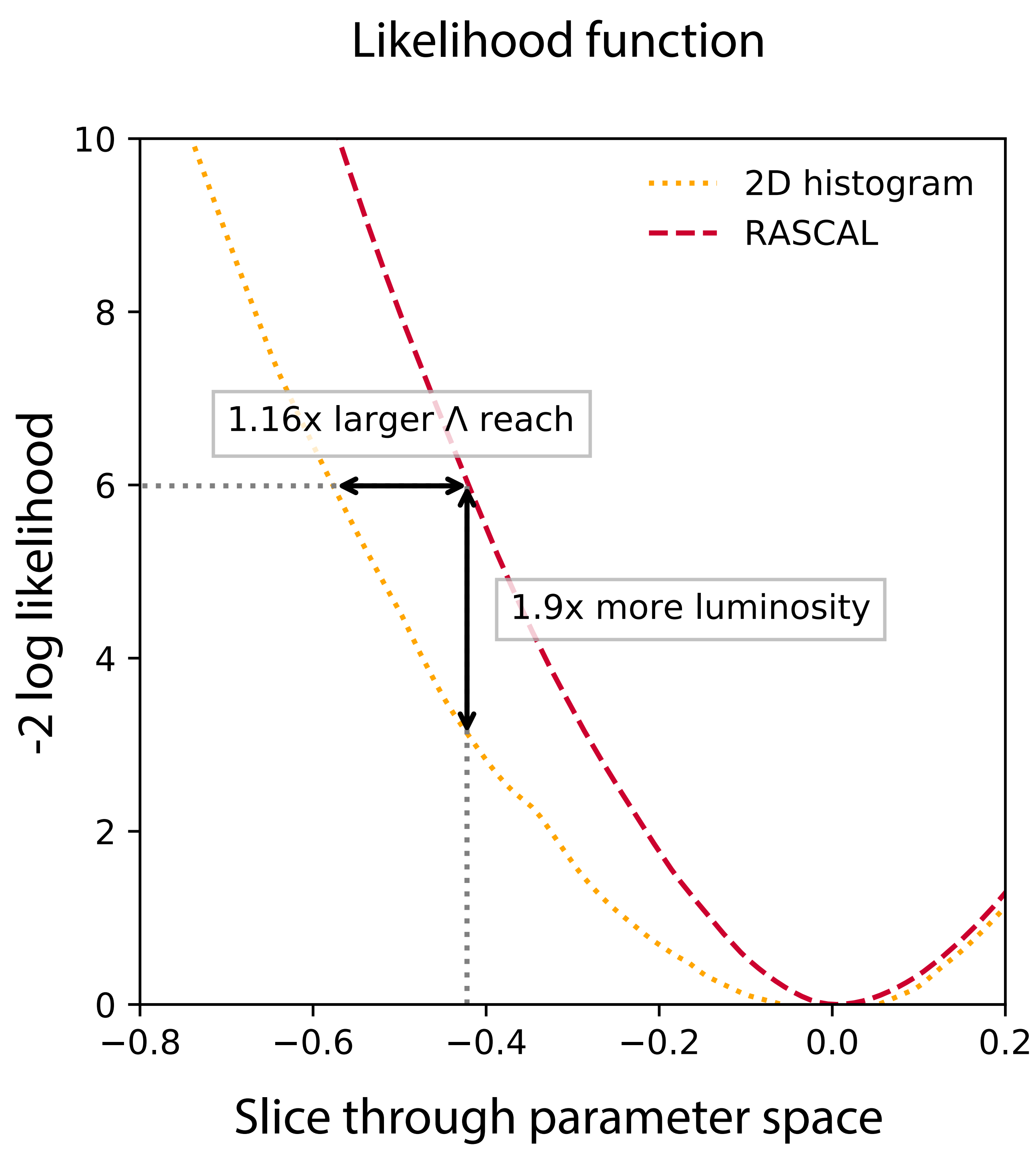
# Less training data needed




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

Traditional histogram

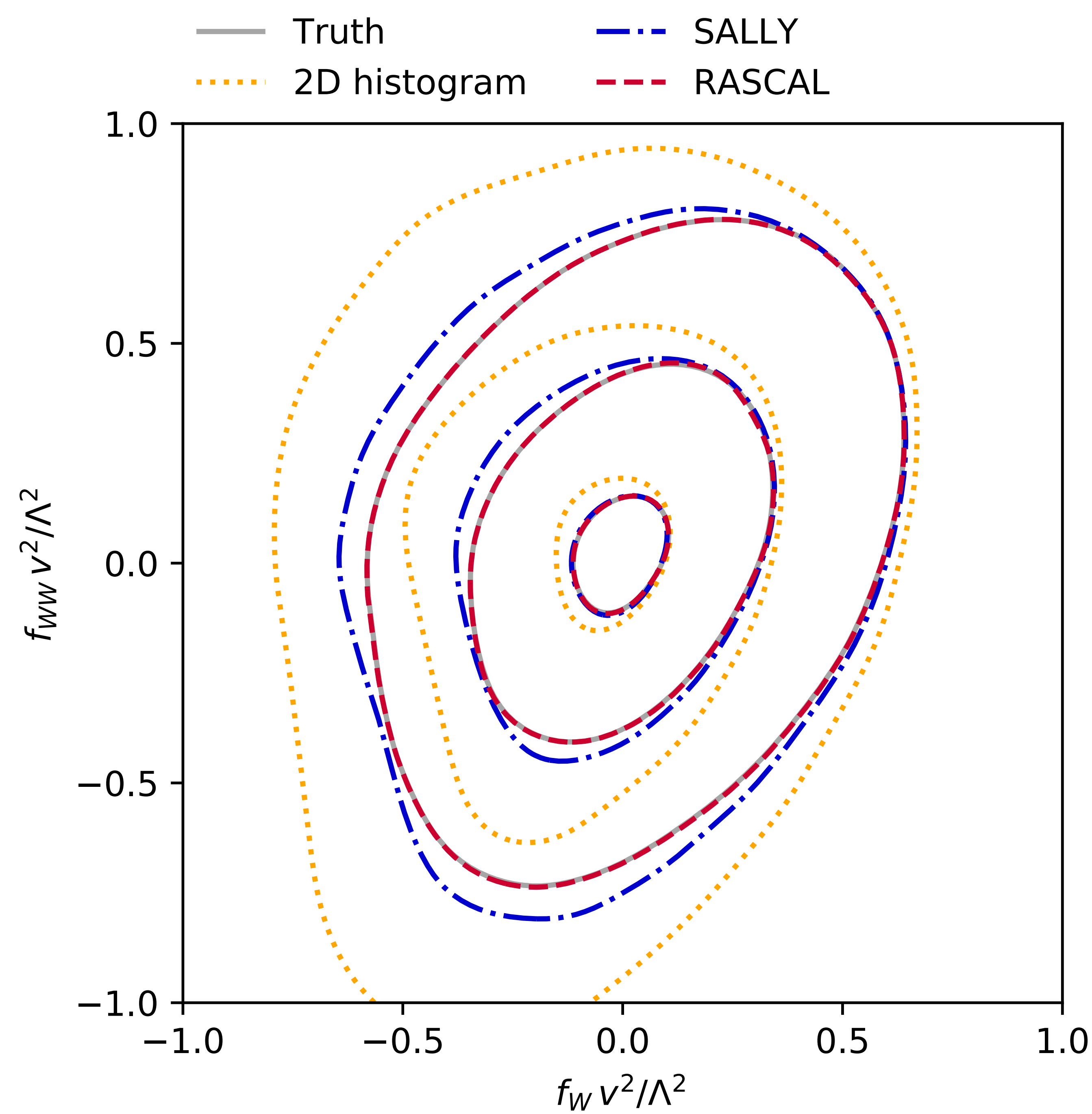
# Better sensitivity




36 events, assuming SM

# Stronger bounds

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

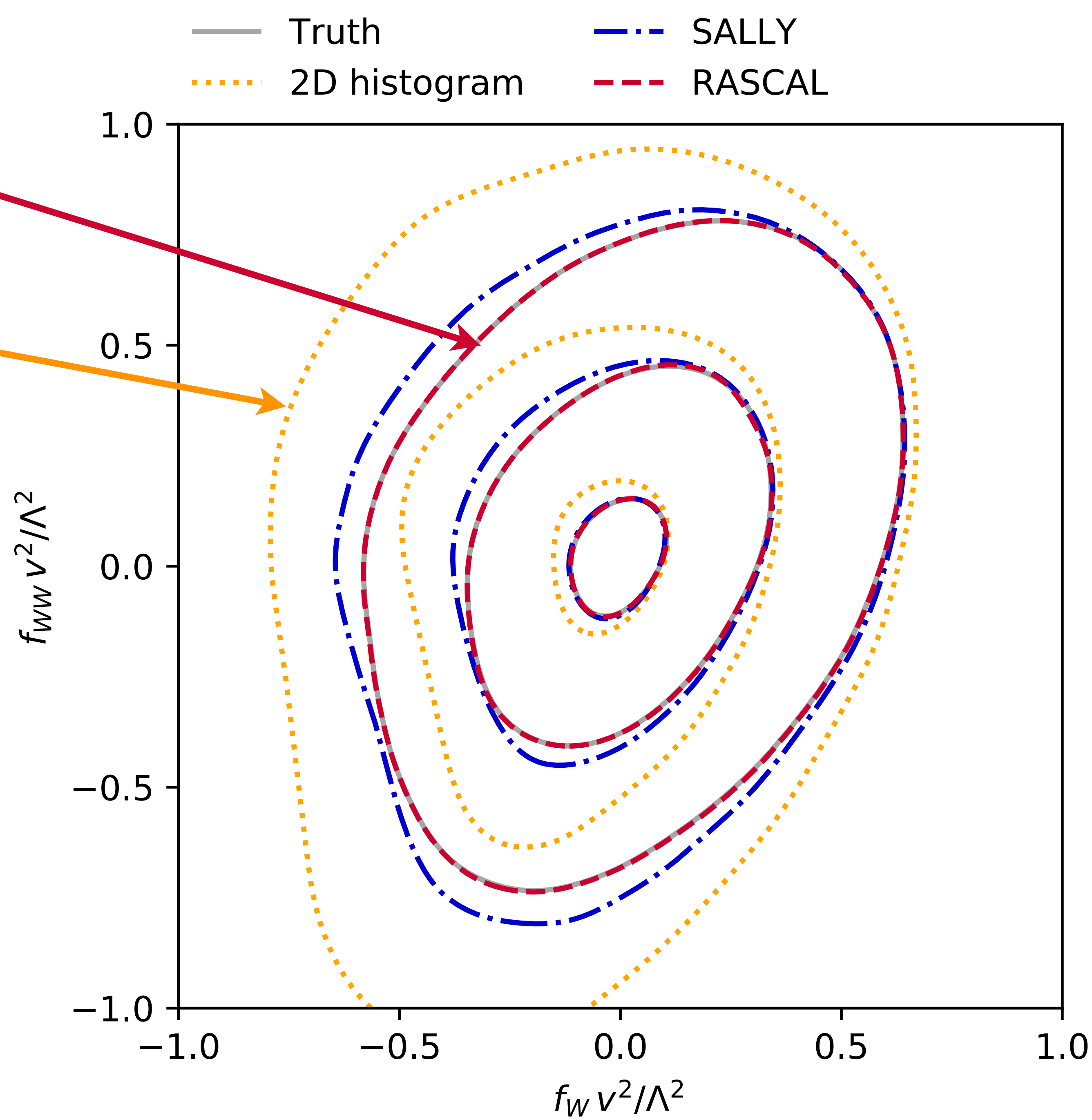



36 events, assuming SM

# Stronger bounds

RASCAL
enables stronger
limits than
traditional histogram

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




36 events, assuming SM

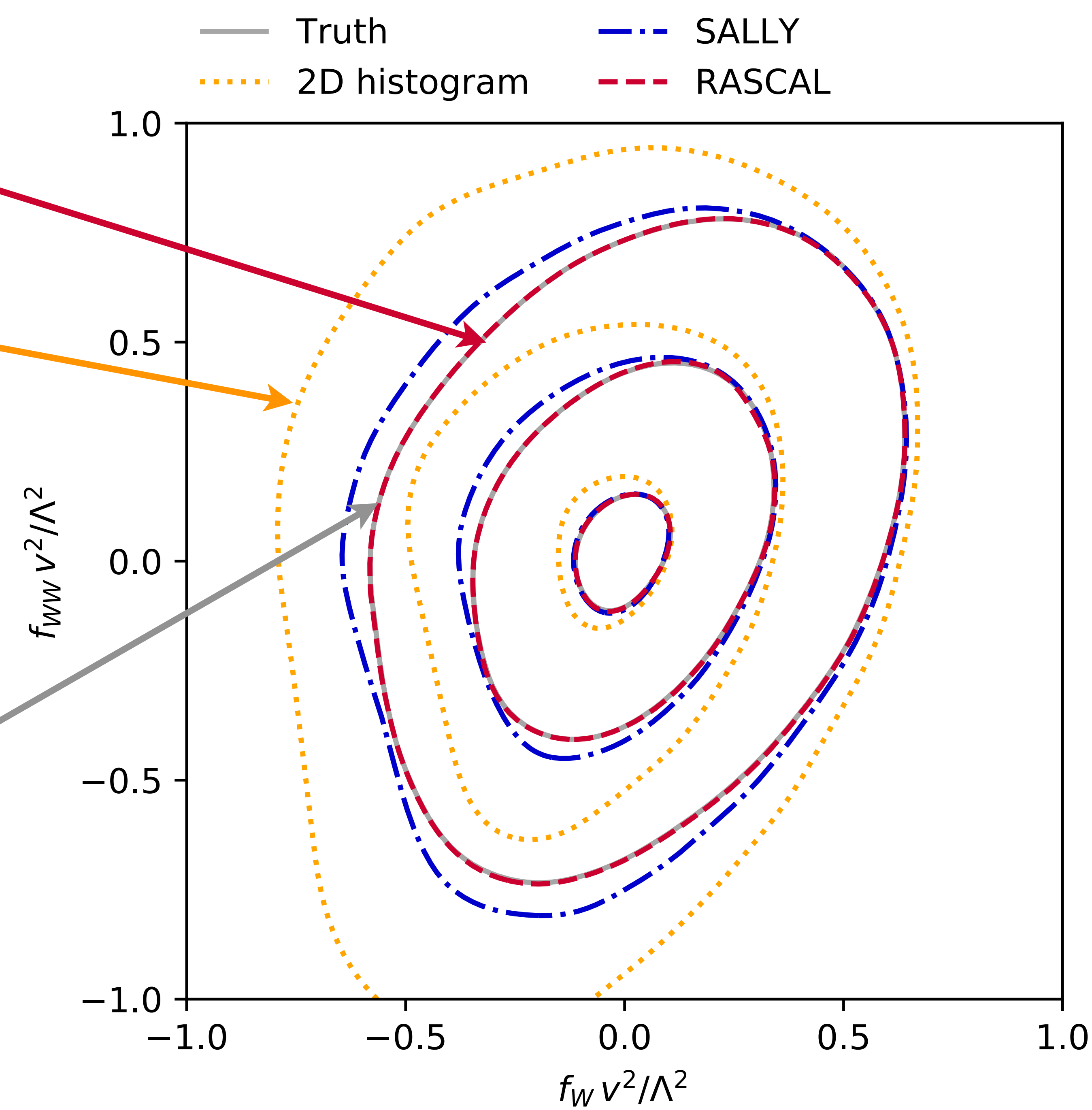
# Stronger bounds

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



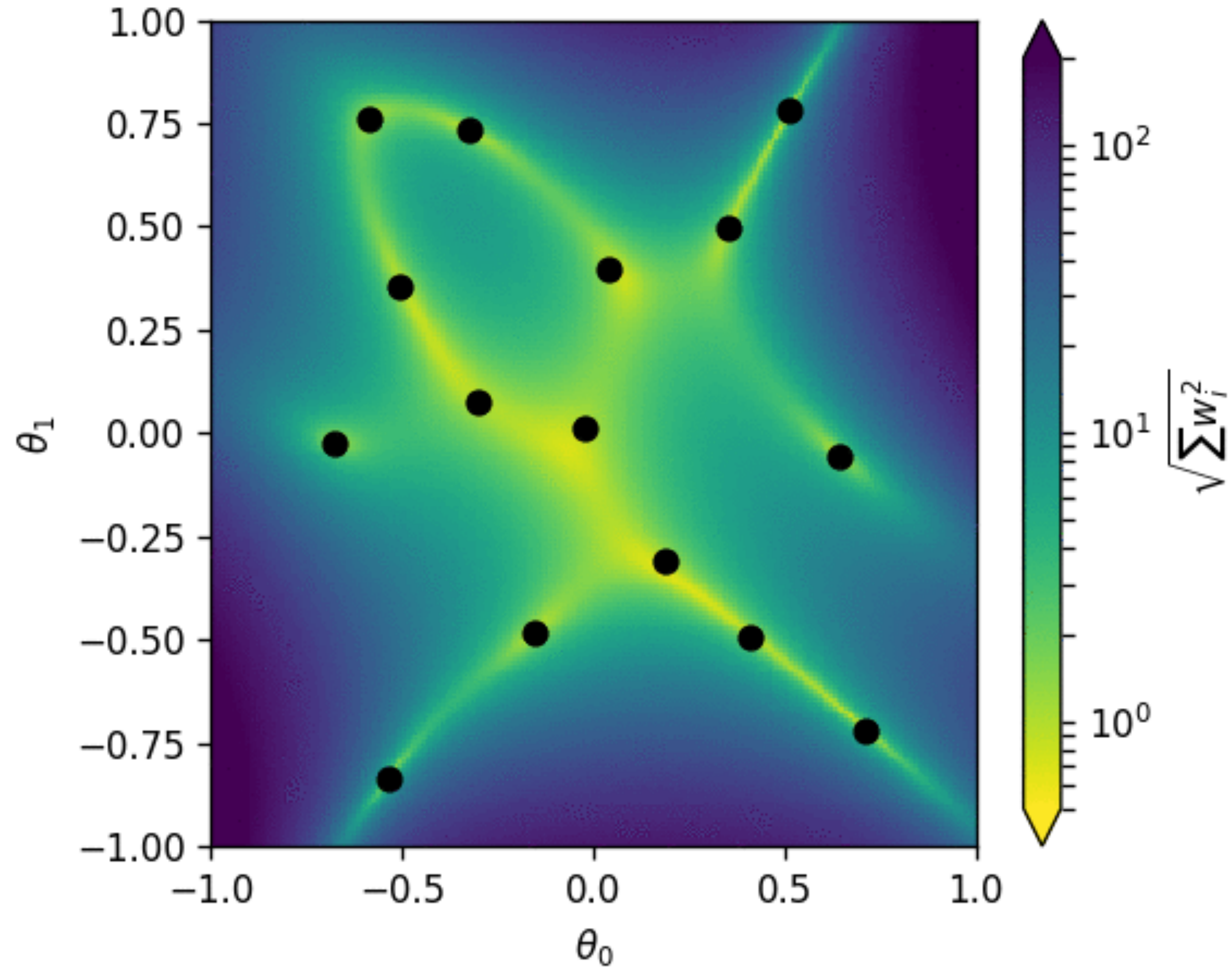

36 events, assuming SM

# MadMiner

# Can I use any of this?

Yes! To make that as painless as possible, we're working on the python package MadMiner:

- "Mining gold" from MadGraph + Pythia + detector simulation
- Morphing: reconstruct full dependence on model parameters from few MC runs
- Likelihood ratio estimation with RASCAL and friends
- Calculate Fisher information (truth or reco level)

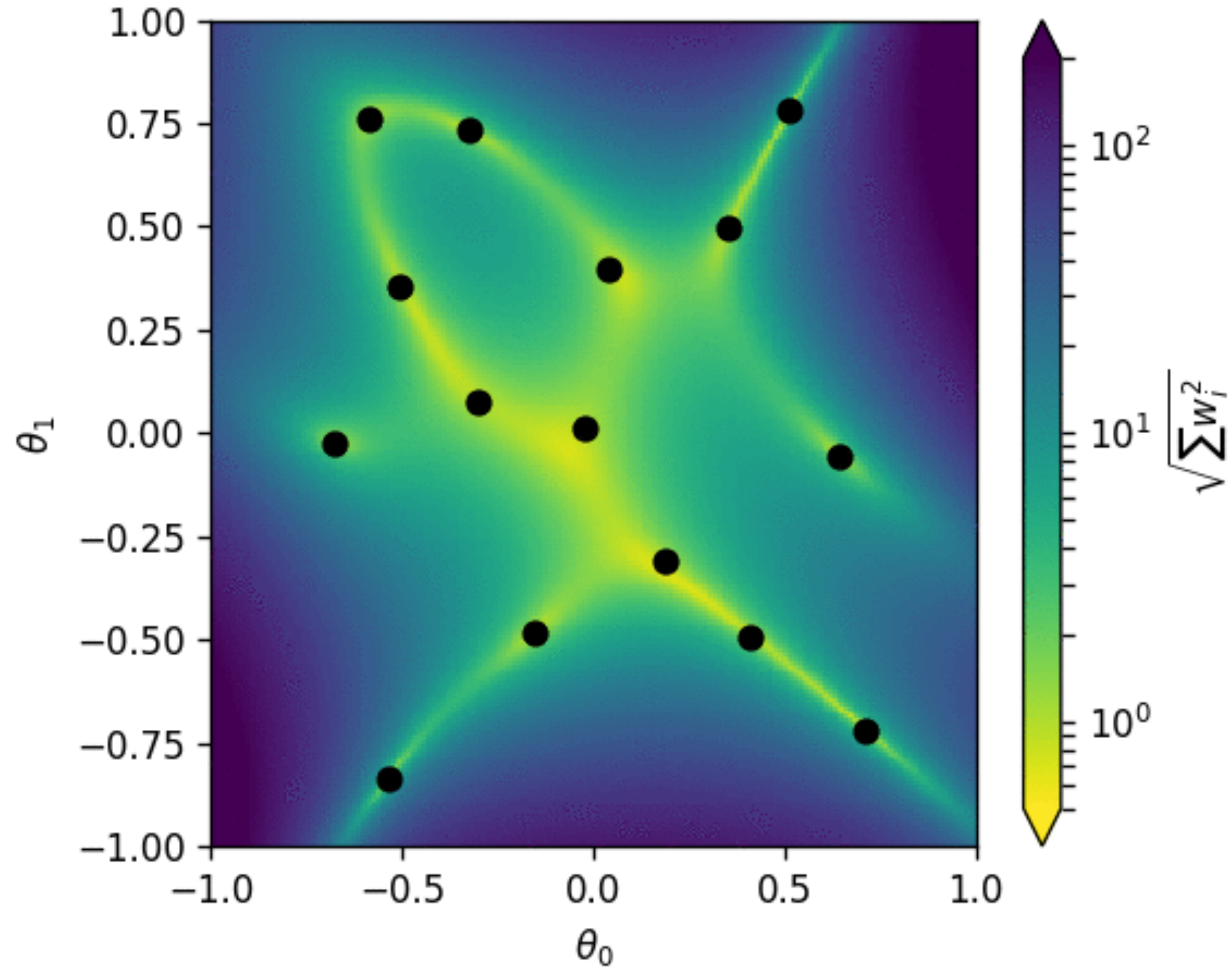



Come visit us at github.com/johannbrehmer/madminer!

# Can I use any of this?

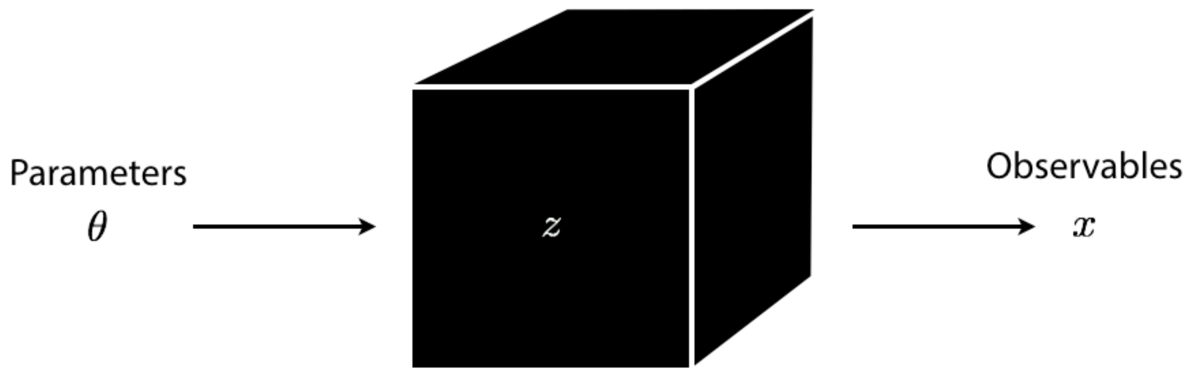
Yes! To make that as painless as possible, we're working on the python package MadMiner:

- "Mining gold" from MadGraph + Pythia + detector simulation
- Morphing: reconstruct full dependence on model parameters from few MC runs
- Likelihood ratio estimation with RASCAL and friends
- Calculate Fisher information (truth or reco level)



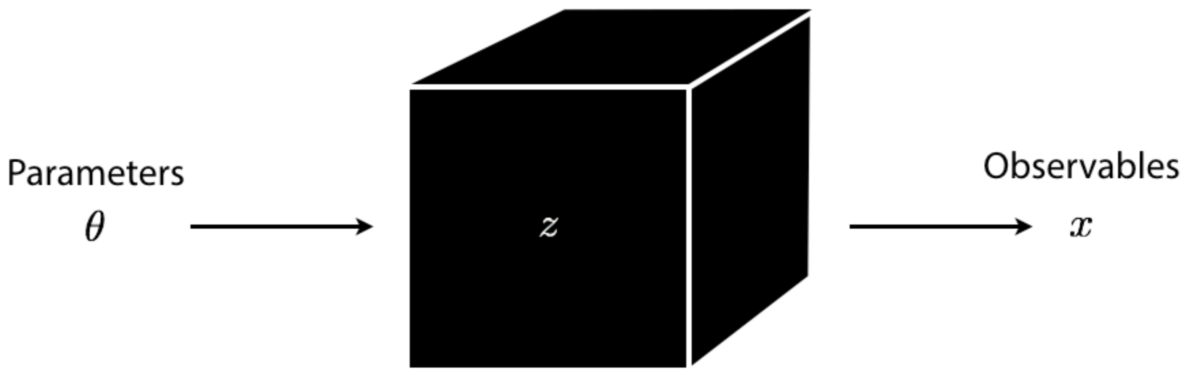

Come visit us at github.com/johannbrehmer/madminer!

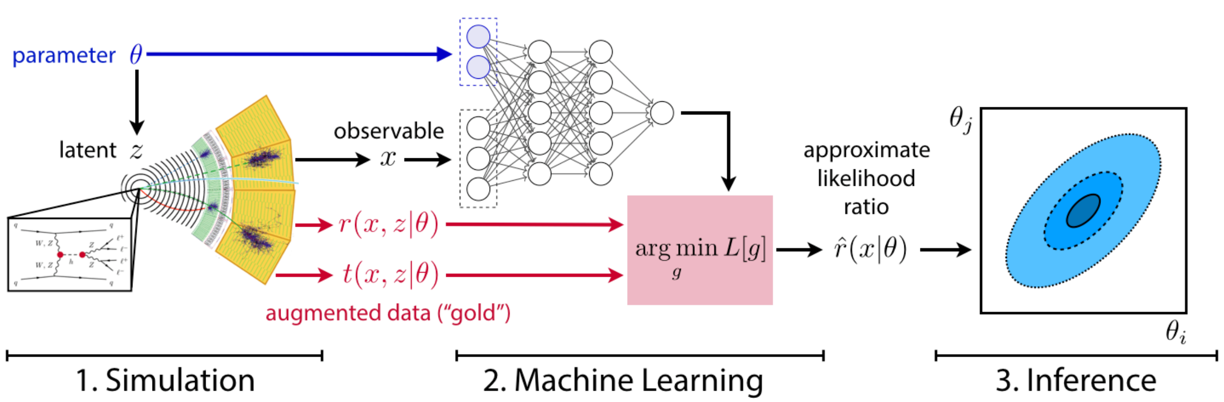
# A new approach to simulator-based inference




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

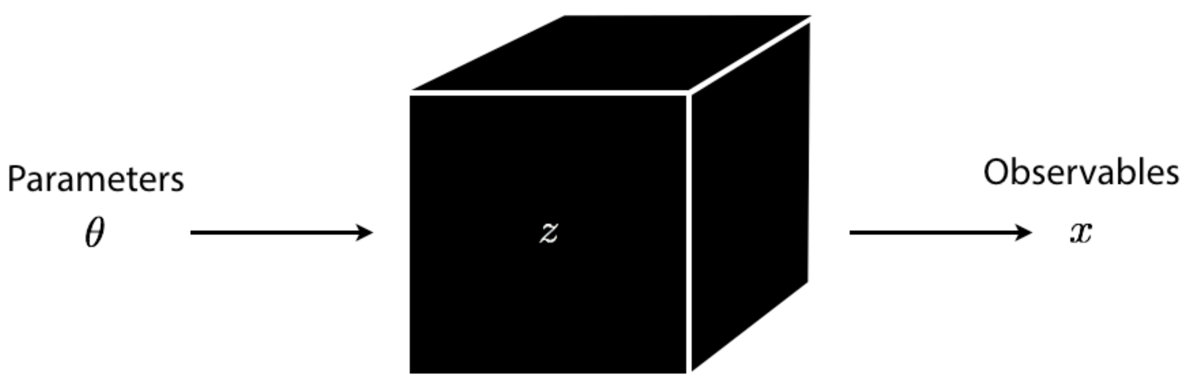
# A new approach to simulator-based inference

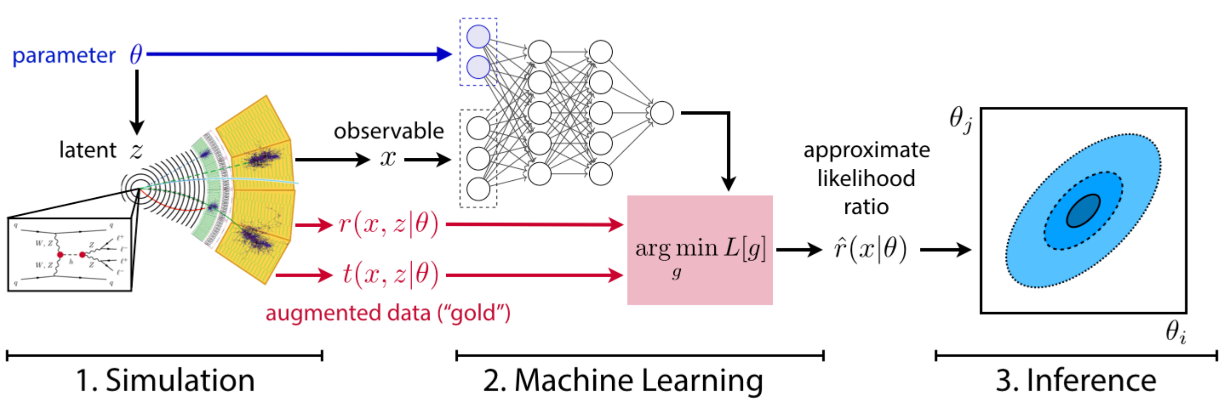


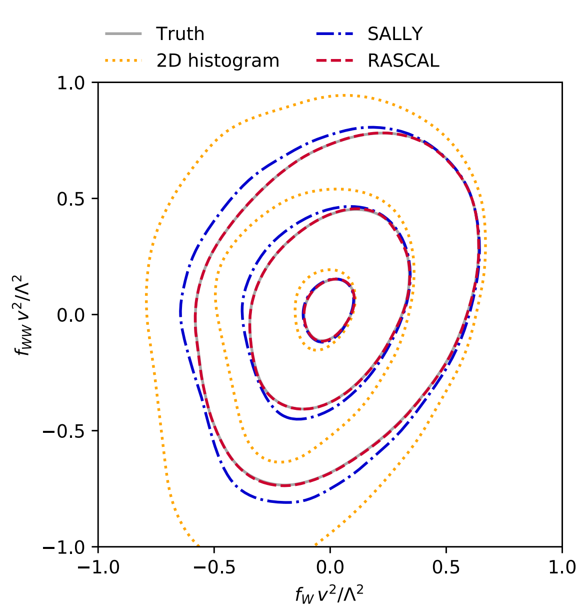




- Many LHC analyses (and much of modern science) are based on simulations, "likelihood-free"
- Established inference methods treat simulator as black box
- New inference techniques: Leverage more information from simulator + power of machine learning

# 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
- The MadMiner package will make this easy to use for everyone

# References

Kyle Cranmer

Gilles Louppe

Juan Pavez

Felix Kling

Markus Stoye

| | | |
|---|---|---|
| JB, KC, GL, JP: | Constraining Effective Field Theories with Machine Learning | [1805.00013] |
| JB, KC, GL, JP: | A Guide to Constraining Effective Field Theories with Machine Learning | [1805.00020] |
| JB, GL, JP, KC: | Mining gold from implicit models to improve likelihood-free inference | [1805.12244] |
| MS, JB, GL, JP, KC: | Likelihood-free inference with an improved cross-entropy estimator | [1808.00973] |
| JB, KC, FK: | MadMiner | In preparation |

Thanks to Kyle and Gilles for inspiring many of my slides!

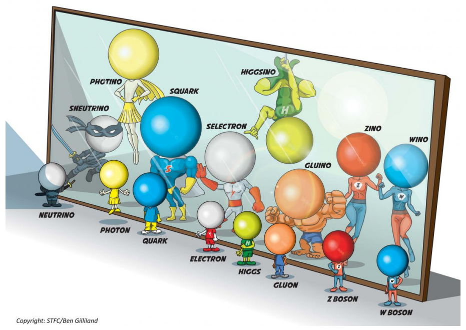
# Bonus material

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

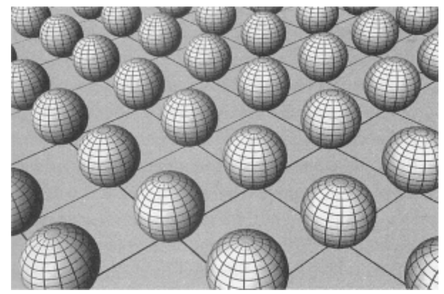
# Effective field theory

Energy

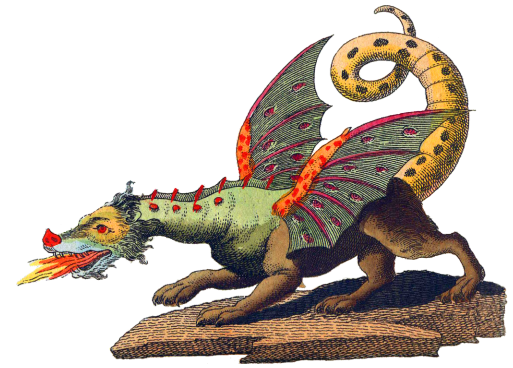
$\Lambda$



Supersymmetry?



Extra dimensions?



Dragons?


Composite Higgs?

…

Matching

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

Effective field theory

Measurements

$E \ll \Lambda$

LHC Higgs

Other LHC measurements

LEP

Flavour

…

# SMEFT (Standard Model Effective Field Theory)

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

## 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

$$\mathcal{O}_{\phi,1} = (D_\mu \phi)^\dagger \phi\, \phi^\dagger D^\mu \phi \qquad \mathcal{O}_{GG} = (\phi^\dagger \phi)\, G^a_{\mu\nu}\, G^{\mu\nu\, a}$$

$$\mathcal{O}_{\phi,2} = \frac{1}{2}\, \partial_\mu (\phi^\dagger \phi)\, \partial^\mu (\phi^\dagger \phi) \qquad \mathcal{O}_{BB} = -\frac{g'^2}{4}\, (\phi^\dagger \phi)\, B_{\mu\nu}\, B^{\mu\nu}$$

$$\mathcal{O}_{\phi,3} = \frac{1}{3} (\phi^\dagger \phi)^3 \qquad \mathcal{O}_{WW} = -\frac{g^2}{4}\, (\phi^\dagger \phi)\, W^a_{\mu\nu}\, W^{\mu\nu\, a}$$

$$\mathcal{O}_{\phi,4} = (\phi^\dagger \phi)\, (D_\mu \phi)^\dagger D^\mu \phi \qquad \mathcal{O}_{BW} = -\frac{g\, g'}{4}\, (\phi^\dagger \sigma^a \phi)\, B_{\mu\nu}\, W^{\mu\nu\, a}$$

$$\mathcal{O}_B = \frac{\mathrm{i} g'}{2}\, (D^\mu \phi)^\dagger D^\nu \phi\, B_{\mu\nu}$$

$$\mathcal{O}_W = \frac{\mathrm{i} g}{2}\, (D^\mu \phi)^\dagger \sigma^a D^\nu \phi\, W^a_{\mu\nu}$$

[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 $E^2 / \Lambda^2$

**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

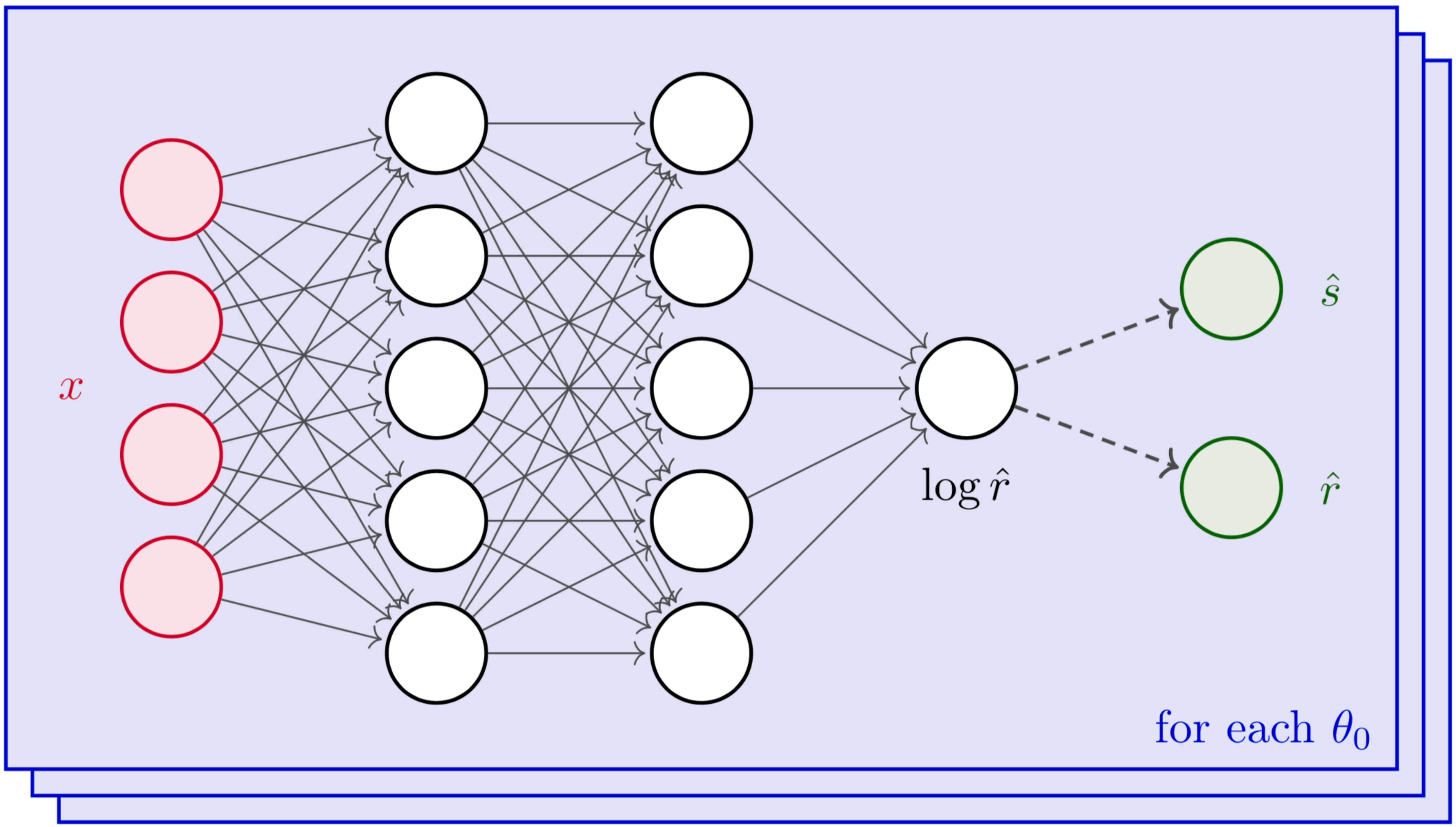
$$L[\hat{g}(x)] = \int \mathrm{d}x\, \mathrm{d}z\; p(x,z|\theta)\; |g(x,z) - \hat{g}(x)|^2$$

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

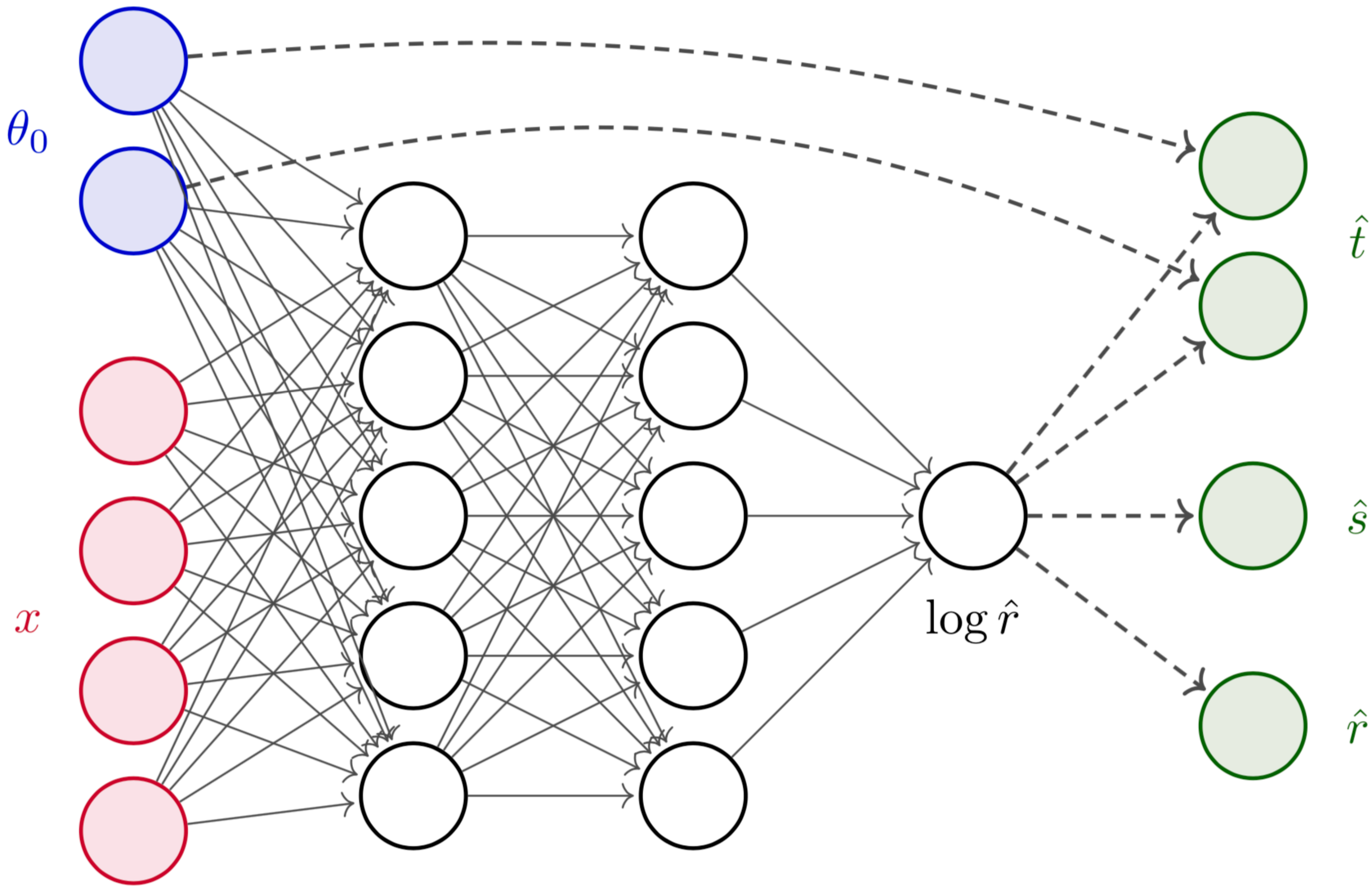
$$0 = \left.\frac{\delta F}{\delta \hat{g}}\right|_{g^*} = 2\hat{g}\; \underbrace{\int \mathrm{d}z\; p(x,z|\theta)}_{=p(x|\theta)} - 2 \int \mathrm{d}z\; p(x,z|\theta)\, g(x,z)$$

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

# Point by point

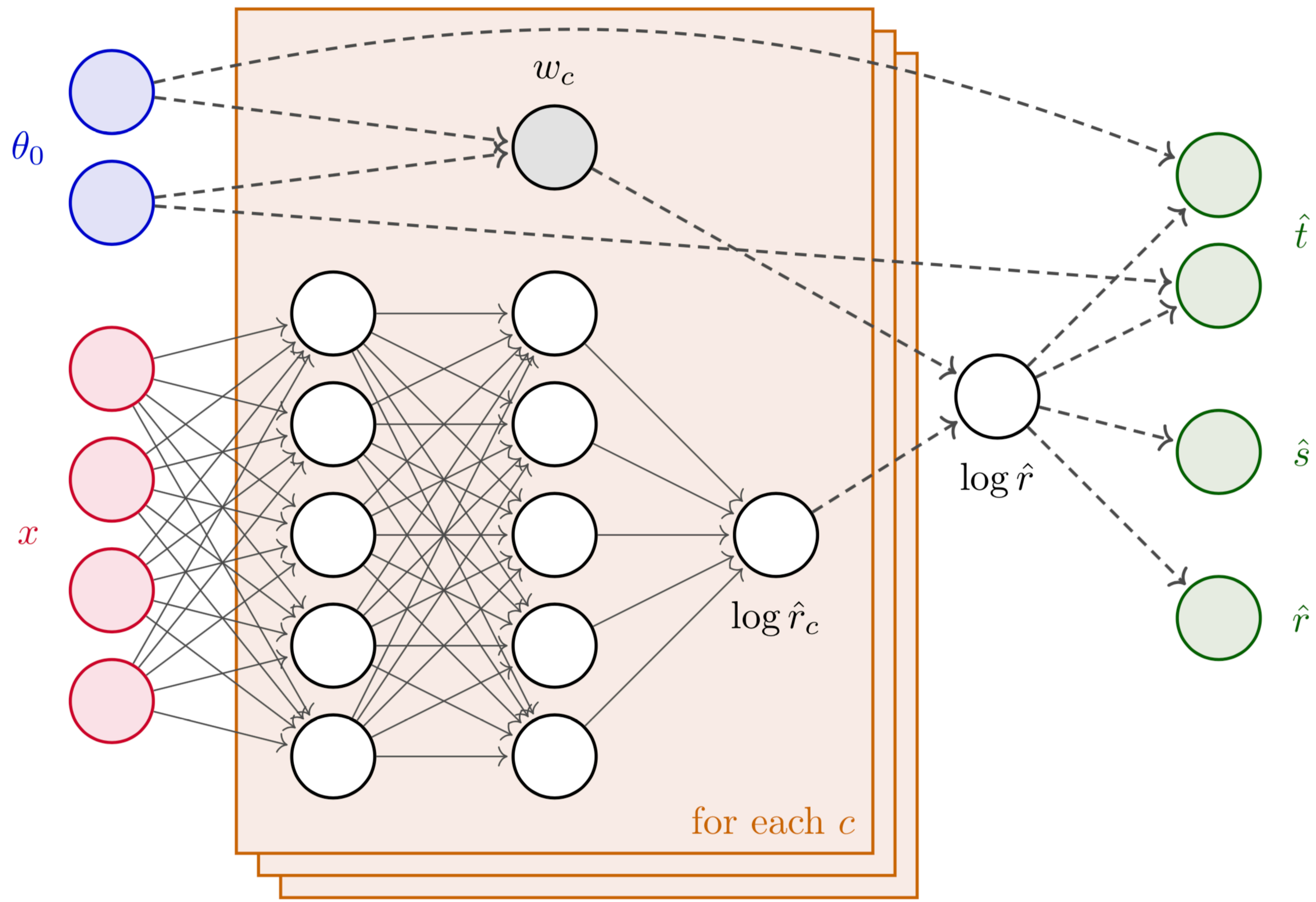


# (Agnostic) parameterized estimators

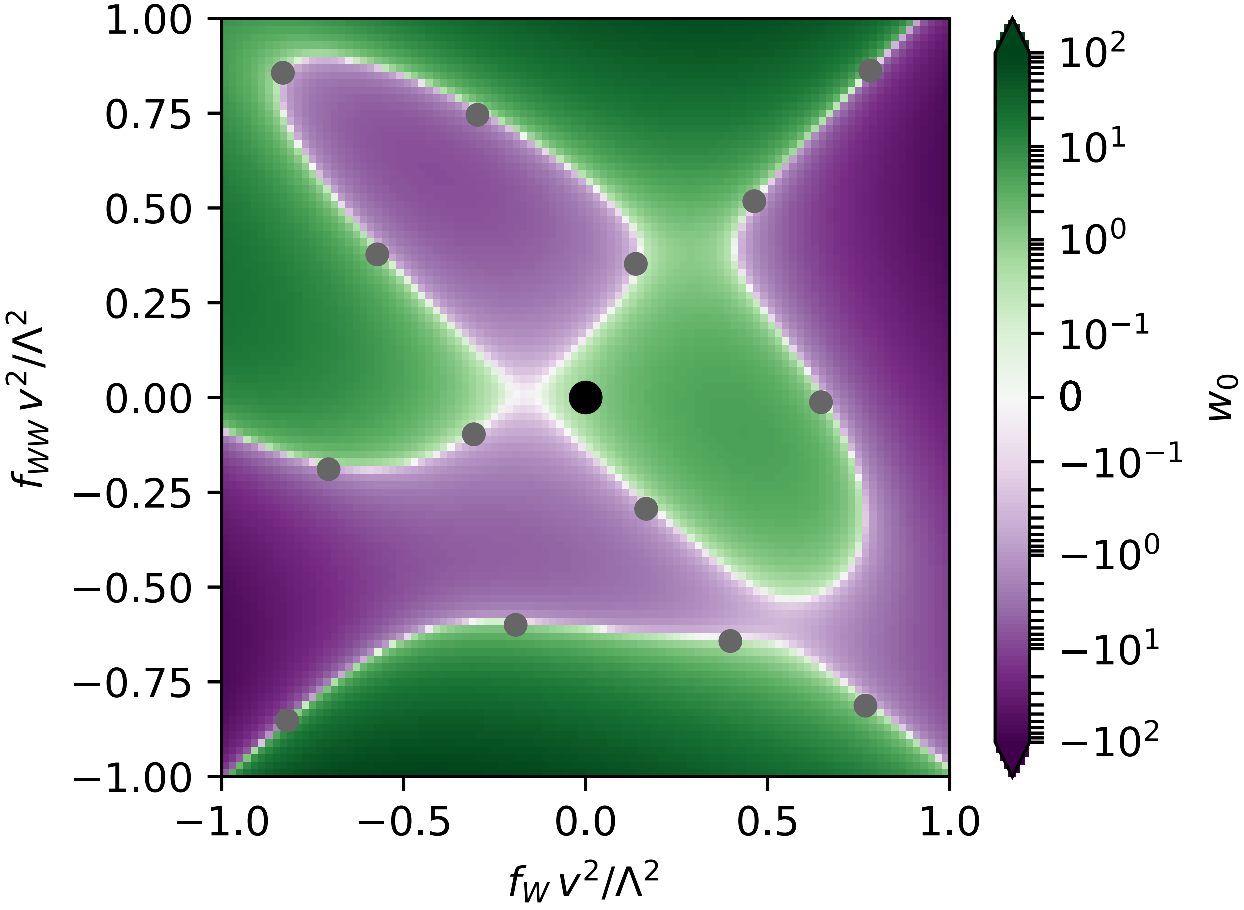


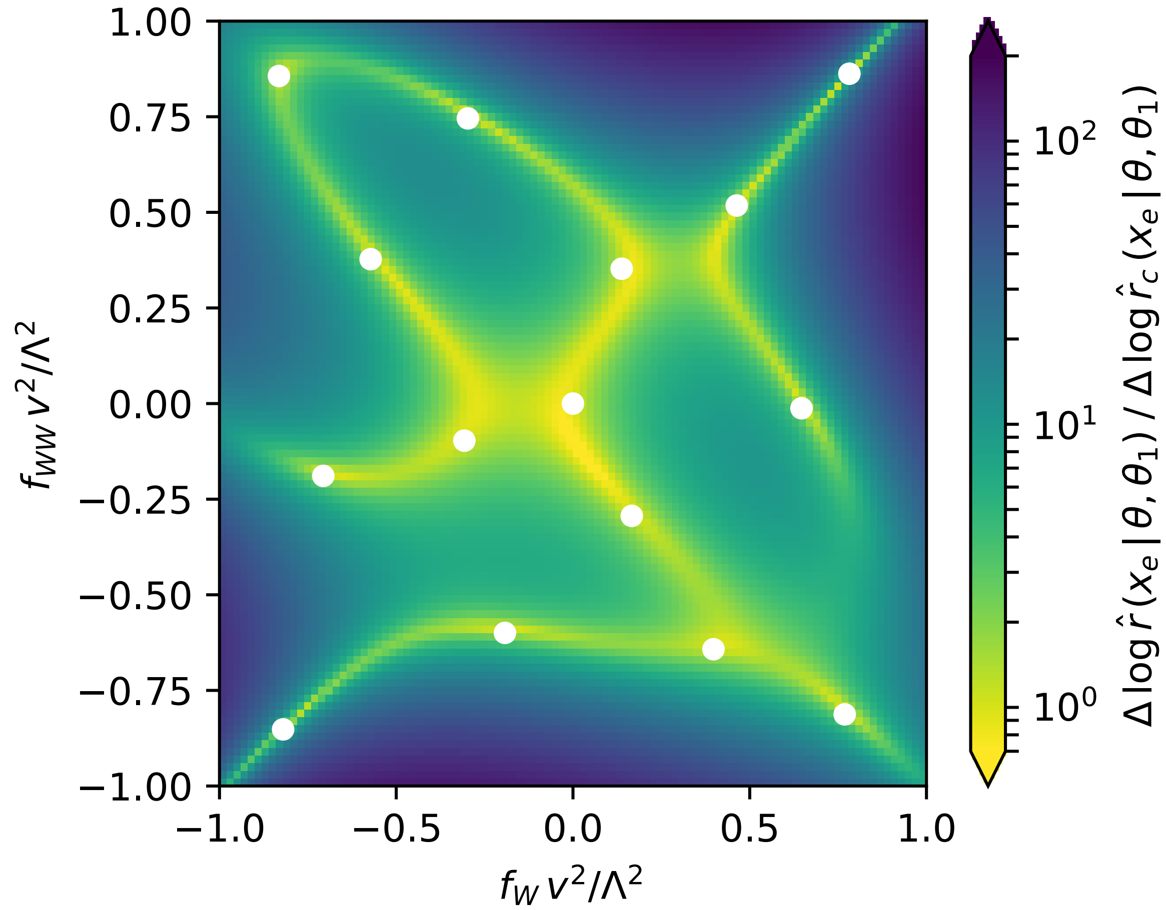
# Morphing-aware parameterized estimators

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

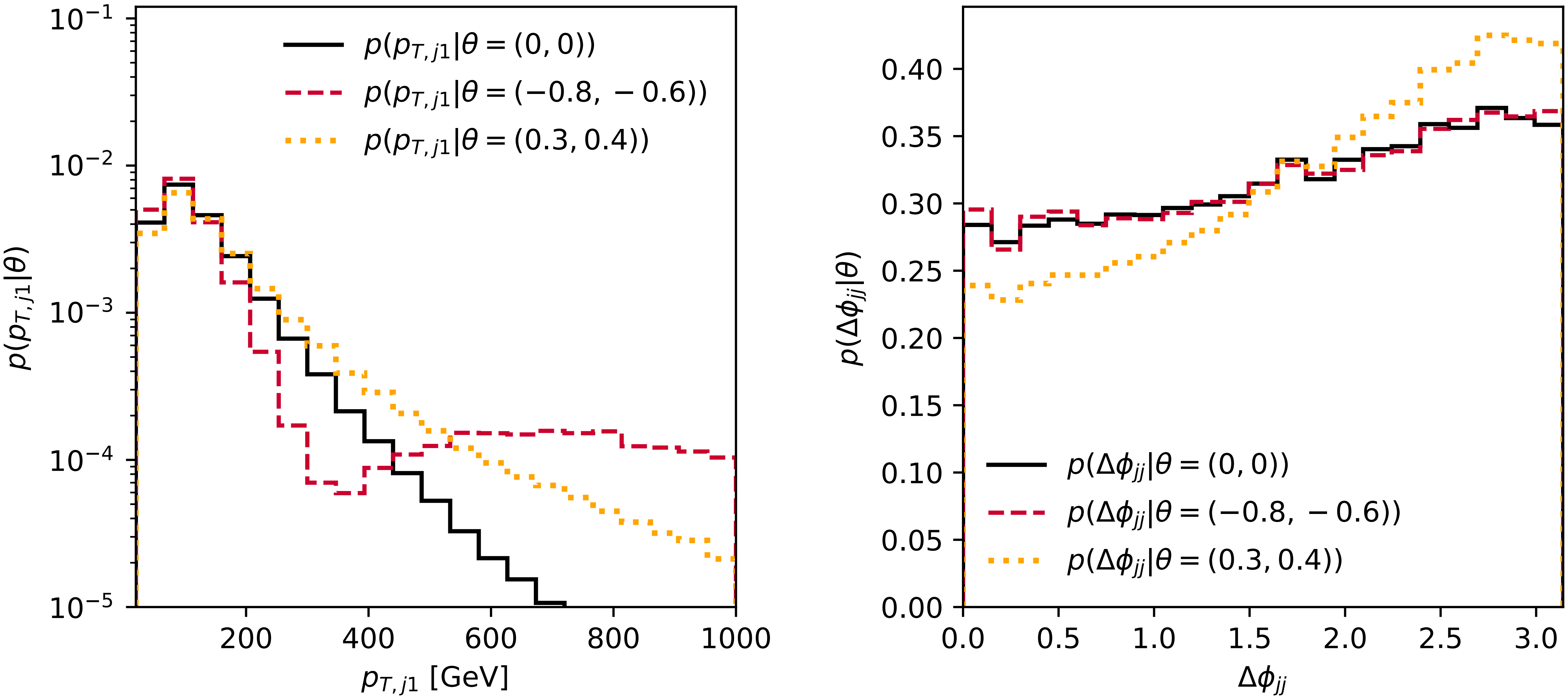


# Morphing coefficients

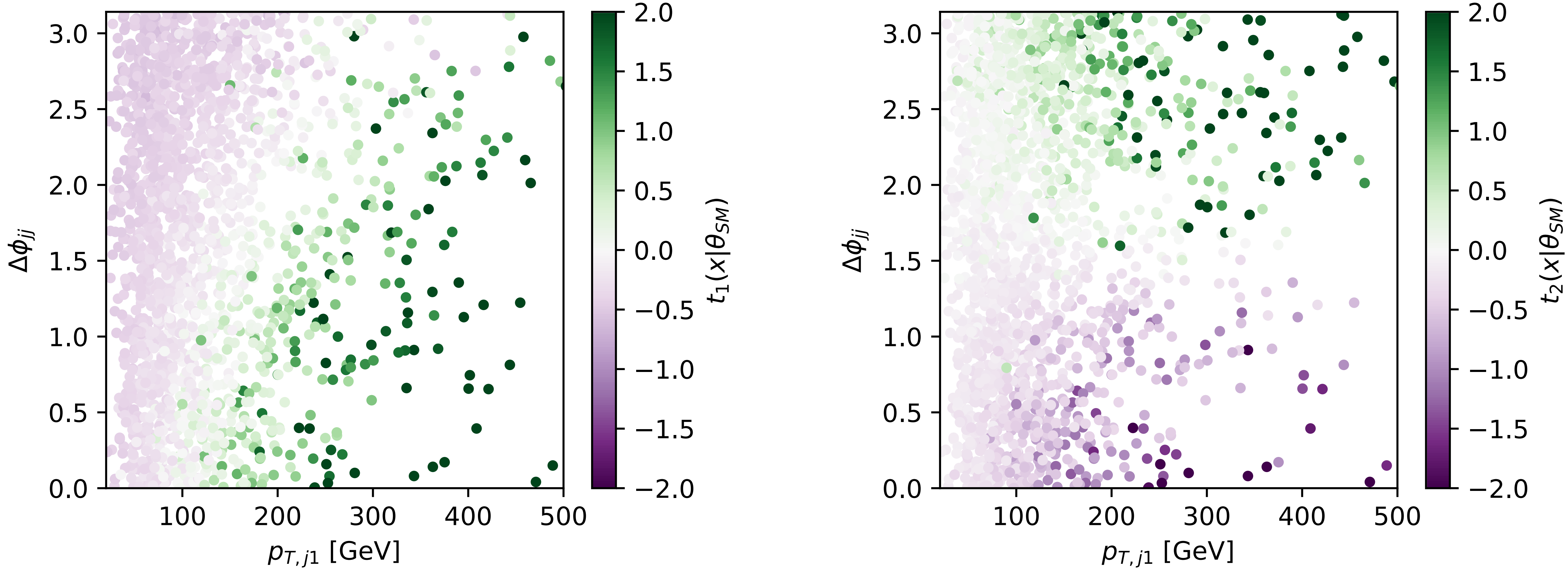


# Distributions



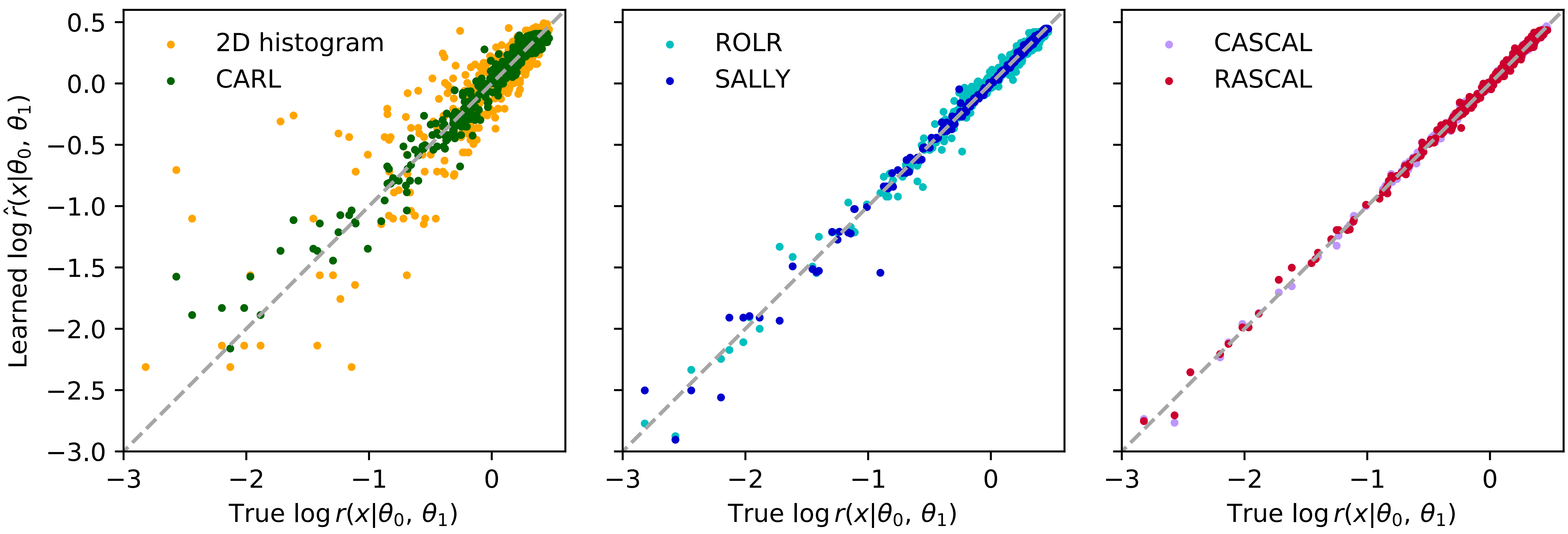
# Score vs observables



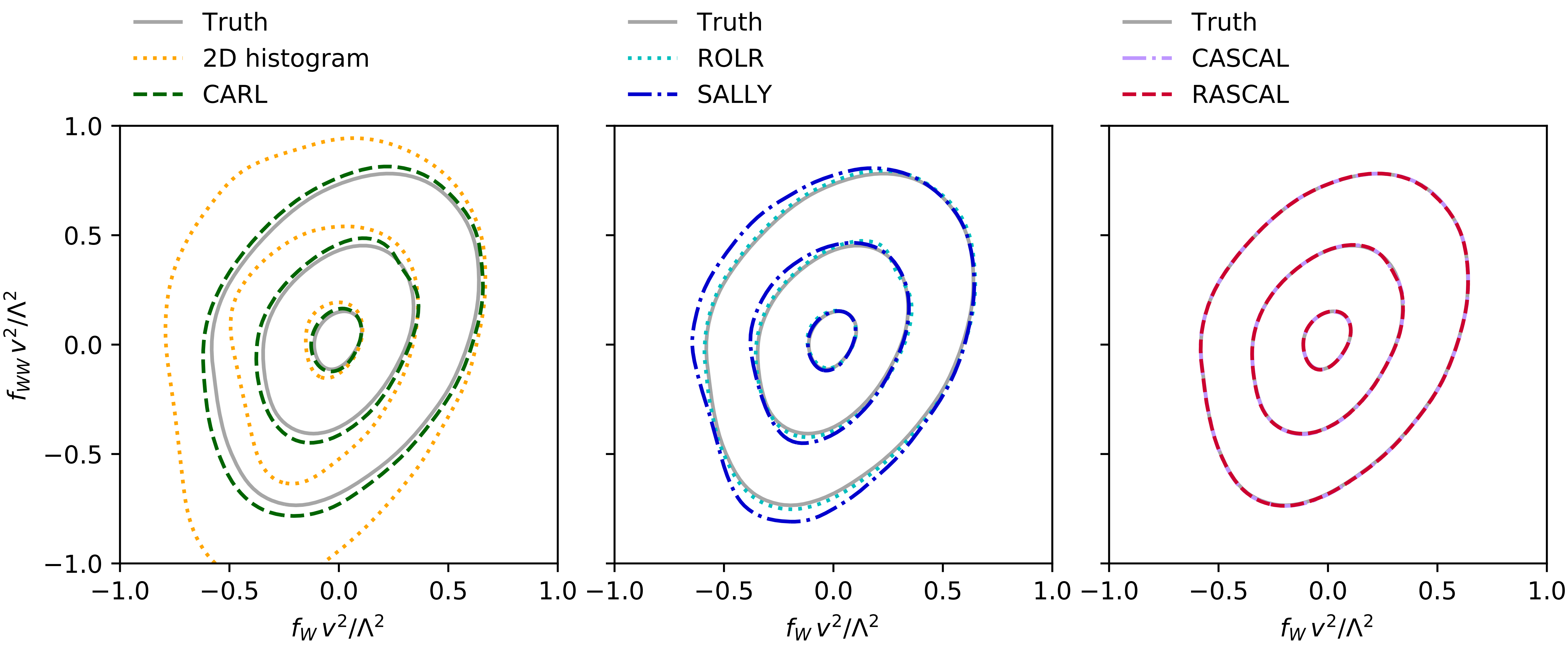
# Results

| 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 $\theta$ | **0.012** | 0.0028 | |
| CARL (morphing-aware) | Baseline | 0.076 | 0.0200 | Fig. 12 |
| | Random $\theta$ | 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 $\theta$ | **0.003** | **0.0014** | |
| ROLR (morphing-aware) | Baseline | 0.024 | 0.0063 | |
| | Random $\theta$ | 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 $\theta$ | 0.001 | 0.0002 | |
| CASCAL (morphing-aware) | Baseline | 0.136 | 0.0427 | |
| | Random $\theta$ | 0.092 | 0.0268 | |
| | Morphing basis | 0.040 | 0.0081 | |
| RASCAL (parameterized) | Baseline | 0.001 | 0.0004 | ✓ |
| | Random $\theta$ | **0.001** | **0.0004** | |
| RASCAL (morphing-aware) | Baseline | 0.125 | 0.0514 | |
| | Random $\theta$ | 0.132 | 0.0539 | |
| | Morphing basis | 0.031 | 0.0072 | |

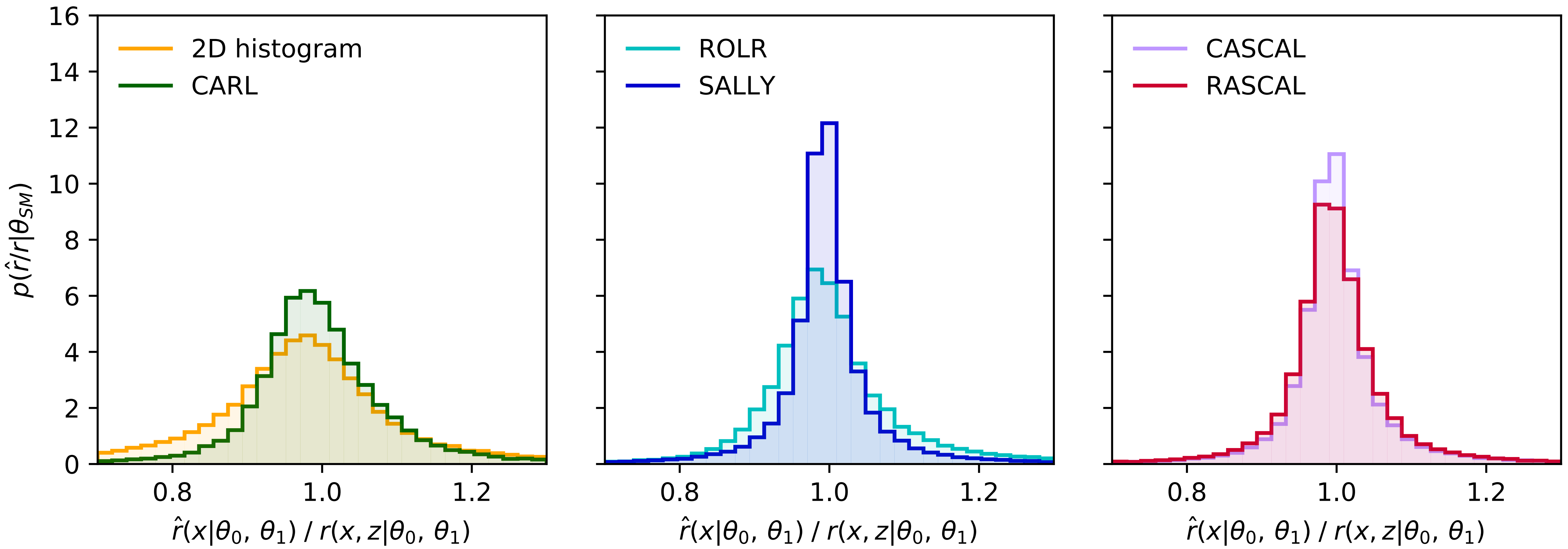
# Precision of estimates



# Expected limits (truth level)



# Detector effects



# Expected limits (with detector effects)