Applications and Potentials of Normalising Flows

Gregor Kasieczka
gregor.kasieczka@uni-hamburg.de
Twitter: @GregorKasieczka

Wiehl Workshop

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

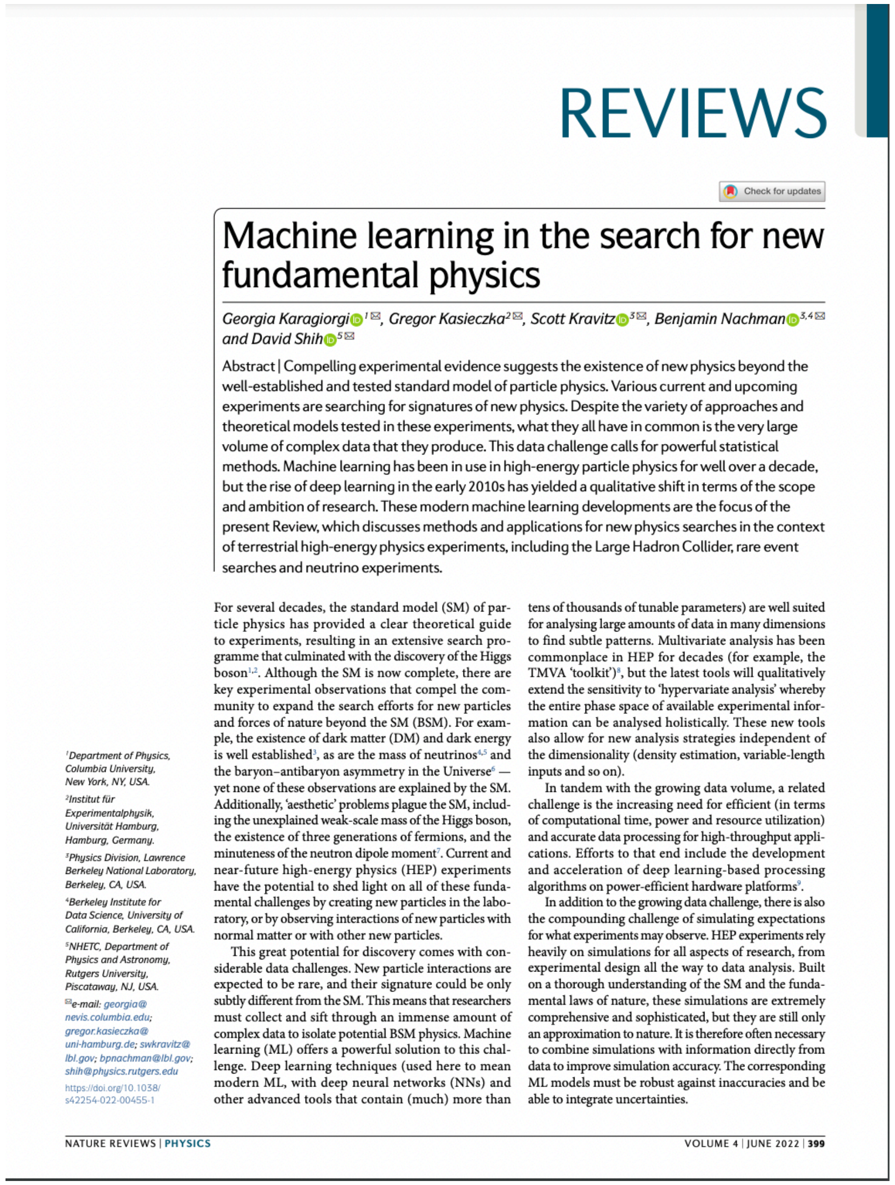
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

PIER
Partnership of Universität Hamburg and DESY

Bundesministerium für Bildung und Forschung

Ressources



REVIEWS

Check for updates

Machine learning in the search for new fundamental physics

Georgia Karagiorgi[1], Gregor Kasieczka[2], Scott Kravitz[3], Benjamin Nachman[3,4] and David Shih[5]

Abstract | Compelling experimental evidence suggests the existence of new physics beyond the well-established and tested standard model of particle physics. Various current and upcoming experiments are searching for signatures of new physics. Despite the variety of approaches and theoretical models tested in these experiments, what they all have in common is the very large volume of complex data that they produce. This data challenge calls for powerful statistical methods. Machine learning has been in use in high-energy particle physics for well over a decade, but the rise of deep learning in the early 2010s has yielded a qualitative shift in terms of the scope and ambition of research. These modern machine learning developments are the focus of the present Review, which discusses methods and applications for new physics searches in the context of terrestrial high-energy physics experiments, including the Large Hadron Collider, rare event searches and neutrino experiments.

For several decades, the standard model (SM) of particle physics has provided a clear theoretical guide to experiments, resulting in an extensive search programme that culminated with the discovery of the Higgs boson[1,2]. Although the SM is now complete, there are key experimental observations that compel the community to expand the search efforts for new particles and forces of nature beyond the SM (BSM). For example, the existence of dark matter (DM) and dark energy is well established[3], as are the mass of neutrinos[4,5] and the baryon–antibaryon asymmetry in the Universe[6] — yet none of these observations are explained by the SM. Additionally, 'aesthetic' problems plague the SM, including the unexplained weak-scale mass of the Higgs boson, the existence of three generations of fermions, and the minuteness of the neutron dipole moment[7]. Current and near-future high-energy physics (HEP) experiments have the potential to shed light on all of these fundamental challenges by creating new particles in the laboratory, or by observing interactions of new particles with normal matter or with other new particles.

This great potential for discovery comes with considerable data challenges. New particle interactions are expected to be rare, and their signature could be only subtly different from the SM. This means that researchers must collect and sift through an immense amount of complex data to isolate potential BSM physics. Machine learning (ML) offers a powerful solution to this challenge. Deep learning techniques (used here to mean modern ML, with deep neural networks (NNs) and other advanced tools that contain (much) more than tens of thousands of tunable parameters) are well suited for analysing large amounts of data in many dimensions to find subtle patterns. Multivariate analysis has been commonplace in HEP for decades (for example, the TMVA 'toolkit')[8], but the latest tools will qualitatively extend the sensitivity to 'hypervariate analysis' whereby the entire phase space of available experimental information can be analysed holistically. These new tools also allow for new analysis strategies independent of the dimensionality (density estimation, variable-length inputs and so on).

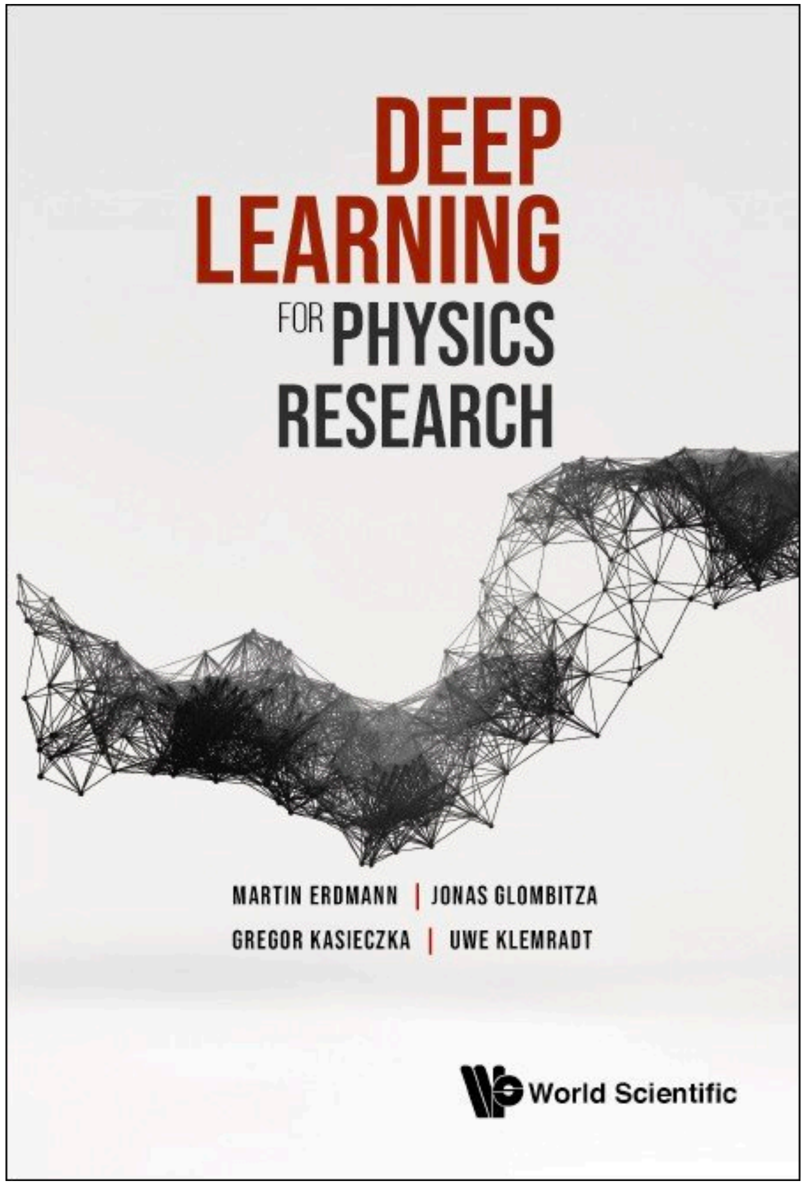
In tandem with the growing data volume, a related challenge is the increasing need for efficient (in terms of computational time, power and resource utilization) and accurate data processing for high-throughput applications. Efforts to that end include the development and acceleration of deep learning-based processing algorithms on power-efficient hardware platforms[9].

In addition to the growing data challenge, there is also the compounding challenge of simulating expectations for what experiments may observe. HEP experiments rely heavily on simulations for all aspects of research, from experimental design all the way to data analysis. Built on a thorough understanding of the SM and the fundamental laws of nature, these simulations are extremely comprehensive and sophisticated, but they are still only an approximation to nature. It is therefore often necessary to combine simulations with information directly from data to improve simulation accuracy. The corresponding ML models must be robust against inaccuracies and be able to integrate uncertainties.

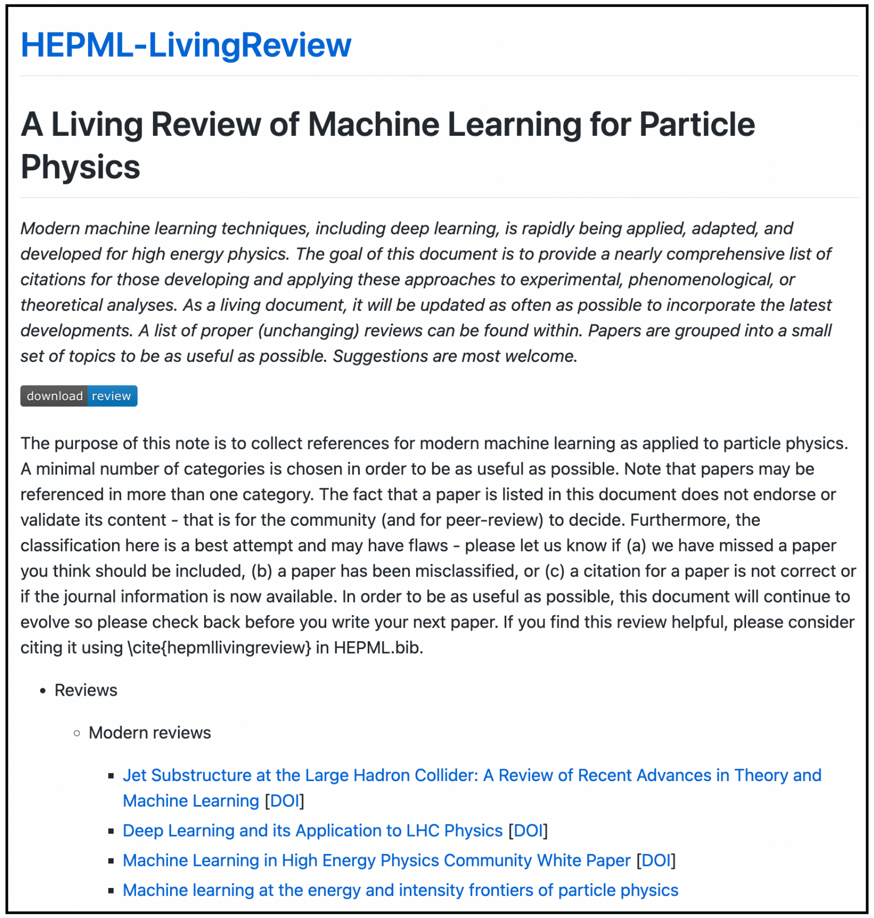
[1]Department of Physics, Columbia University, New York, NY, USA.
[2]Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany.
[3]Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[4]Berkeley Institute for Data Science, University of California, Berkeley, CA, USA.
[5]NHETC, Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, USA.
e-mail: georgia@nevis.columbia.edu; gregor.kasieczka@uni-hamburg.de; swkravitz@lbl.gov; bpnachman@lbl.gov; shih@physics.rutgers.edu

https://doi.org/10.1038/s42254-022-00455-1

https://arxiv.org/abs/2112.03769



https://www.worldscientific.com/worldscibooks/10.1142/12294#t=aboutBook



HEPML-LivingReview

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

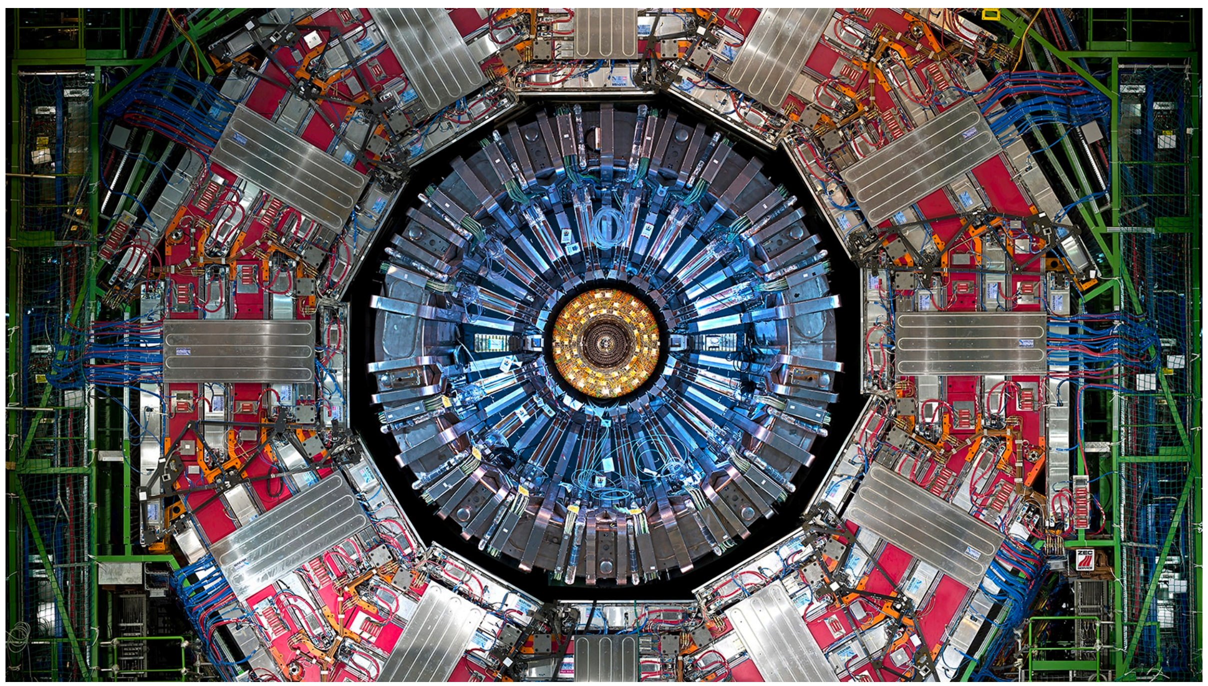
download review

The purpose of this note is to collect references for modern machine learning as applied to particle physics. A minimal number of categories is chosen in order to be as useful as possible. Note that papers may be referenced in more than one category. The fact that a paper is listed in this document does not endorse or validate its content - that is for the community (and for peer-review) to decide. Furthermore, the classification here is a best attempt and may have flaws - please let us know if (a) we have missed a paper you think should be included, (b) a paper has been misclassified, or (c) a citation for a paper is not correct or if the journal information is now available. In order to be as useful as possible, this document will continue to evolve so please check back before you write your next paper. If you find this review helpful, please consider citing it using \cite{hepmllivingreview} in HEPML.bib.

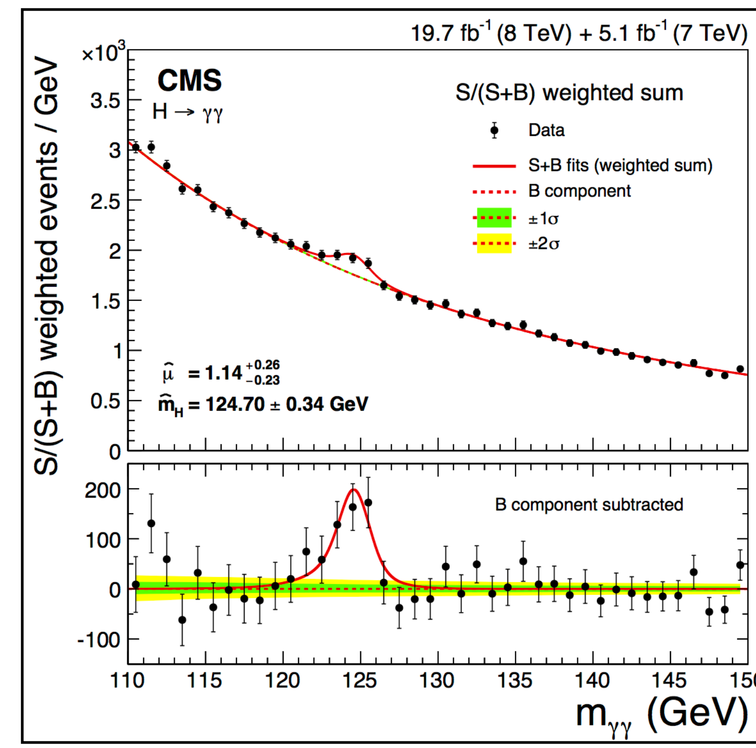
- Reviews
  - Modern reviews
    - Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning [DOI]
    - Deep Learning and its Application to LHC Physics [DOI]
    - Machine Learning in High Energy Physics Community White Paper [DOI]
    - Machine learning at the energy and intensity frontiers of particle physics

https://iml-wg.github.io/HEPML-LivingReview/

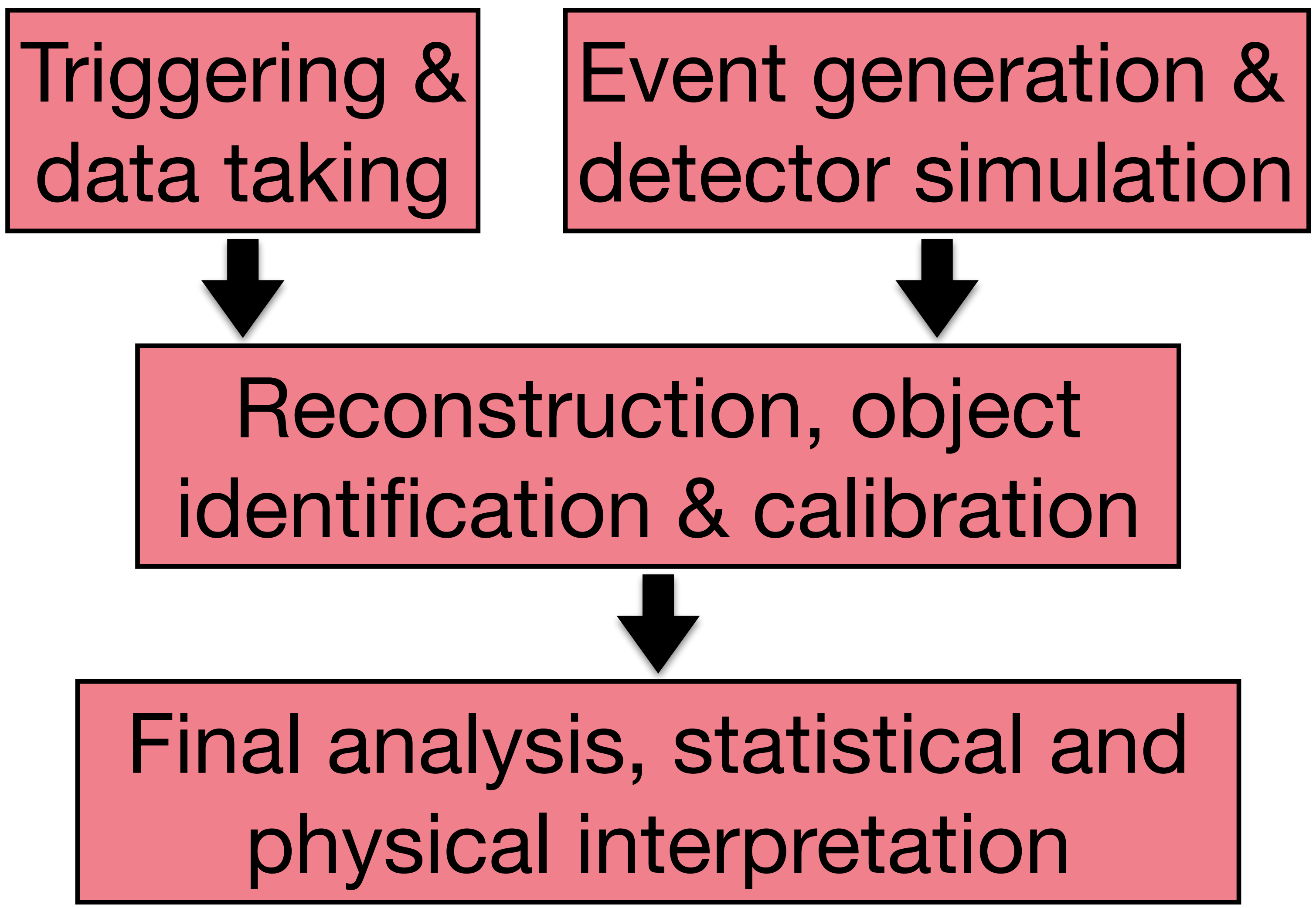
Experimental particle physics workflow

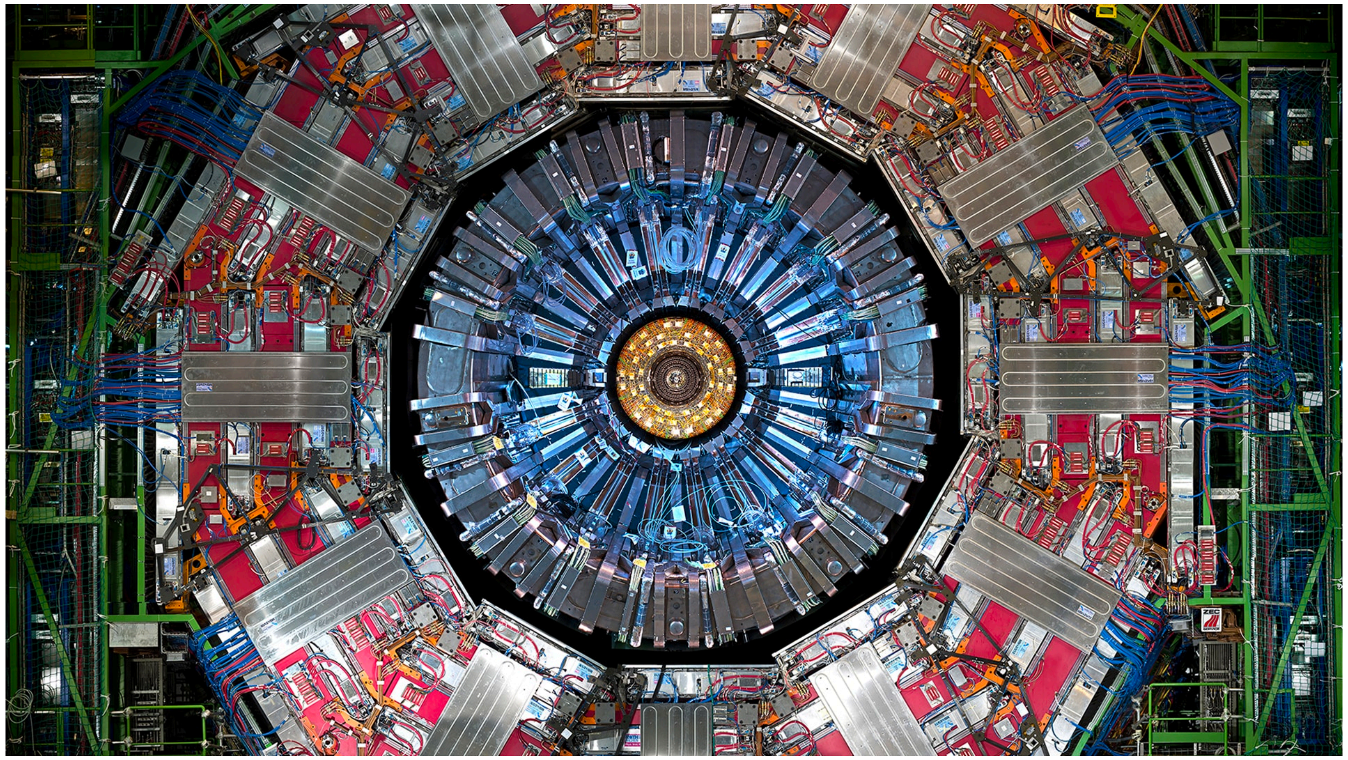


?



Experimental particle physics workflow





Triggering and data taking

Particle collisions happen at a rate of 40 MHz with size ~1 MB/event.

Need to distill to ~1 kHz via lossy, irreversible filtering algorithms (Trigger).

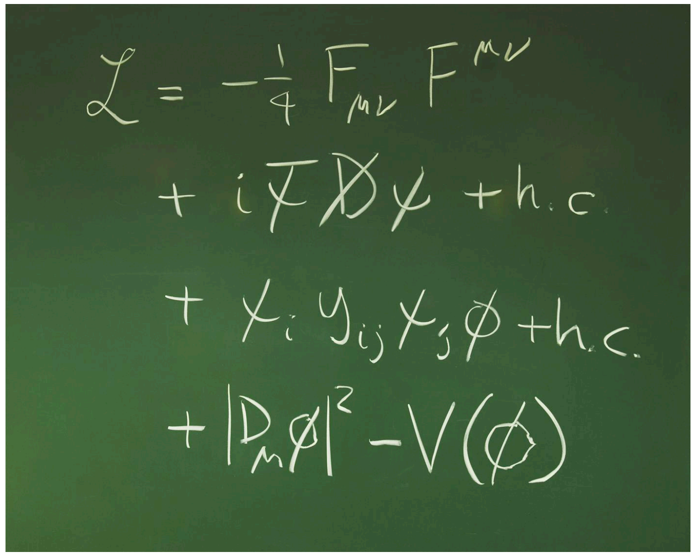
Data is very heterogenous: low-level readouts in ~100M channels.

Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation



&

Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Simulation

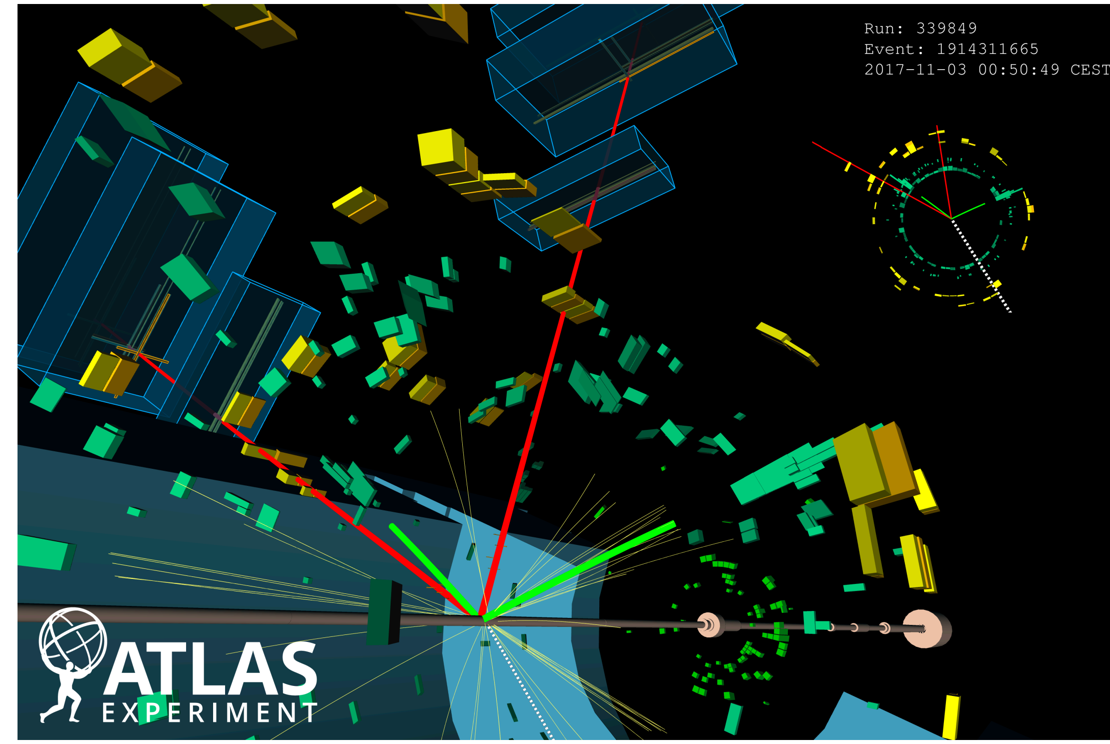
Theoretically well motivated Monte Carlo based simulations of known and hypothetical processes as well as detector responses.

As ~similar amount of simulated and real data is needed, significant compute goes here.

Reconstruction

Build high level objects (particles, leptons, jets, ..) from raw measurements in detectors and identify different particle decays.

Same processing chain for simulation and real data.



Triggering & data taking

Event generation & detector simulation

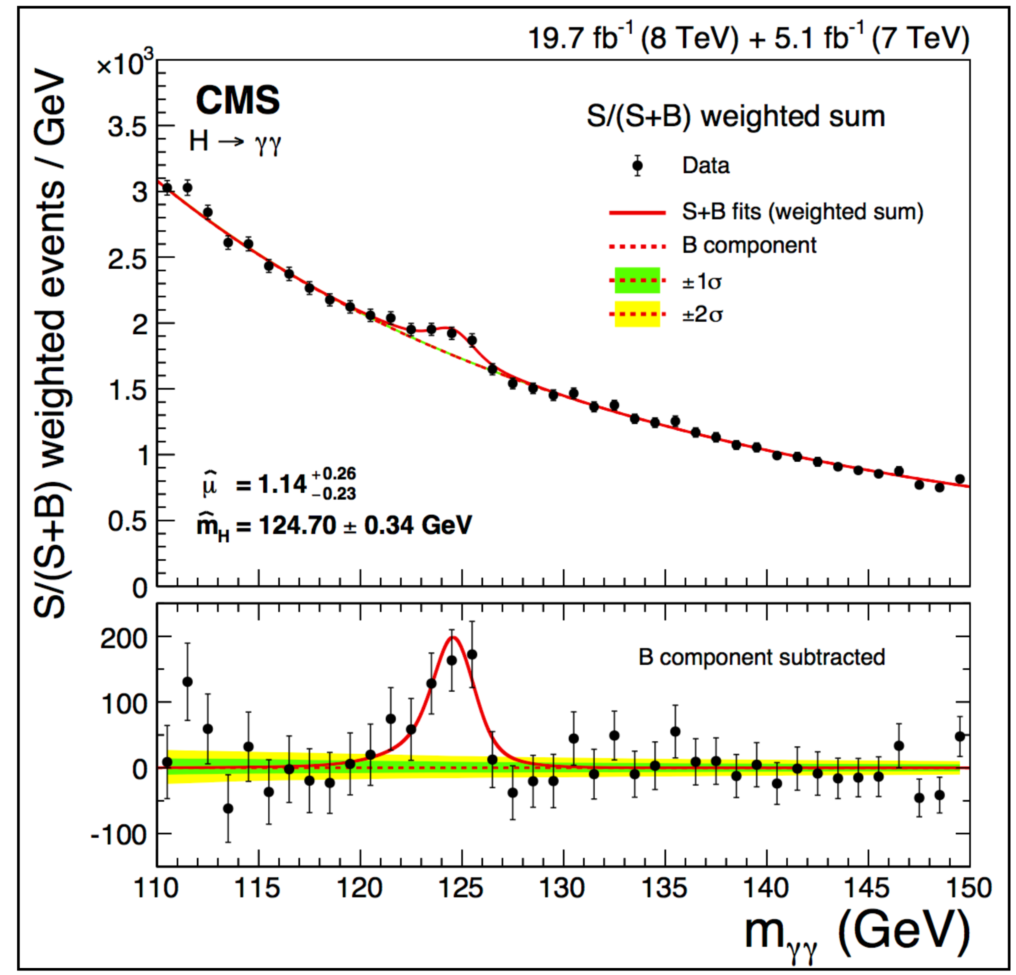
Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Analysis

Select region of phase space that isolates a physical phenomen of interest and perform detailed statistical analysis.

Compares simulation and data, quantifies uncertainties.



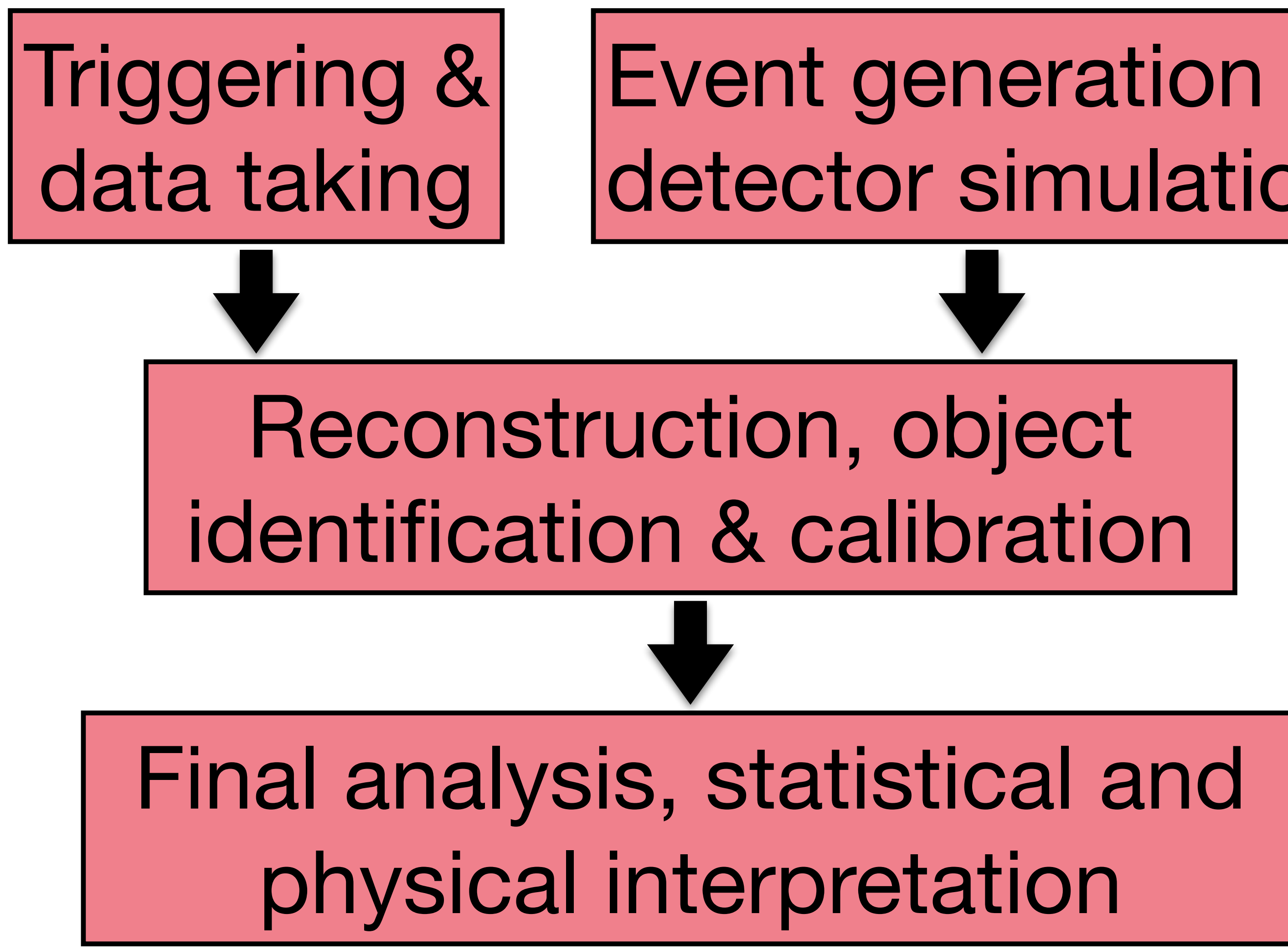
Triggering & data taking

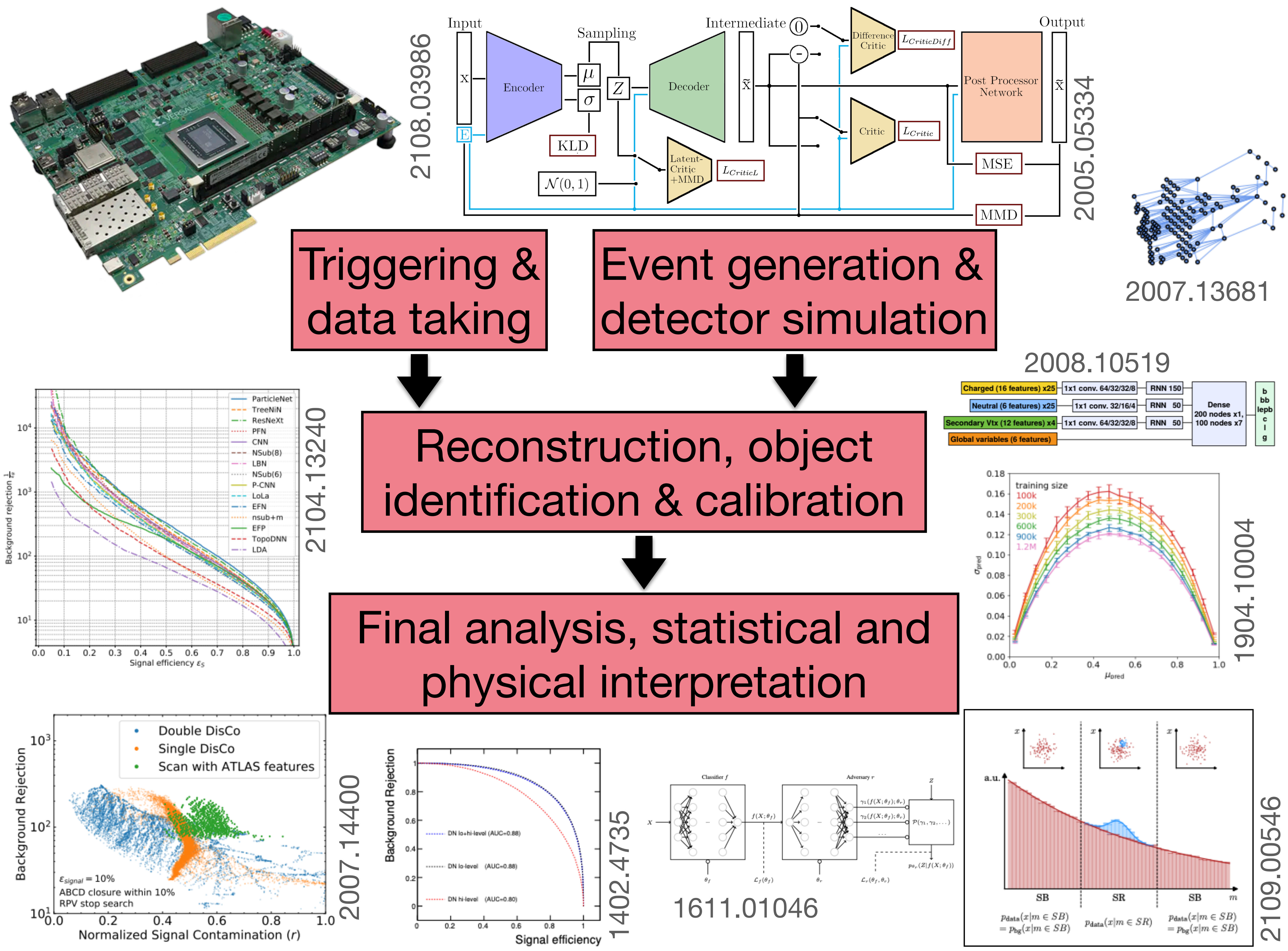
Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

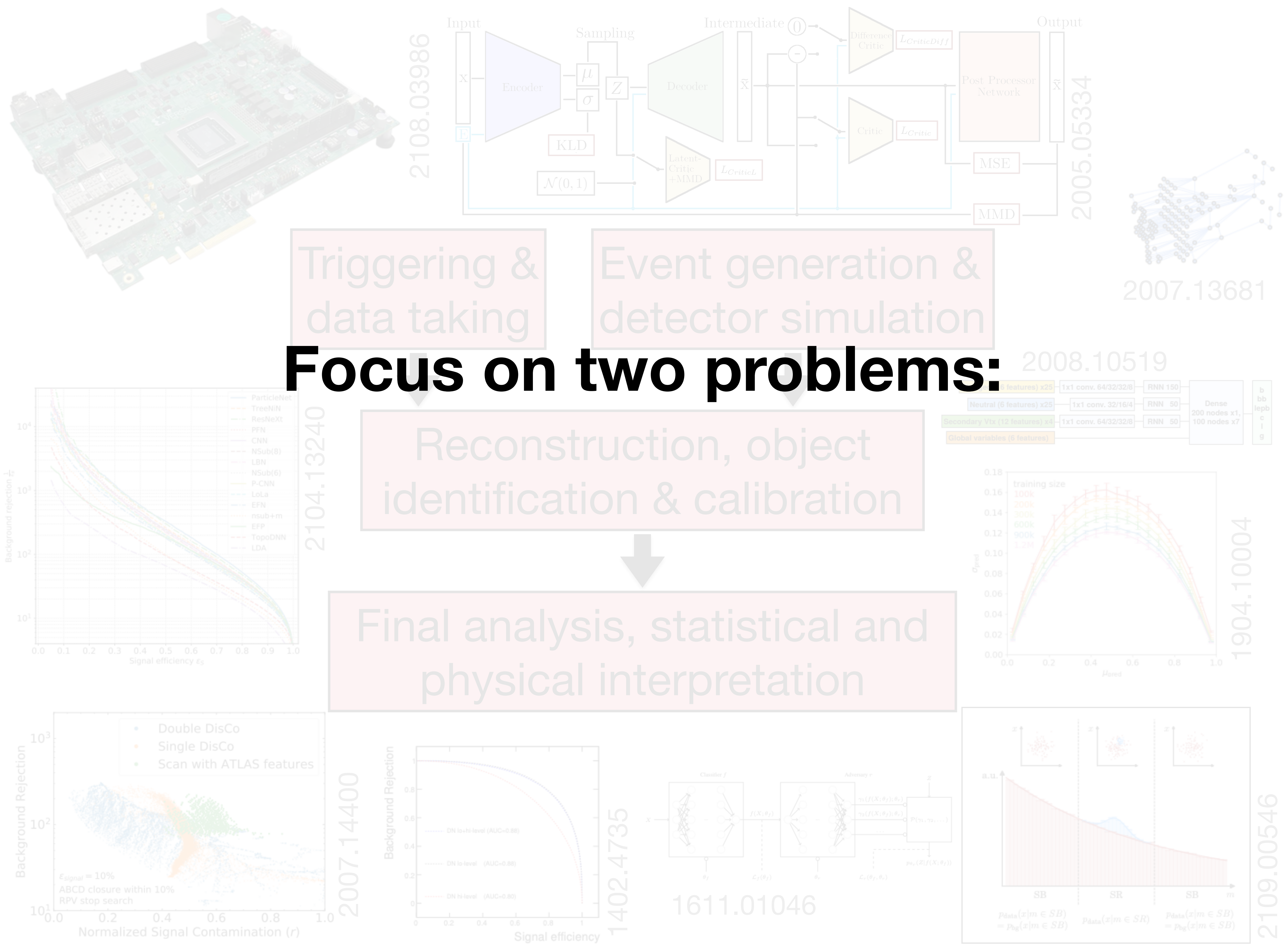
Machine learning plays an increasing role in all of these steps



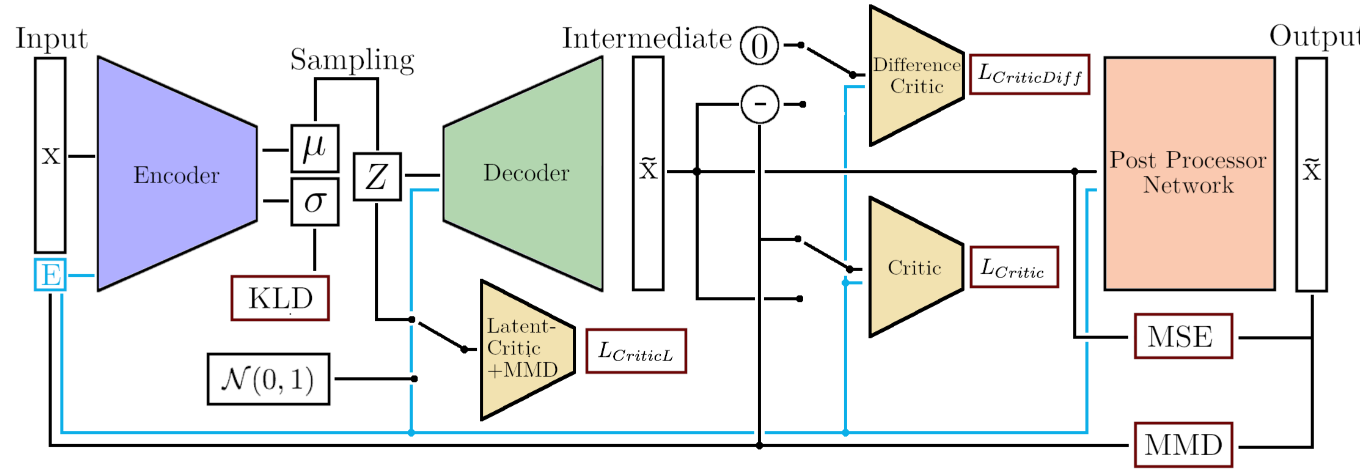


Triggering & data taking
Event generation & detector simulation
Reconstruction, object identification & calibration
Final analysis, statistical and physical interpretation
2108.03986
Input
Sampling
Intermediate
Output
Encoder
Decoder
KLD
Latent-Critic +MMD
Difference Critic
Critic
Post Processor Network
MSE
MMD
2005.05334
2007.13681
2008.10519
Charged (16 features) x25
Neutral (6 features) x25
Secondary Vtx (12 features) x4
Global variables (6 features)
1x1 conv. 64/32/32/8
1x1 conv. 32/16/4
RNN 150
RNN 50
Dense 200 nodes x1, 100 nodes x7
b bb lepb c l g
2104.13240
ParticleNet
TreeNiN
ResNeXt
PFN
CNN
NSub(8)
LBN
NSub(6)
P-CNN
LoLa
EFN
nsub+m
EFP
TopoDNN
LDA
Background rejection
Signal efficiency
1904.10004
training size
100k
200k
300k
600k
900k
1.2M
2007.14400
Double DisCo
Single DisCo
Scan with ATLAS features
Background Rejection
ABCD closure within 10%
RPV stop search
Normalized Signal Contamination (r)
1402.4735
DN lo+hi-level (AUC=0.88)
DN lo-level (AUC=0.88)
DN hi-level (AUC=0.80)
Signal efficiency
1611.01046
Classifier f
Adversary r
2109.00546
a.u.
SB
SR
SB

Focus on two problems:



Simulation



Triggering & data taking

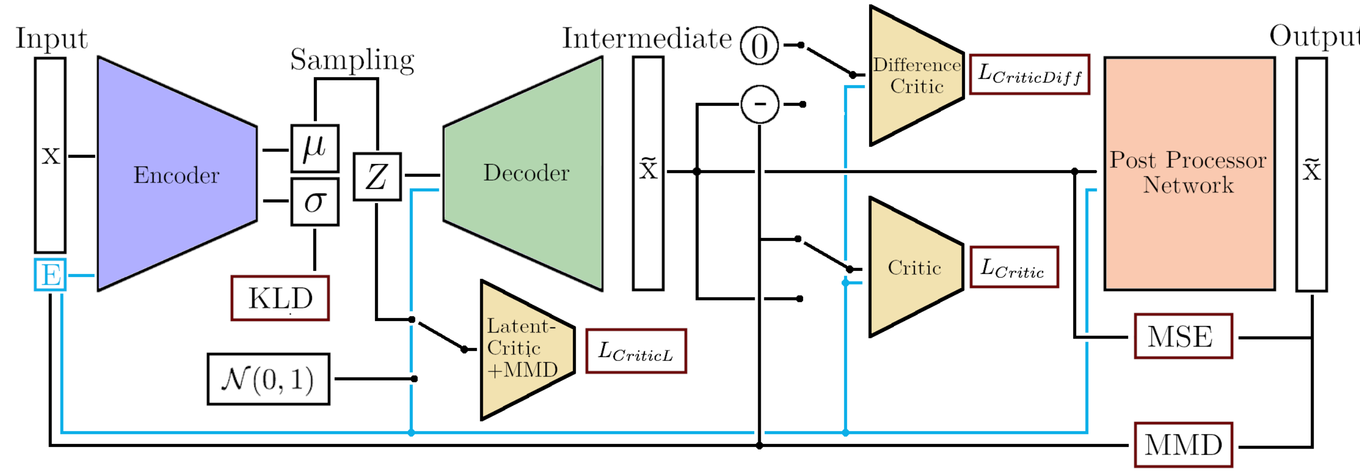
Event generation & detector simulation

Focus on two problems:

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Simulation



Triggering & data taking

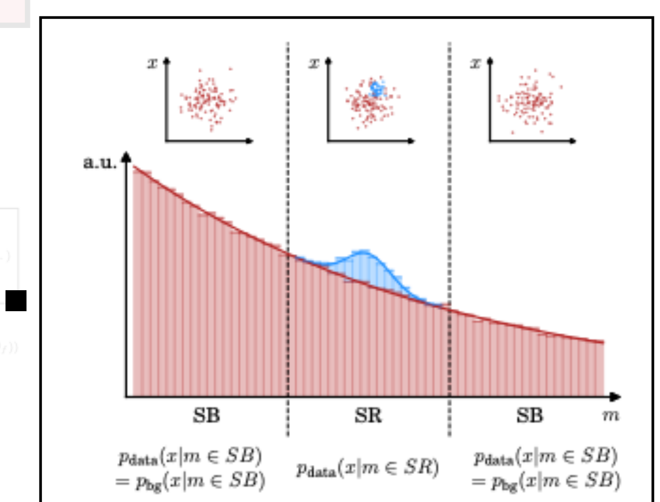
Event generation & detector simulation

Focus on two problems:

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

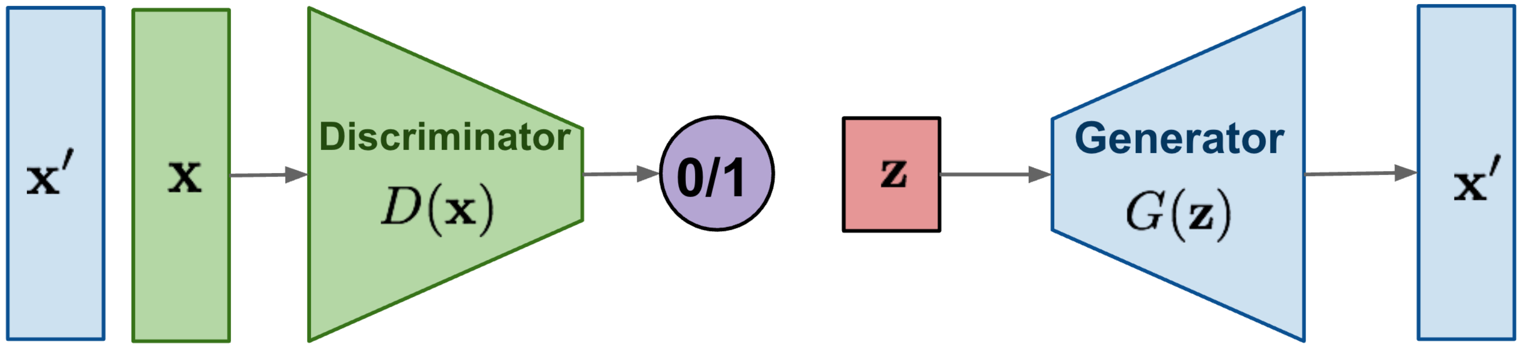
and finding new physics.



Unifying theme:
Generative models

Unifying theme: Generative models

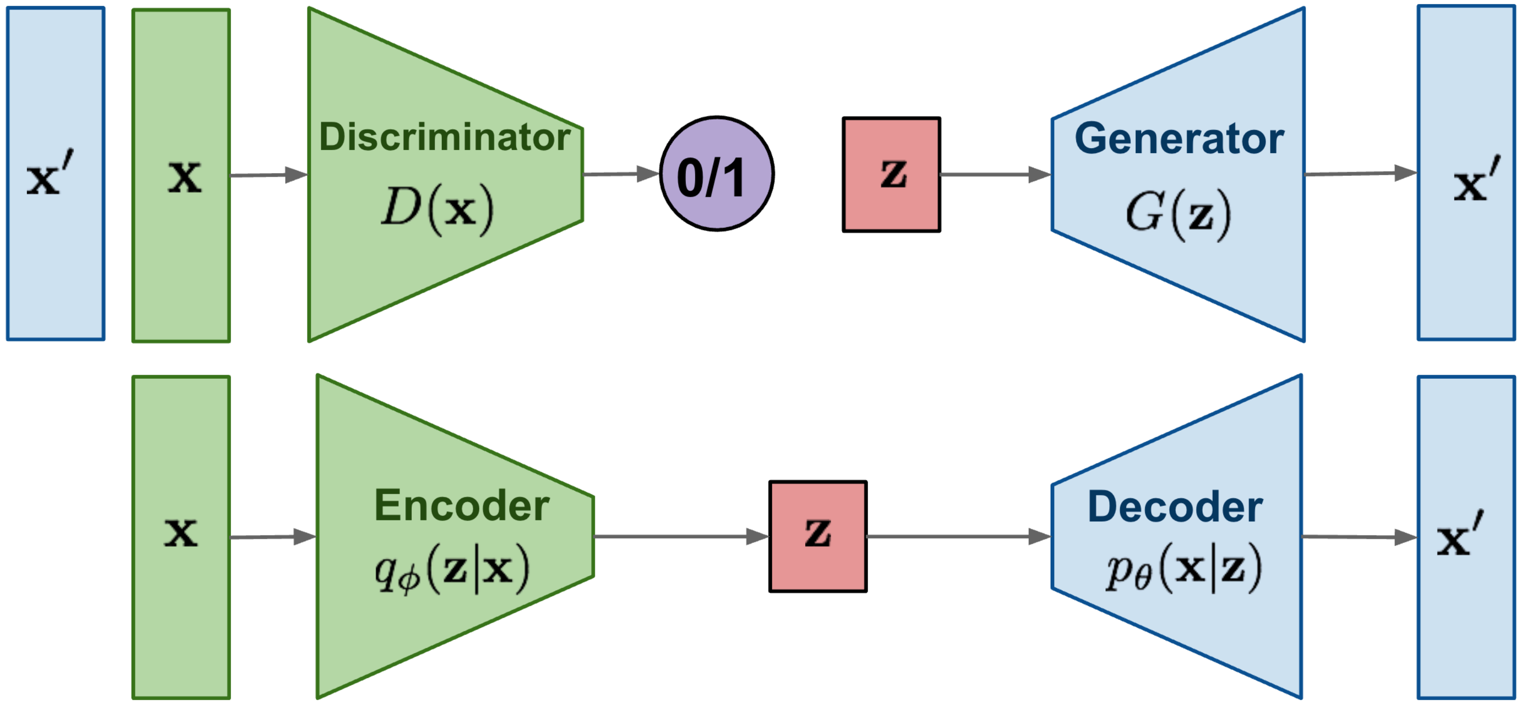
GAN: Adversarial training



Unifying theme: Generative models

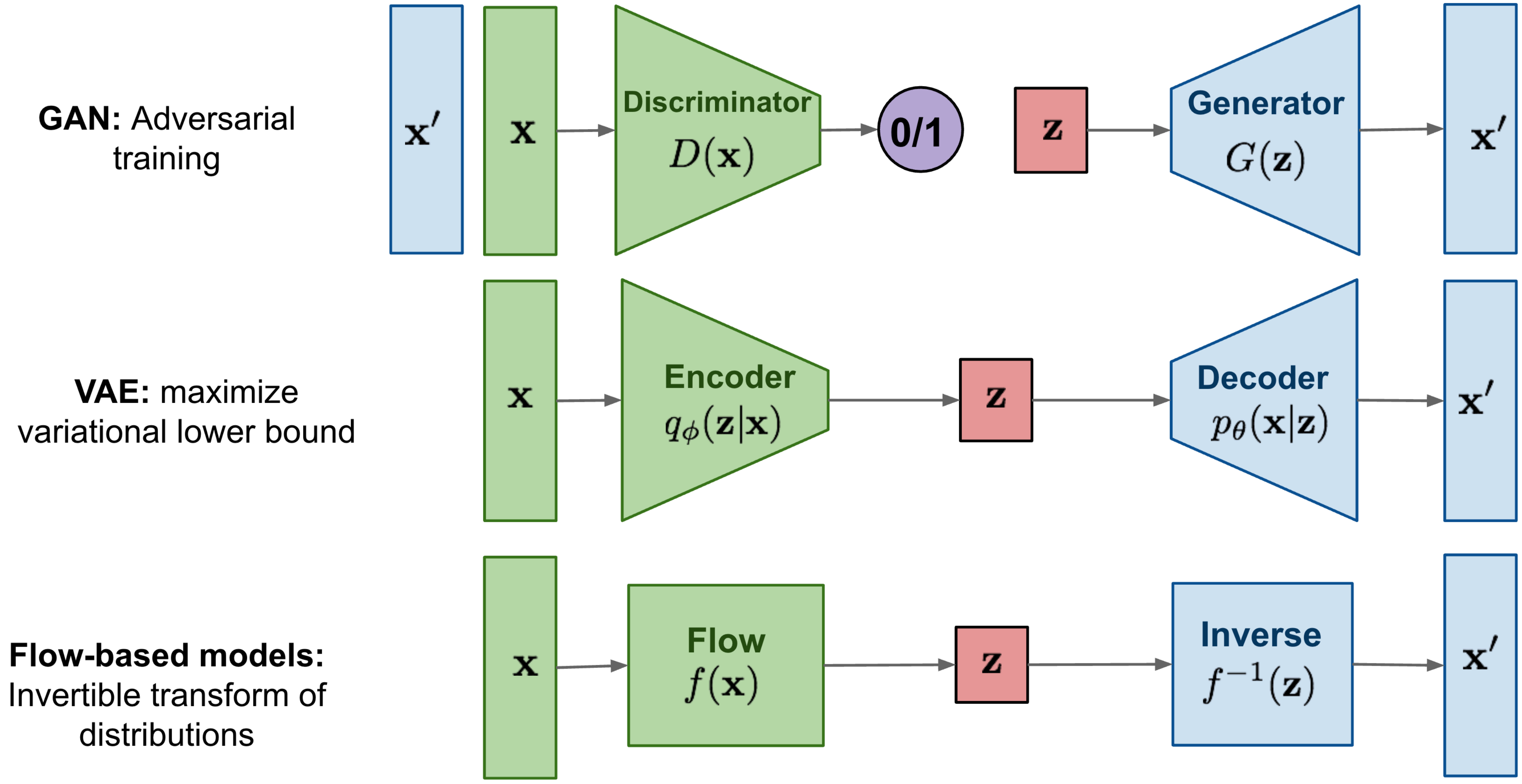
GAN: Adversarial training

VAE: maximize variational lower bound

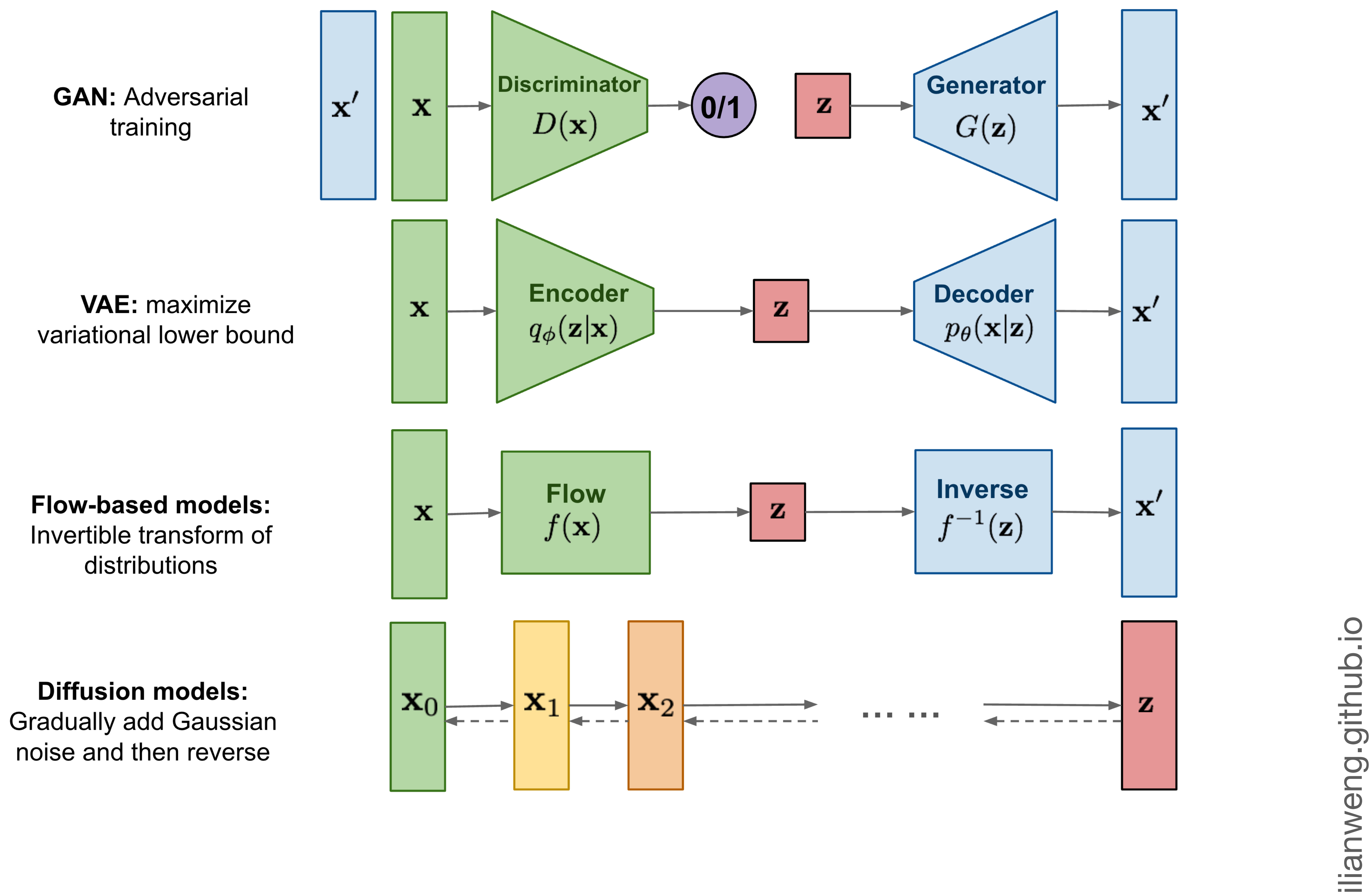


lilianweng.github.io

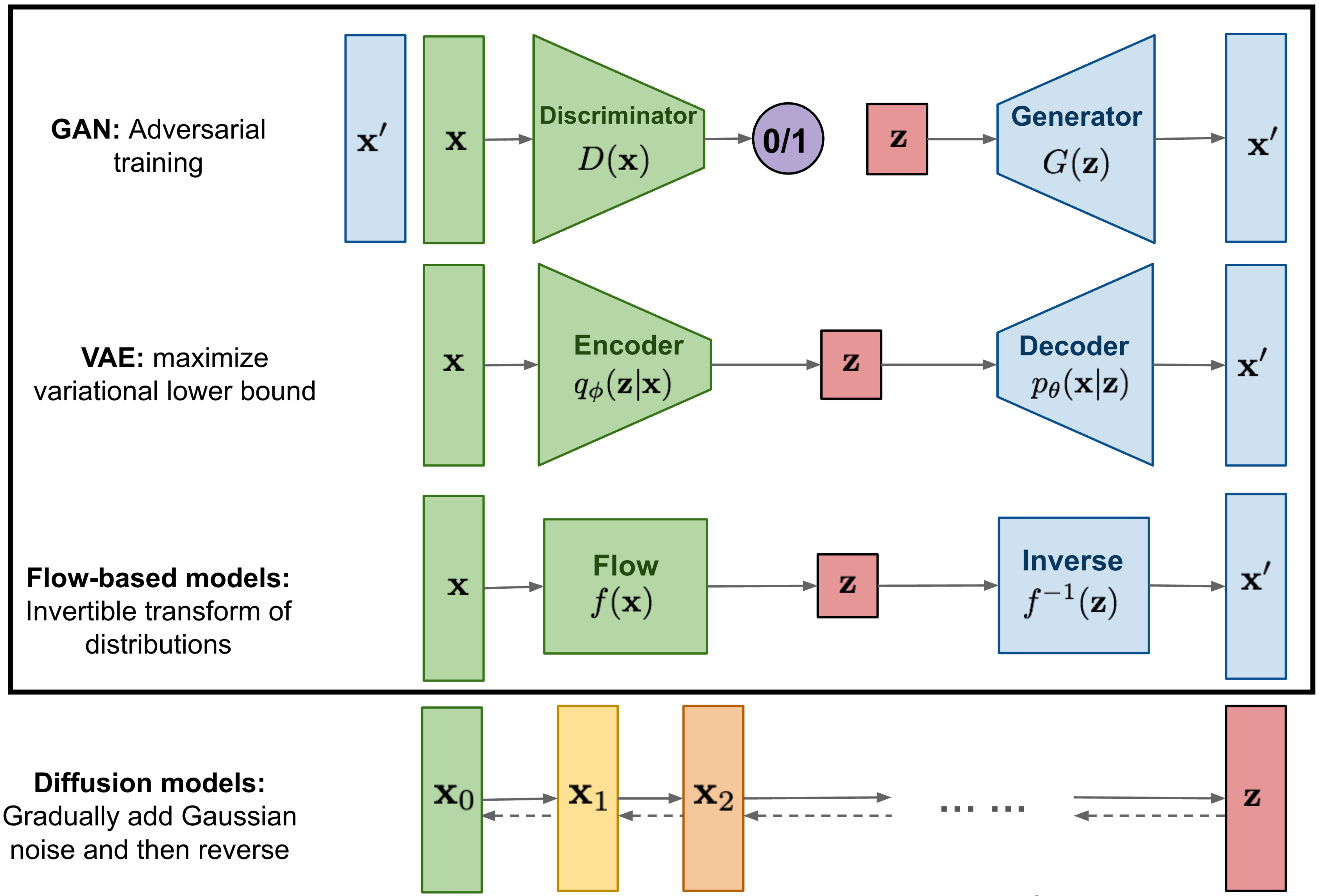
Unifying theme: Generative models



Unifying theme: Generative models



Unifying theme:
Generative models



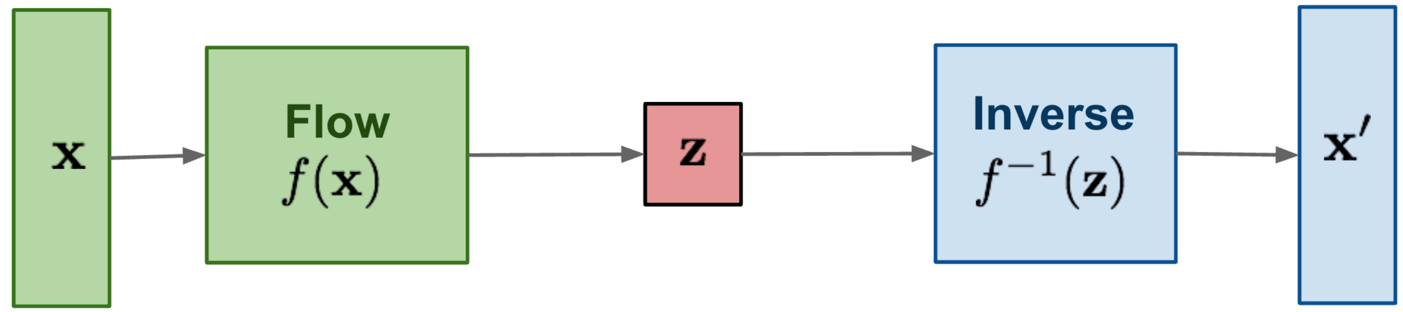
Widely used in HEP

One example so far: 2206.11898

lilianweng.github.io

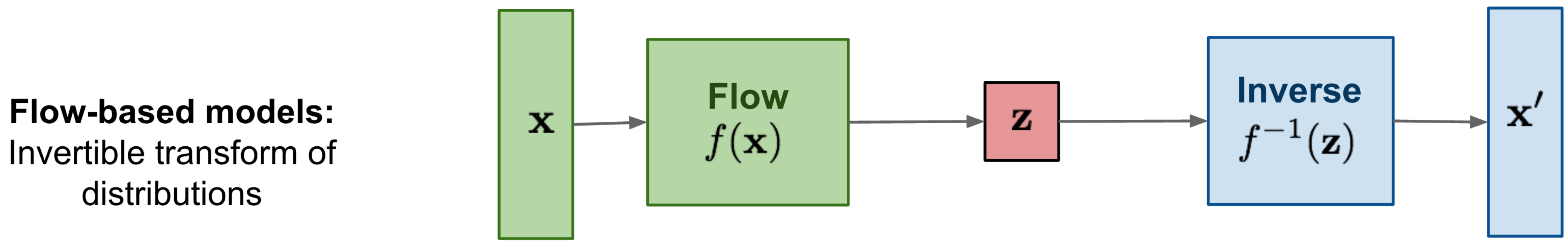
Introduction to NF

Flow-based models:
Invertible transform of distributions



- Basic idea: Learn a mapping between data and an intial latent-space distribution (e.g. Gaussians)
  - Bijective, so that it is invertable
    (f^{-1} is not a learned approximated inversion, but the exact inverse of f by construction)
  - Actually a diffeomorphism
  - Take into account Jacobian determinant (change of prob. variable formula) to evaluate probability density in data space
    (need to construct f to allow easy calculation of Jacobian determinant)

Introduction to NF

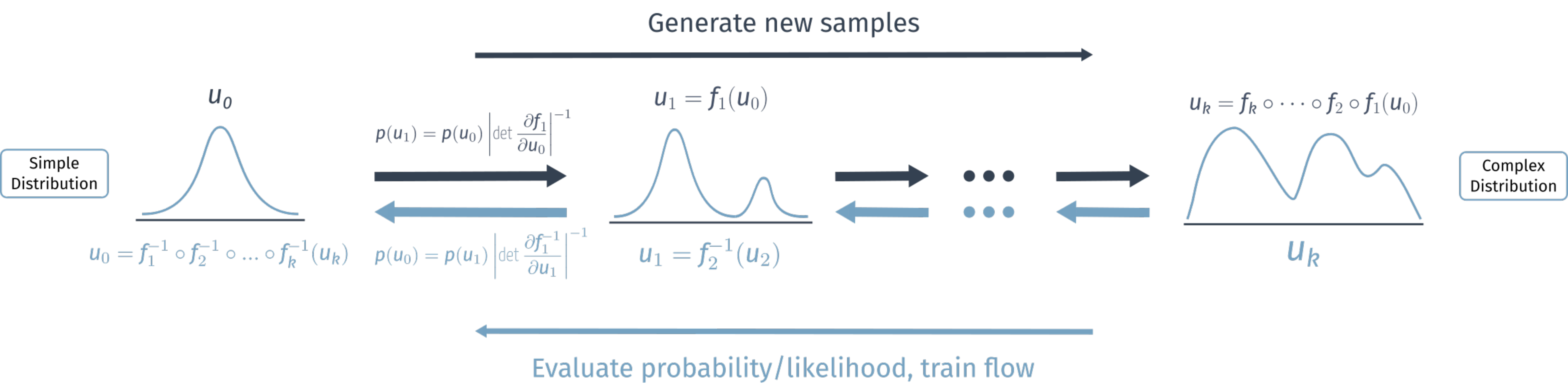


- Why could this be useful?
  - Can sample from latent space and transform with f^{-1} into data space for use as generative model
  - Can assign likelihood to data points by applying f

- Will see some physics applications later

- (See e.g. D. J. Rezende and S. Mohamed, Variational inference with normalizing flows, International Conference on Machine Learning 37, 1530 (2015); I. Kobyzev, S. Prince, and M. Brubaker, Normalizing Flows: An Introduction and Review of Current Methods, IEEE Transactions on Pattern Analysisand Machine Intelligence , 1 (2020))

Introduction to NF



- Goal: assign probability density to each datapoint
- Learn bijective transformation between data and a latent space with tractable probability
- Build from simple invertible transformations with tractable Jacobian

$$p(\boldsymbol{x}) = p(\boldsymbol{f}^{-1}(\boldsymbol{x})) \prod_i \left|\det\left(\frac{\partial \boldsymbol{f}_i^{-1}}{\partial \boldsymbol{x}}\right)\right| = p(\boldsymbol{u}) \prod_i \left|\det\left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{u}}\right)\right|^{-1}$$

Thanks to T. Loesche

Introduction to NF

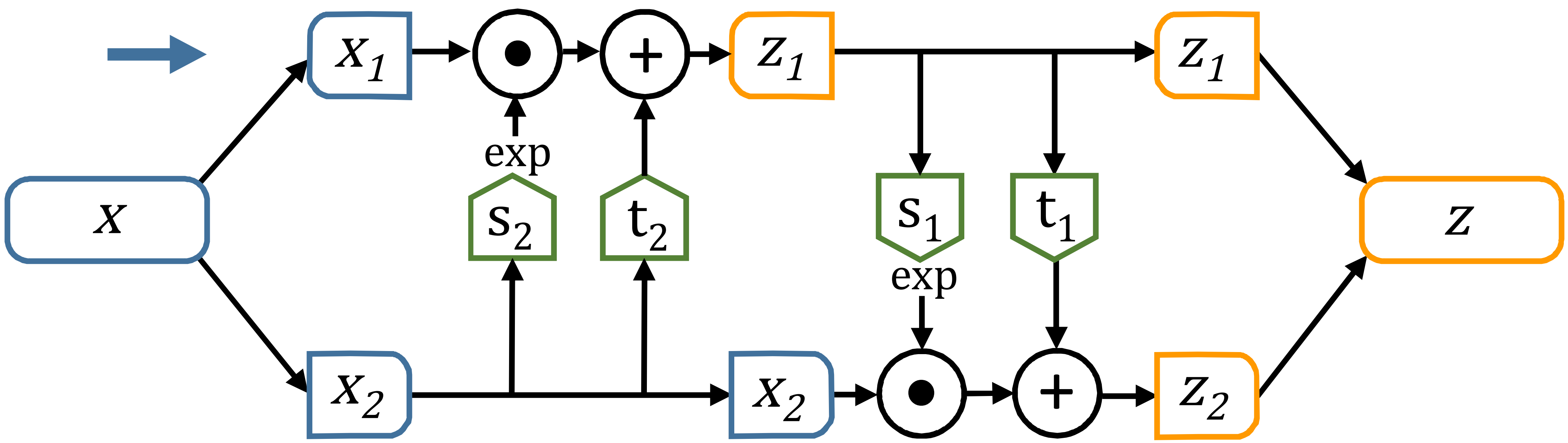
- Goal: assign probability density to each datapoint
- Learn bijective transformation between data and a latent space with tractable probability
- Build from simple invertible transformations with tractable Jacobian

$$p(\boldsymbol{x}) = p(\boldsymbol{f}^{-1}(\boldsymbol{x})) \prod_i \left| \det \left(\frac{\partial \boldsymbol{f}_i^{-1}}{\partial \boldsymbol{x}} \right) \right| =$$

$$p(\boldsymbol{u}) \prod_i \left| \det \left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{u}} \right) \right|^{-1}$$

Thanks to T. Loesche

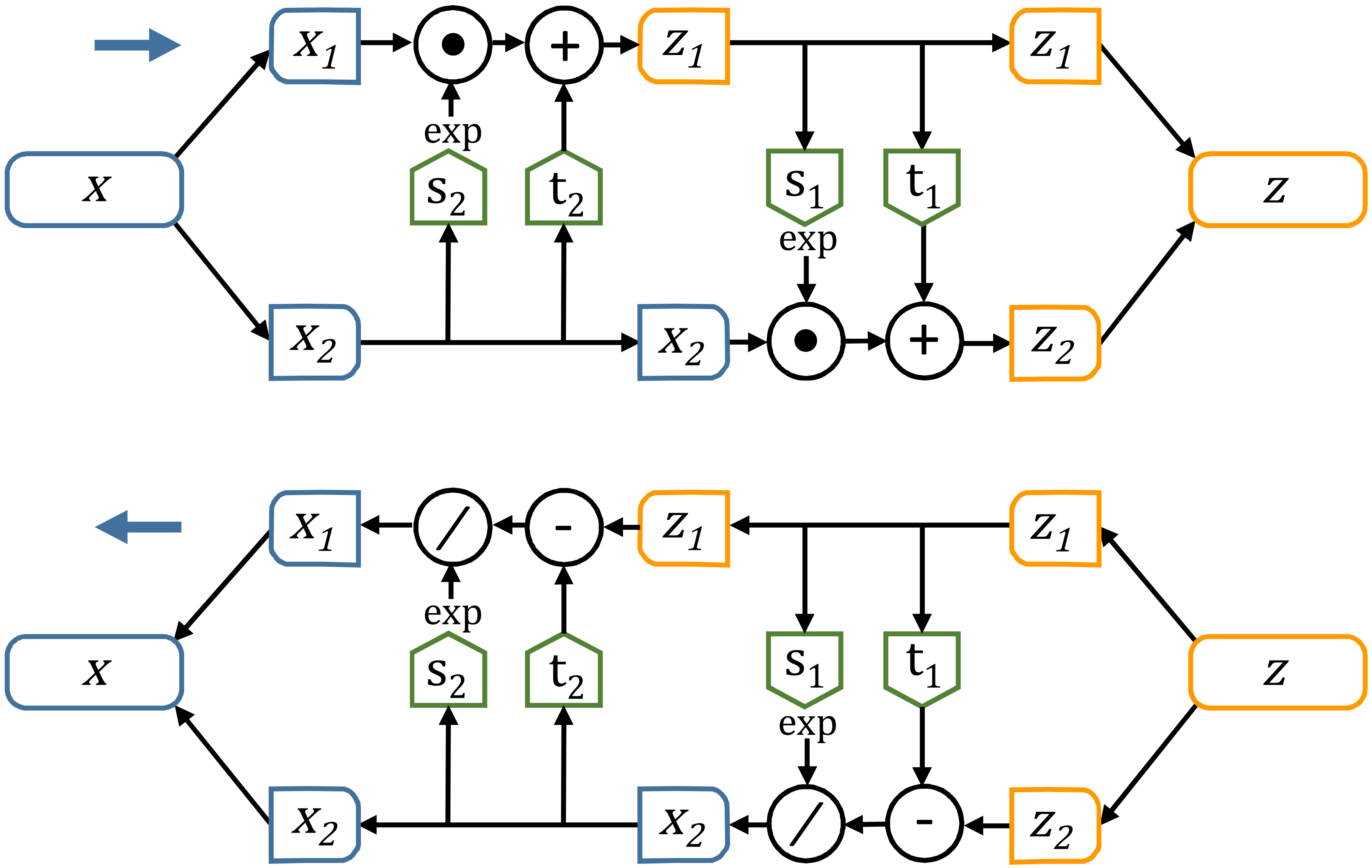
Coupling Flows



- Coupling flows / real NVP
  - Practically not the most widely used flow, but useful for illustration/understanding
  - Will use an alternative (masked autoregressive flows) for exercise

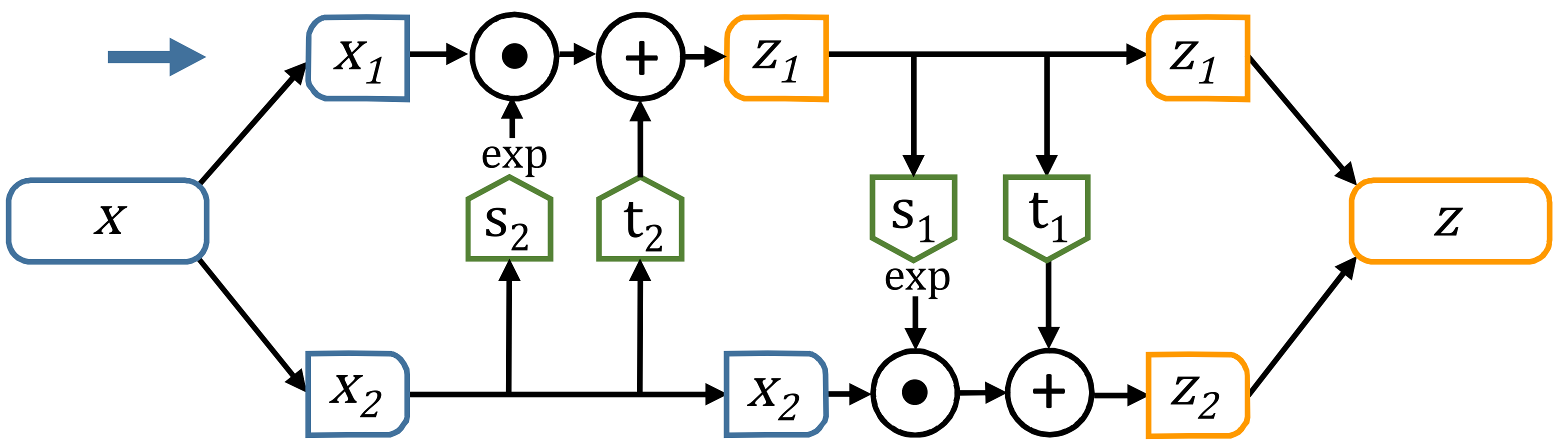
- Forward direction
- *s* and *t* are standard (e.g. fully connected) neural networks

Coupling Flows



- Forward and backward direction
- Can already see invertability
- What about Jacobian determinant?

Calculating Jacobian determinant



$$\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix} \xrightarrow{f_1} \begin{pmatrix}\mathbf{z}_1\\ \mathbf{x}_2\end{pmatrix} \xrightarrow{f_2} \begin{pmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{pmatrix}$$

with

$$\mathbf{x}_1 \xrightarrow{f_1} \mathbf{z}_1 = \mathbf{x}_1 \odot \exp(s_2(\mathbf{x}_2)) + t_2(\mathbf{x}_2)$$

$$\mathbf{x}_2 \xrightarrow{f_1} \mathbf{x}_2 .$$

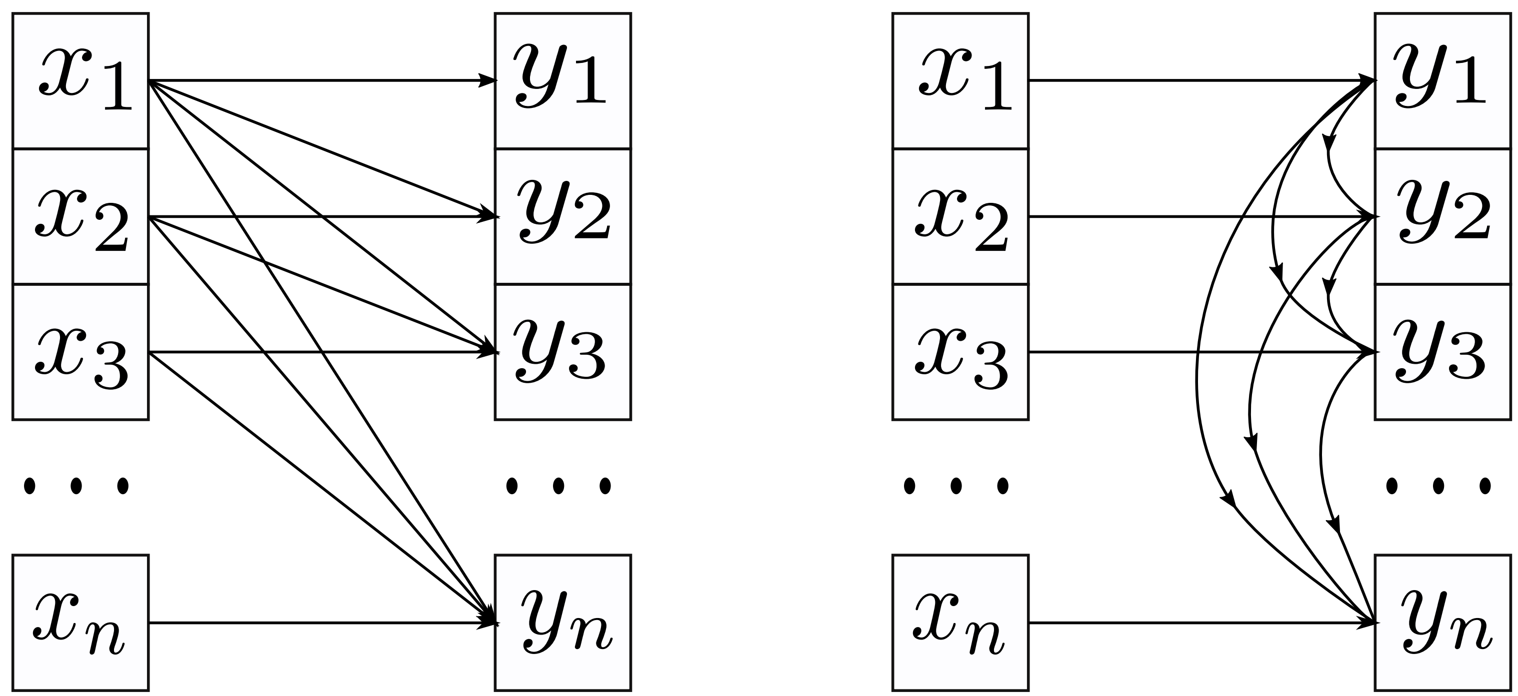
$$\mathbf{J_1} = \begin{pmatrix} \frac{\partial \mathbf{z}_1}{\partial \mathbf{x}_1} & \frac{\partial \mathbf{z}_1}{\partial \mathbf{x}_2} \\ \frac{\partial \mathbf{x}_2}{\partial \mathbf{x}_1} & \frac{\partial \mathbf{x}_2}{\partial \mathbf{x}_2} \end{pmatrix} = \begin{pmatrix} \mathrm{diag}(\exp(s_2(\mathbf{x}_2))) & \frac{\partial \mathbf{z}_1}{\partial \mathbf{x}_2} \\ 0 & \mathbb{1} \end{pmatrix}$$

Triangular matrix by construction

$$\det \mathbf{J_1} = \prod \exp(s_2(\mathbf{x}_2)) = \exp\left(\sum s_2(\mathbf{x}_2)\right)$$

Similarly simple for J_2. Composition of functions means multiplying their det J.

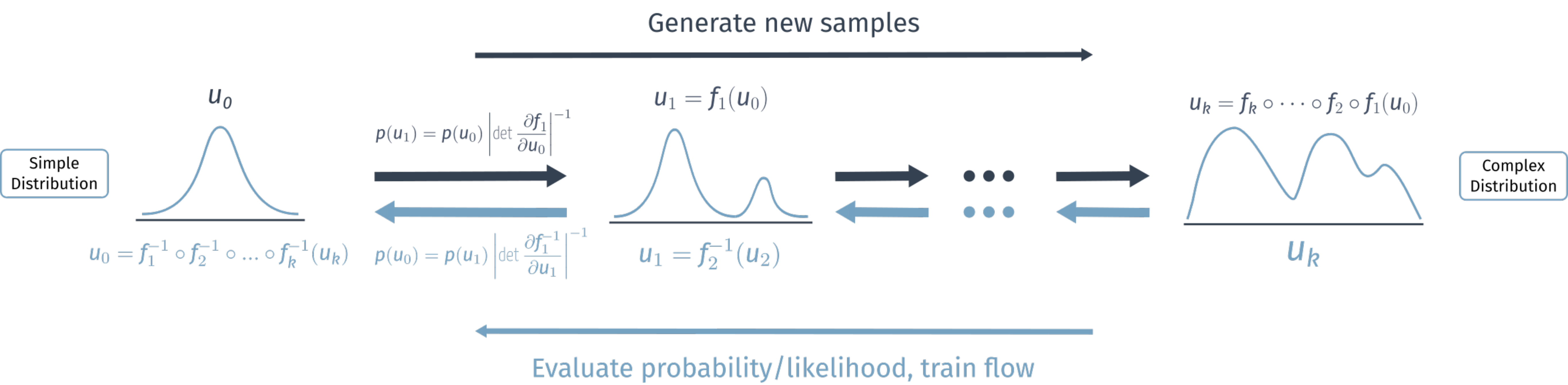
Autoregressive Flows



- Autoregressive property: Outputs conditioned on previous inputs
- Again, leads to simple Jacobian and invertible functions
- MAF: Masked Autoregressive Flow
  - Forward direction (data->latent) fast, backward slow
- IAF: Inverse Autoregressive Flow
  - Sampling direction (latent->data) fast
- **Many** other constructions exist as well (1908.09257 for an overview)

1908.09257

How to train NF?



- Loss is the negative log likelihood, assume Gaussian latent space distribution
- Sample points from the training dataset
- Transform into latent space using flow (and keep track of det J)

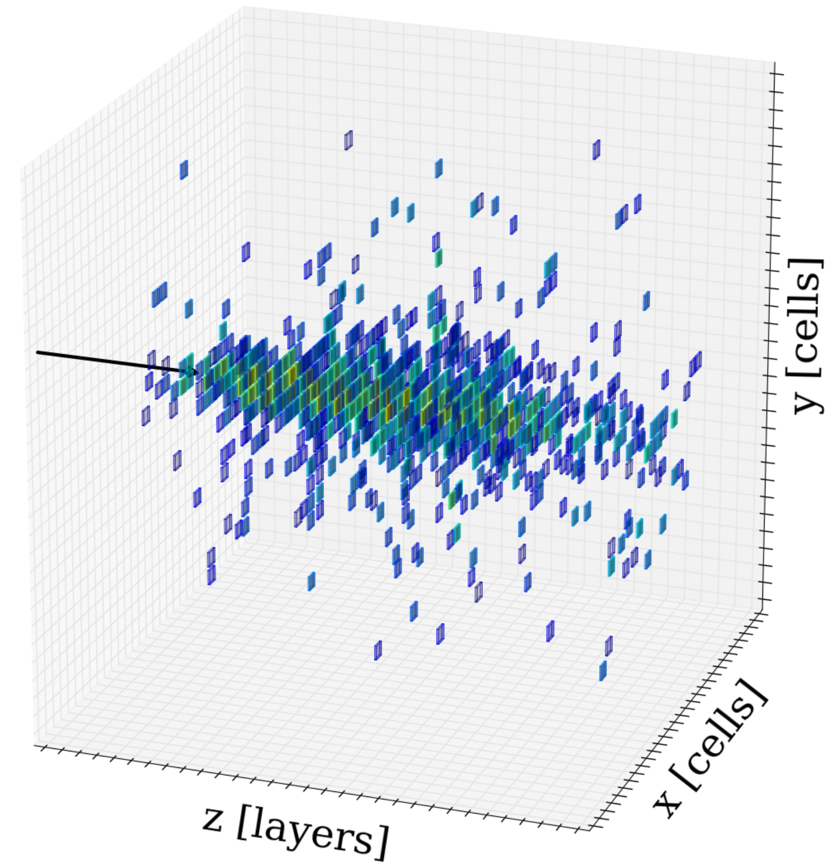
$$\mathcal{L} = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}} \left[-\frac{1}{2} ||f(\mathbf{x}))||_2^2 + \sum s(\mathbf{x}) \right]$$

Thanks to T. Loesche

Back to physics applications

Triggering & data taking

Event generation & detector simulation



Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Context

- Roughly speaking, “simulation” consists of two steps:
  - Event generation
    Model short-lived physics of high energy particle collision and resulting shower

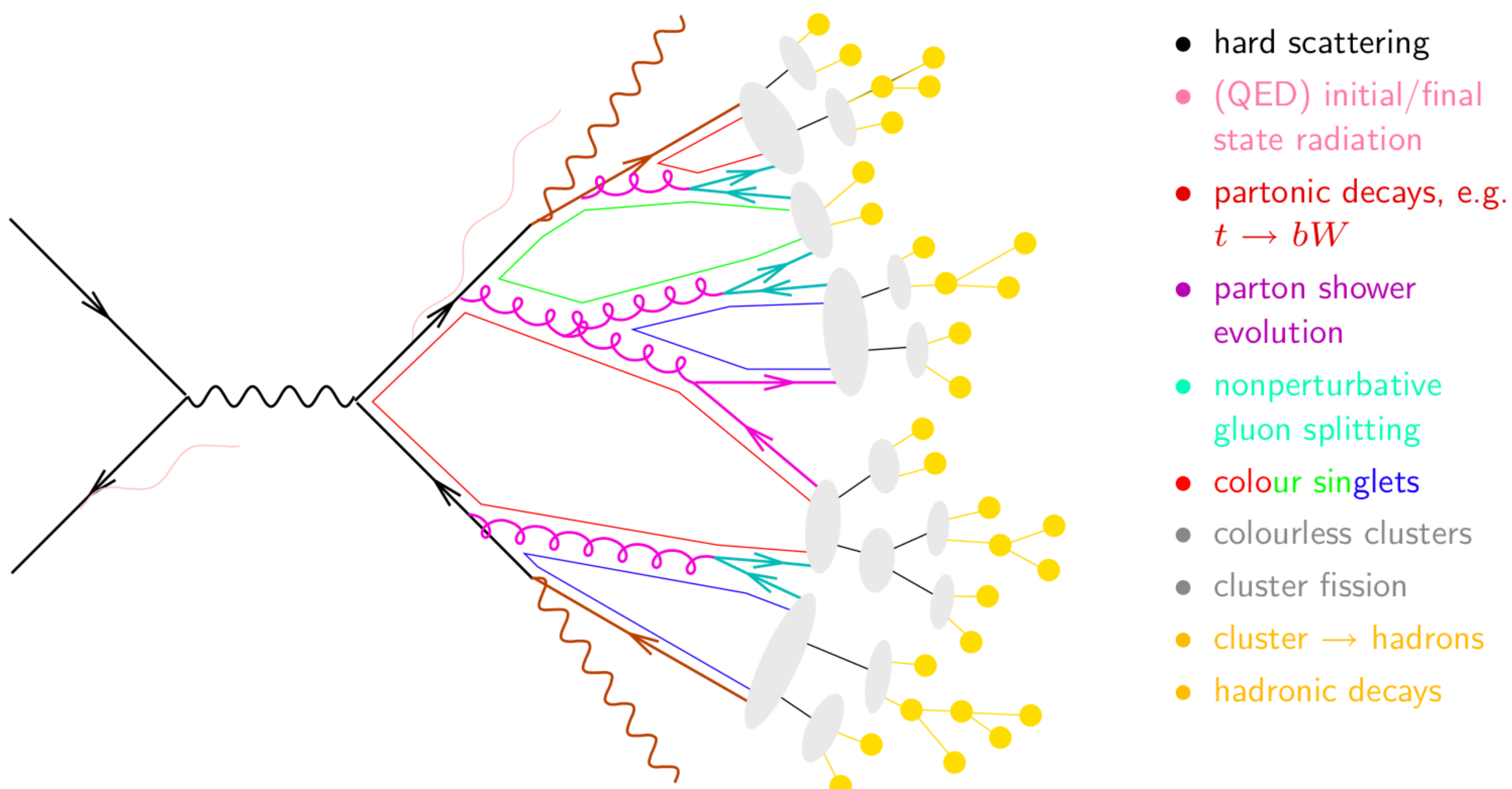
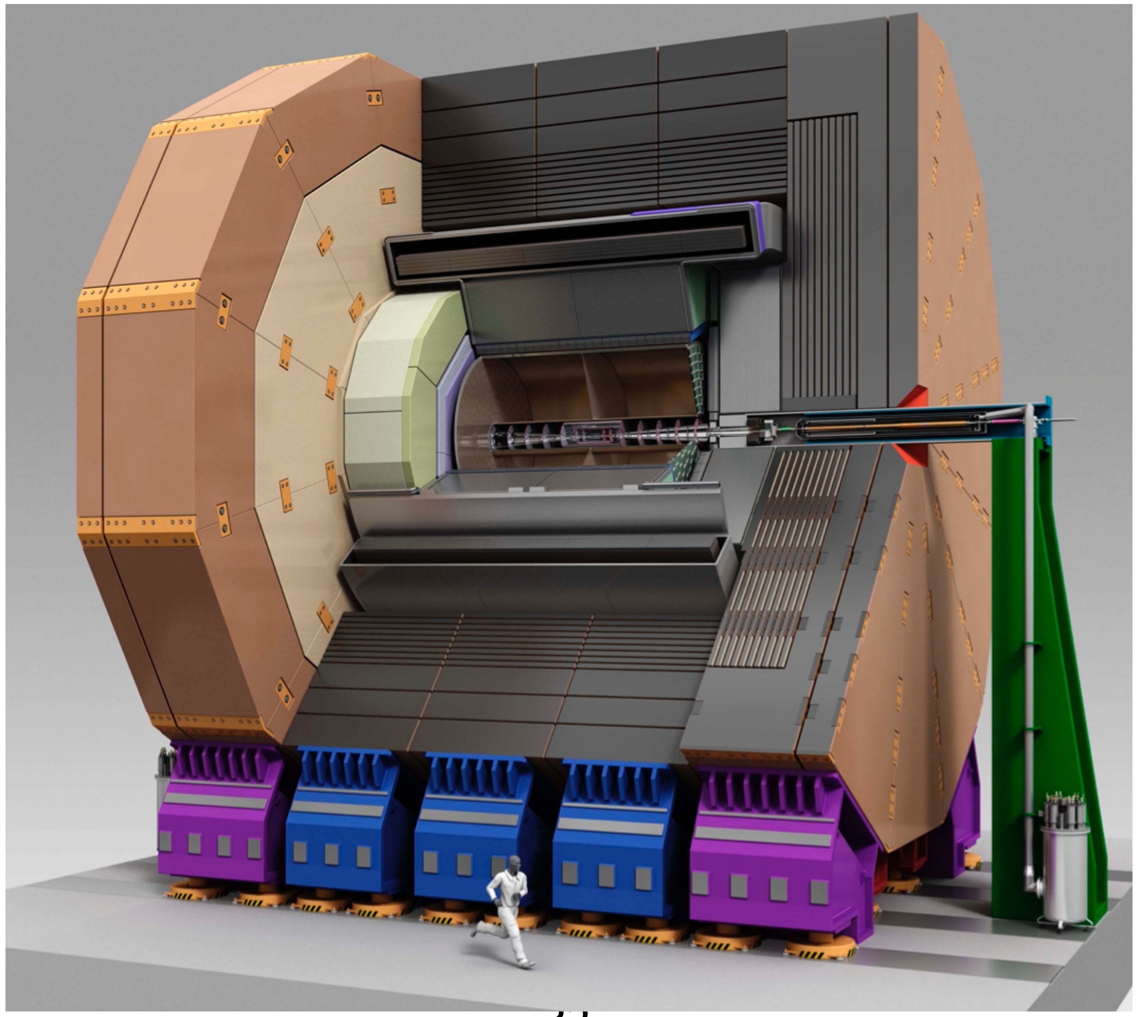


Figure from D. Zeppenfeld.

Context

- Roughly speaking, “simulation” consists of two steps:
  - Event generation
  - Detector simulation
    Describe interaction of particle shower with various detector components on a microscopic level



Context

- Roughly speaking, “simulation” consists of two steps:
  - Event generation
  - Detector simulation

- Both are computationally expensive, performed by a multitude of specific software packackes and ML-based efforts exist to replace/augment them

- Potential benefits:
  - Reduce ressource consumption (details in JRs talk)
  - On-the-fly data generation
  - Simulation trained directly on data (reduce modelling uncertainty?)
  - New analysis techniques utilising fully differentiable (invertible?) generators

Context

- Roughly speaking, “simulation” consists of two steps:
  - Event generation
  - Detector simulation

- Both are computationally expensive, performed by a multitude of specific software packackes and ML-based efforts exist to replace/augment them

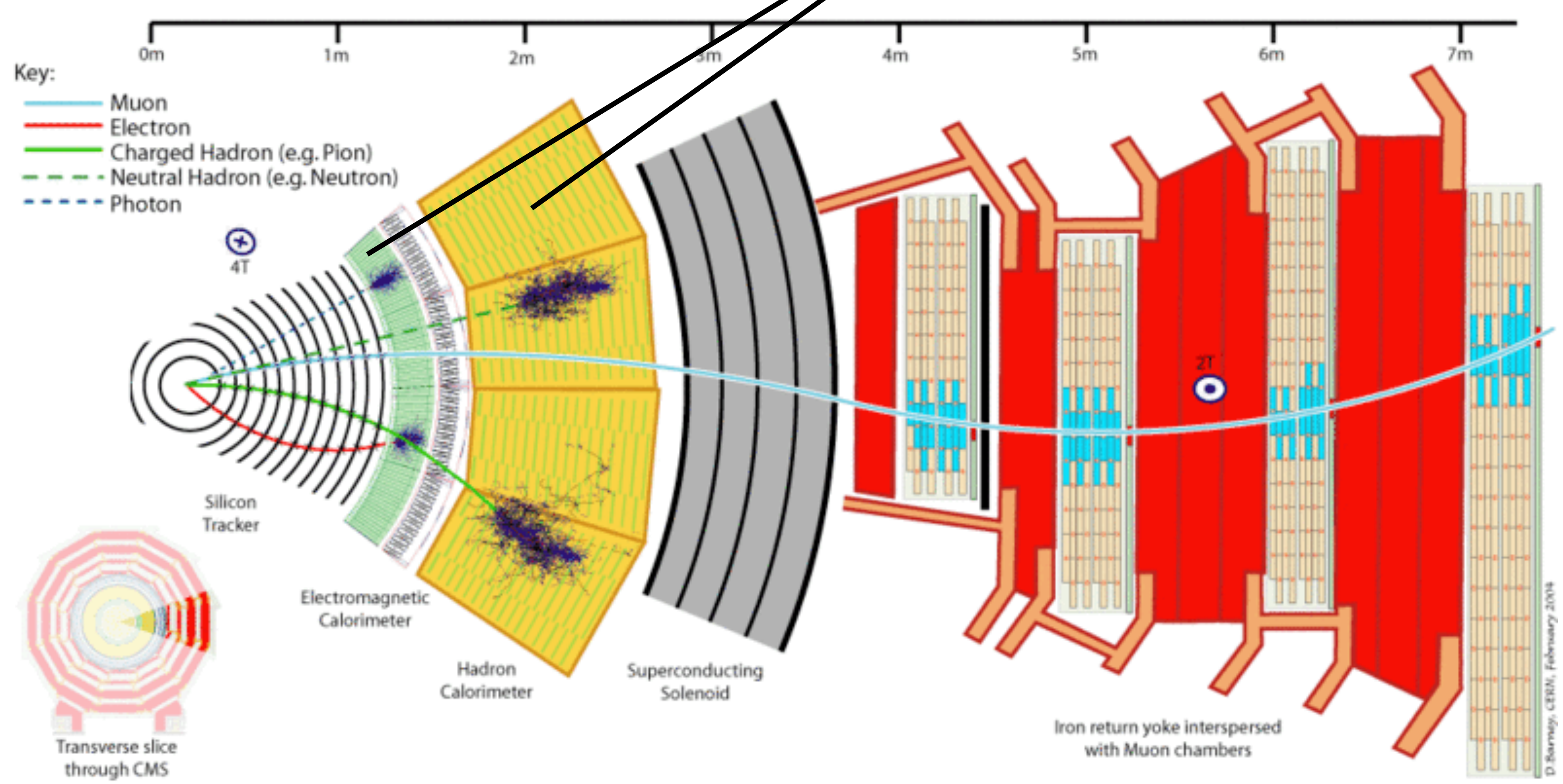
- Potential benefits:
  - Reduce ressource consumption (details in JRs talk)
  - On-the-fly data generation
  - Simulation trained directly on data (reduce modelling uncertainty?)
  - New analysis techniques utilising fully differentiable (invertible?) generators

- **Focus on detector (calorimeter) simulation in the following**

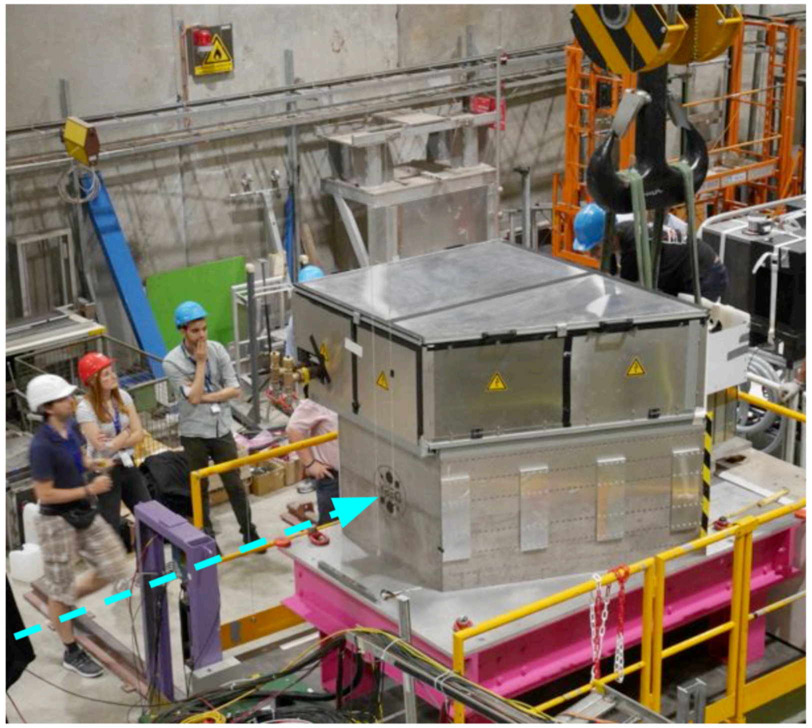
Calorimeter Showers

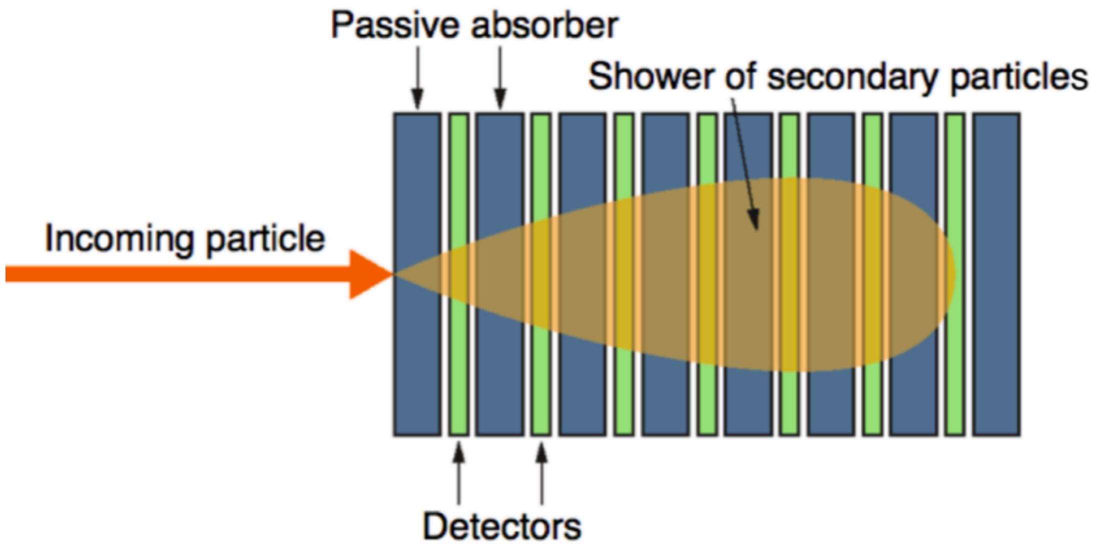
- Calorimeters aim to fully stop incoming particles, and measure their energy in the process
- Due to large amount of classical simulation time spent on calorimeters, good target for ML-based simulation.
- Started by 1712.10321, MANY results since

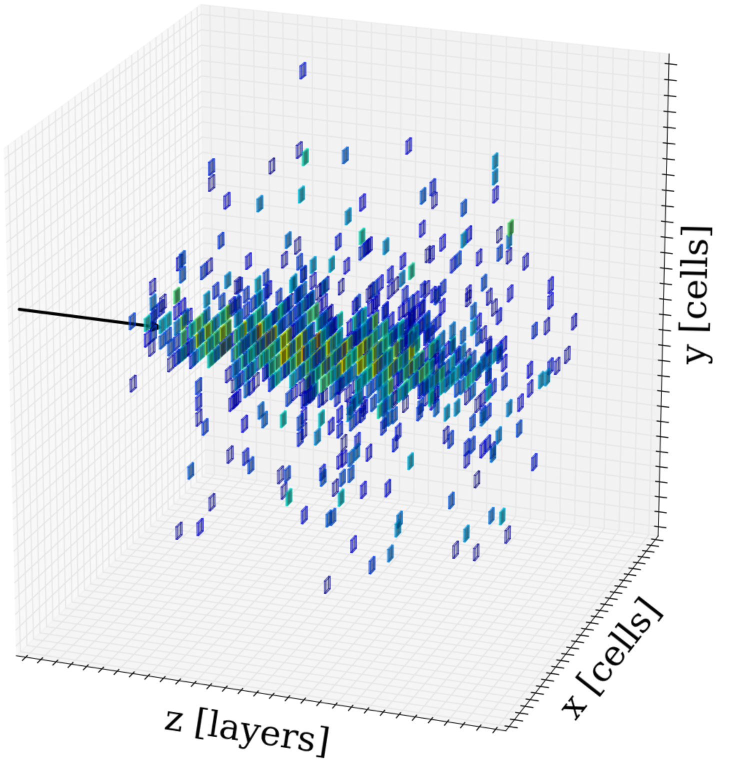
Calorimeters of the CMS detector



Calorimeter Showers







CALICE AHCal testbeam.
(Slightly different detector, but close enough)

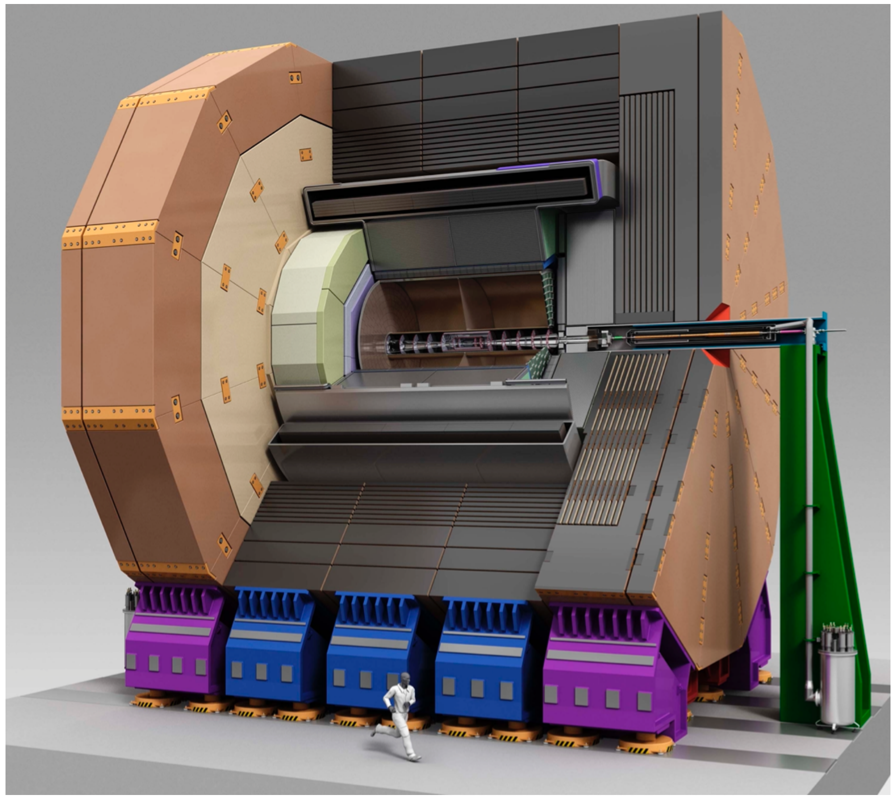
Illustration of particle shower in a sampling calorimeter.

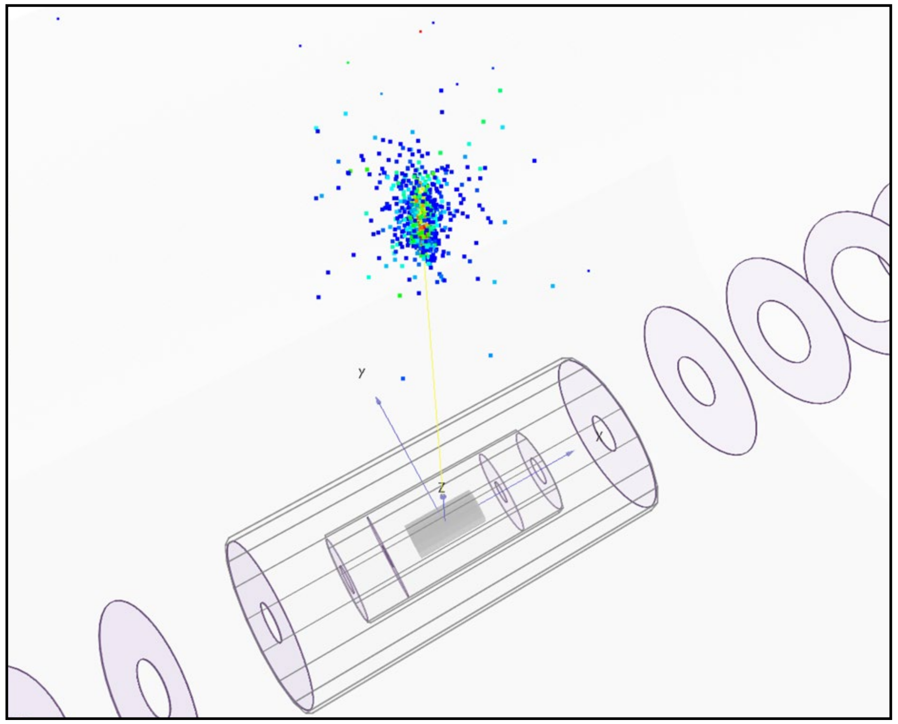
One data example.

Concrete Problem

Describe photon showers in high granularity calorimeter segment

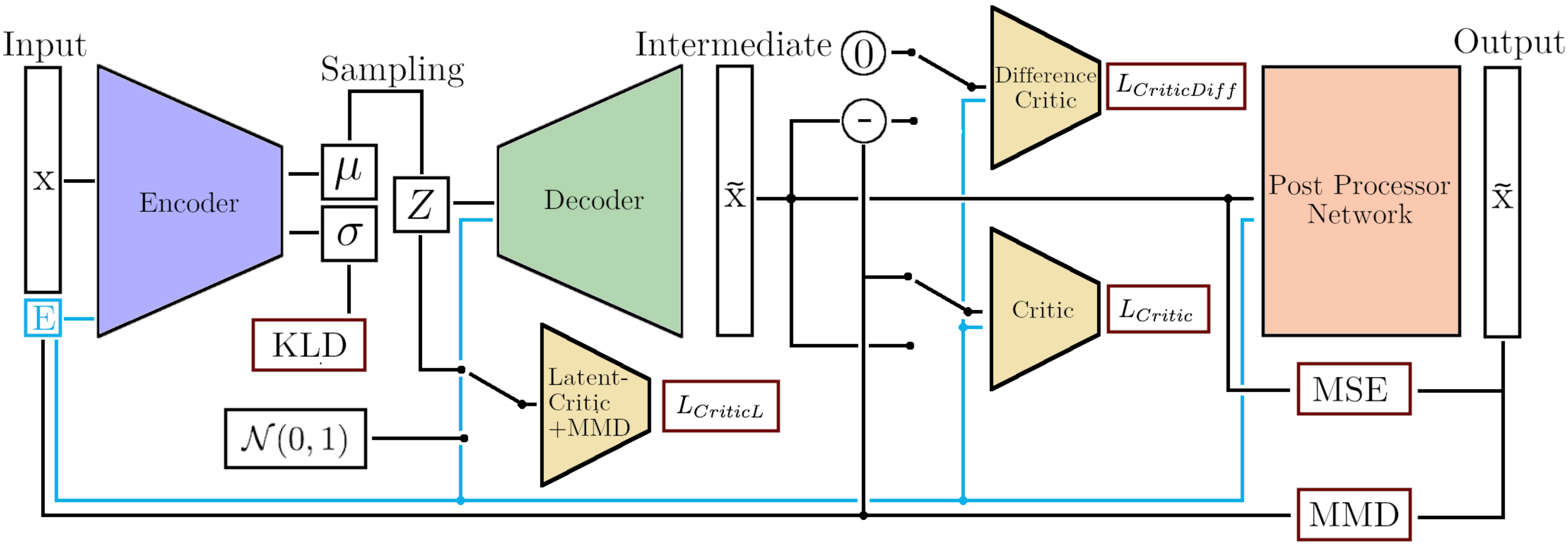
- Model energy in 30x30x30 (=27k) cells (pixels): *grayscale images*
- Incoming photon energies from 10 to 100 GeV: *need to condition on this*
- Consider fixed geometric area of detector
- Use ~1M examples from classical simulation as training data





2005.05334

Architecture

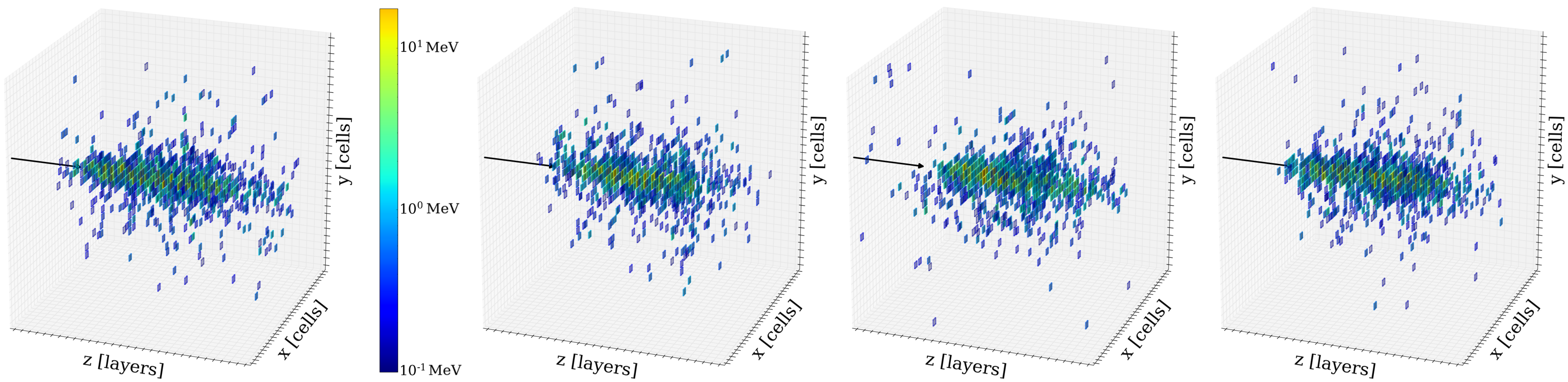


- Bounded Information Bottleneck Autoencoder (BIB-AE, based on 1912.00830)
- Unifies features of GAN and VAE
- 71M trainable parameters

$$
\begin{aligned}
L_{\text{BIB-AE}} = &-\beta_{C_L} \cdot \mathbb{E}[C_L(E(x))] \\
&-\beta_C \cdot \mathbb{E}[C(D(E(x)))] \\
&-\beta_{C_D} \cdot \mathbb{E}[C_D(D(E(x)) - x)] \\
&+\beta_{\text{KLD}} \cdot \text{KLD}(E(x)) \\
&+\beta_{\text{MMD}} \cdot \text{MMD}(E(x), \mathcal{N}(0,1)).
\end{aligned}
$$

2005.05334

Results

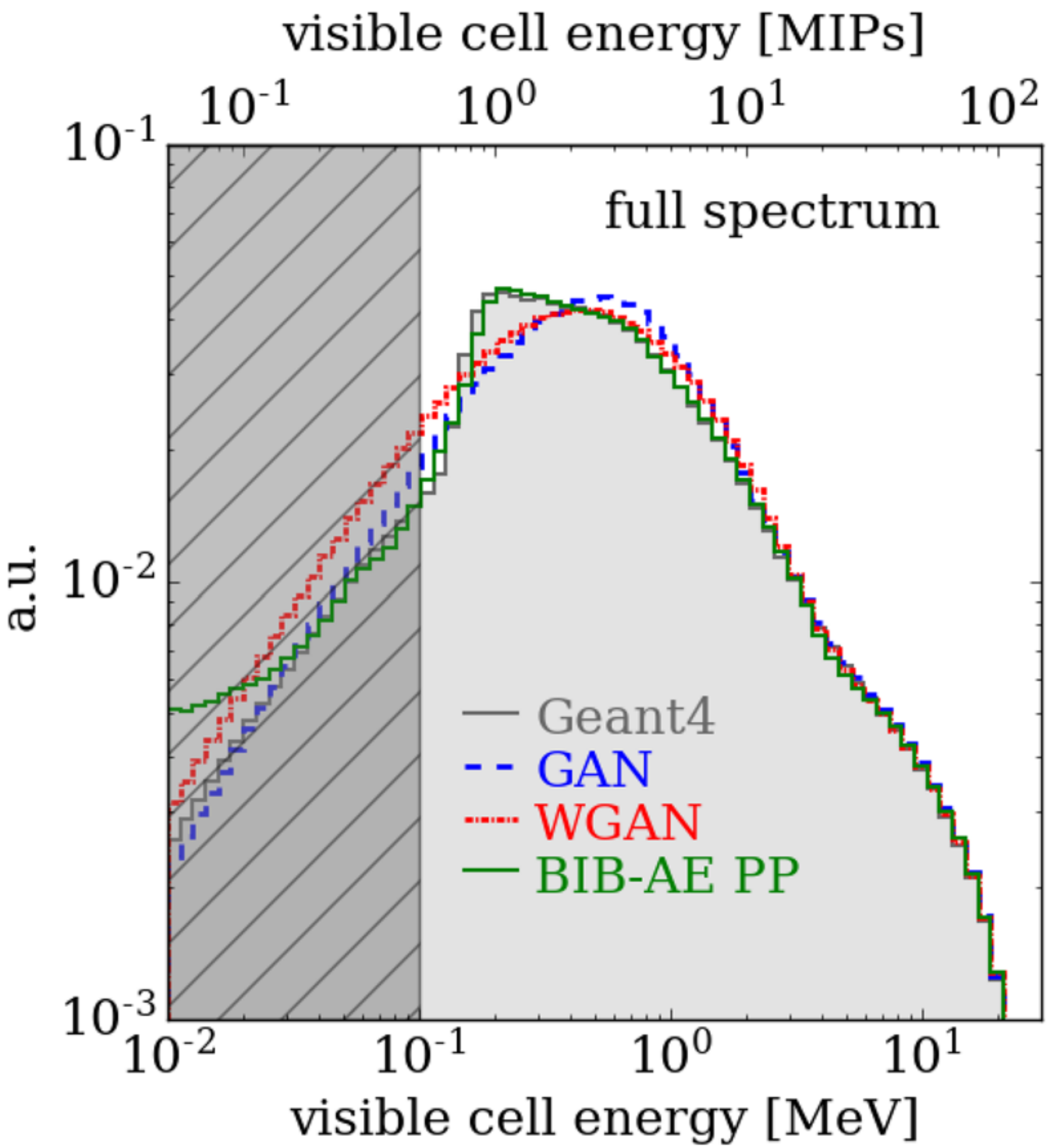


Ground Truth **GAN** **Wasserstein GAN** **BIB-AE**

Individual shower images very hard to judge *per-eye*

2005.05334

Results



- Different from e.g. photographs, there is a number physically meaningful quantities
- Use to judge quality of simulated data
- BIB-AE first model to correctly model cell-energy distribution *(histogram of pixel values)* correctly
- (And of course other marginal distributions and correlations)

2005.05334

Generative Frontiers

- Good progress in various directions

Generative Frontiers

- Good progress in various directions
- Still many issues to be solved:
  - Experimental integration of simulation for high-granularity calorimeters
  - Multi-dimensional conditioning
  - Whole calorimeter simulation
  - Irregular geometries
  - Benchmarking
  - …

Generative Frontiers

- Good progress in various directions

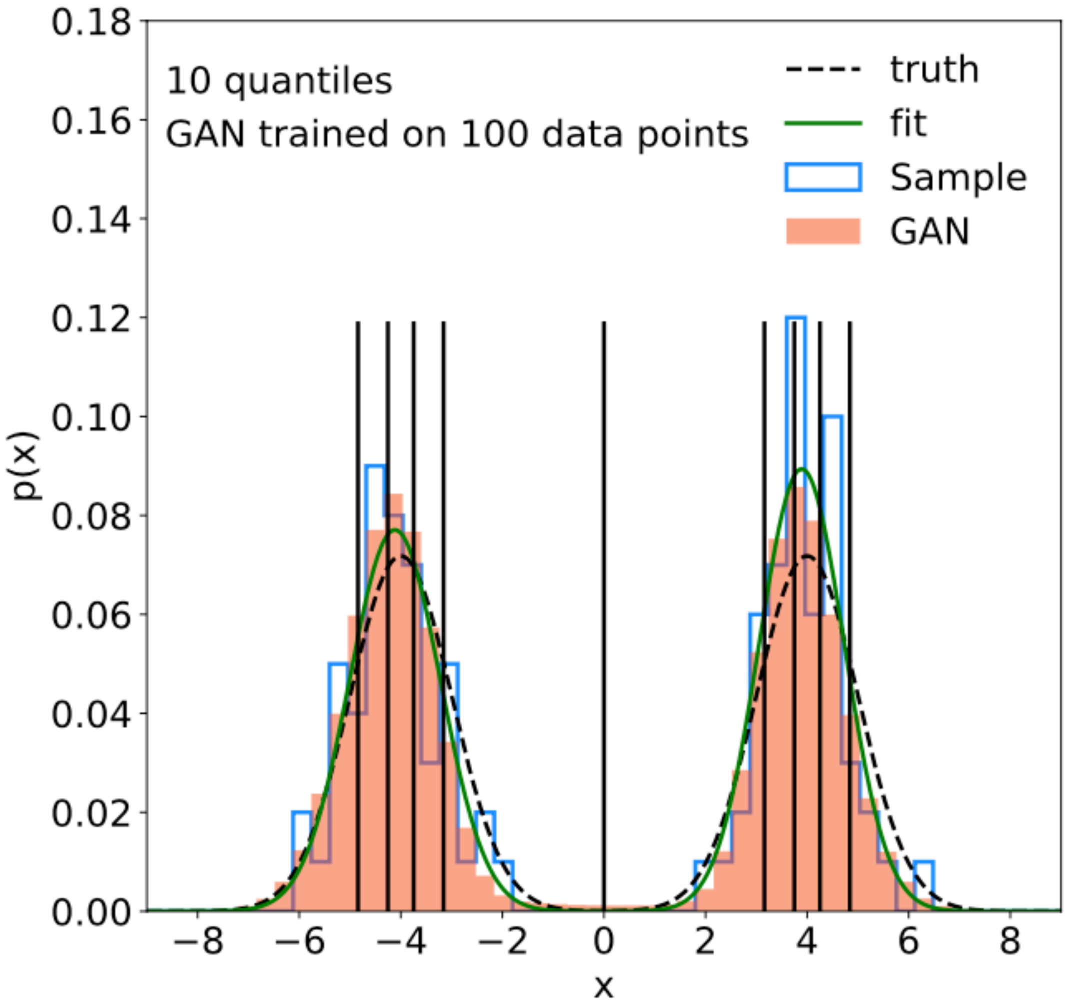
- Still many issues to be solved:
  - Experimental integration of simulation for high-granularity calorimeters
  - Multi-dimensional conditioning
  - Whole calorimeter simulation
  - Irregular geometries
  - Benchmarking
  - …

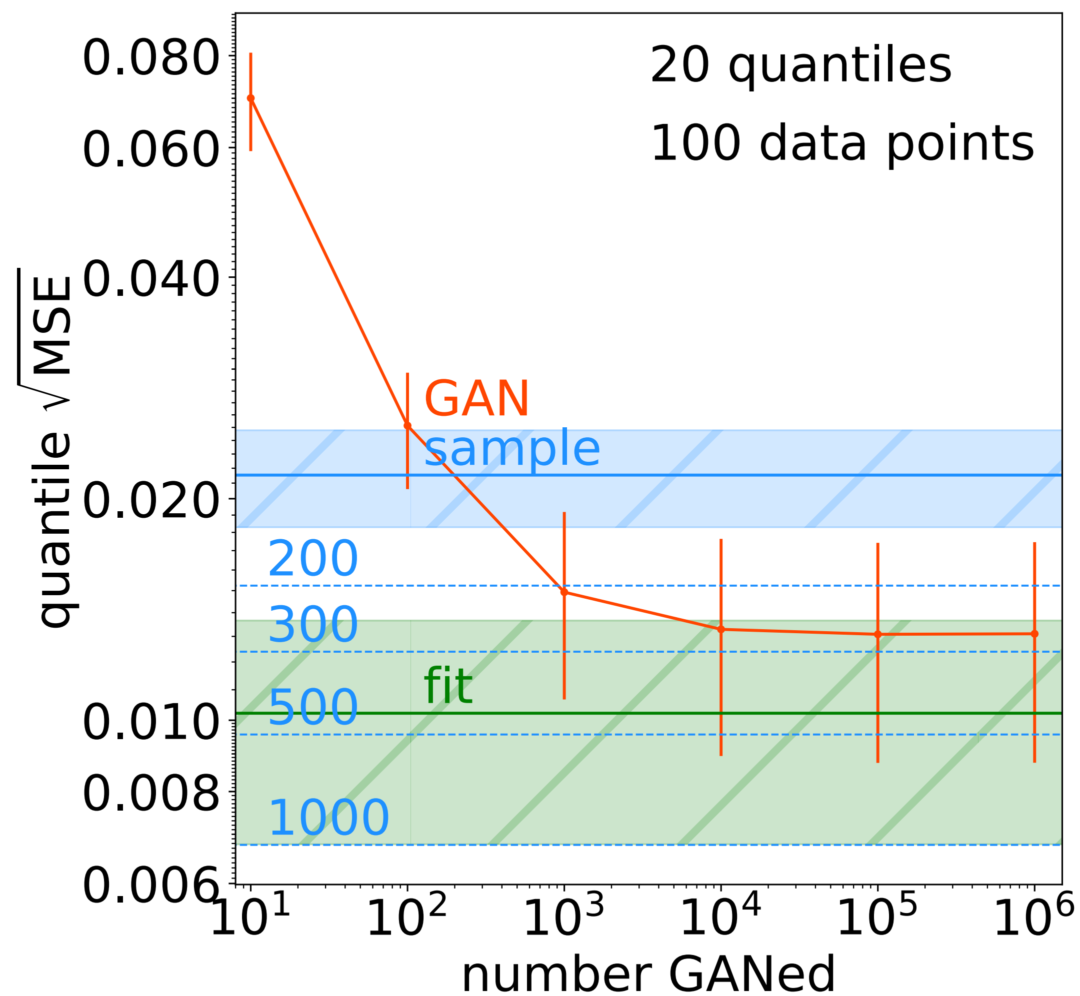
- Statistics?

Statistics

If we train a generator on N data points, and use it to produce M>>N examples, what is the statistical power of the M points?

Compare (known) truth distribution to sample and oversampled data from GAN

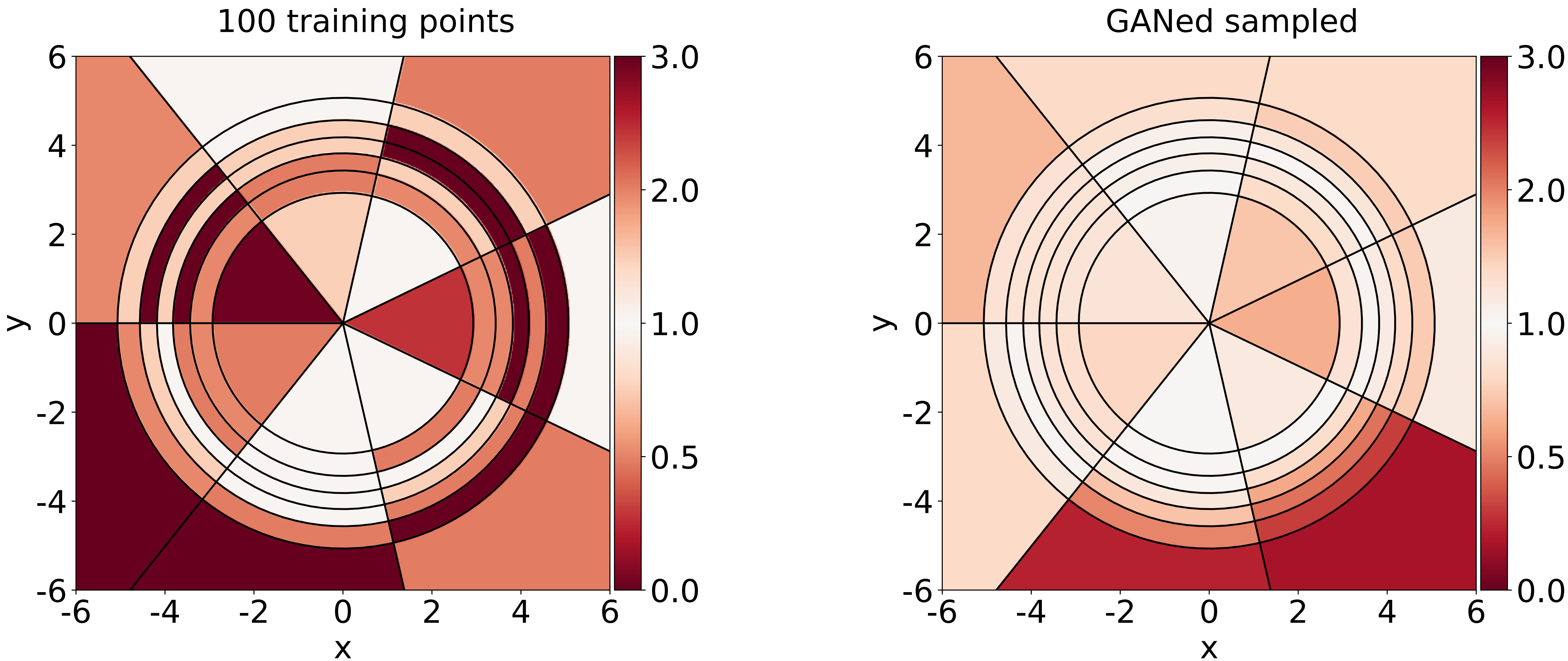




$$\mathrm{MSE} = \frac{1}{N_{\mathrm{quant}}} \sum_{j=1}^{N_{\mathrm{quant}}} \left(x_j - \frac{1}{N_{\mathrm{quant}}} \right)^2$$

2008.06545

Statistics - 2D

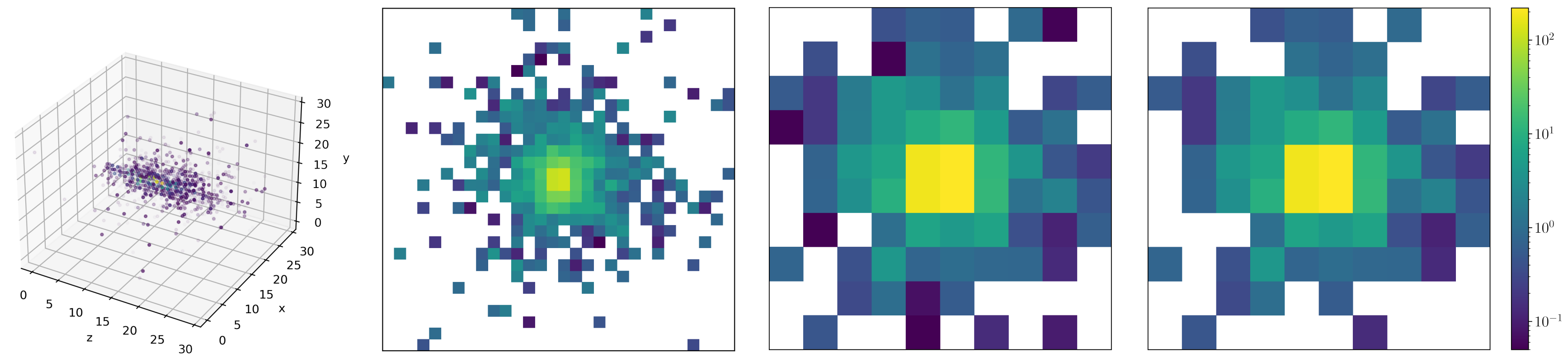


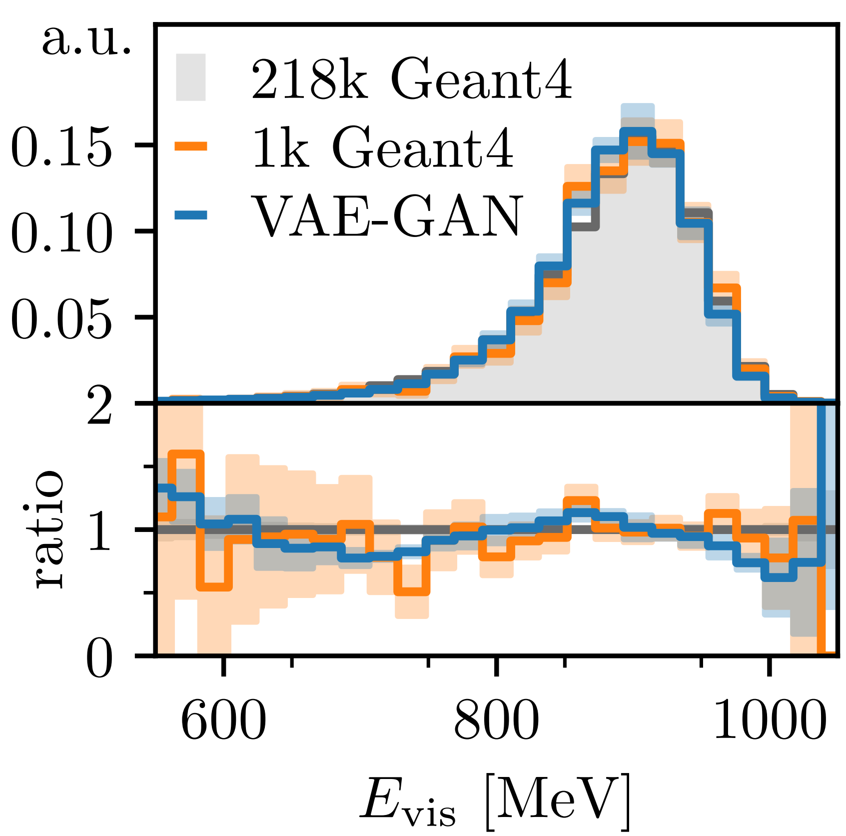
Relative deviation from Gaussian ring distribution

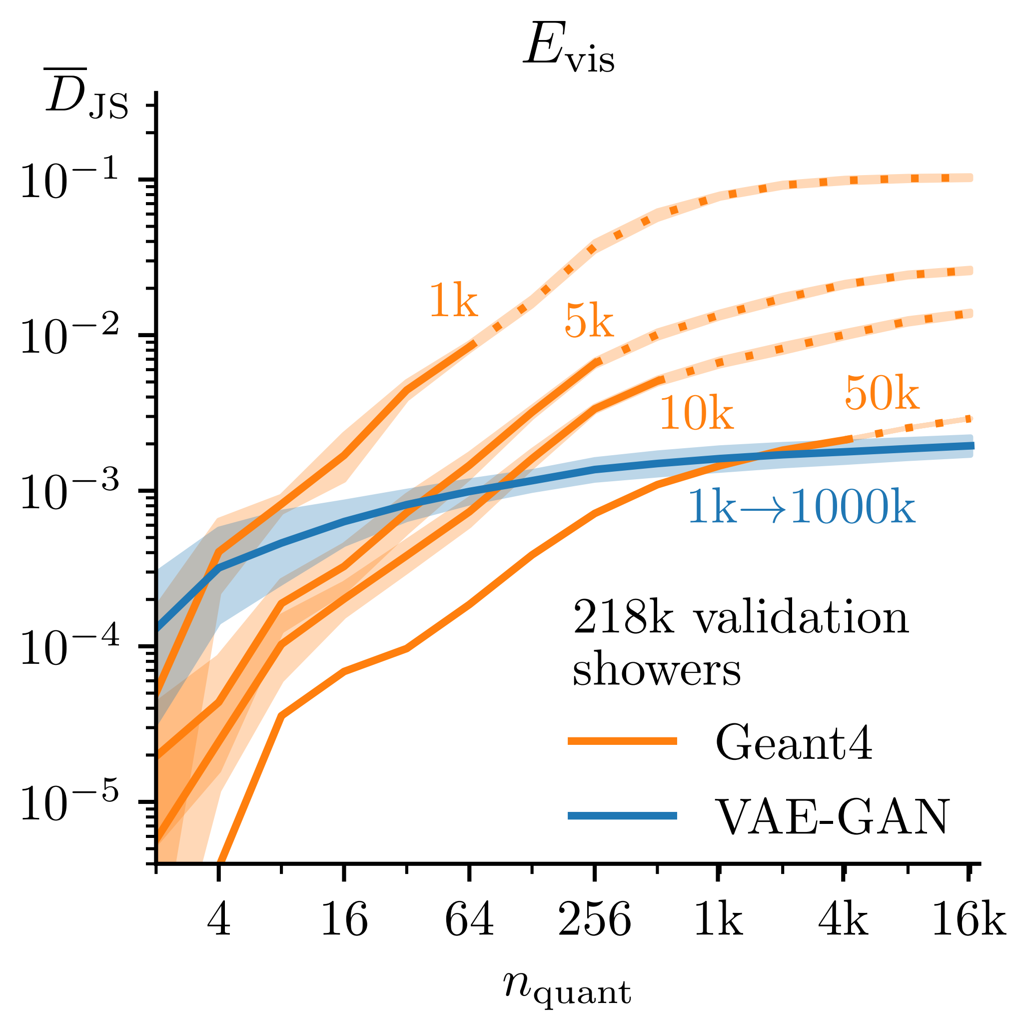
2008.06545

Statistics - Physics

Test the statistical properties of simplified calorimeter showers.







Scaling of difference to ground truth with resolution again better for the generative model.

2202.07352

Generative Frontiers

- Good progress in various directions
- Still many issues to be solved:
  - Experimental integration of simulation for high-granularity calorimeters
  - Multi-dimensional conditioning
  - Whole calorimeter simulation
  - Irregular geometries
  - Benchmarking
  - …
- Statistics
- Quality of simulation

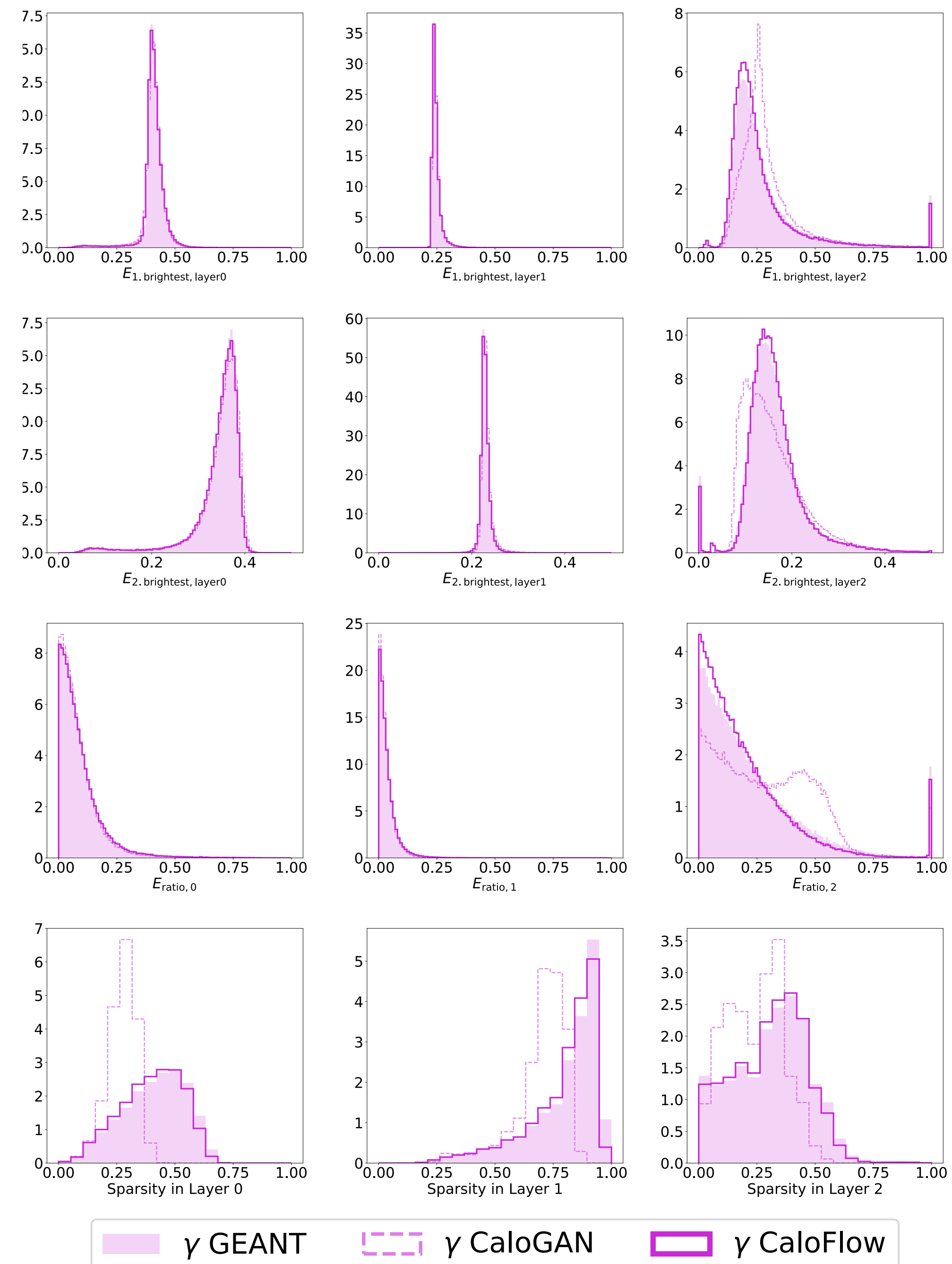
Flows for generation

- So far, only discussed GAN/VAE based approaches to calorimeter simulation

- Can also attempt to simulate with flows

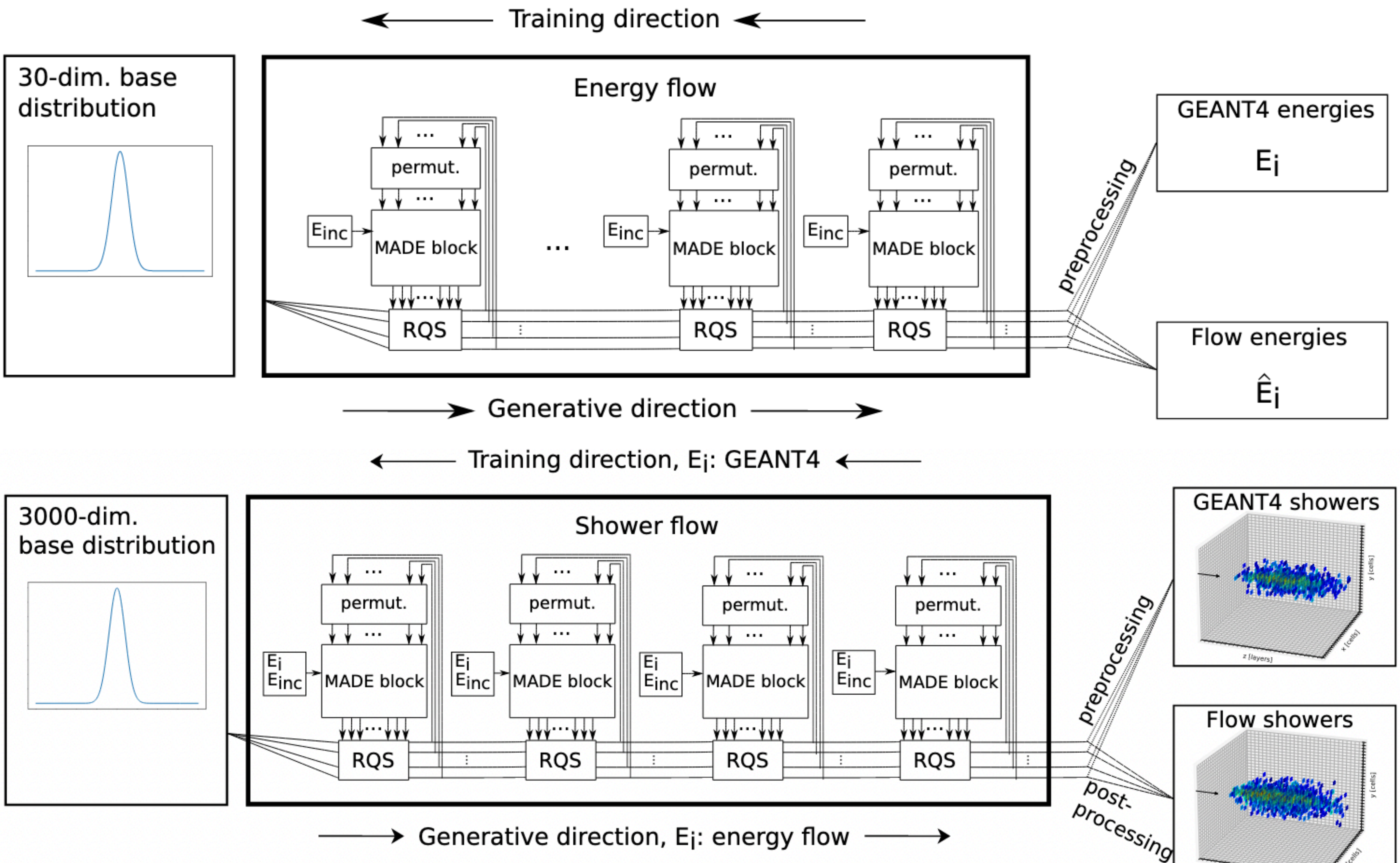
- Issue: As the flows are bijective, dim(latent space) = dim(data space)
- *This is bad*

- CaloFlow improves the performance on simple calorimeter data (1712.10321) by training a two-step MAF-based density estimator: Flow 1 learns energy/layer, Flow 2 learns to distribute this energy
- CaloFlow II speeds up evaluation by training another flow type
  - Student/teacher training an IAF (inverse autoregressive flow) on the MAF
  - Sampling from the IAF



2106.05285, 2110.11377 (Krause & Shih)

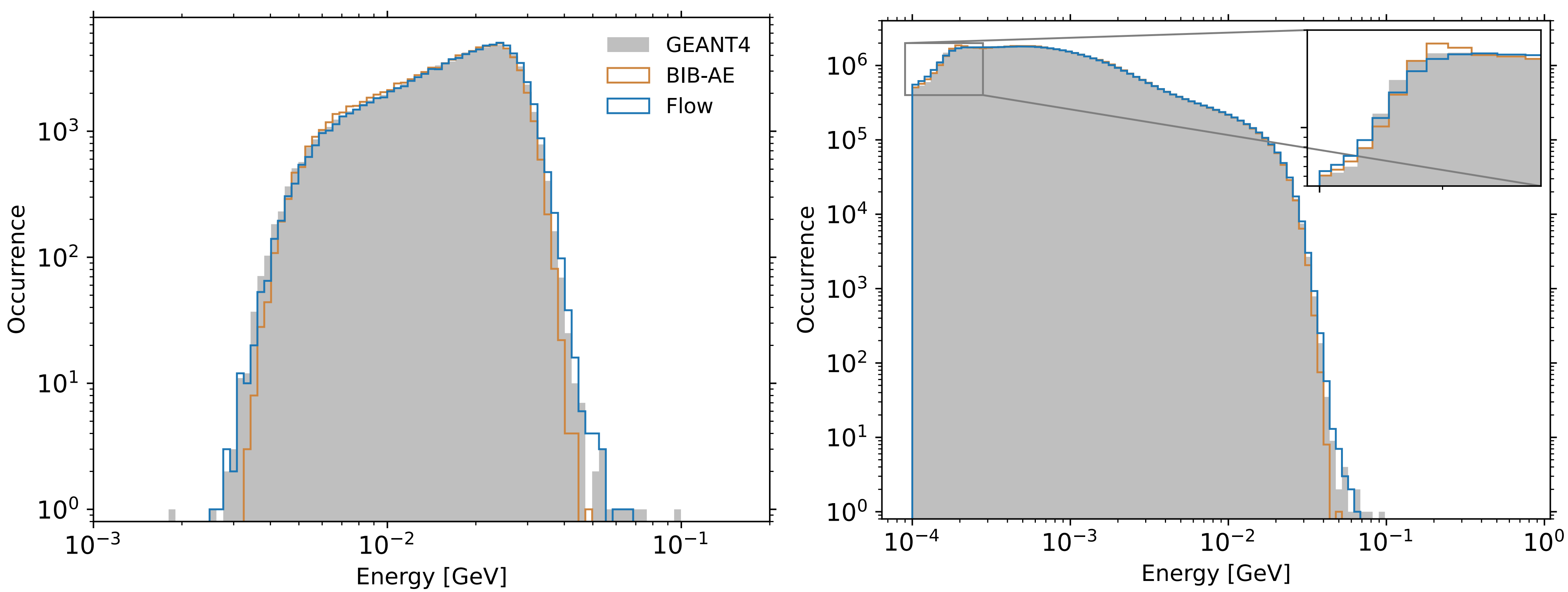
Flows



Further challenges when extending to higher dimensions..

Diefenbacher, Kasieczka, Krause, **Shekhzadeh**, Shih – coming soon

Flows



… but results look very promising.

Diefenbacher, Kasieczka, Krause, **Shekhzadeh**, Shih — coming soon

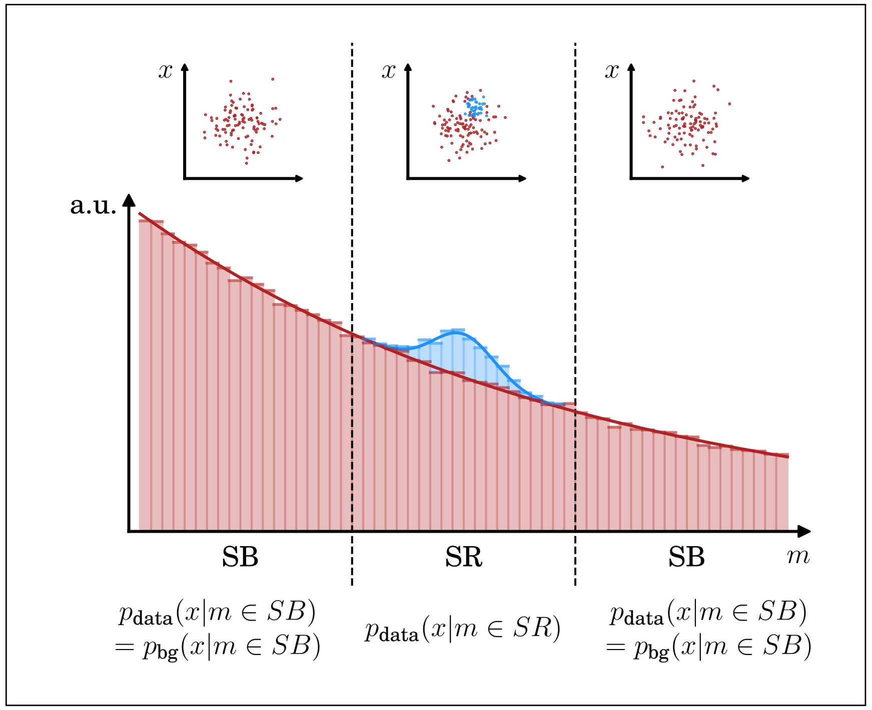
Anomaly Detection

Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

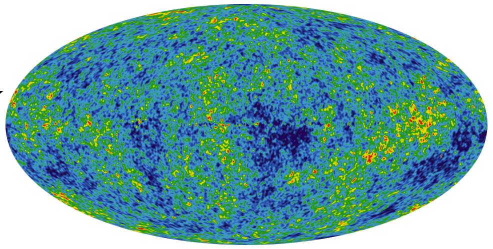


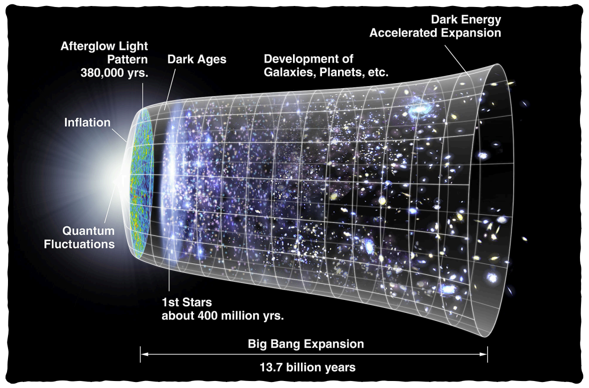
Motivation

- Theoretical and experimental reasons to expect new physics beyond the Standard Model

Why are neutrinos massive?

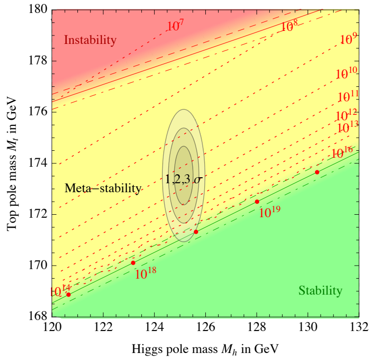
What is the nature of dark matter & dark energy?





Why is there more matter than anti-matter?

What are the origins of the LHCb flavour anomaly?



Why is there more matter than anti-matter?

Is the electroweak vacuum stable?

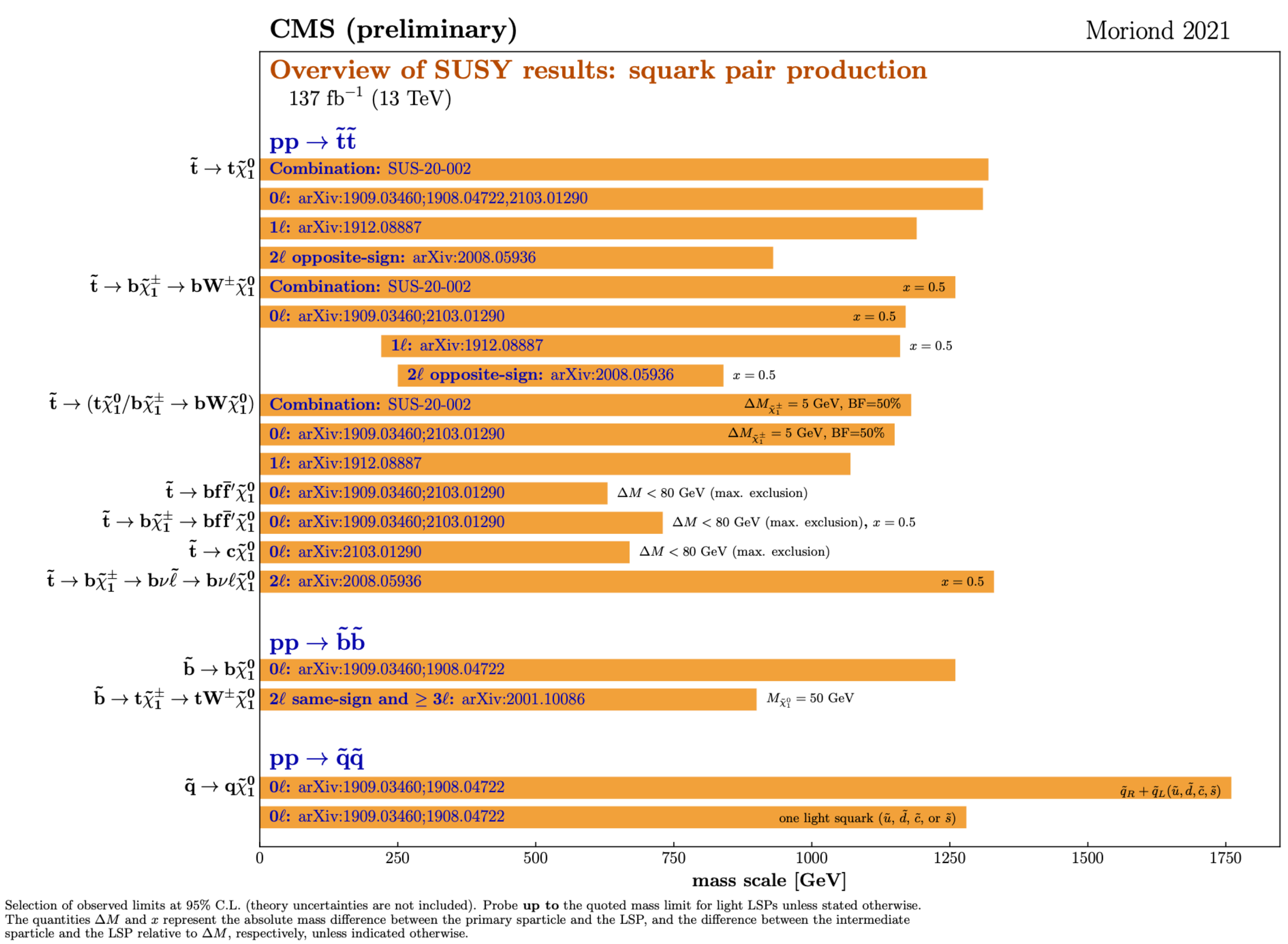
How can the Higgs boson be light when the mass receives large quantum corrections?

What are the details of cosmic inflation?

Motivation

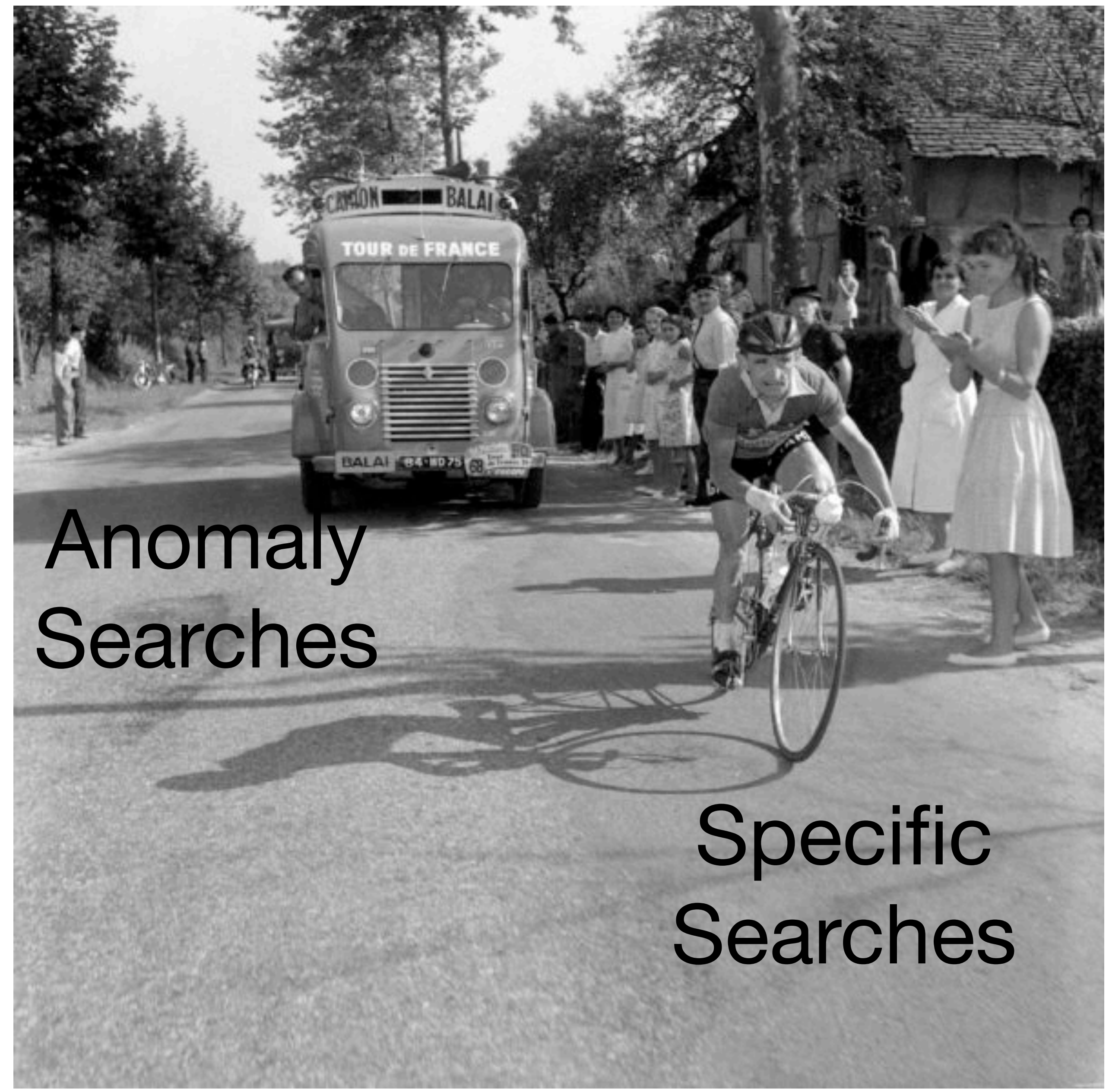
- Theoretical and experimental reasons to expect new physics beyond the Standard Model

- However, so far only negative results in direct (model driven) searches



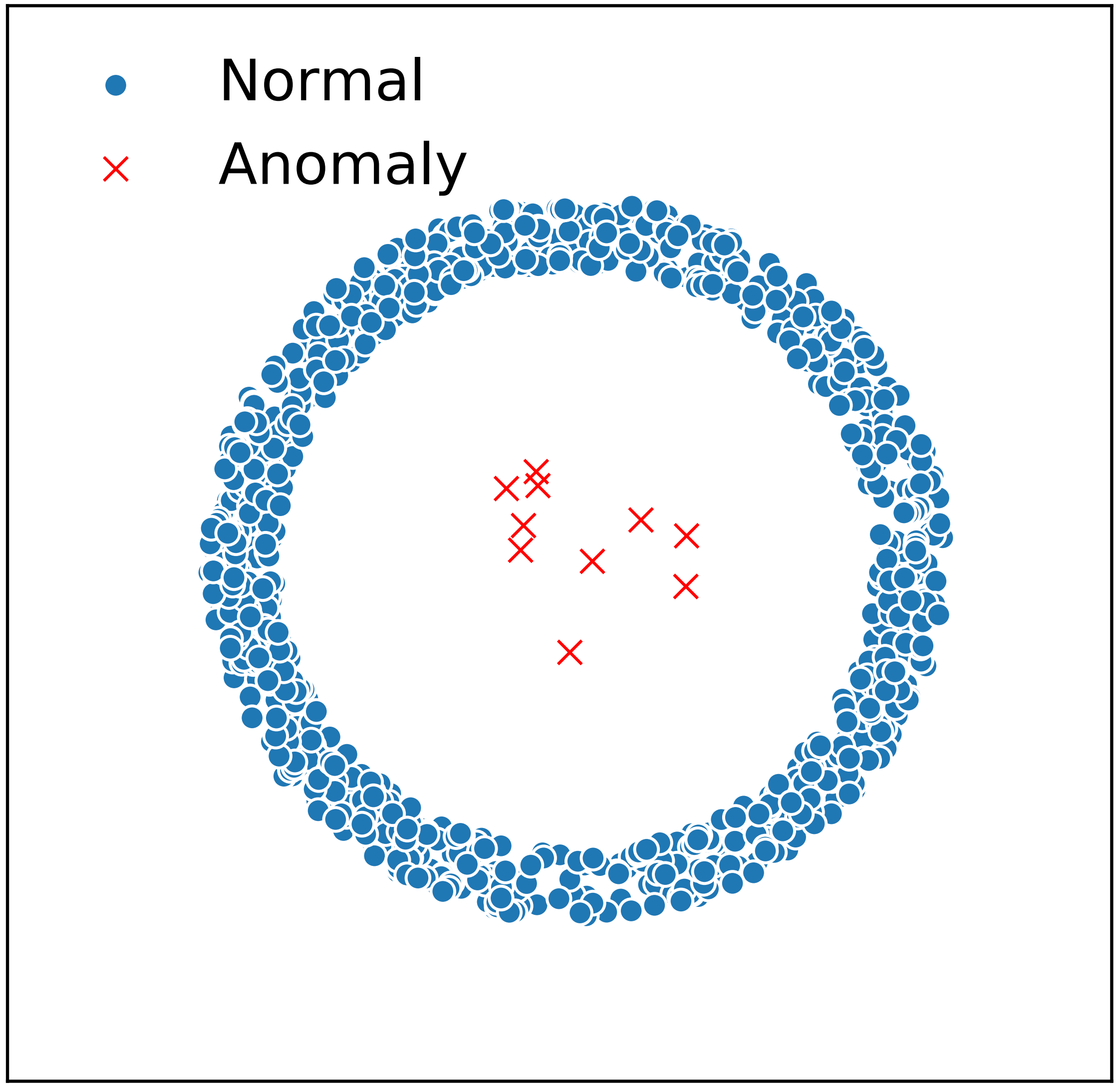
Motivation

- Theoretical and experimental reasons to expect new physics beyond the Standard Model
- However, so far only negative results in direct (model driven) searches
- Make sure that we do not miss potential discoveries at the LHC →**Anomaly detection**



Types of anomalies

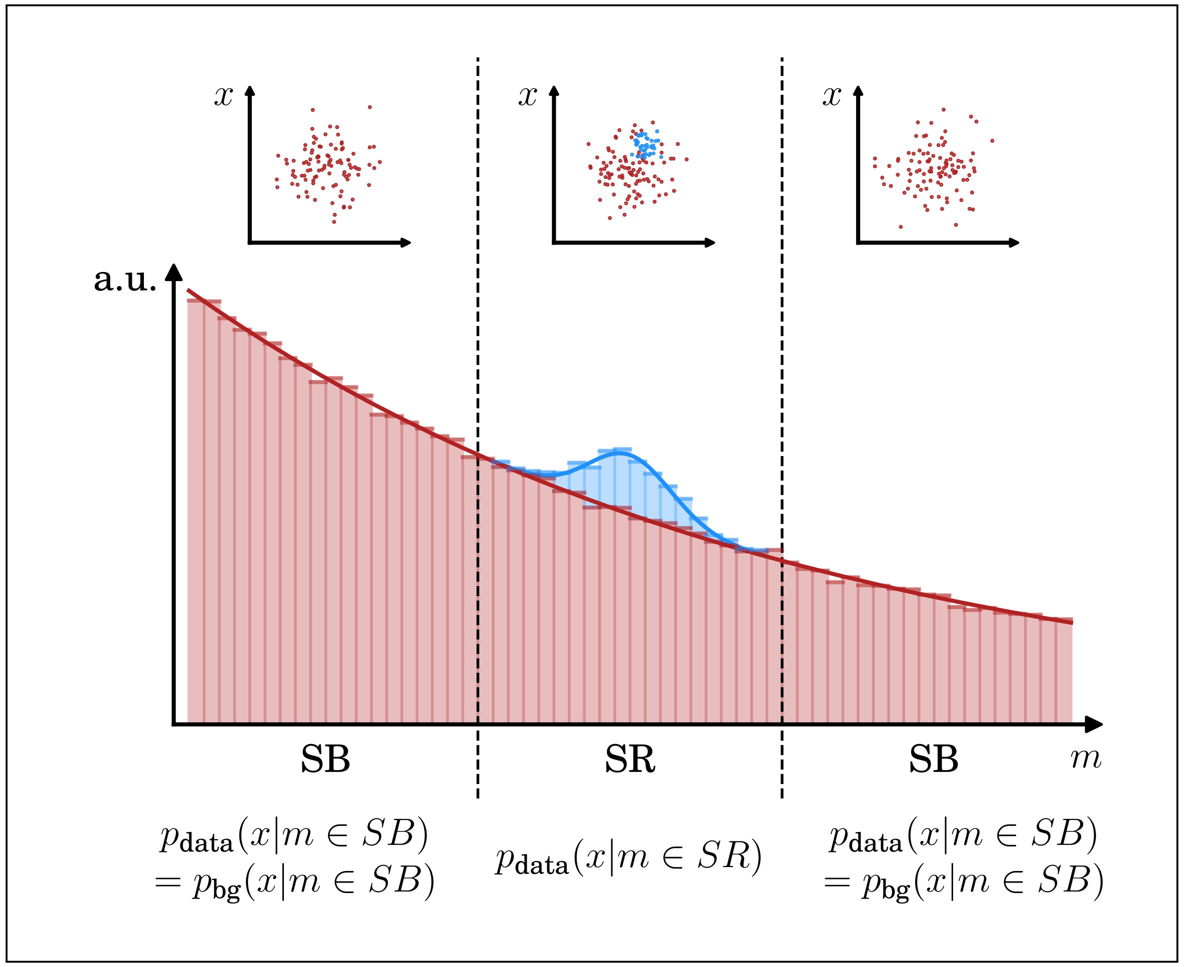
- **Outliers/Point anomalies**: Datapoints far away from regular distribution
- Examples:
  - Detector malfunctions
  - Background-free search



Types of anomalies

- **Outliers/Point anomalies**: Datapoints far away from regular distribution
- Examples:
  - Detector malfunctions
  - Background-free search

- **Group anomlies**: Individual examples not interesting,
but signal is an overdensity with respect to background
- Examples:
  - Resonance searches
  - Transient signals in time series

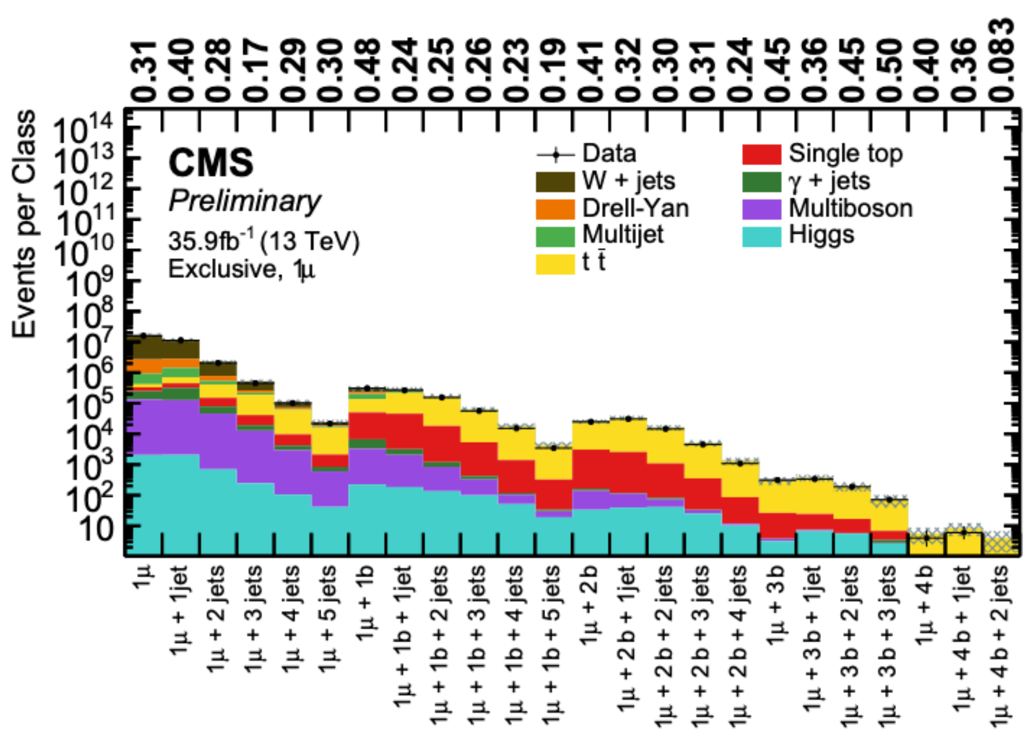


Approaches

Use classical simulation to estimate backgrounds?

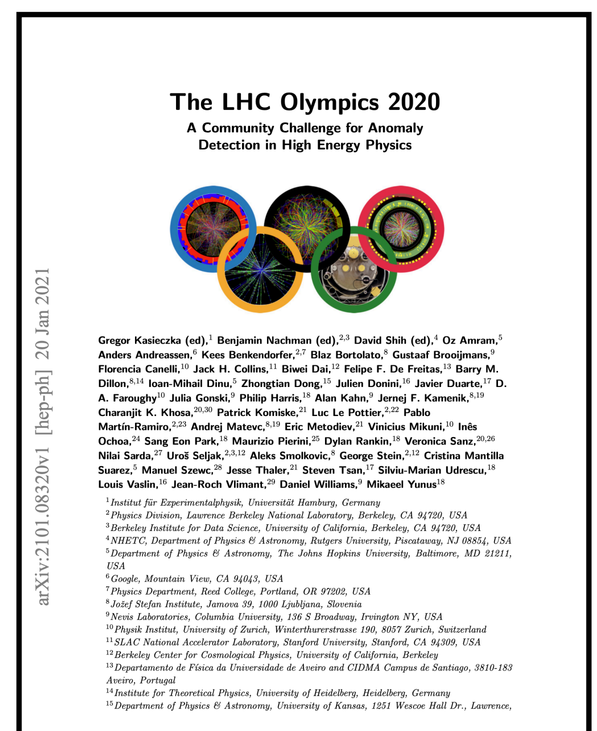
Yes

- Systematically compare simulation and recorded data, look for differences
- Con: Relies on imperfect simulation, Maximally background model dependent
- Pro: Sensitive to all types of anomalies



No

- Estimate background from data
- Con: Need to make assumptions about signal model
- Pro: No reliance on simulation



The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

arXiv:2101.08320v1 [hep-ph] 20 Jan 2021

Gregor Kasieczka (ed),[1] Benjamin Nachman (ed),[2,3] David Shih (ed),[4] Oz Amram,[5] Anders Andreassen,[6] Kees Benkendorfer,[2,7] Blaz Bortolato,[8] Gustaaf Brooijmans,[9] Florencia Canelli,[10] Jack H. Collins,[11] Biwei Dai,[12] Felipe F. De Freitas,[13] Barry M. Dillon,[8,14] Ioan-Mihail Dinu,[5] Zhongtian Dong,[15] Julien Donini,[16] Javier Duarte,[17] D. A. Faroughy[10] Julia Gonski,[9] Philip Harris,[18] Alan Kahn,[9] Jernej F. Kamenik,[8,19] Charanjit K. Khosa,[20,30] Patrick Komiske,[21] Luc Le Pottier,[2,22] Pablo Martín-Ramiro,[2,23] Andrej Matevc,[8,19] Eric Metodiev,[21] Vinicius Mikuni,[10] Inês Ochoa,[24] Sang Eon Park,[18] Maurizio Pierini,[25] Dylan Rankin,[18] Veronica Sanz,[20,26] Nilai Sarda,[27] Uroš Seljak,[2,3,12] Aleks Smolkovic,[8] George Stein,[2,12] Cristina Mantilla Suarez,[5] Manuel Szewc,[28] Jesse Thaler,[21] Steven Tsan,[17] Silviu-Marian Udrescu,[18] Louis Vaslin,[16] Jean-Roch Vlimant,[29] Daniel Williams,[9] Mikaeel Yunus[18]

[1] *Institut für Experimentalphysik, Universität Hamburg, Germany*
[2] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[3] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[4] *NHETC, Department of Physics & Astronomy, Rutgers University, Piscataway, NJ 08854, USA*
[5] *Department of Physics & Astronomy, The Johns Hopkins University, Baltimore, MD 21211, USA*
[6] *Google, Mountain View, CA 94043, USA*
[7] *Physics Department, Reed College, Portland, OR 97202, USA*
[8] *Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*
[9] *Nevis Laboratories, Columbia University, 136 S Broadway, Irvington NY, USA*
[10] *Physik Institut, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[11] *SLAC National Accelerator Laboratory, Stanford University, Stanford, CA 94309, USA*
[12] *Berkeley Center for Cosmological Physics, University of California, Berkeley*
[13] *Departamento de Física da Universidade de Aveiro and CIDMA Campus de Santiago, 3810-183 Aveiro, Portugal*
[14] *Institute for Theoretical Physics, University of Heidelberg, Heidelberg, Germany*
[15] *Department of Physics & Astronomy, University of Kansas, 1251 Wescoe Hall Dr., Lawrence,*

2010.02984

Approaches

Use classical simulation to estimate backgrounds?

Yes ***No***

& many ideas in between!

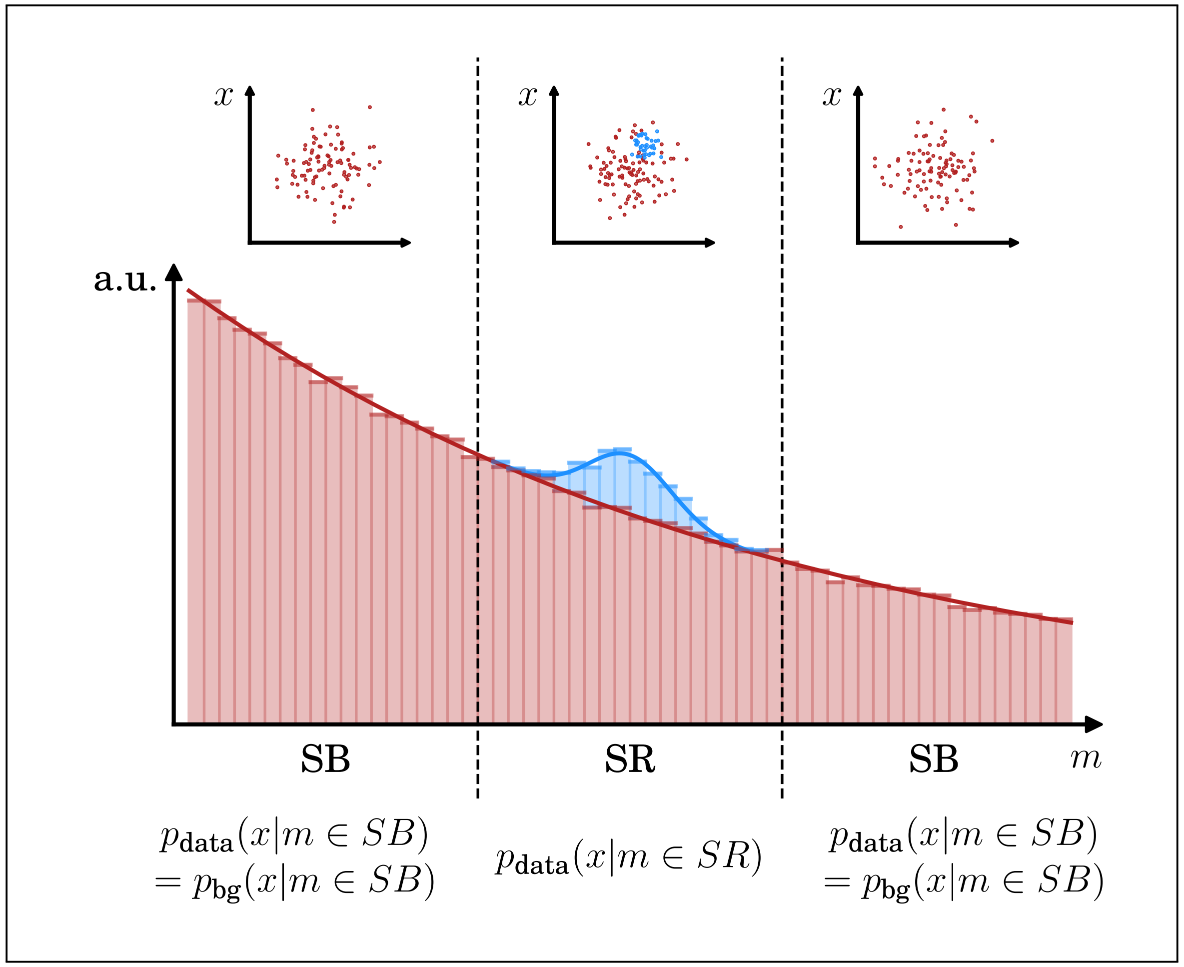
Much more anomaly detection throughout this workshop.

Assumptions

Rarity: Pr(anomaly) ≪ Pr(normal)
Overlap:
max $_x$ $p(x|\text{anomaly})/p\,(x|\text{normal}) < \infty$

Resonance: Pr($|m - m0| > \delta$|anomaly) ≈ 0 for some feature m (often a mass) and fixed $m0$, δ

Smoothness: $p\,(x|m$, normal) varies slowly with m so that one can use data with $|m - m0| > \delta$ to estimate $p(x|m$,normal) for $|m - m0| < \delta$



2101.08320

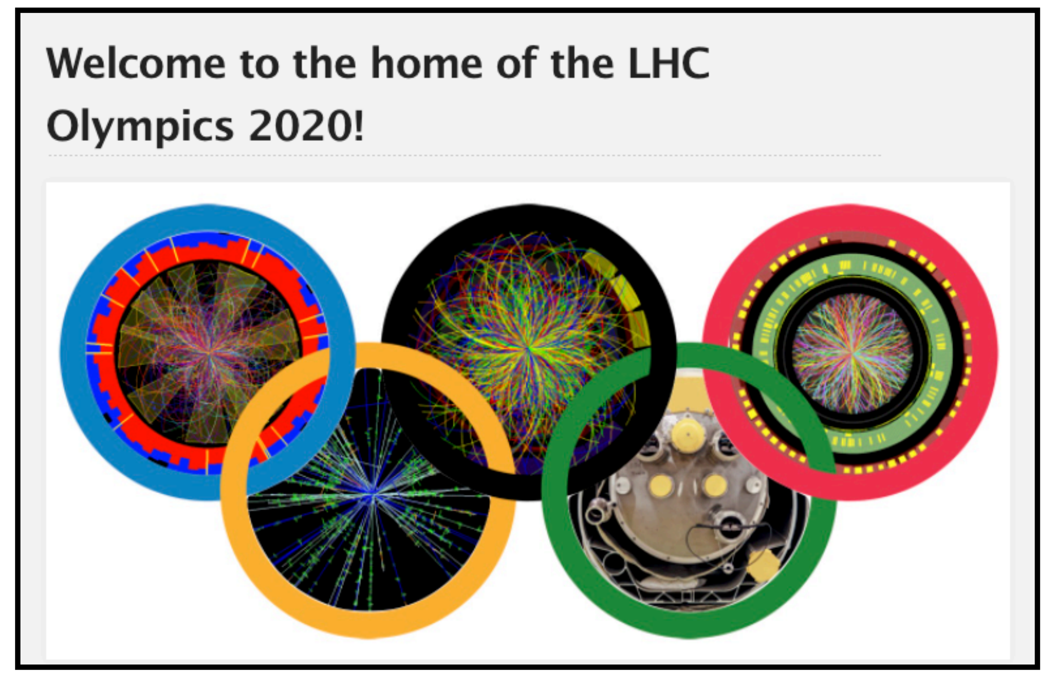
Introducing: LHC Olympics

- Encourage development and comparison of model-agnostic search strategies
    - Focus on group anomalies, data-driven searches
    - Use for a convenient overview of space of techniques
    - Complementary to 2105.14027

- Provide a complete package, balance details vs accessiblity

- Datasets:
    - One R&D dataset for algorithm development
    - Three black box datasets (BB1-BB3)
        - Unblinded over time

- Timeline:
    - Spring 2019: Release R&D dataset (link)
    - Autumn 2019: Release BB datasets (link)
    - January 2020: Winter Olympics as part of ML4Jets, unblinding of BB1 (link)
    - July 2020: (Virtual) Summer Olympics, unblinding of BB2 and BB3 (link)
    - LHC Olympics paper (https://arxiv.org/abs/2101.08320) public



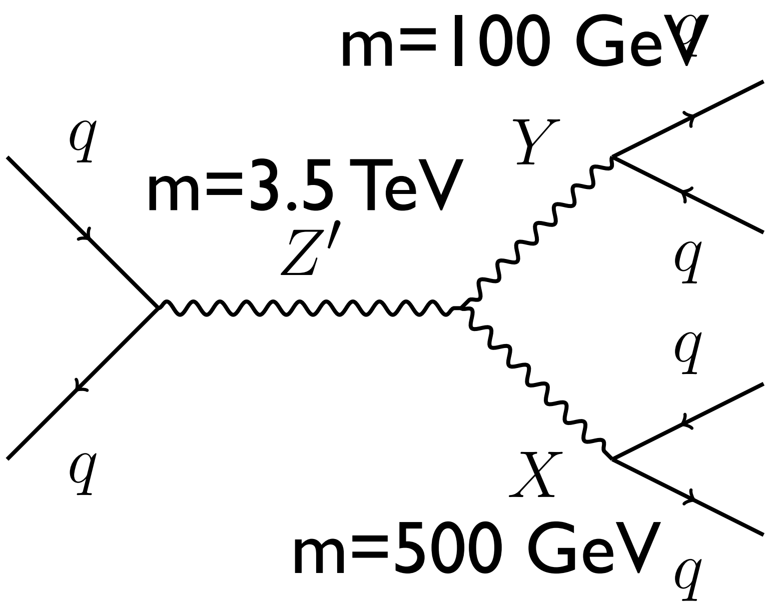
https://lhco2020.github.io/homepage/

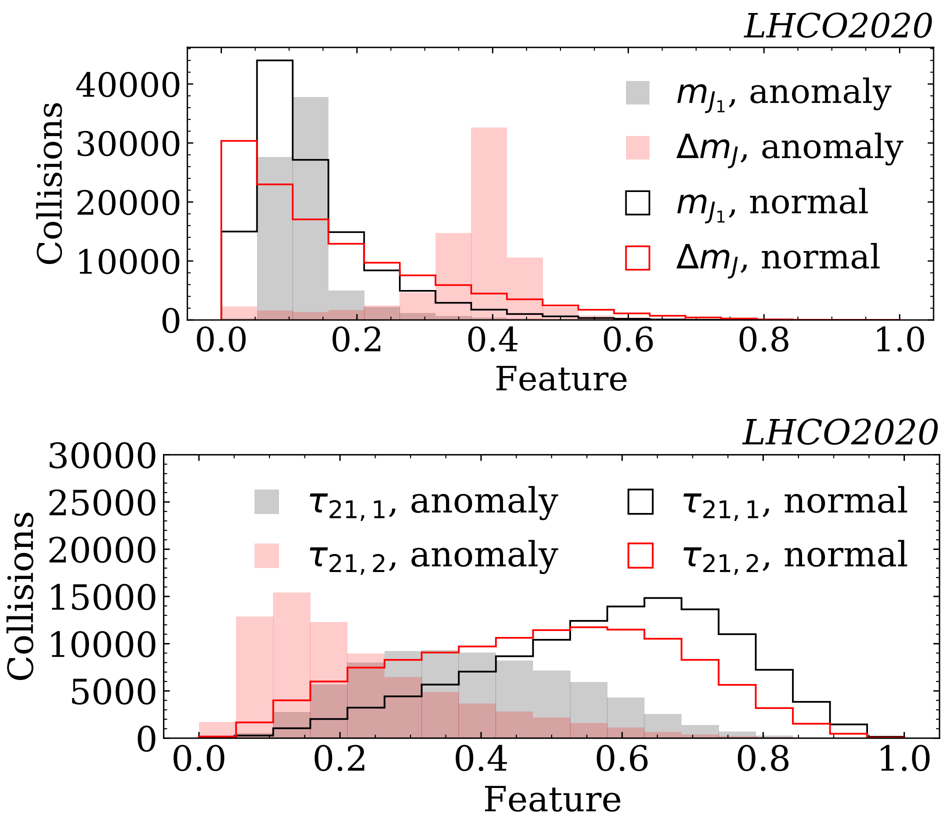
2101.08320

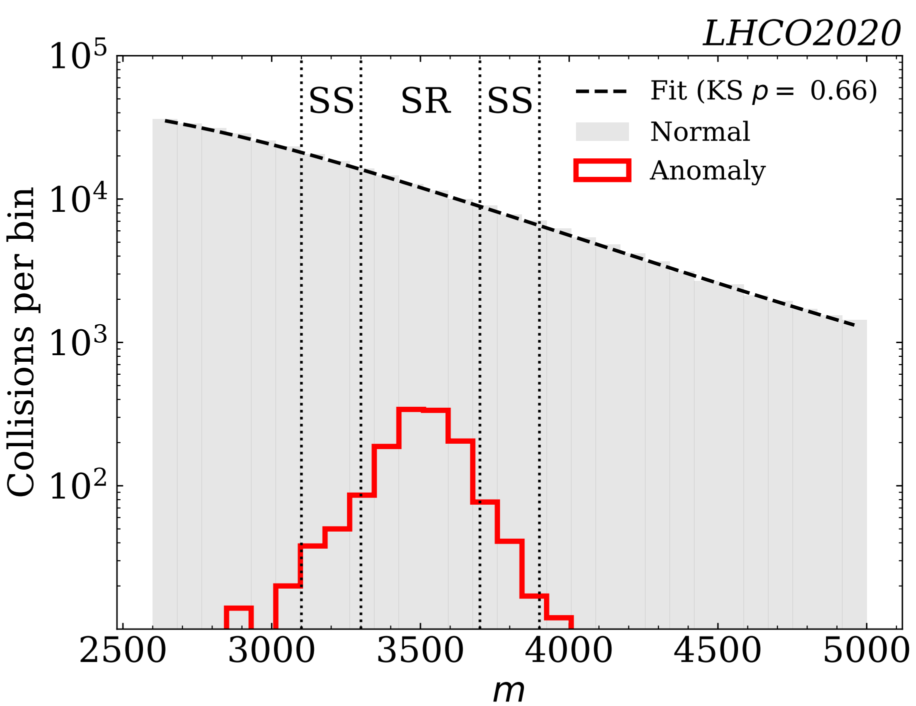
R&D dataset

- For building and testing methods
- 1M background examples (Standard Model), 100k signal examples (signal, see Feynman diagram on the right)
- Labels provided

- Relatively simple signal
  - Known to differ in previously mentioned features from background distribution
- Unrealistically high S/B

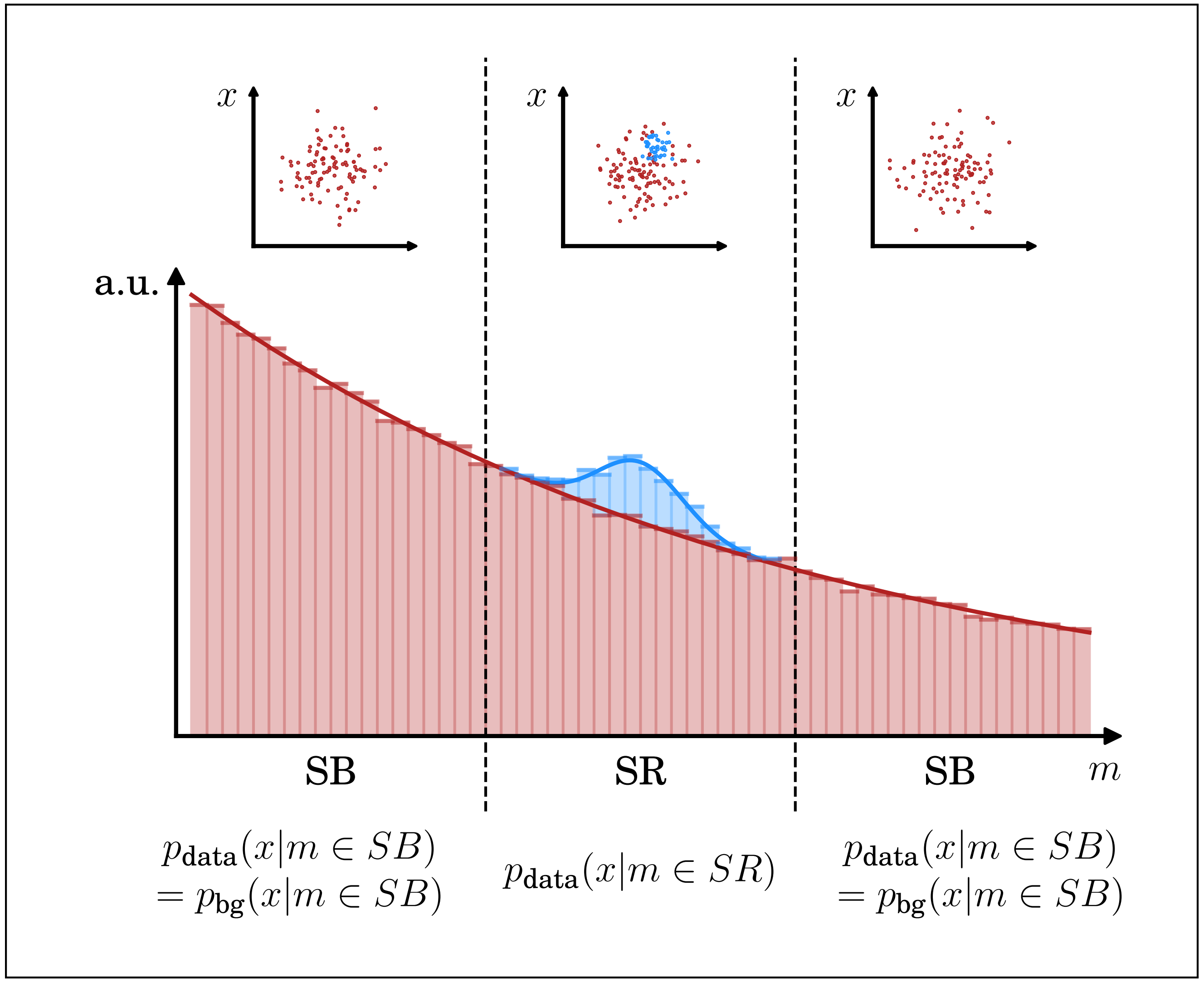






2107.02821

Generative models for anomaly detection

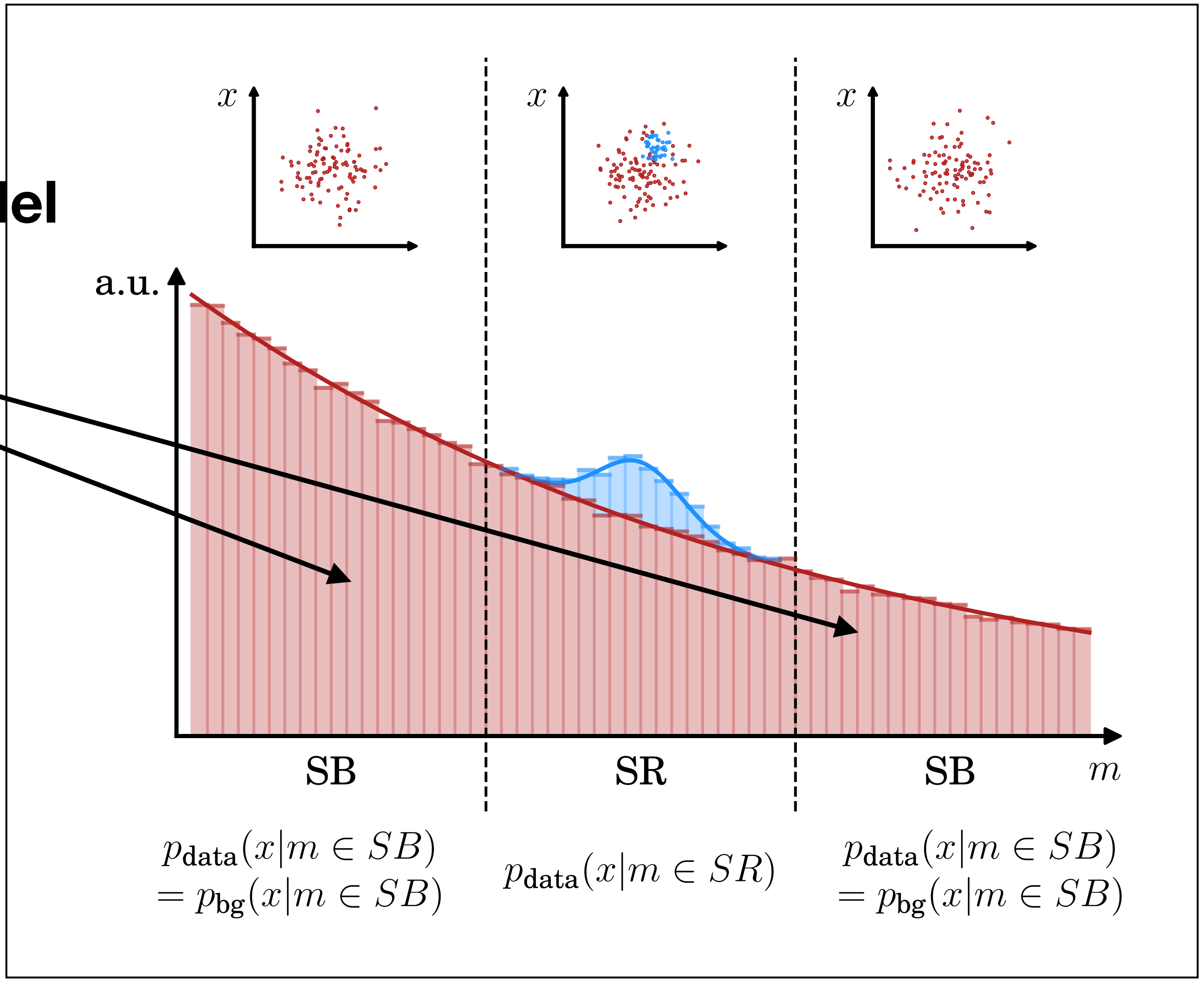


- 1): Choose one feature (m) in which to search for resonances

2109.00546

Generative models for anomaly detection

Train generative model here

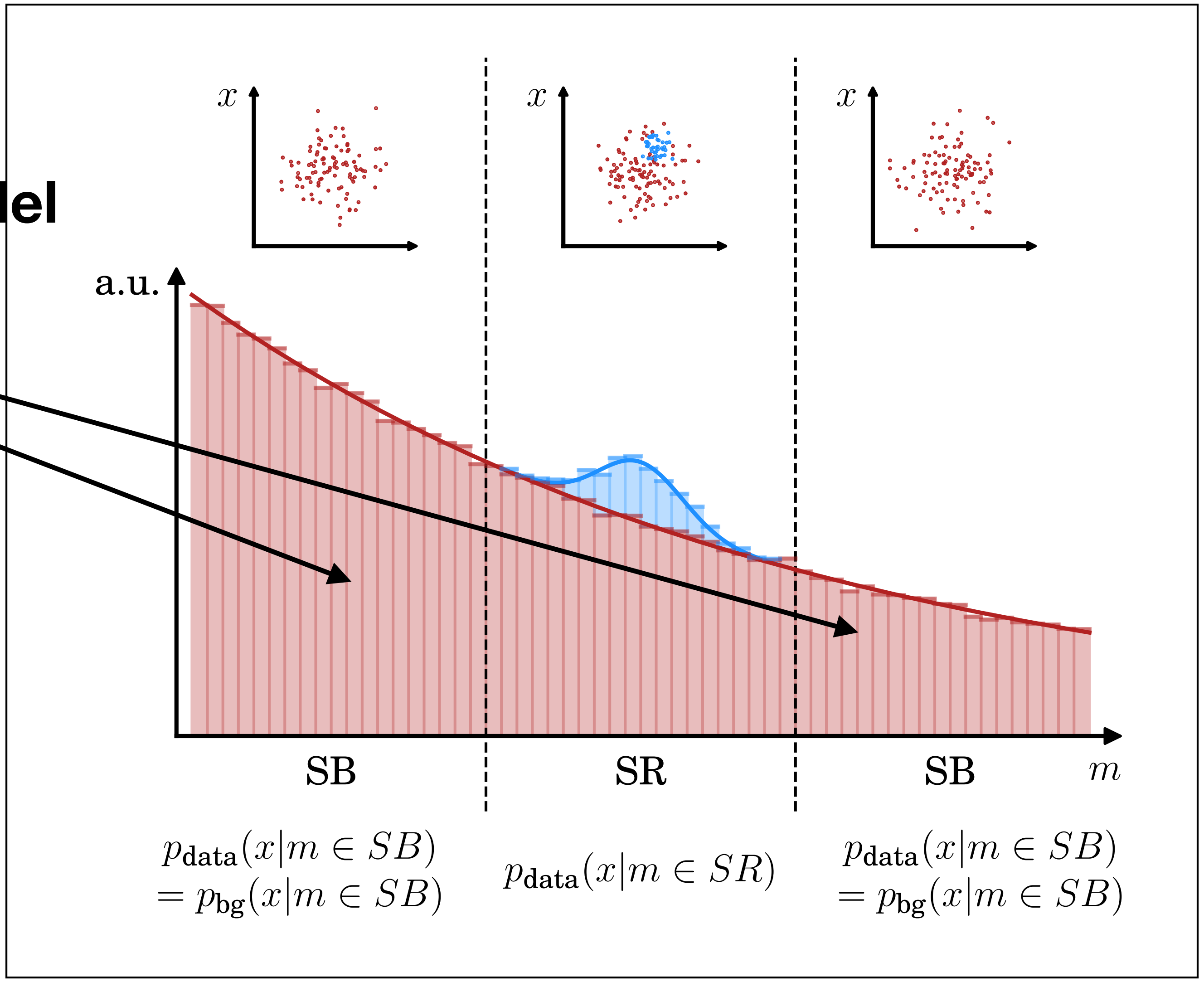


- 1): Choose one feature (m) in which to search for resonances
- 2): Use m divide spectrum into non-overlapping regions. Designate one as signal region (SR), others as sidebands (SB). Repeat the following for all choices of SR

2109.00546

Generative models for anomaly detection

Train generative model here

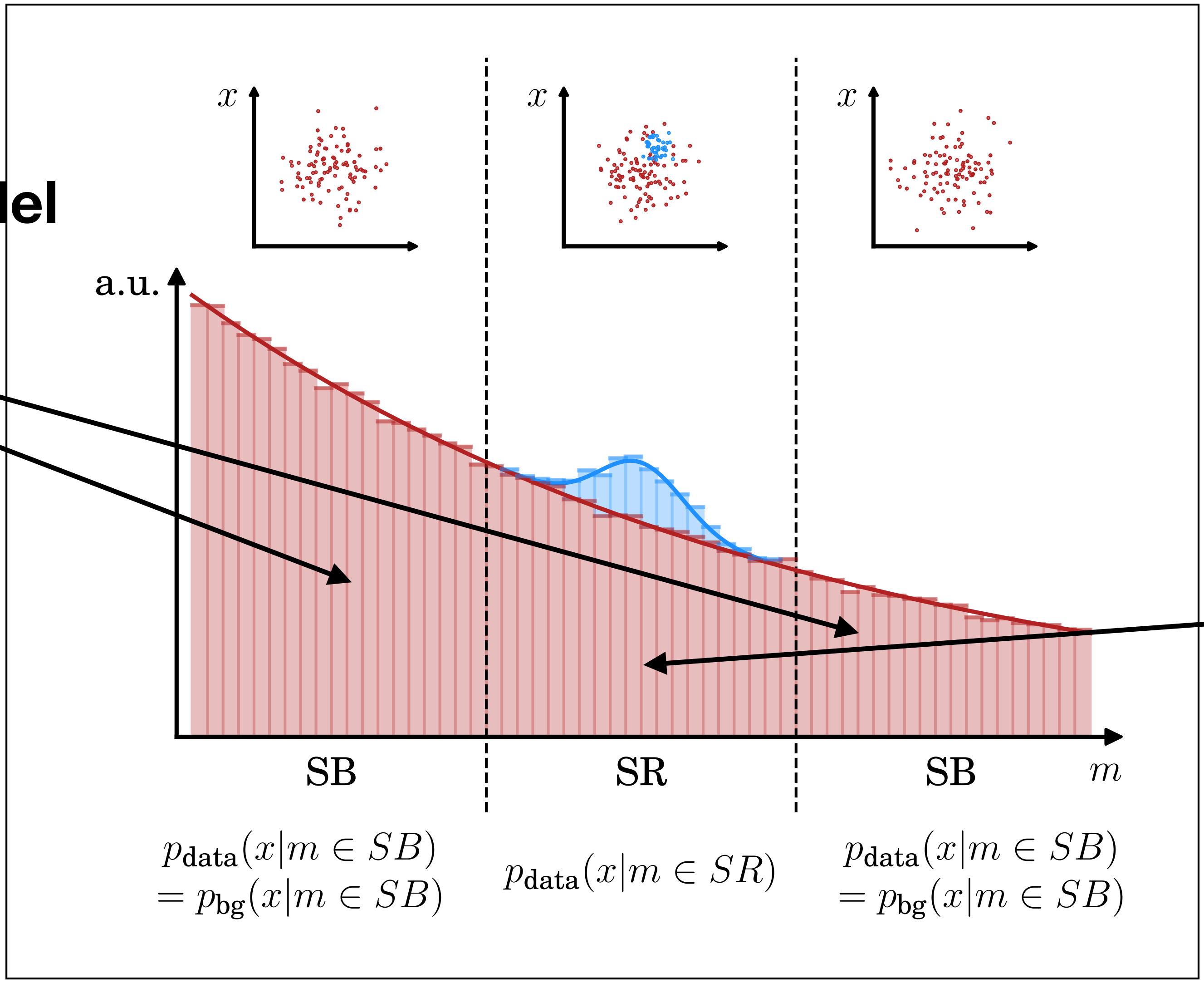


- 3) Train a generative model p(x|m) on auxiliary features in SB (used MAF, other choices including GAN/VAE possible as well)

2109.00546

Generative models for anomaly detection

Train generative model here

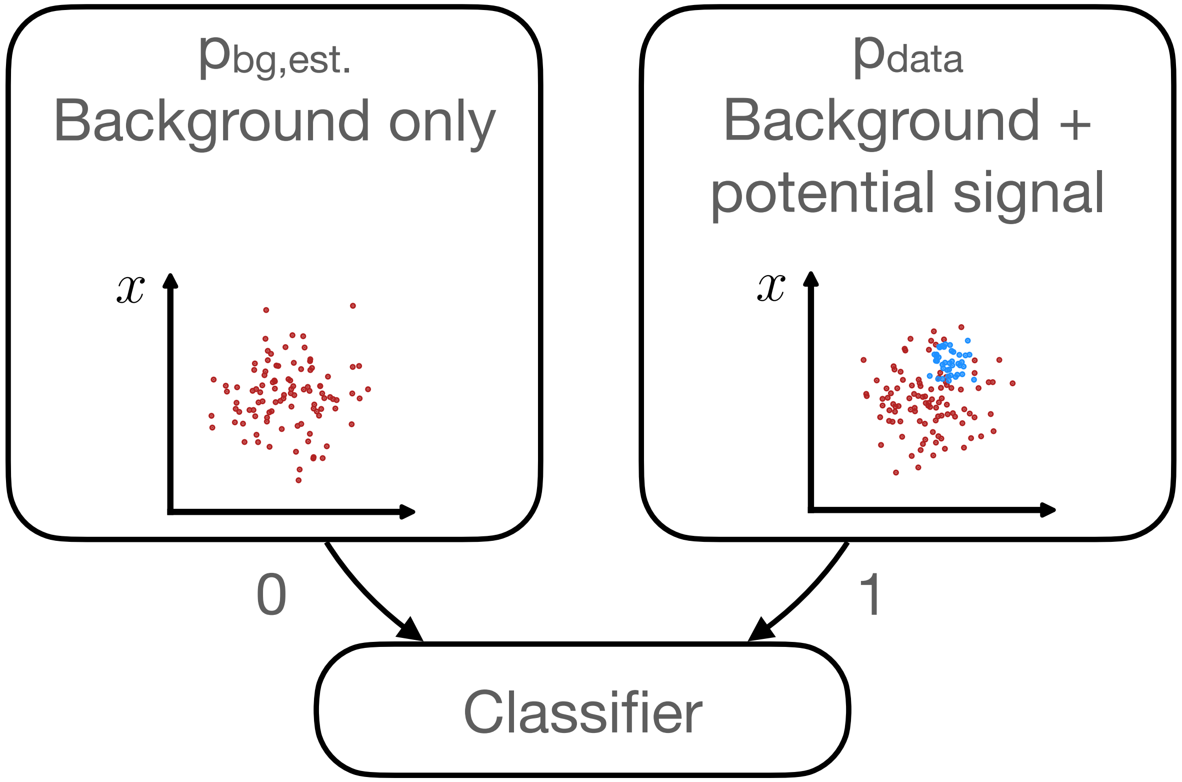


and sample here

- 3) Train a generative model p(x|m) on auxiliary features in SB
- 4) Sample from p(x|m) in SR. Designate as $p_{bg,est}$.

2109.00546

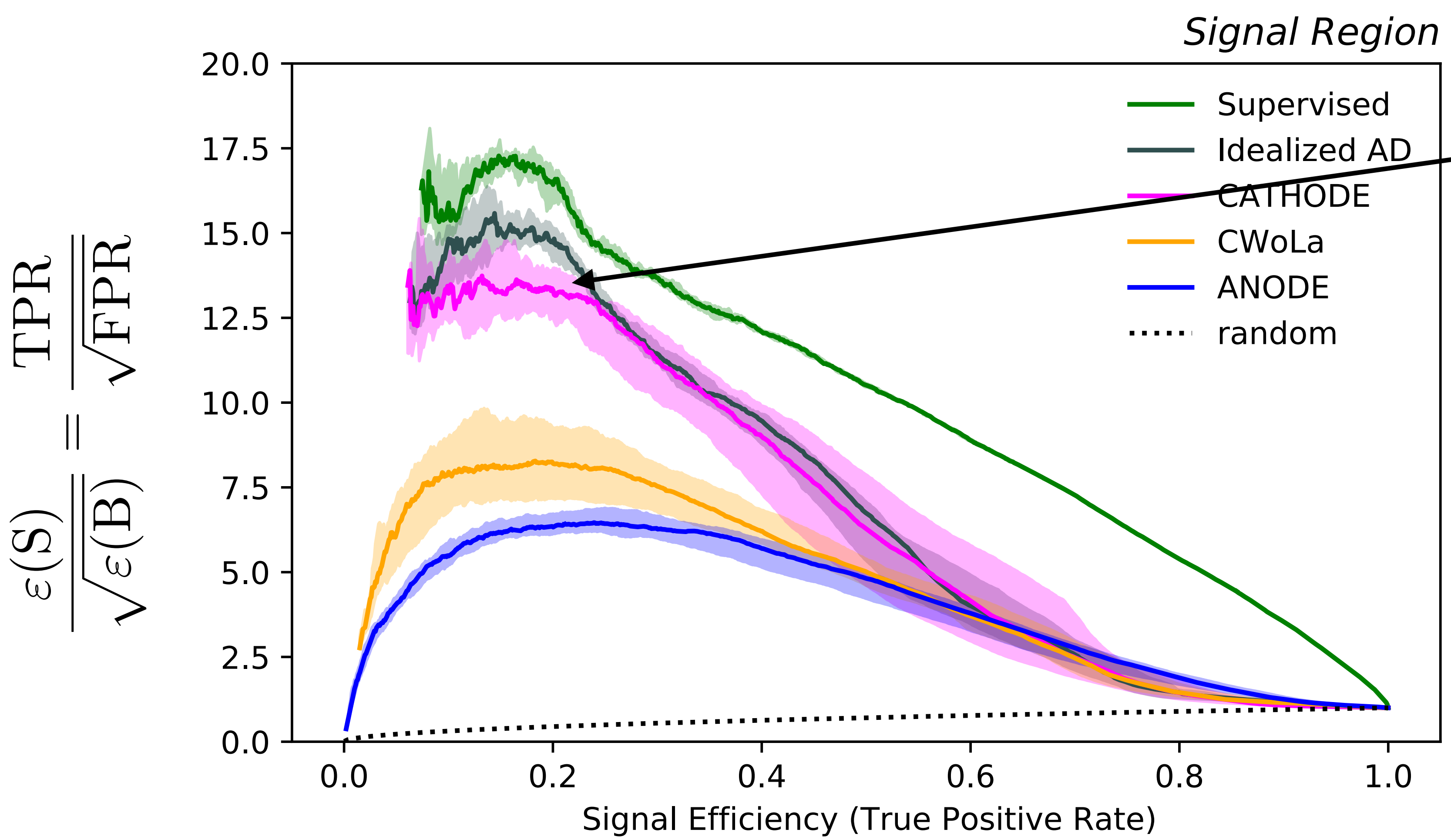
Generative models for anomaly detection



- 3) Train a generative model p(x|m) on auxiliary features in SB
- 4) Sample from p(x|m) in SR. Designate as $p_{bg,est}$
- 5) Train binary classifer between p_{data} and $p_{bg,est.}$

2109.00546

Generative models for anomaly detection



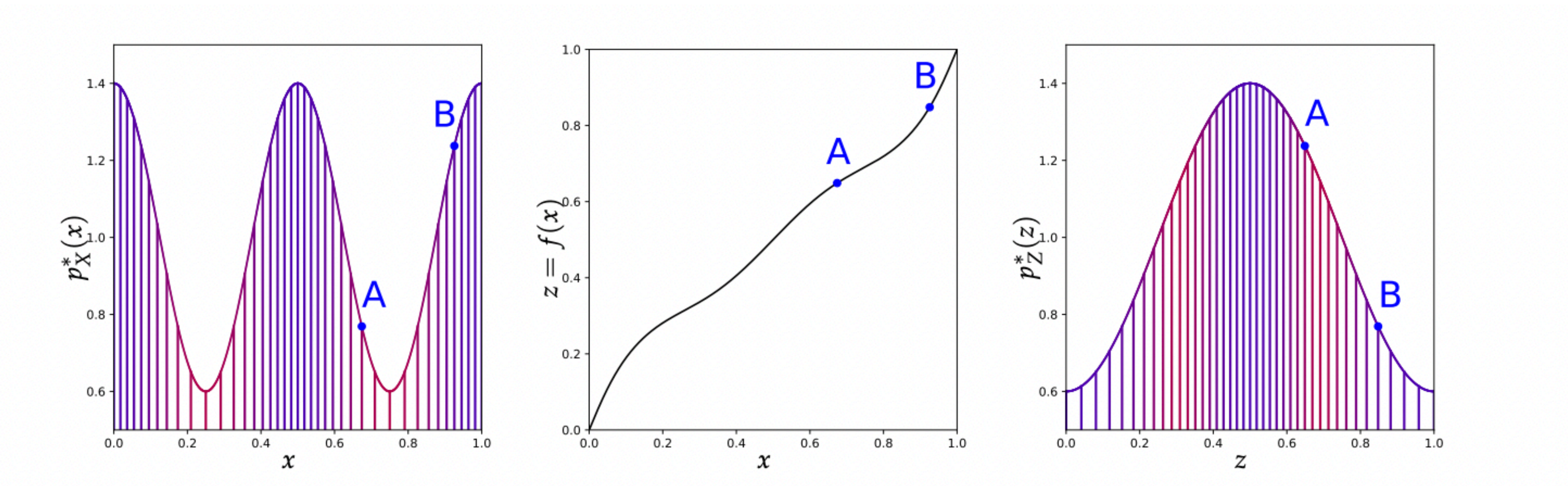
ours
CATHODE:
Classifying Anomalies THrough Outer Density Estimation

- 3) Train a generative model p(x|m) on auxiliary features in SB
- 4) Sample from p(x|m) in SR. Designate as $p_{bg,est.}$
- 5) Train binary classifer between p_{data} and $p_{bg,est.}$ (mixed sample classifer)
- 6) Cut on high classifier scores to enrich sample with anomalies
  (and perform statistical analysis)

2109.00546

Comments on anomaly detection

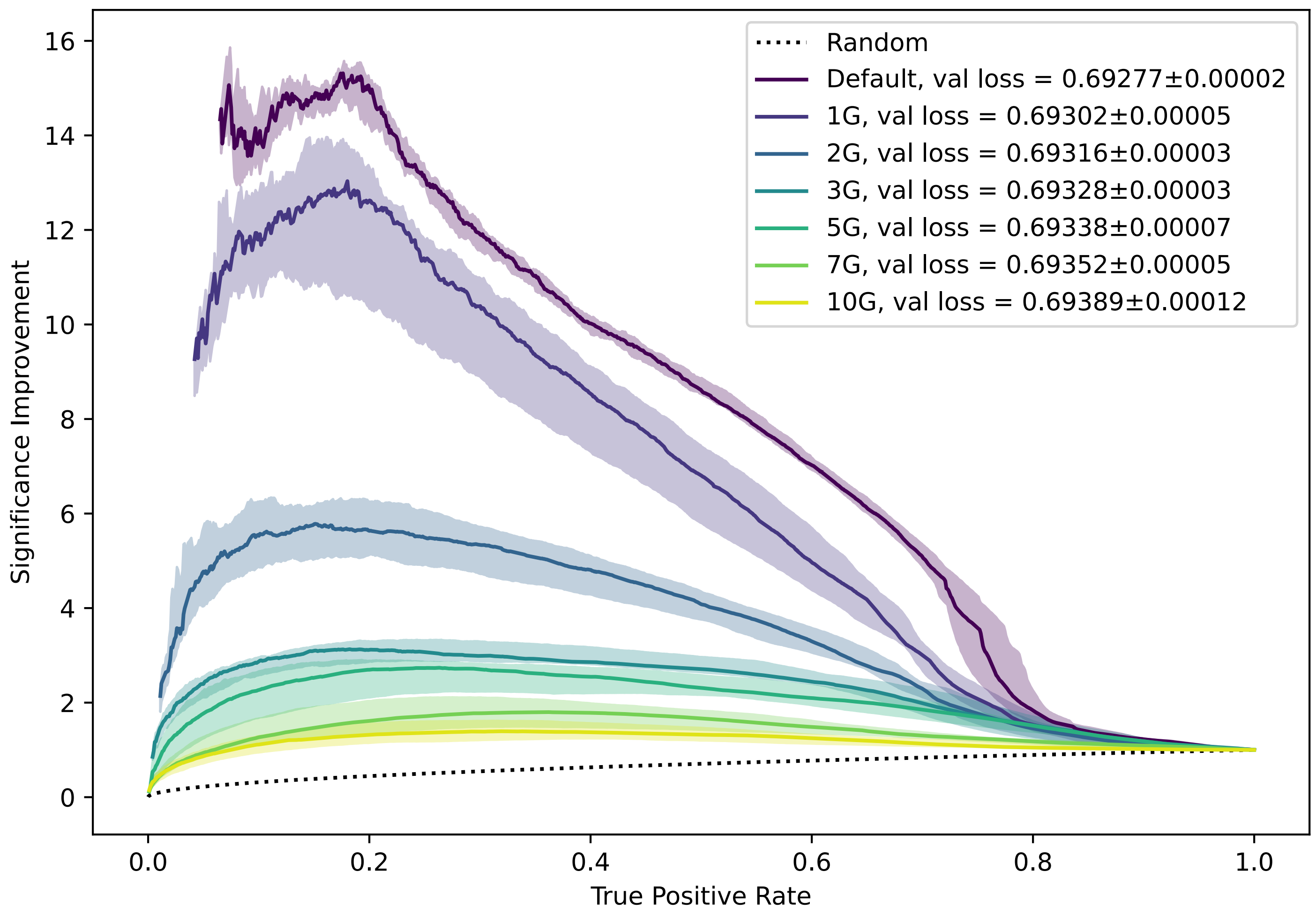
- As CATHODE approximates a likelihood ratio, it should be robust compared to methods that only use $p_{\mathrm{Background}}$ (e.g. autoencoders)



2012.03808 (Lan & Dinh)

Comments on anomaly detection

- As CATHODE approximates a likelihood ratio, it should be robust compared to methods that only use $p_{\text{Background}}$ (e.g. autoencoders)
- However, still can be sensitive to choice of input features



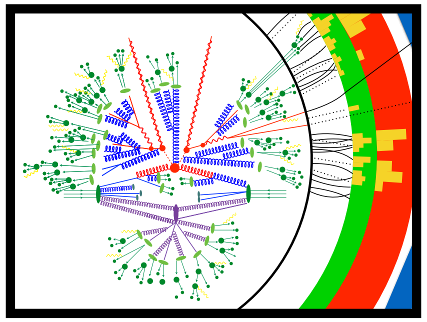
MSc thesis work of P. Prangchaikul

Comments on anomaly detection

- As CATHODE approximates a likelihood ratio, it should be robust compared to methods that only use $p_{\mathrm{Background}}$ (e.g. autoencoders)
- However, still can be sensitive to choice of input features

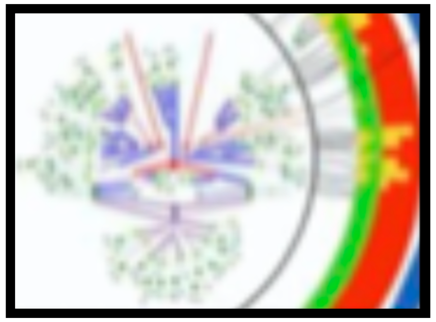
- Need also consider
  - Shaping of distributions under tigher anomaly detection cuts
  - Cost of signal-injection in training on data
  - How to efficiently estimate / compare / communicate sensitive regions of different anomaly detection algorithms
  - Make data-based anomaly detection more flexible

No compression



Many numbers per **event**

Compress per event



Small set of numbers per **event**

Compress entire dataset

$$p(x\,|\,\theta)$$

Small set of numbers per **dataset**

Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

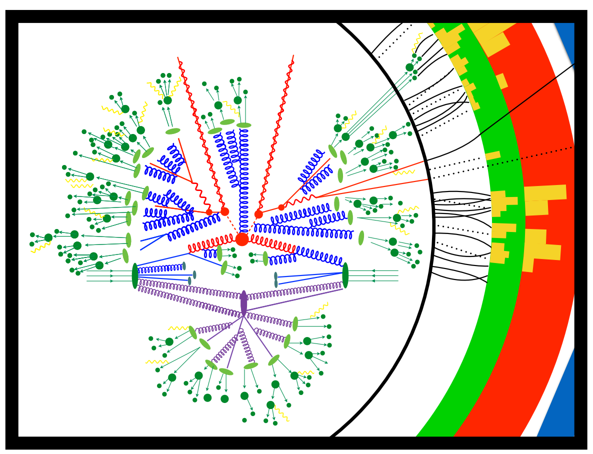
Final analysis, statistical and physical interpretation

What else can we do?

Emphemeral Learning

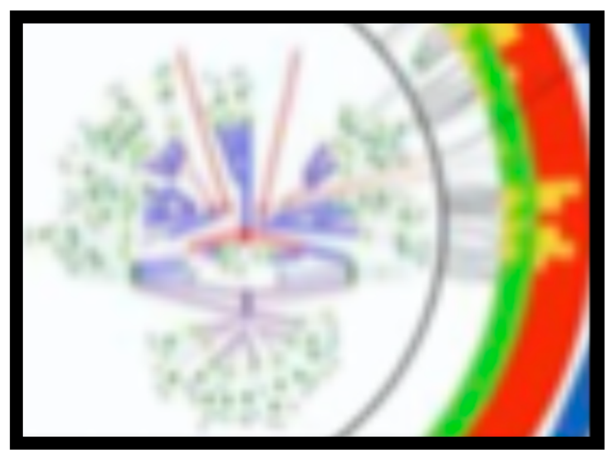
- Remember triggering:
  - Only able to store a subset (<1 in 10.000) of events
- Possible (wild) alternative:
  - Train a generative model online during data taking

No compression



Many numbers per **event**

Compress per event



Small set of numbers per **event**

Compress entire dataset

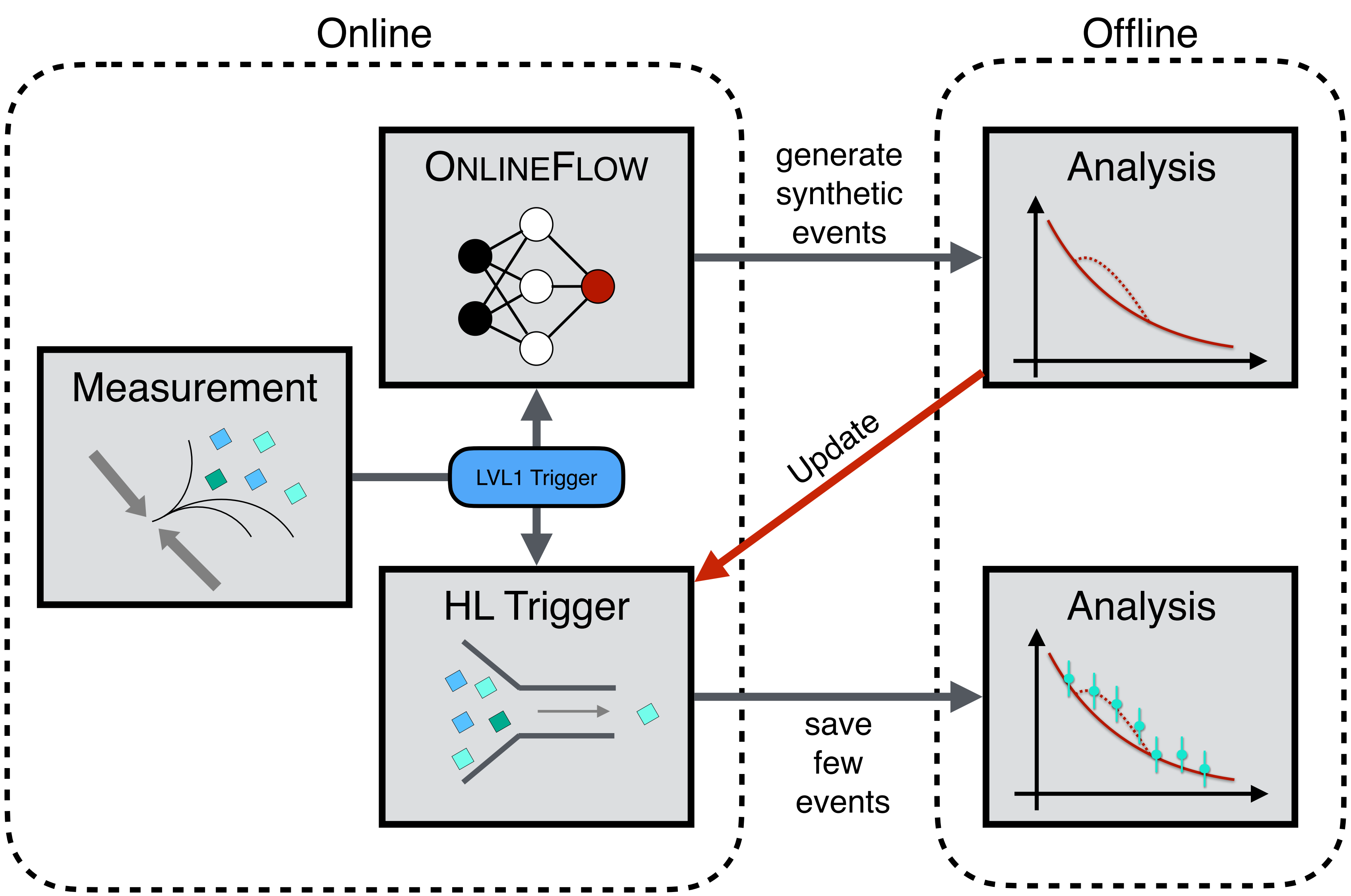
$$p(x\,|\,\theta)$$

Small set of numbers per **dataset**

- Fixed size, independent of training data amount
- Radically different format from usual way of storing data, but might open up new approaches

2202.0937

OnlineFlows



Schematic of proposed approach.

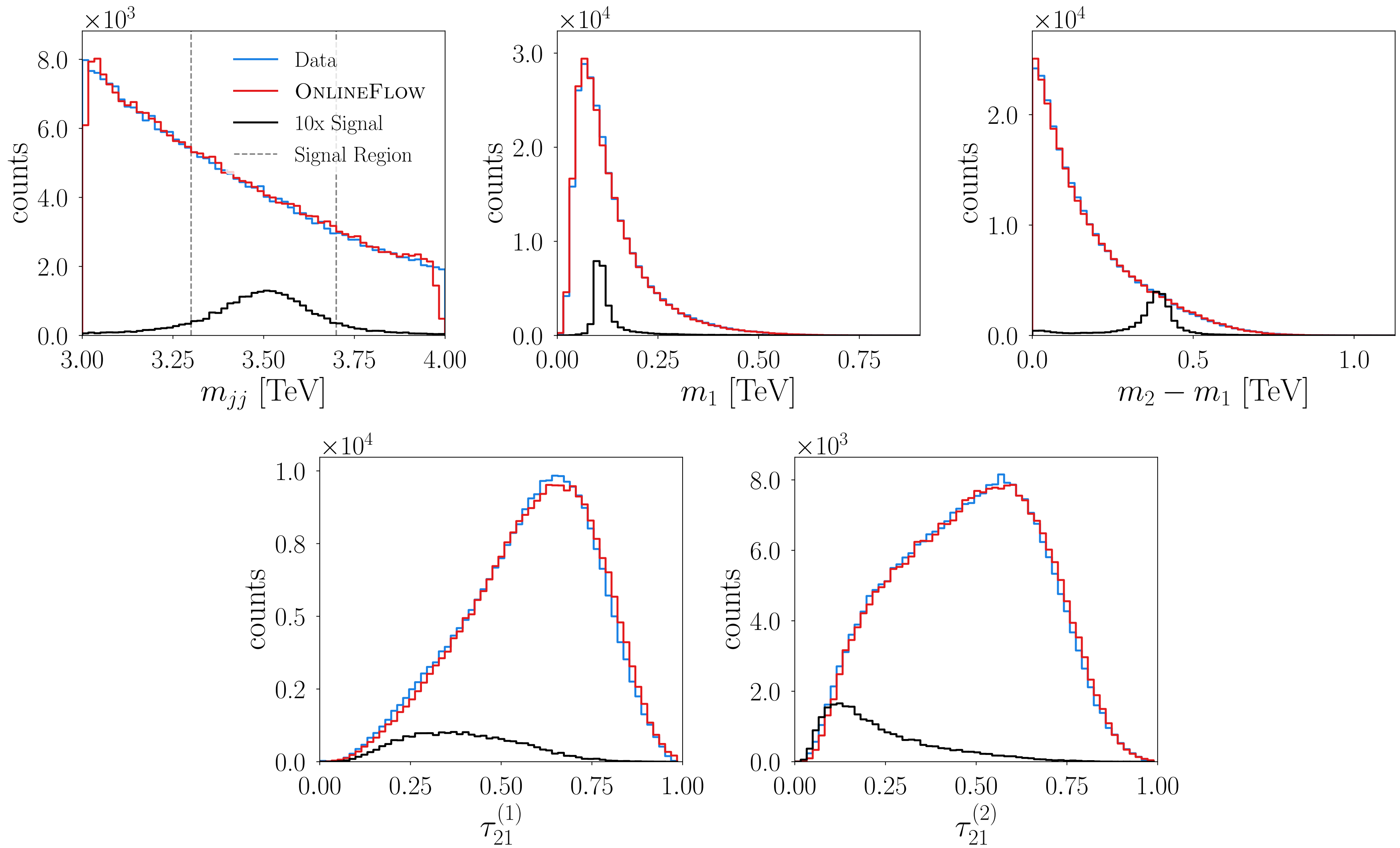
Focus on HLT, more technical challenges for use in hardware Trigger.

Main problem:
How to make training work if each event is only available for short time?

2202.0937

Proof of concept

Use LHCO dataset, train on high-level features on a mixture of background (99%) and signal (1%).



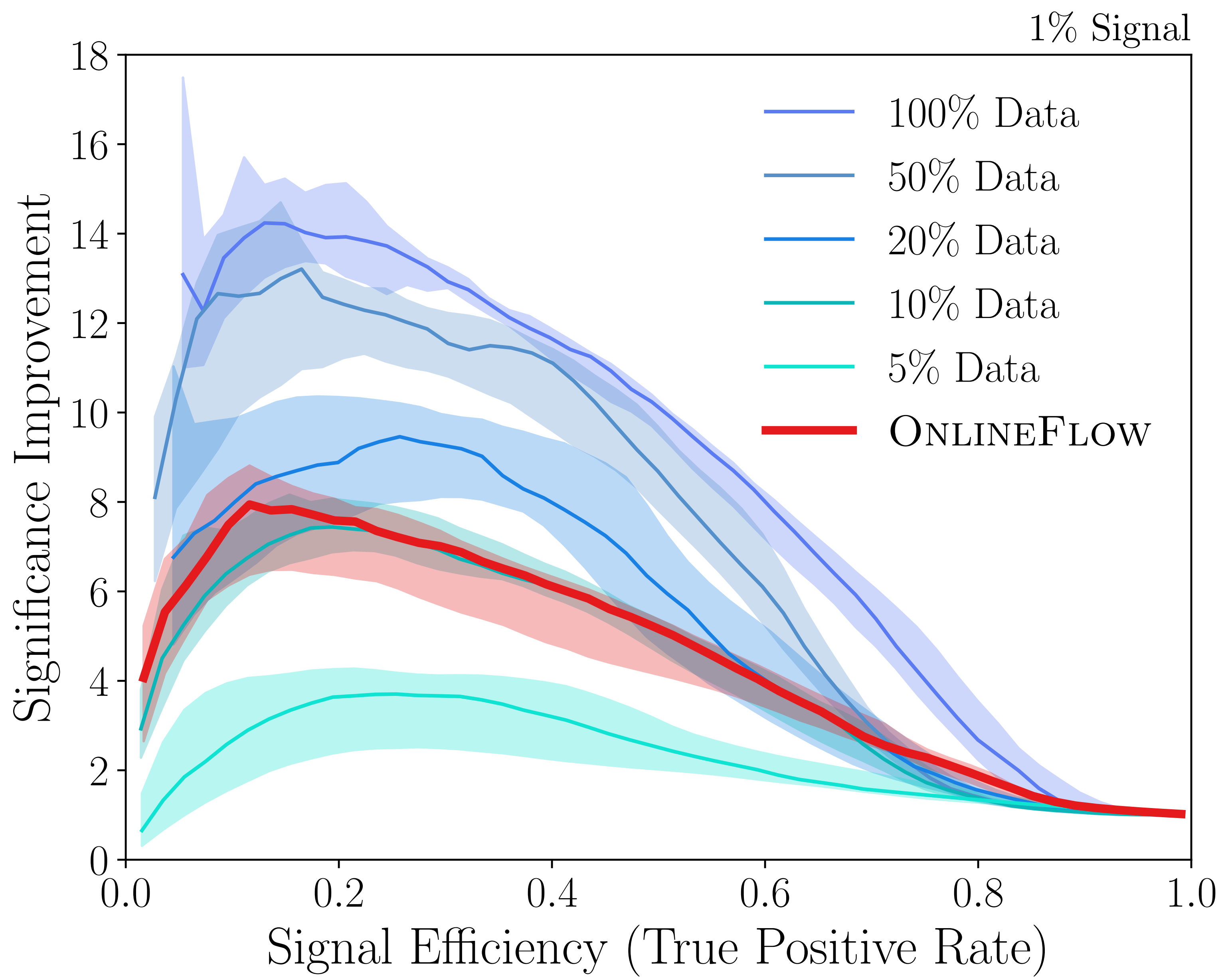
2202.0937

Proof of concept

Use LHCO dataset, train on high-level features on a mixture of background (99%) and signal (1%).

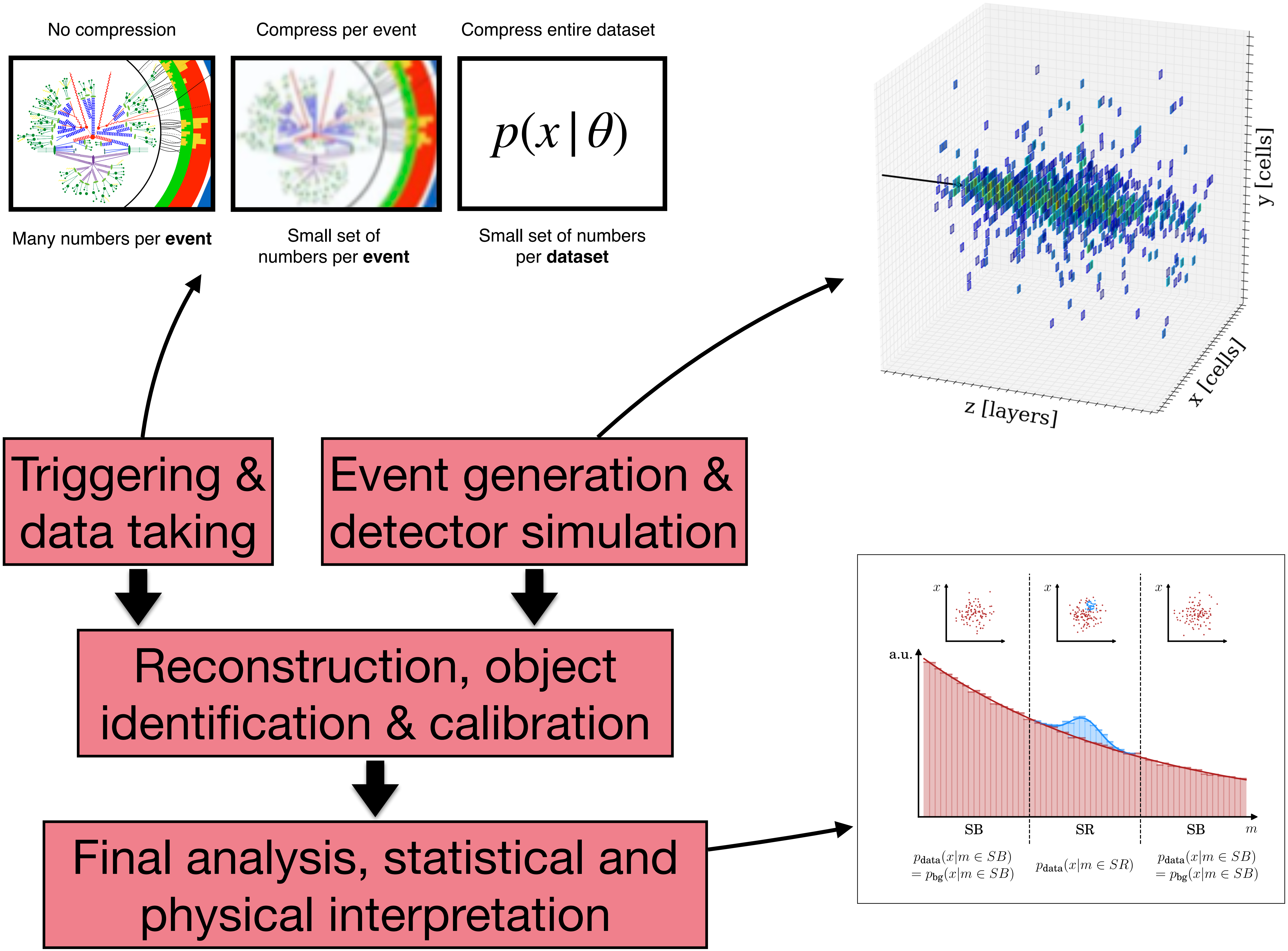
Train classifier to distinguish a signal region and sideband (CWoLA appaorach)

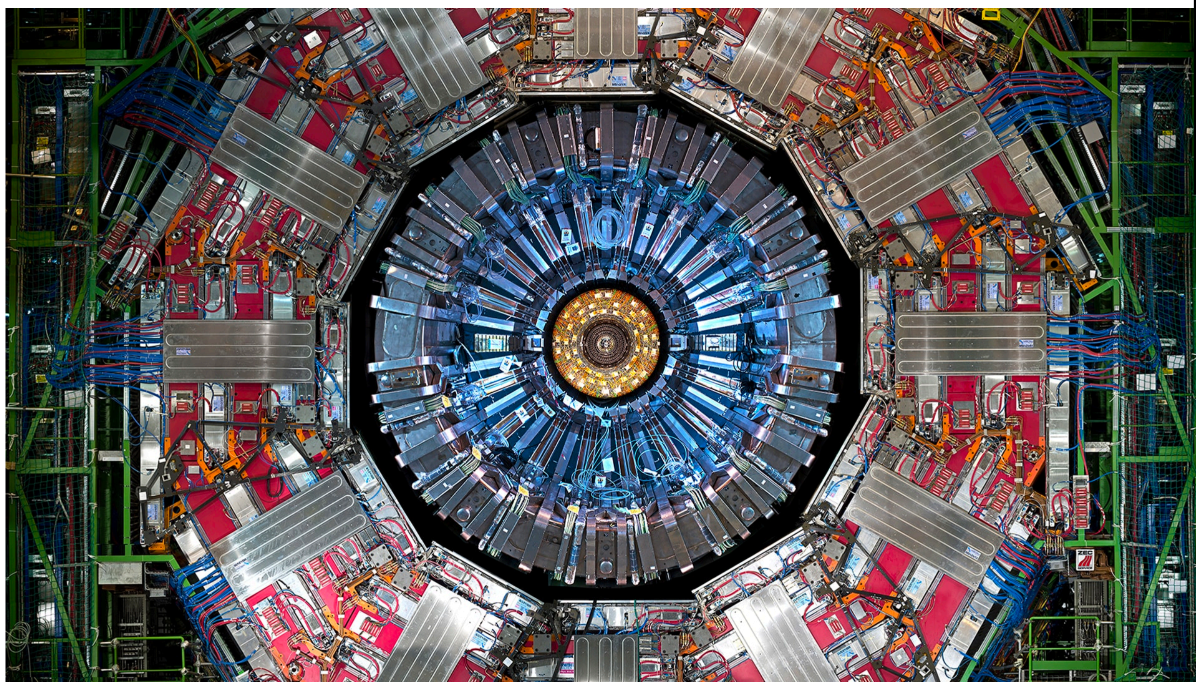
Compare procedure directly carried out on data with output of flow.

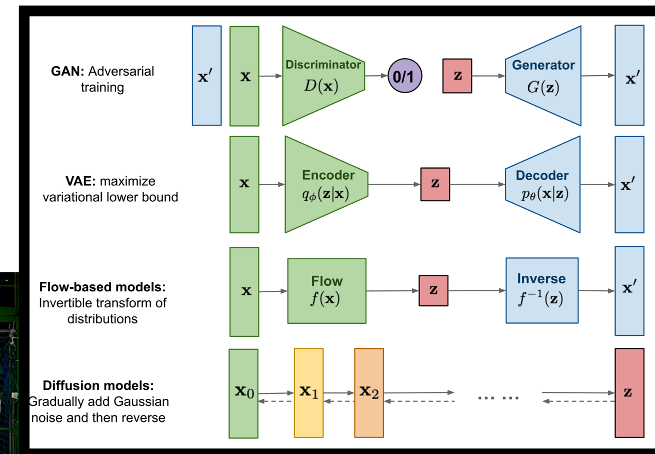


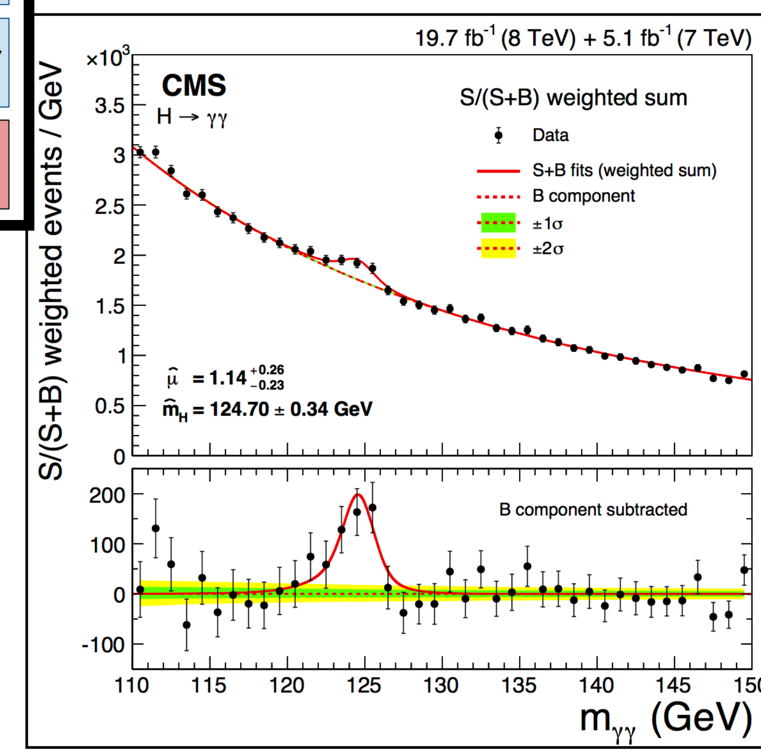
2202.0937

Wrapping up


No compression
Compress per event
Compress entire dataset
$p(x\,|\,\theta)$
Many numbers per **event**
Small set of numbers per **event**
Small set of numbers per **dataset**
y [cells]
x [cells]
z [layers]
Triggering & data taking
Event generation & detector simulation
Reconstruction, object identification & calibration
Final analysis, statistical and physical interpretation
a.u.
SB
SR
SB
m
$p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB)$
$p_{\text{data}}(x|m \in SR)$
$p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB)$




GAN: Adversarial training
Discriminator $D(\mathbf{x})$
0/1
Generator $G(\mathbf{z})$
VAE: maximize variational lower bound
Encoder $q_\phi(\mathbf{z}|\mathbf{x})$
Decoder $p_\theta(\mathbf{x}|\mathbf{z})$
Flow-based models: Invertible transform of distributions
Flow $f(\mathbf{x})$
Inverse $f^{-1}(\mathbf{z})$
Diffusion models: Gradually add Gaussian noise and then reverse


19.7 fb^{-1} (8 TeV) + 5.1 fb^{-1} (7 TeV)
CMS
H → γγ
S/(S+B) weighted sum
Data
S+B fits (weighted sum)
B component
±1σ
±2σ
$\hat{\mu}$ = 1.14$^{+0.26}_{-0.23}$
$\hat{m}_H$ = 124.70 ± 0.34 GeV
B component subtracted
S/(S+B) weighted events / GeV
$m_{\gamma\gamma}$ (GeV)

What else

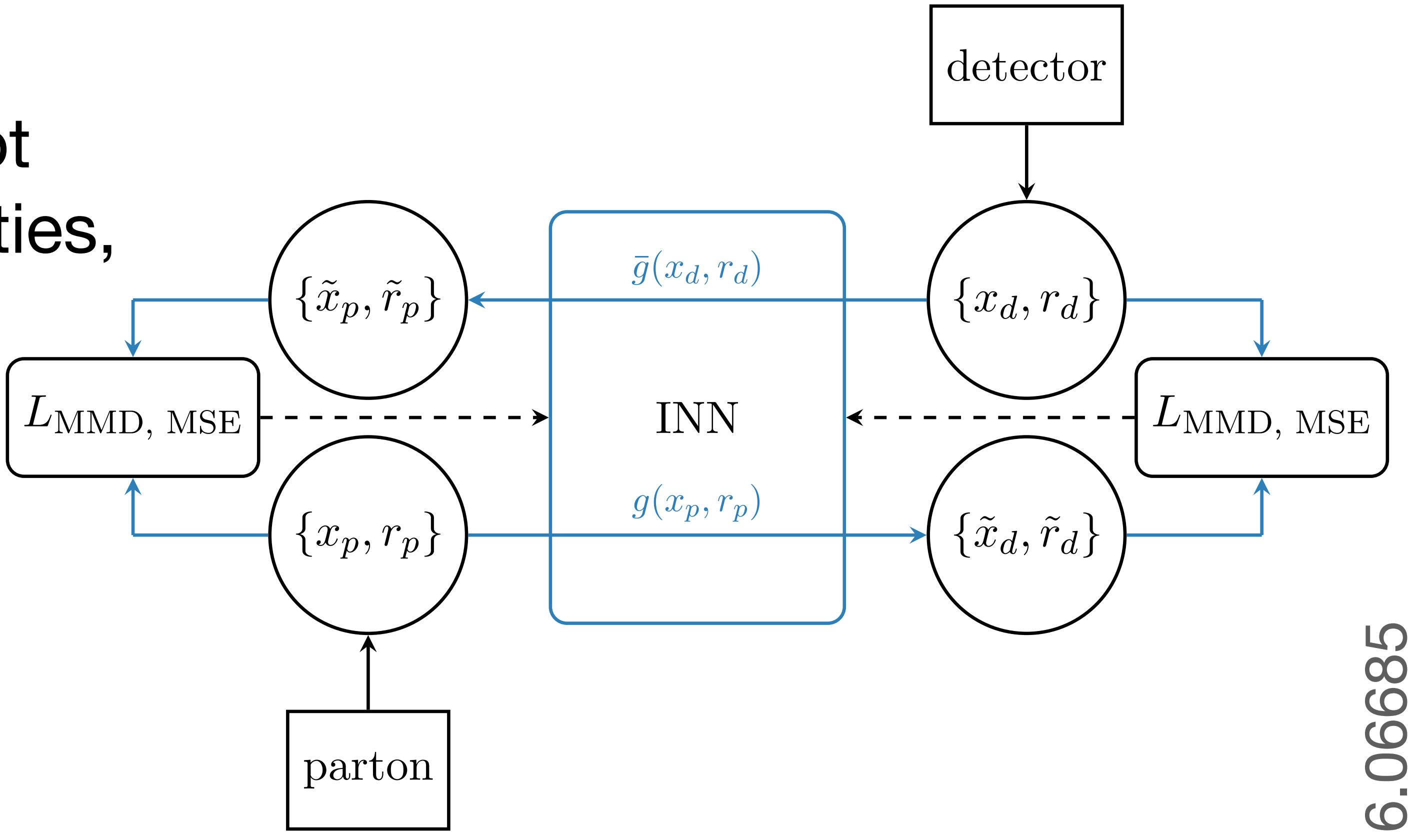
- Notice something?

What else

- Notice something?
  - All examples of flows use the fact that flows are good and easily trainable generative models
  - But none use the fact that we can access a per-example likelihood
    - Might have useful applications by itself?

What else

- Notice something?
  - All examples of flows use the fact that flows are good and easily trainable generative models
  - But none use the fact that we can access a per-example likelihood
    - Might have useful applications by itself?
- Can also use examples where an invertable network does not invert onto the physics quantities, but is parametrised by them



2006.06685

What else

- Notice something?
  - All examples of flows use the fact that flows are good and easily trainable generative models
  - But none use the fact that we can access a per-example likelihood
    - Might have useful applications by itself?
- Can also use examples where an invertable network does not invert onto the physics quantities, but is parametrised by them
- Also uses in other domains, e.g. lattice QCD

Closing

- Unsupersvised learning in the form of density estimators is quickly becoming a key instrument in our toolbox
    - Learning of actual densitities not yet widely exploited
- Advances in the power of these models and the quality of learned distributions opens new doors for physics analysis
- Excited for the future:
    - What can we do with fully differentiable surrogate models with tractable probabilities for *all (ErUM) physics?*

Thank you!

Comments on anomaly detection

- As CATHODE approximates a likelihood ratio, it should be robust compared to methods that only use $p_{Background}$ (e.g. autoencoders)
- However, still can be sensitive to choice of input features

- Need also consider
  - Shaping of distributions under tigher anomaly detection cuts
  - Cost of signal-injection in training on data
  - How to efficiently estimate / compare / communicate sensitive regions of different anomaly detection algorithms
  - Make data-based anomaly detection more flexible

Bonus Slides

Outline

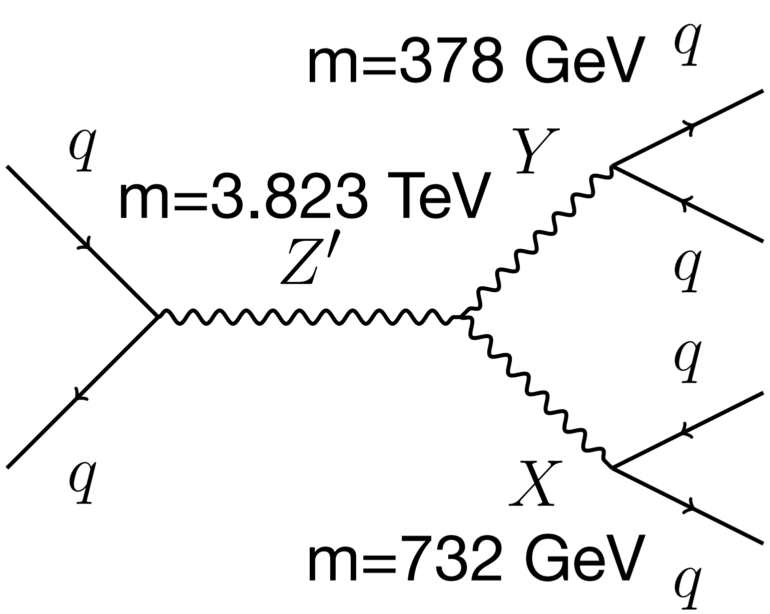
- Introduction
- Applications of generative models
  - Calorimeter simulation
  - Statistical properties
- Anomaly detection
  - Overview
  - CATHODE
  - Current challenges
- Synthesis
  - New approaches
  - Problems / opportunities
- Closing

Challenge datasets

- All contain total of 1M examples; might contain signal;
  no labels provided during ‘content’ phase (labels available no)
- All used different simulation parameters for background (to avoid unrealistic exploits)

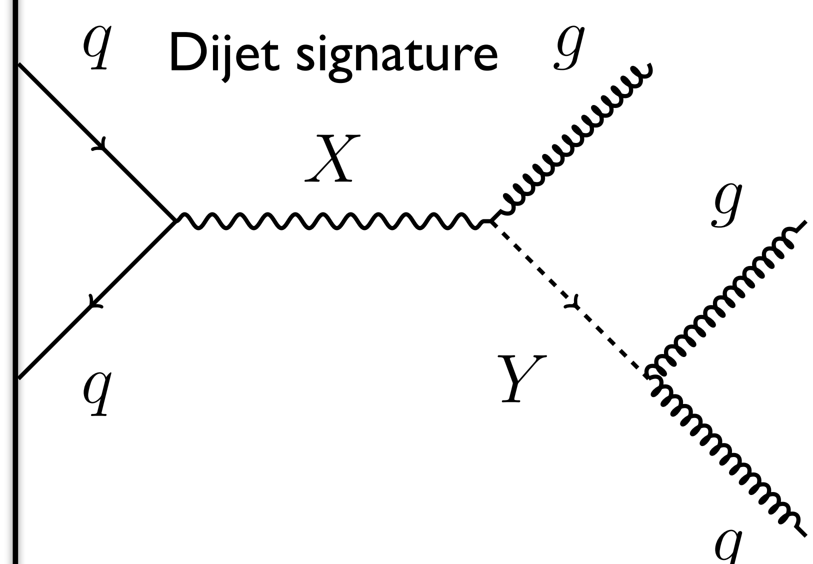
BB1: 834 signal examples
Same event topology as R&D dataset, different masses

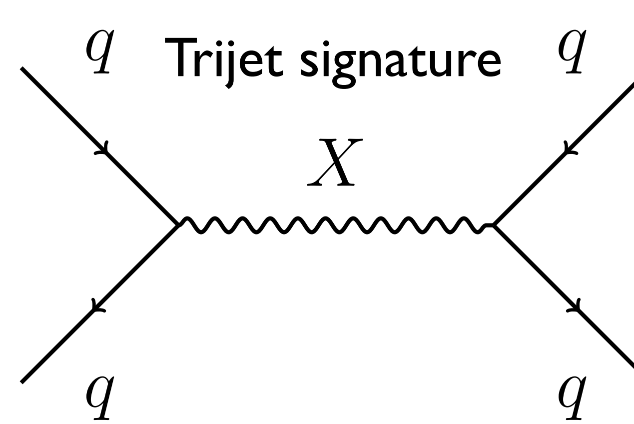
might be easy?



BB2: empty

BB3:





🐨 & Friends

- Situation seems better for density ratio based techniques (CWola, ANODE, CATHODE,..)

Per Neyman-Pearson: Likelihood-ratio is optimal test statistic
Unfortunatly, p(x|anomaly) is not available

$$L_{S/B} = \frac{p(x|\text{anomaly})}{p(x|\text{normal})}$$

Build data/background ratio:

$$L_{D/B} = \frac{p(x)}{p(x|\text{normal})}$$

Approximate background density using control measurement (e.g. sideband)

$$L_{D/B} \approx \frac{p(x)}{\tilde{p}(x|\text{normal})}$$

Expand $$p(x) = f_{\text{normal}}\, p(x|\text{normal}) + f_{\text{anomaly}}\, p(x|\text{anomaly})$$

And insert: $$L_{D/B} \approx f_{\text{normal}} + f_{\text{anomaly}} \frac{p(x|\text{anomaly})}{\tilde{p}(x|\text{normal})}$$

- However…