Unsupervised Learning for Discovery and Simulation

Gregor Kasieczka

(gregor.kasieczka@uni-hamburg.de)

HEP@VUB Seminar

2020-11-05

CLUSTER OF EXCELLENCE

QUANTUM UNIVERSE

Emmy Noether-Programm

Deutsche Forschungsgemeinschaft

DFG

UHH Universität Hamburg

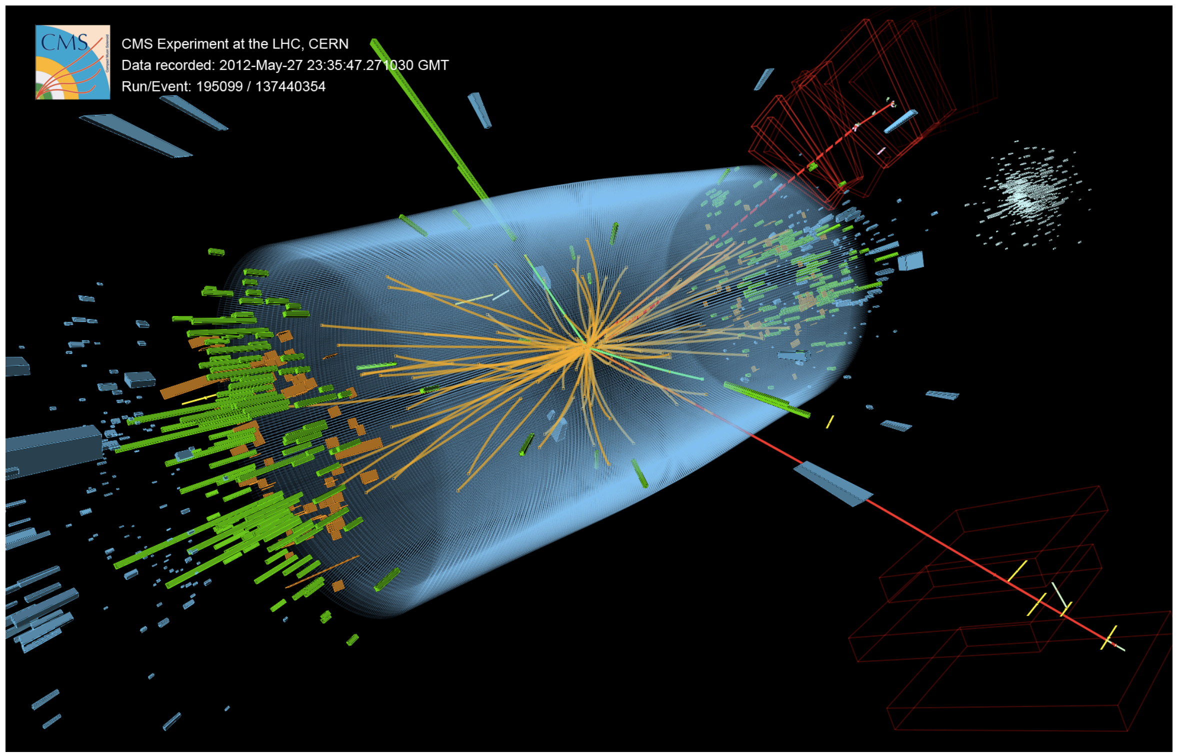
DER FORSCHUNG | DER LEHRE | DER BILDUNG

PIER

Partnership of Universität Hamburg and DESY

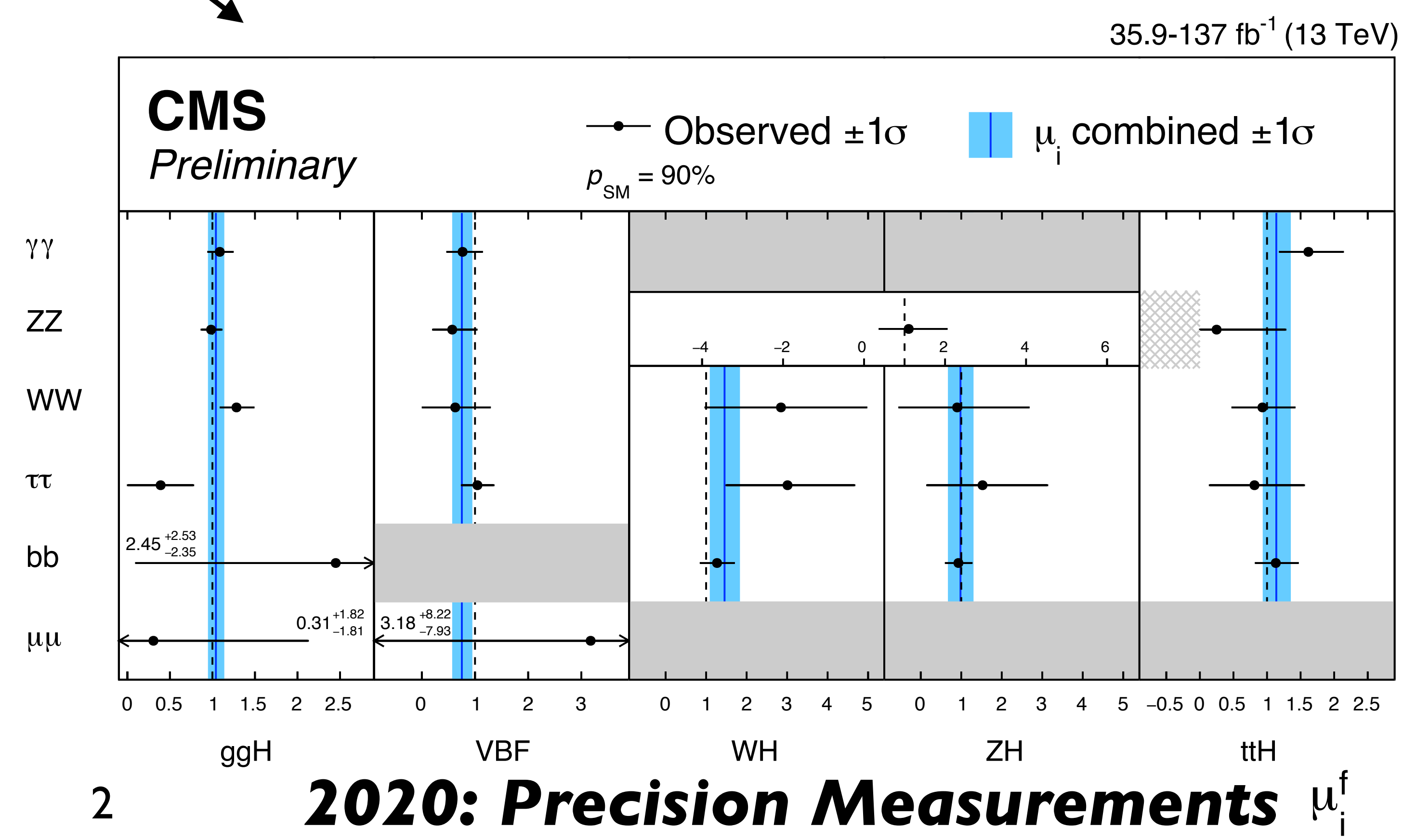
Bundesministerium für Bildung und Forschung

Higgs Boson: Discovery to Precision…



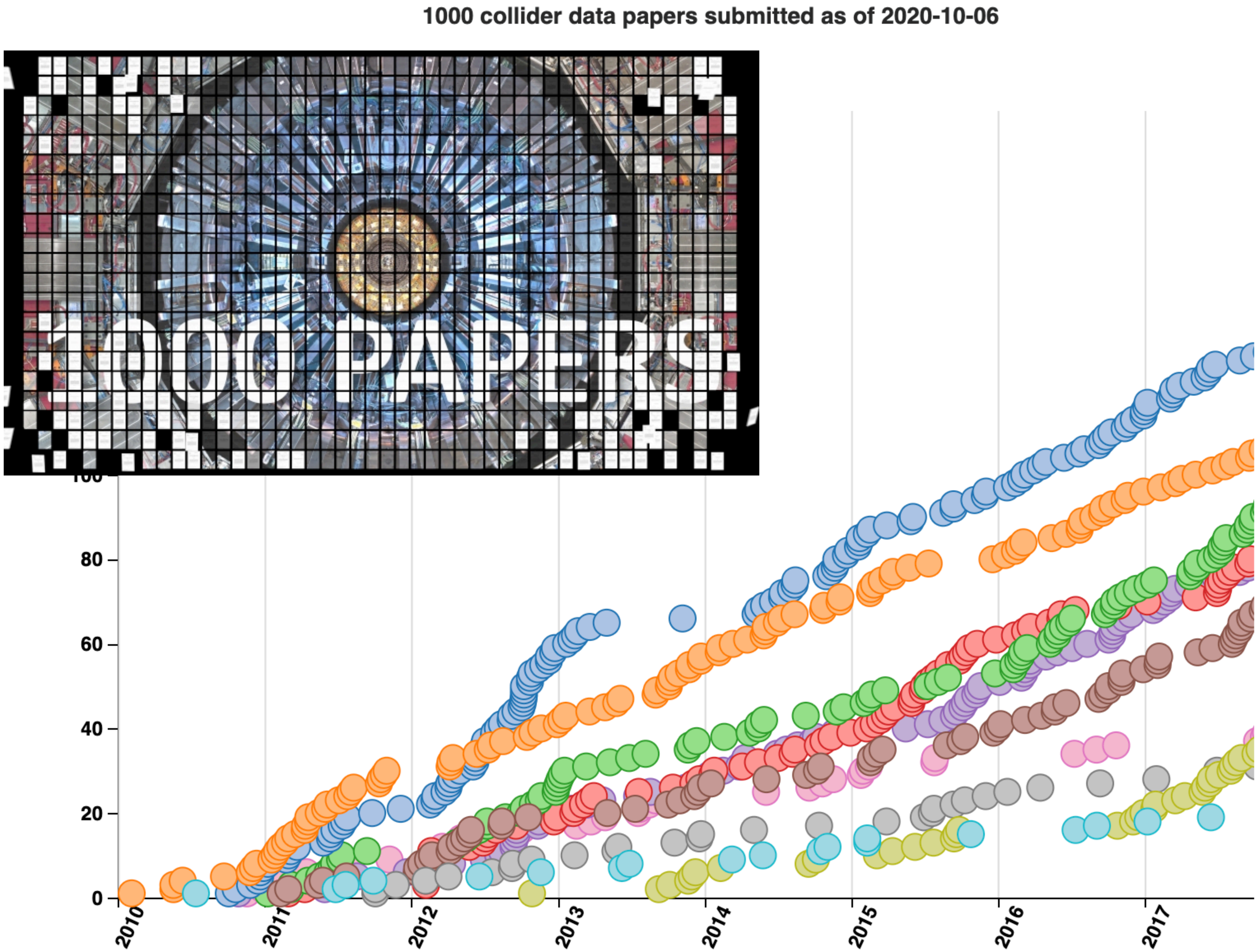
2012: Discovery of the Higgs boson

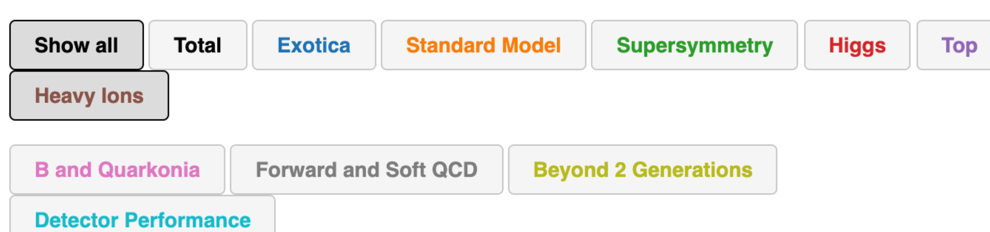
Now



2020: Precision Measurements

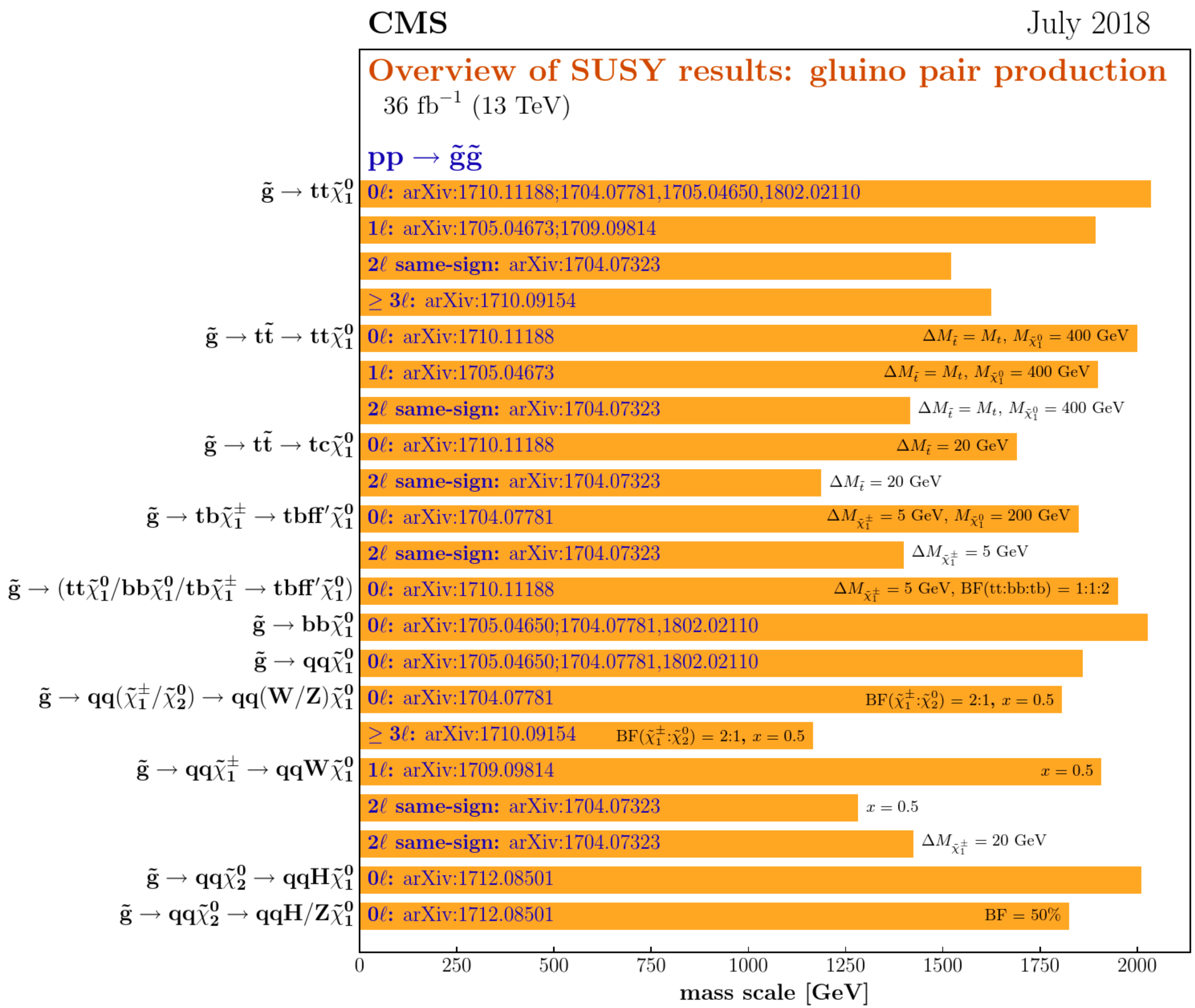
Many results…





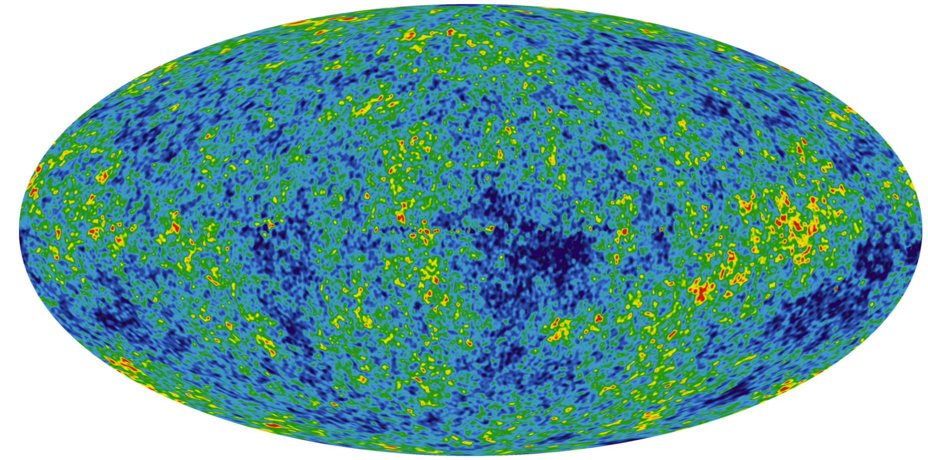
http://cms.web.cern.ch/org/physics-papers-timeline

…but no new physics so far



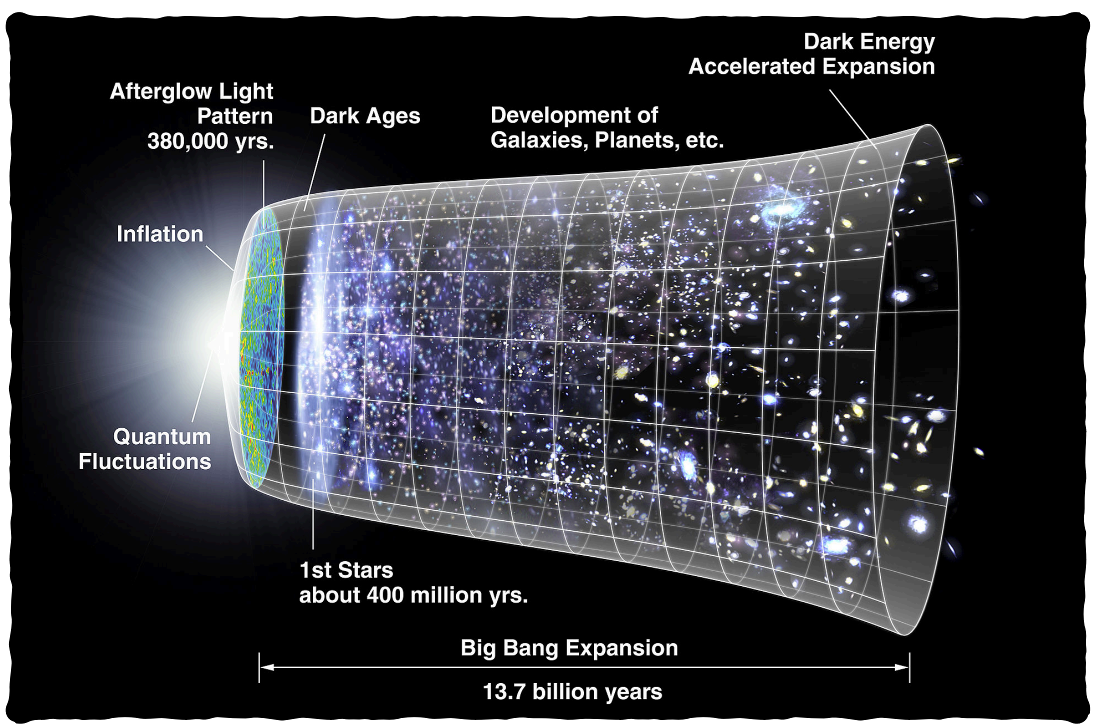
Selection of observed limits at 95% C.L. (theory uncertainties are not included). Probe **up to** the quoted mass limit for light LSPs unless stated otherwise. The quantities ΔM and x represent the absolute mass difference between the primary sparticle and the LSP, and the difference between the intermediate sparticle and the LSP relative to ΔM, respectively, unless indicated otherwise.

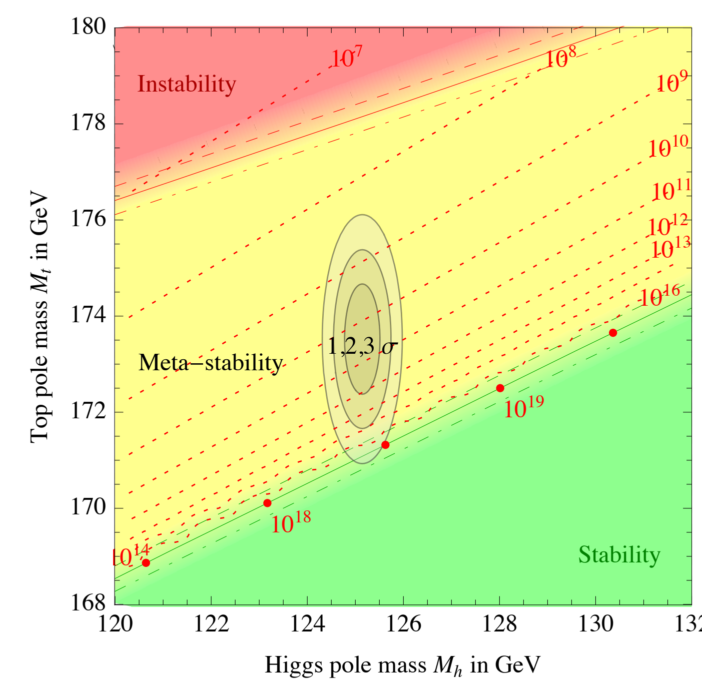
Why are neutrinos massive?



What are the origins of the LHCb flavour anomaly?

What is the nature of dark matter & dark energy?





Is the electroweak vacuum stable?

Why is there more matter than anti-matter?

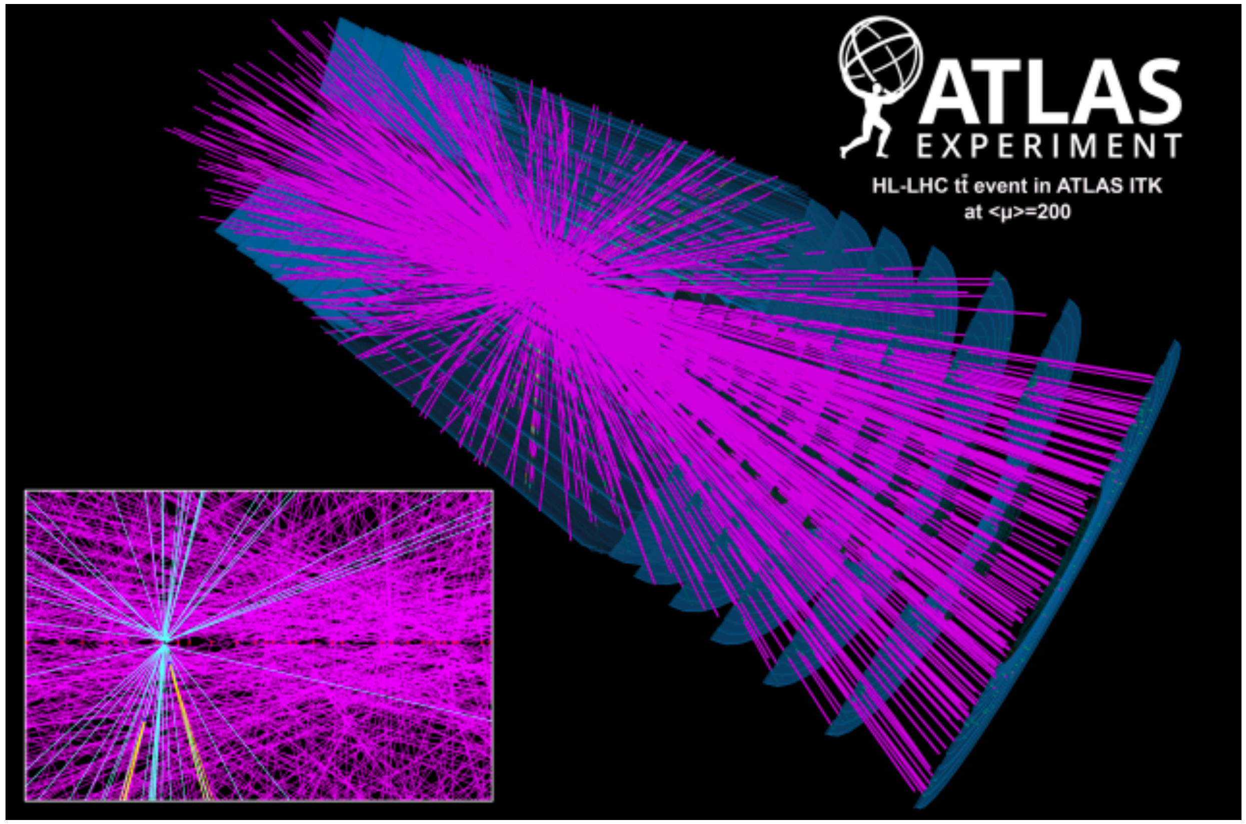
How can the Higgs boson be light when the mass receives large quantum corrections?

What are the details of cosmic inflation?

What next?

- Precision measurements and searches for new physics need
    - better tools to identify known particles and processes
    - higher accuracy and speed
- Finding unknown signatures needs
    - new ways of analysing data
- Future data taking with higher collision rates needs:
    - faster reconstruction and triggering
    - faster simulation and event generation

(a) promising answer: **Deep Learning**



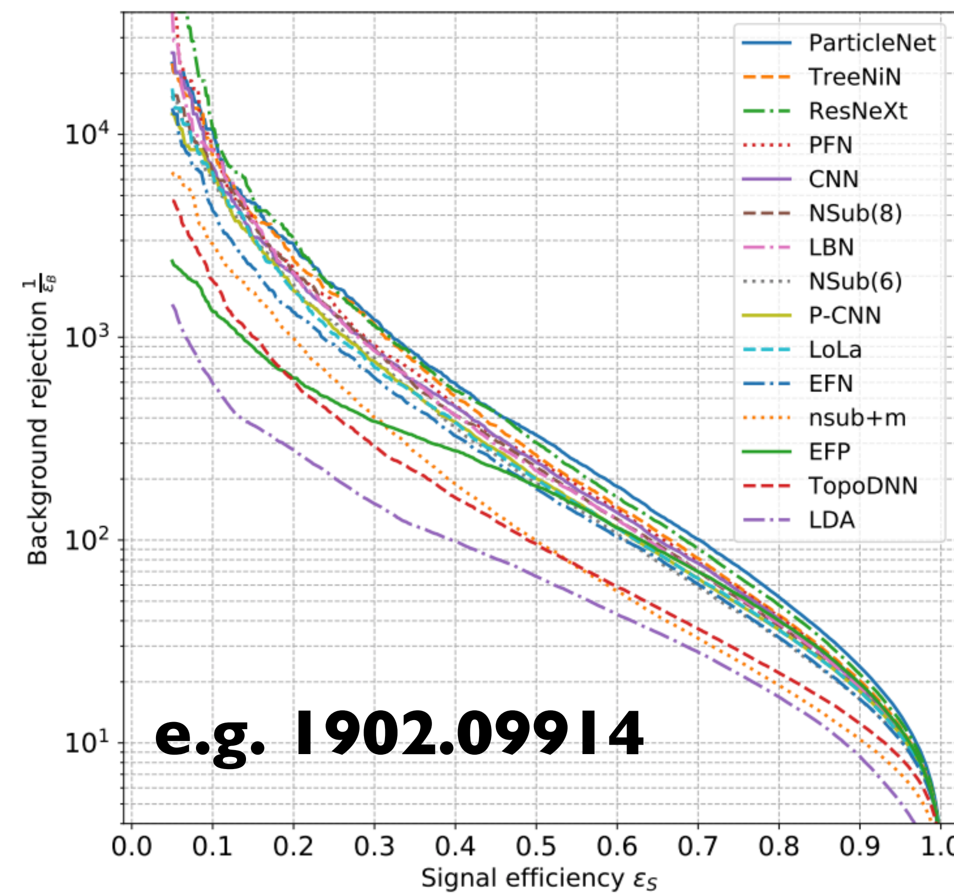
Menu

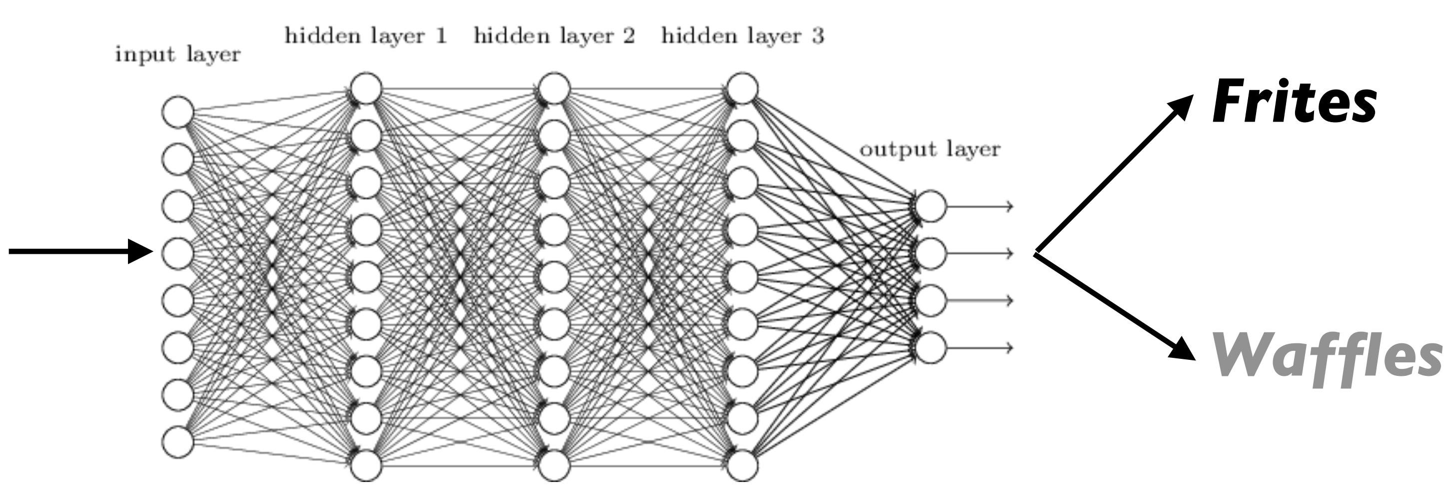
Learning generative models

Data-driven anomaly detection

Supervised Learning

- Attempt to learn some target: classification or regression tasks
- Need to have a dataset with known targets (typically from simulation)
- Examples:
  - Flavour tagging
  - Heavy resonance tagging
  - Signal vs background discrimination

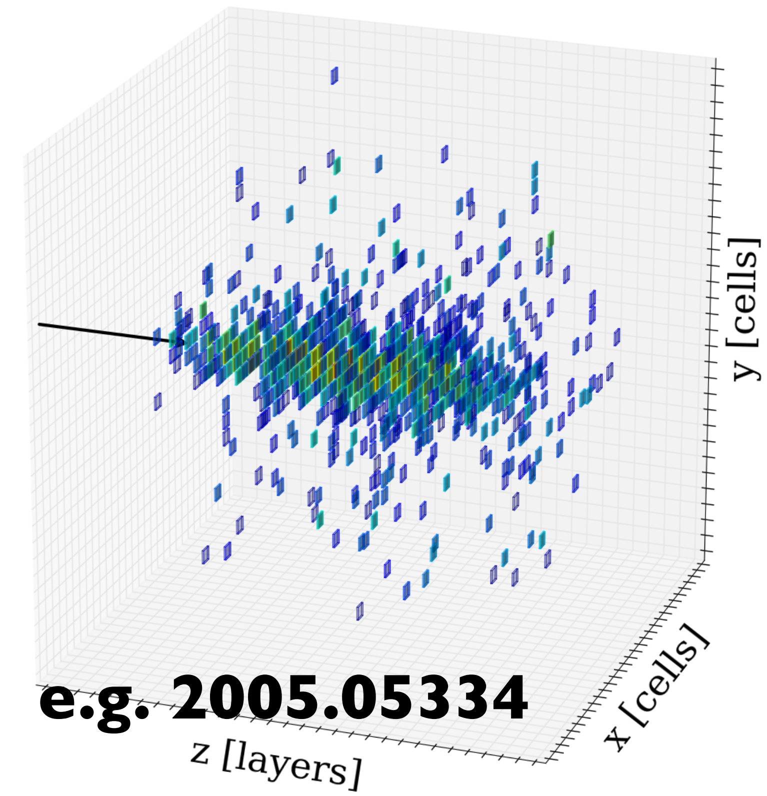




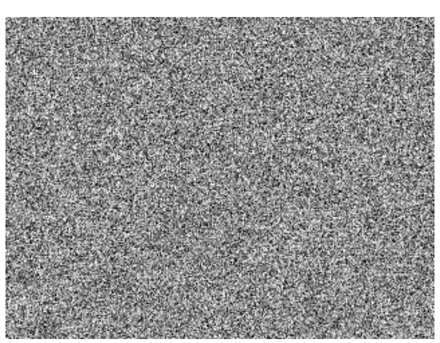
Classification network

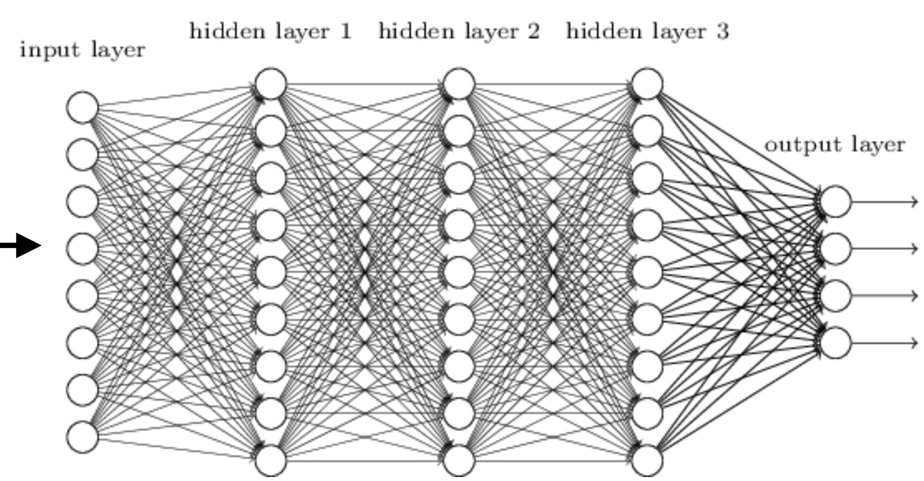
Unsupervised learning

- Learn the probability distribution
- No target needed, train directly on data
- Useful for:
  - Generative models
  - Anomaly detection
  - …?



e.g. 2005.05334



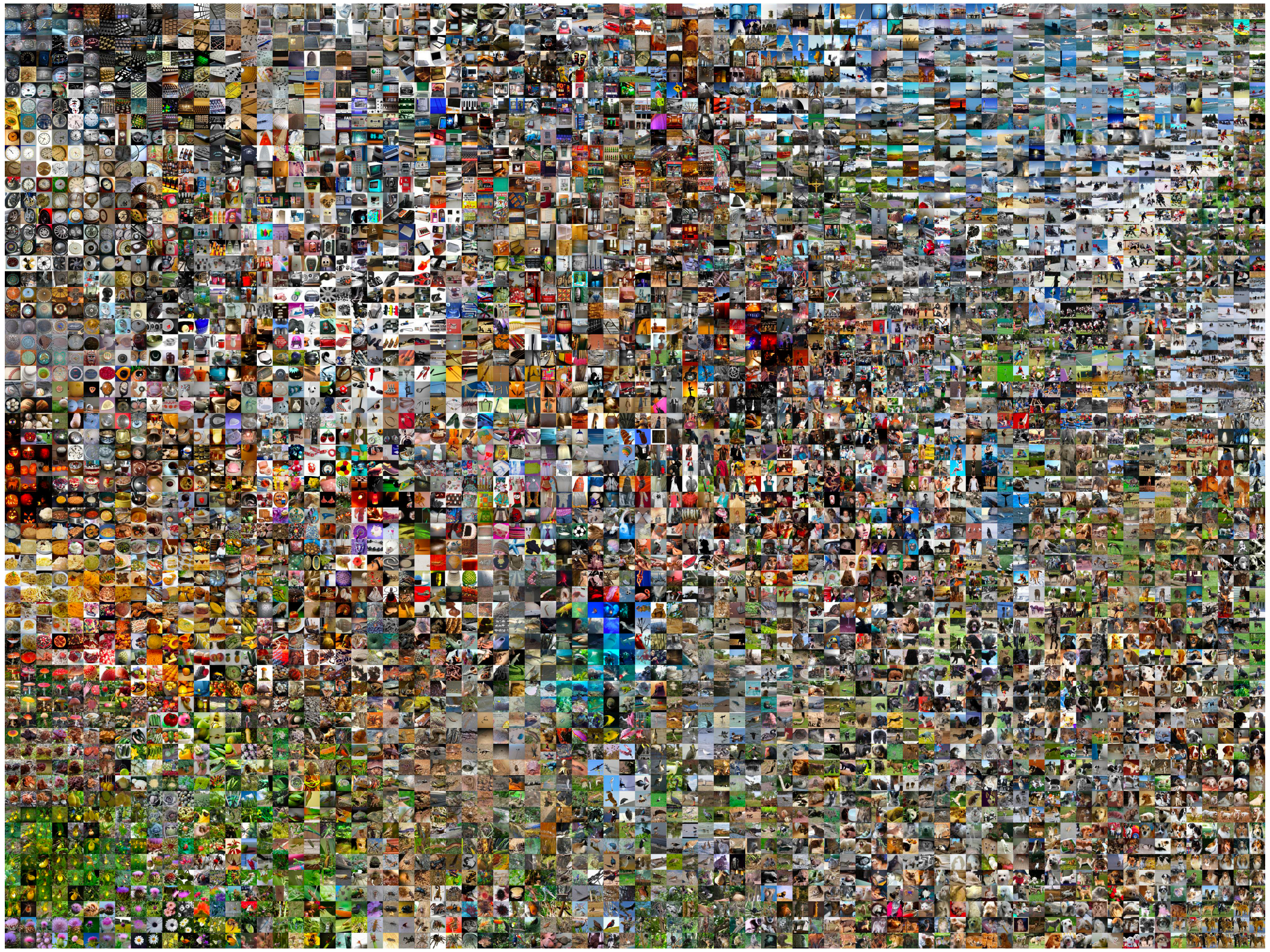


Generative network

Fast generative models

Generators

We **have**:
many images
(or collision events,
or detector readouts, …)



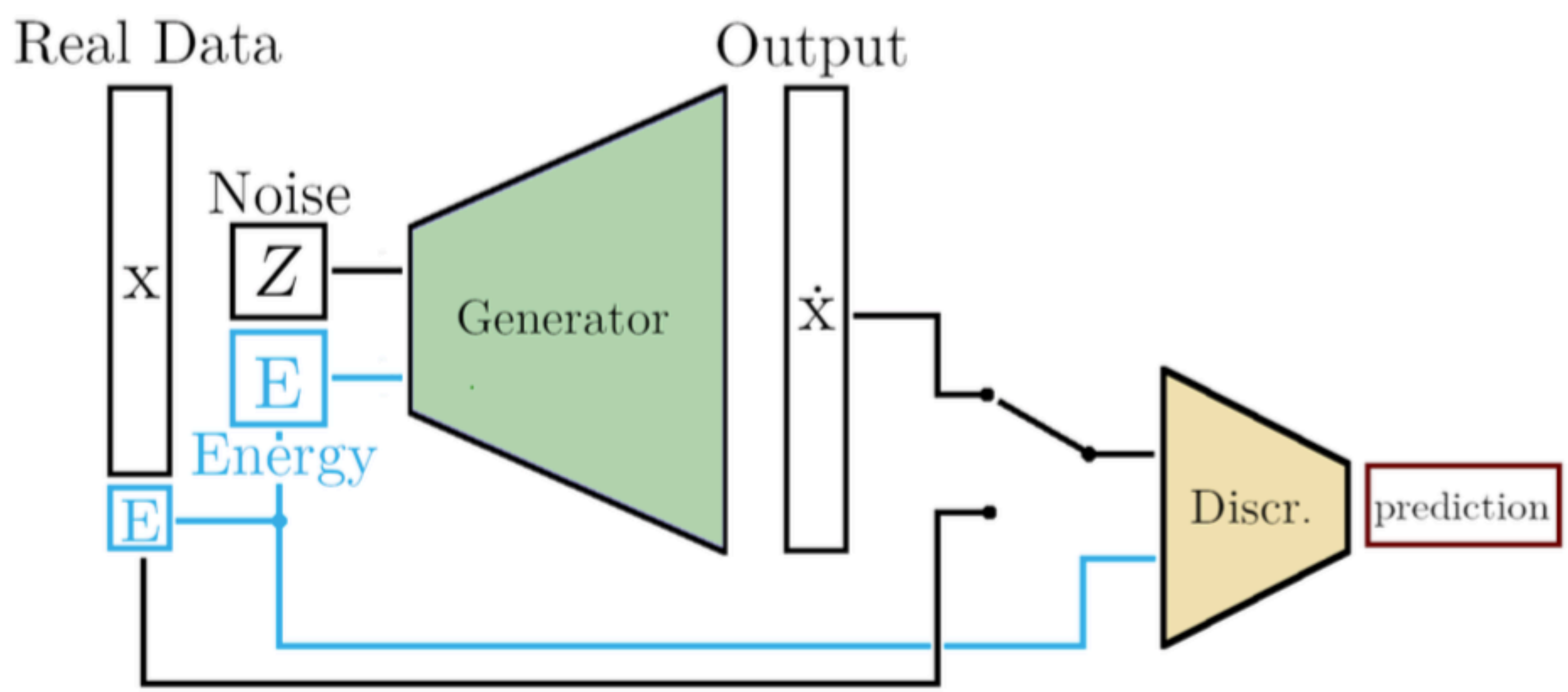
wired.com

We **want**: more images.

(Specifically: New examples that are similar to the examples, but not exact copies)

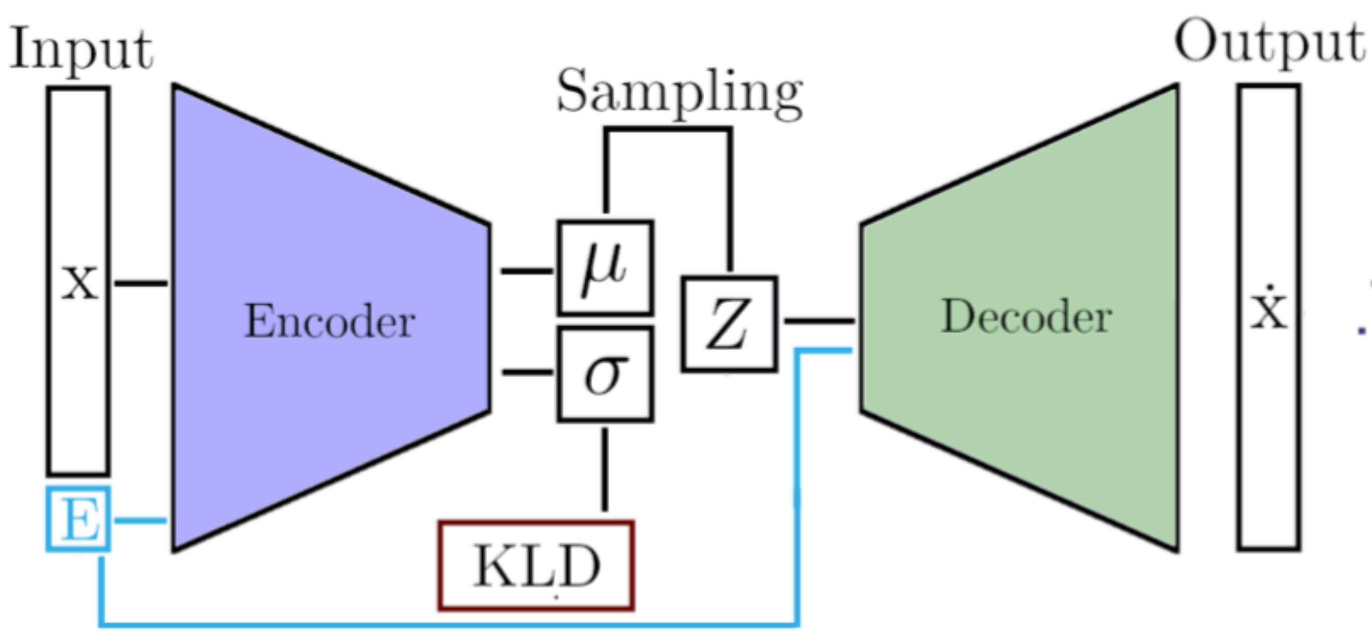
How to encode in neural net?

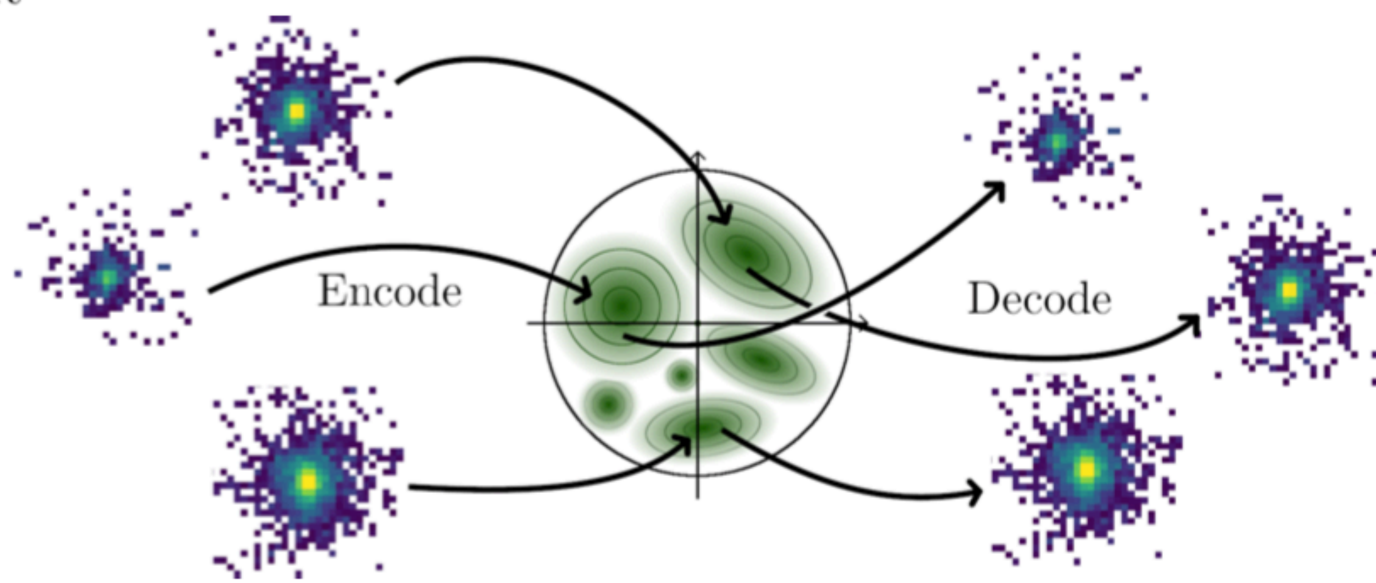
GAN



- **G**enerative **A**dversarial **N**etwork
- Generator generates new fake images from noise
- Second network (discriminator) learns to distinguish fake from real images
- Training via mutual feedback

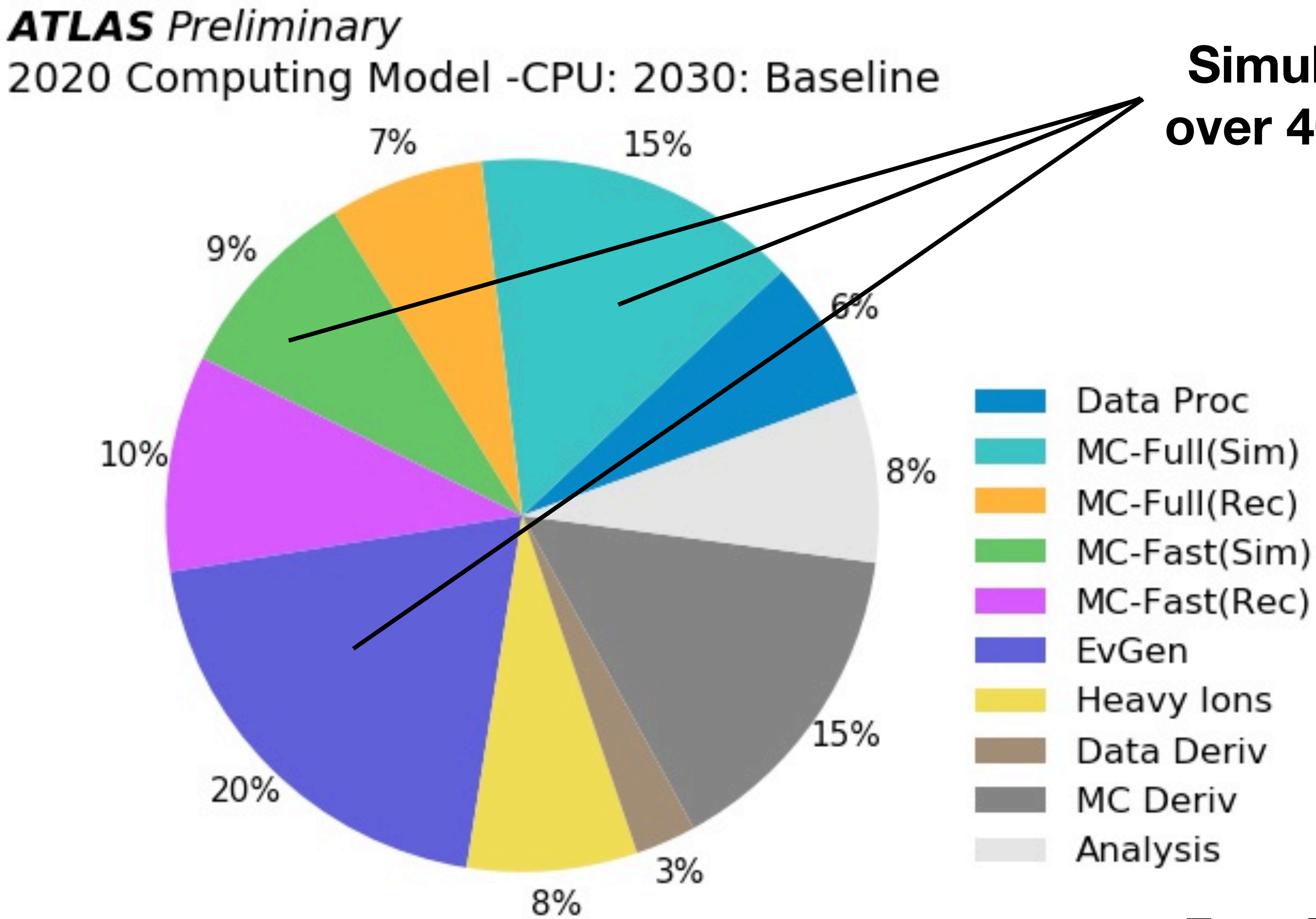
VAE





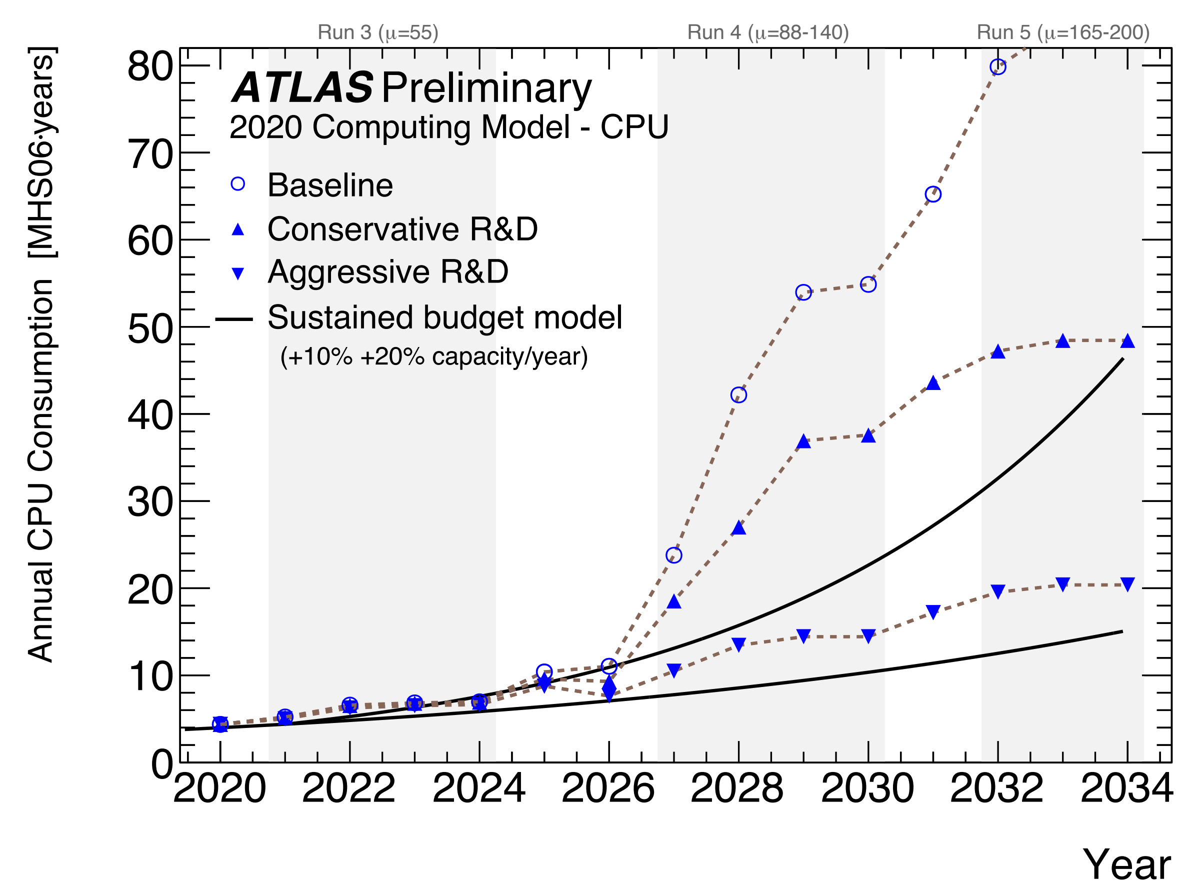
- **V**ariational **A**uto**e**ncoder
- Encode examples into latent space of network
- Sample from latent space to produce new examples

https://thispersondoesnotexist.com/



Simulation and Generation steps over 40% of ATLAS compute effort

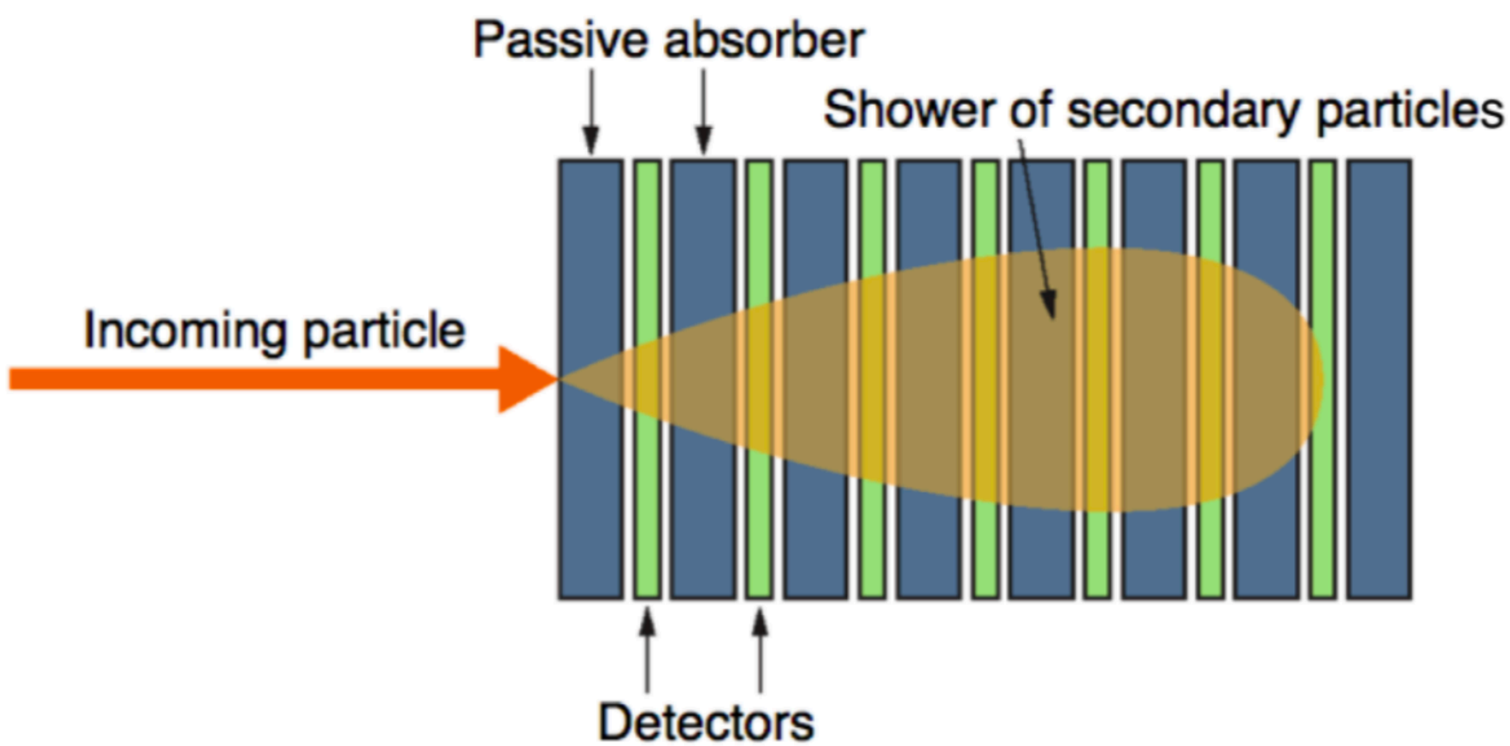
Compute needs measured in million-years of 2006 reference computer

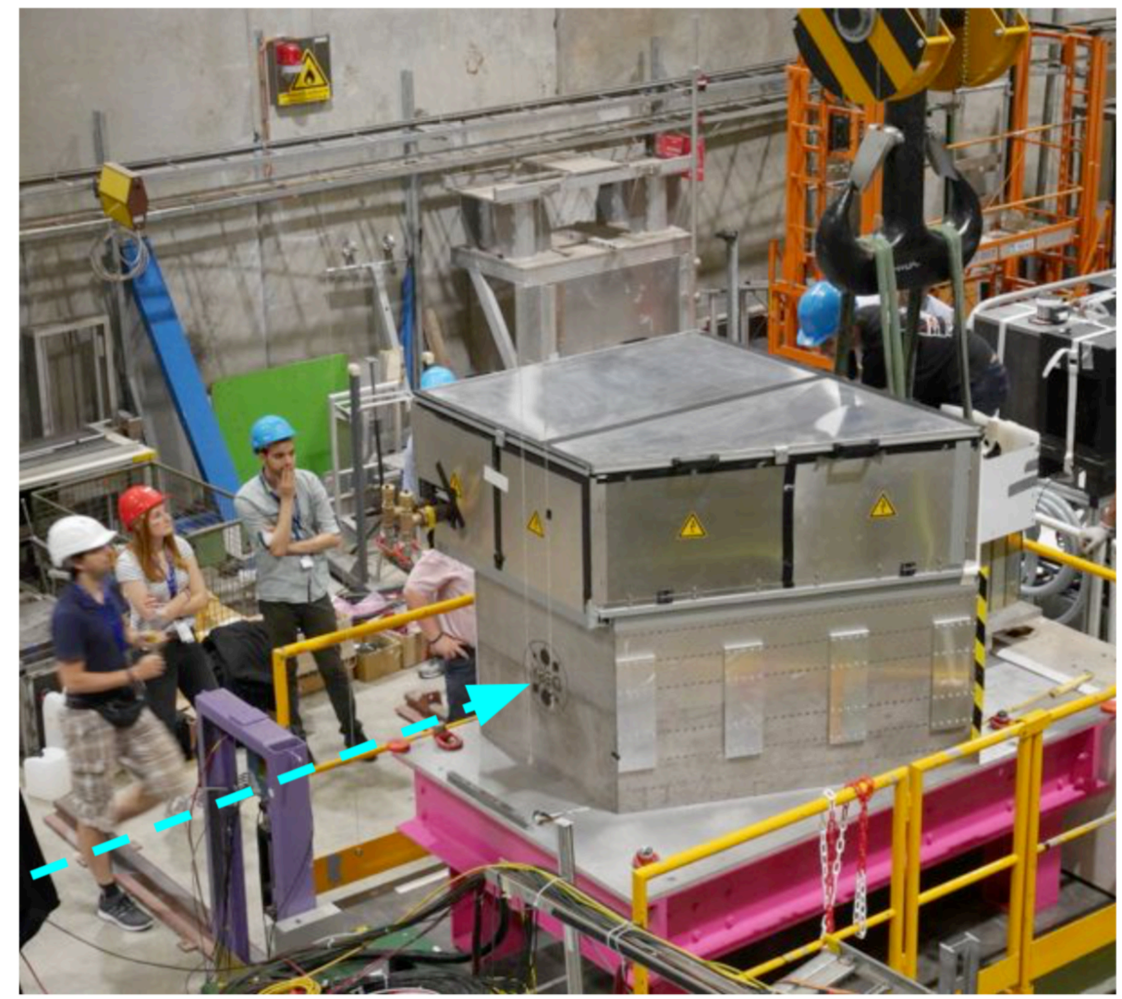


https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults
http://w3.hepix.org/benchmarking

Particle Showers

Main motivation:
Fast simulation of interaction between particles and detector material
Initial proposal: CaloGAN (1705.02355)

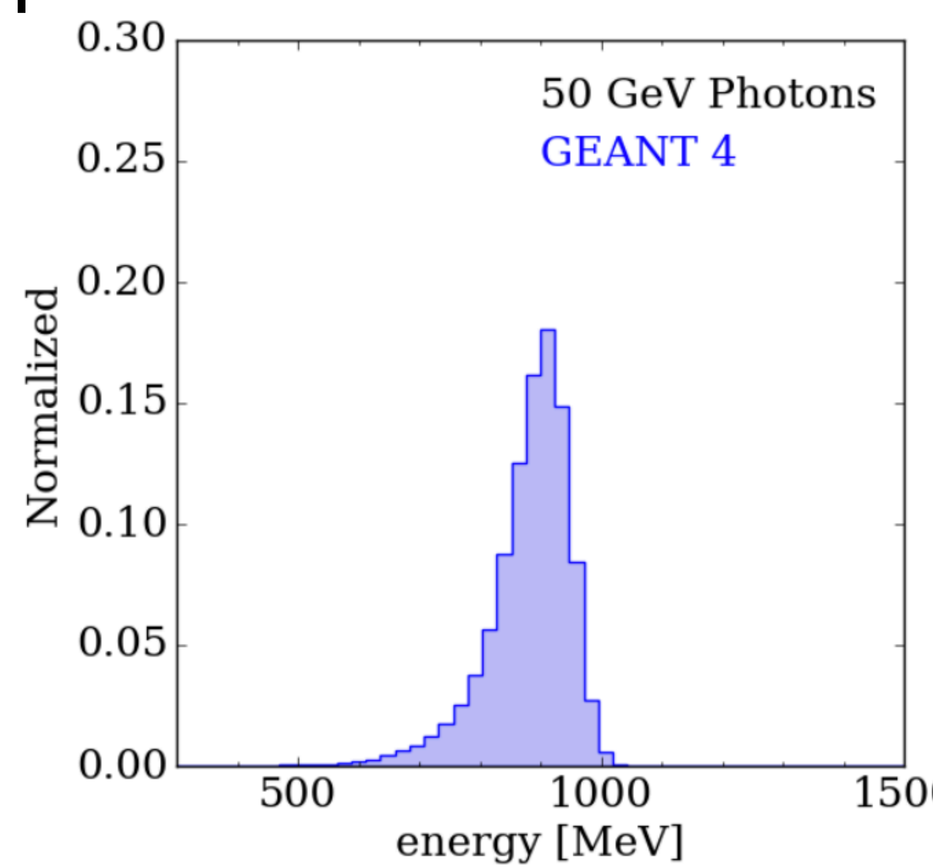




Generative models are also applied to:
phase space integration and sampling, event generation, ….

Additional Challenges

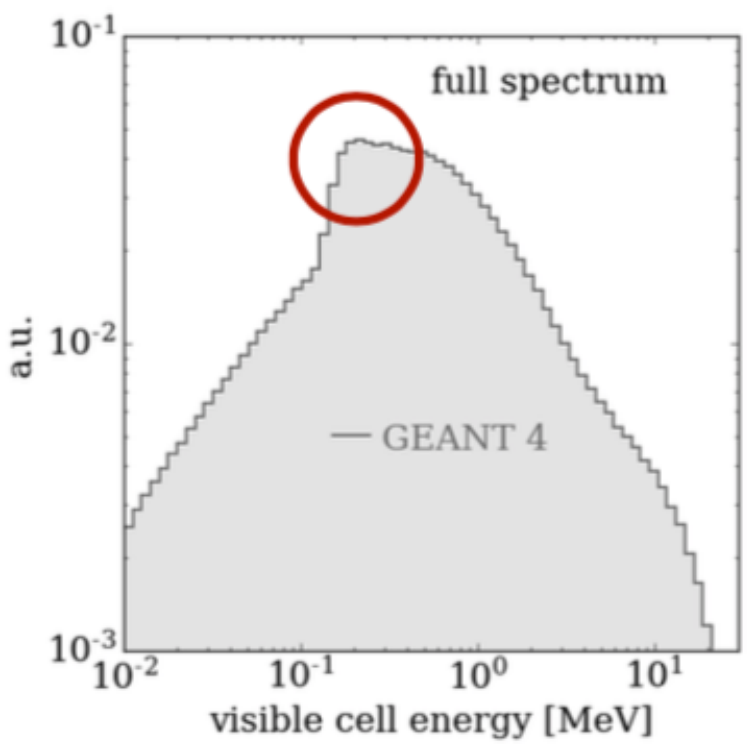
- How to evaluate convergence of models?
- Correctly model differential distributions
- Condition on a large number of quantities (energy, particle type, impact position, angle, …)
- Other considerations:
  Coverage (do I produce example for all phase space?)
  Saliency (is this a good example of the desired type of event)
  Mode collapse
  Overfitting

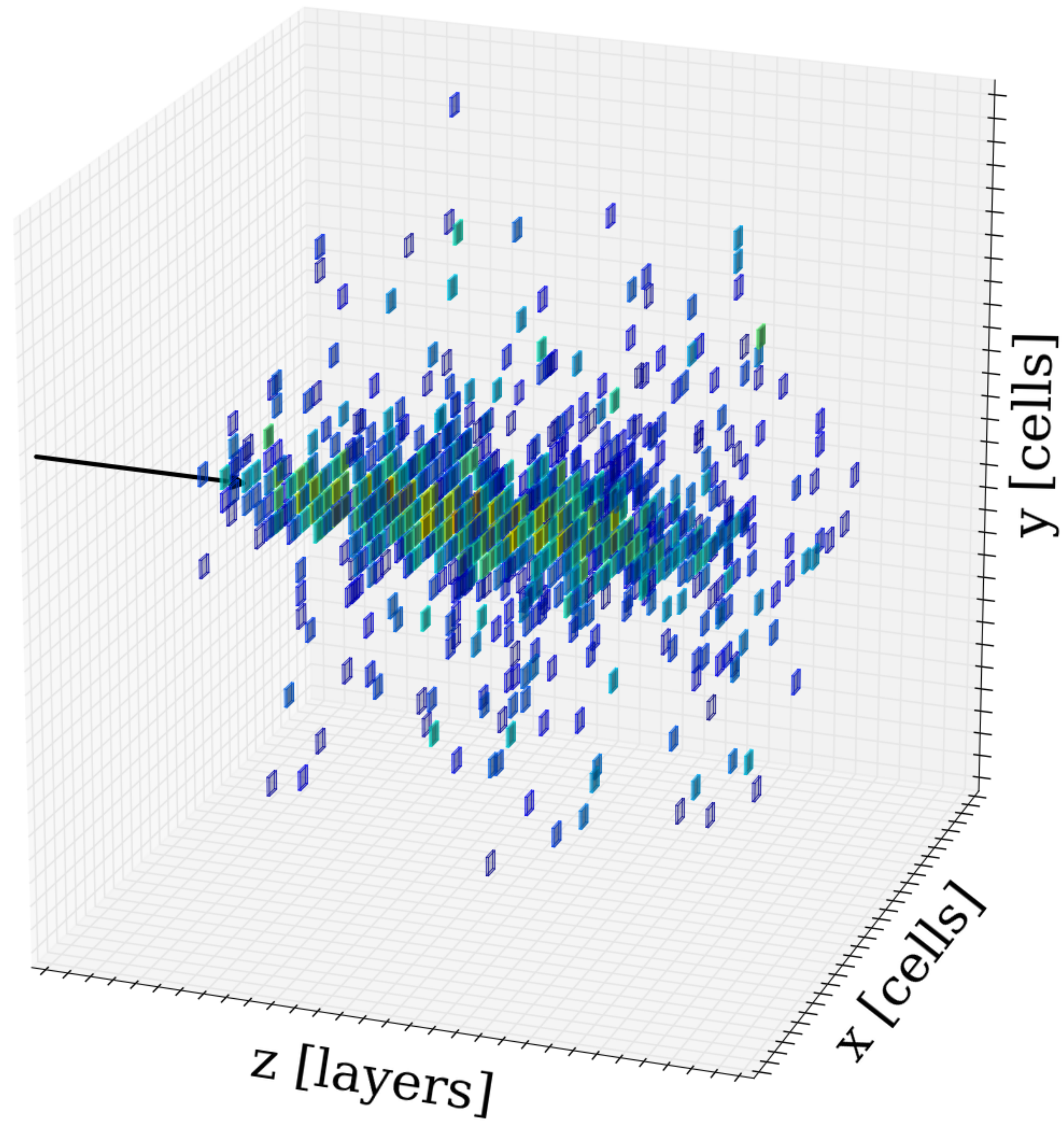


Concrete Problem

Describe photon showers in high granularity calorimeter prototype

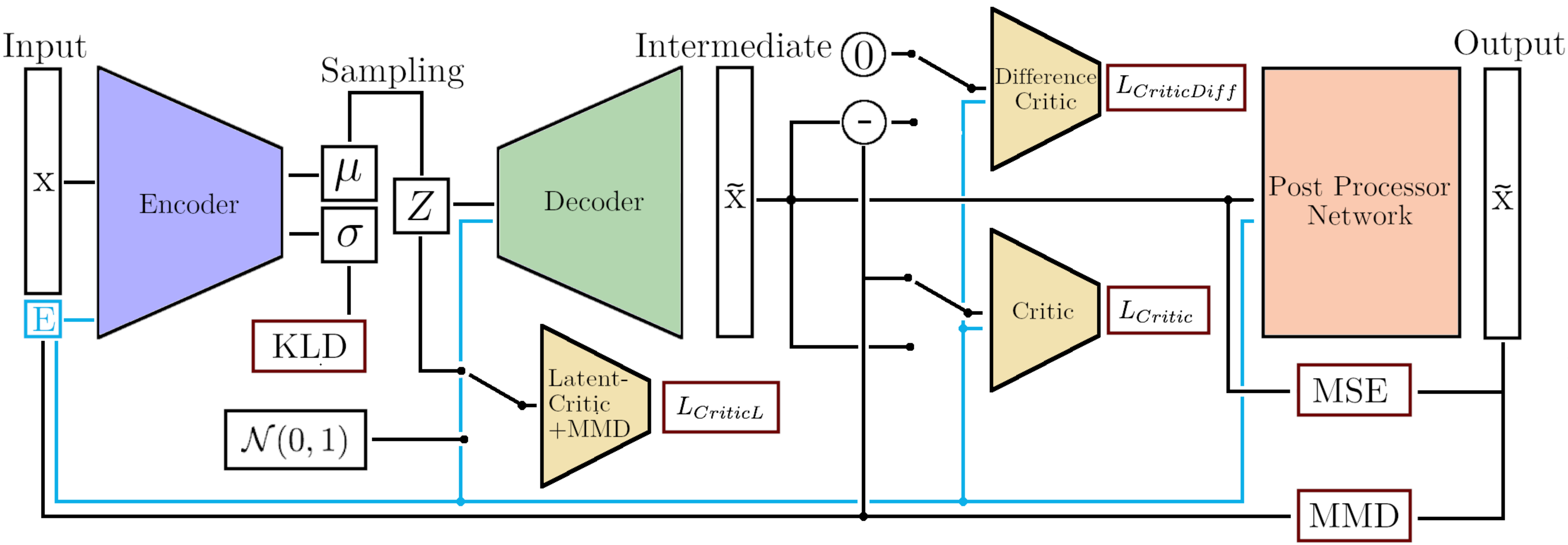
- 30x30x30 cells (Si-W)
- Photon energies from 10 to 100 GeV
- Use 950k examples (uniform in energy) created with GEANT4 to train





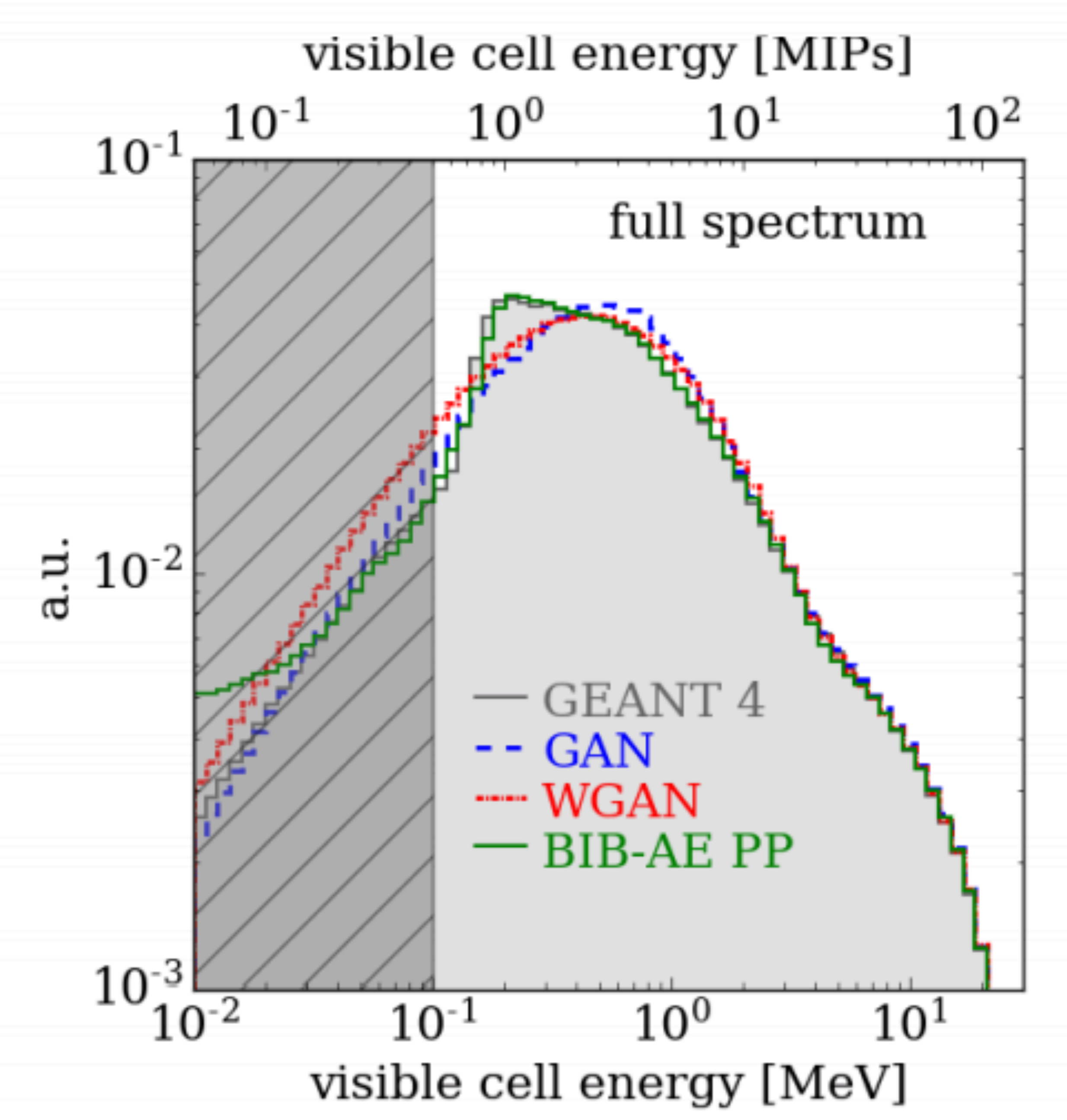
- Not only model individual images but also **differential distributions**

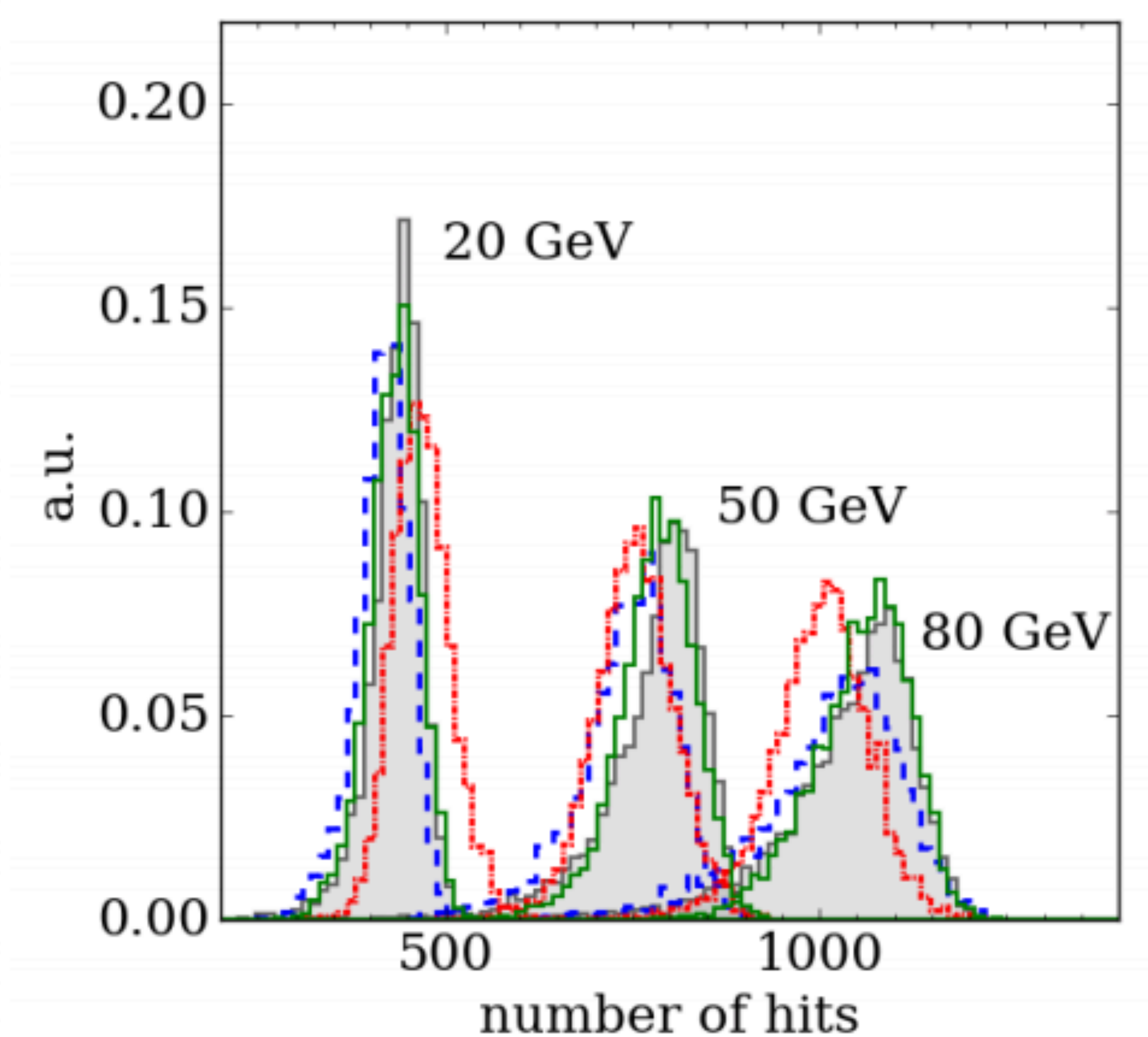
Architecture



- BIB-AE (based on 1912.00830) with added post-processing
- Unifies features of GAN and VAE
- 71M trainable parameters

Result





Can now learn differential distributions
Still room to improve

Potential Limitations

- Generative models are powerful in quickly producing more examples, still need training examples
- Machine learning is great at interpolation, but it cannot do magic
- Expect to simulate typical examples, do not trust the tails of distributions without verification
- Can networks **amplify?**

Amplification: Setup

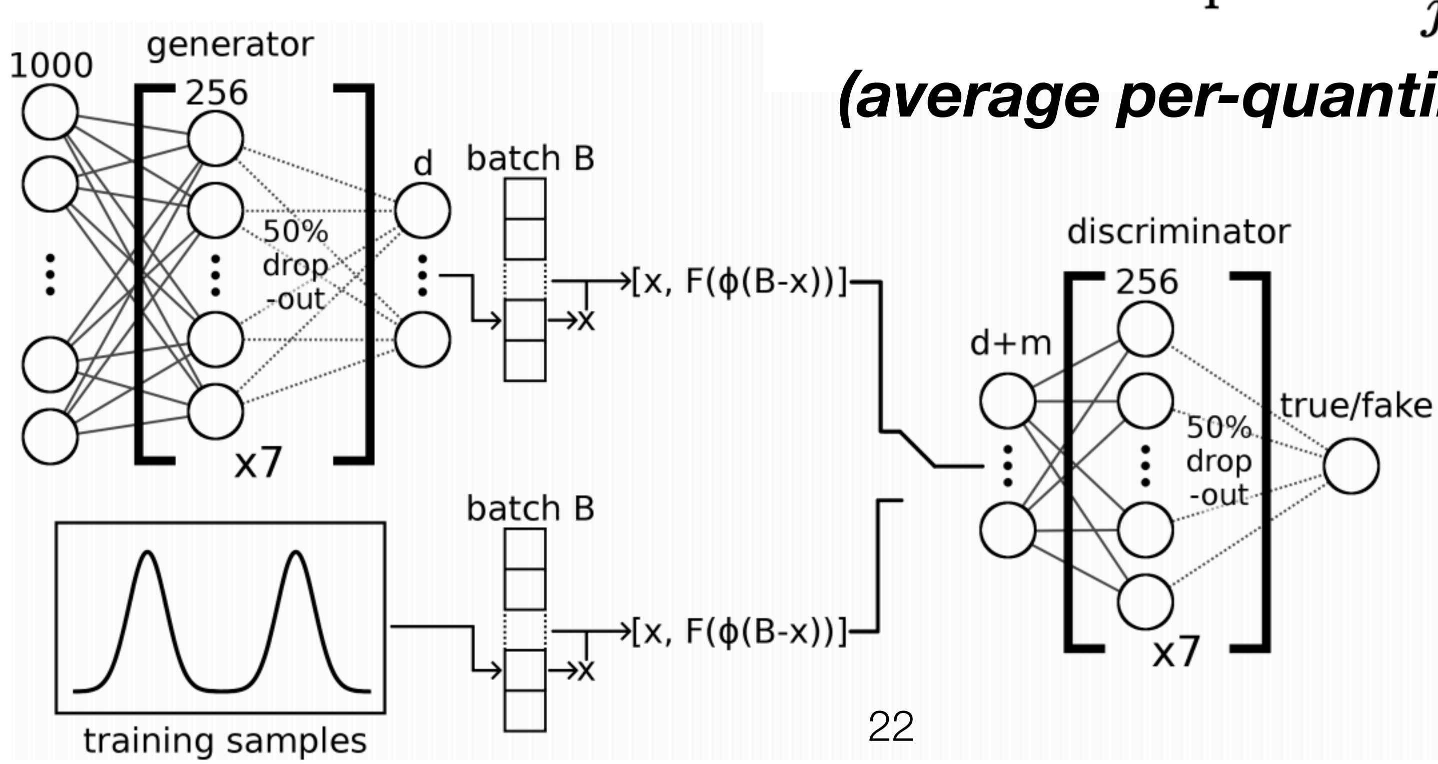
- Setup:
  - Draw N examples from known *truth* function
  - Use to train GAN
  - Sample M>N events from GAN
- Compare per-quantile difference to truth between
  - Initial N examples
  - M GANed examples
  - Fit

$$P(x) = \frac{N_{-4,1}(x) + N_{4,1}(x)}{2}$$

(double Gaussian)

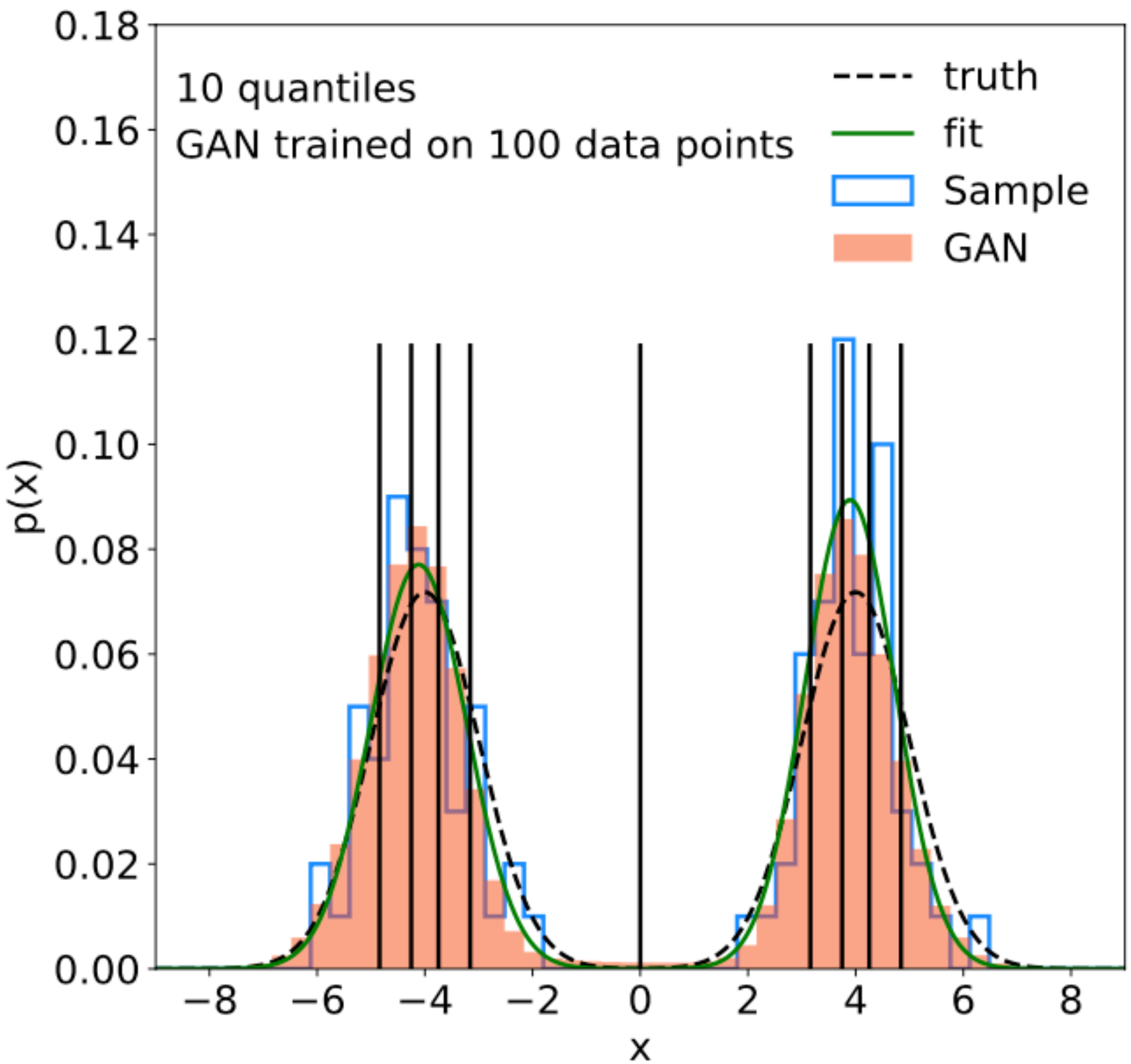
$$\mathrm{MSE} = \frac{1}{N_{\mathrm{quant}}} \sum_{j=1}^{N_{\mathrm{quant}}} \left(x_j - \frac{1}{N_{\mathrm{quant}}} \right)^2$$

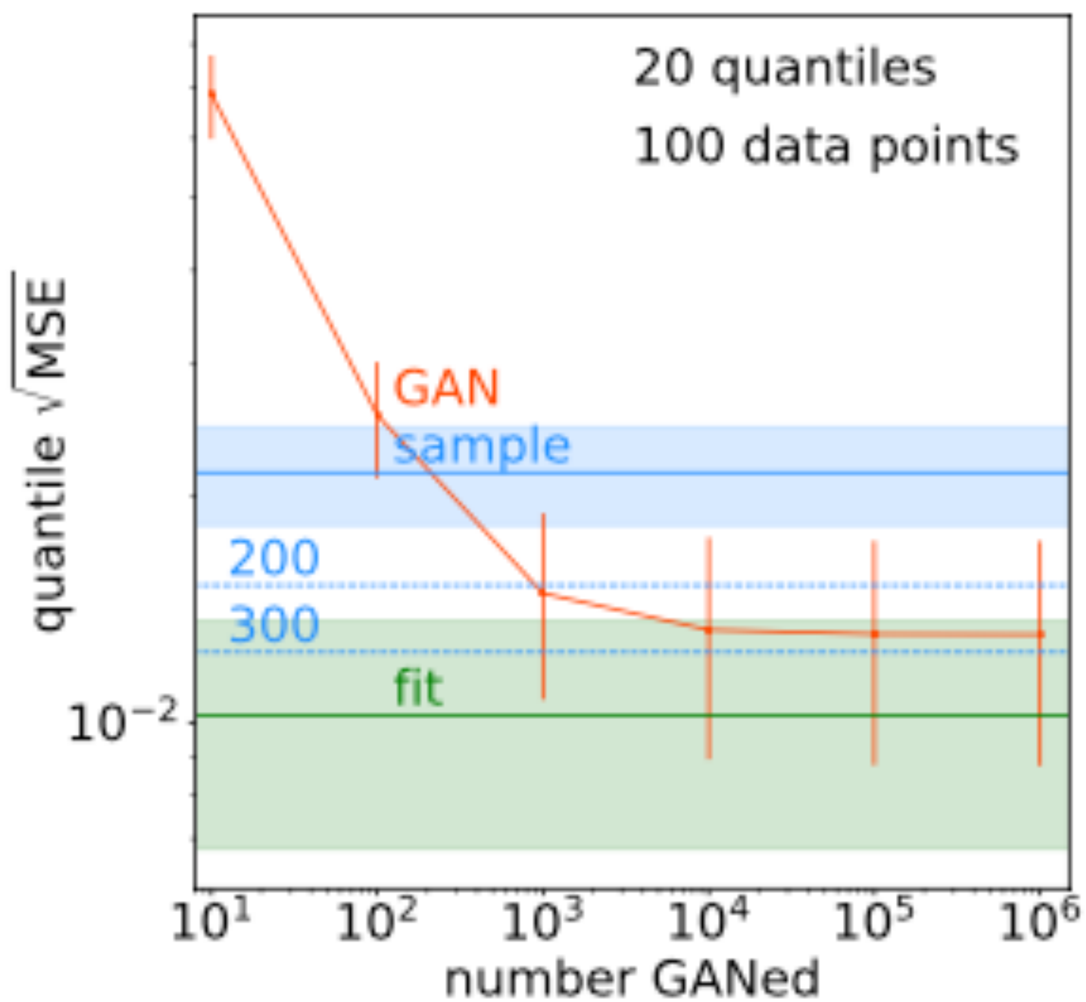
(average per-quantile difference to truth)



2008.06545

Amplification 1D



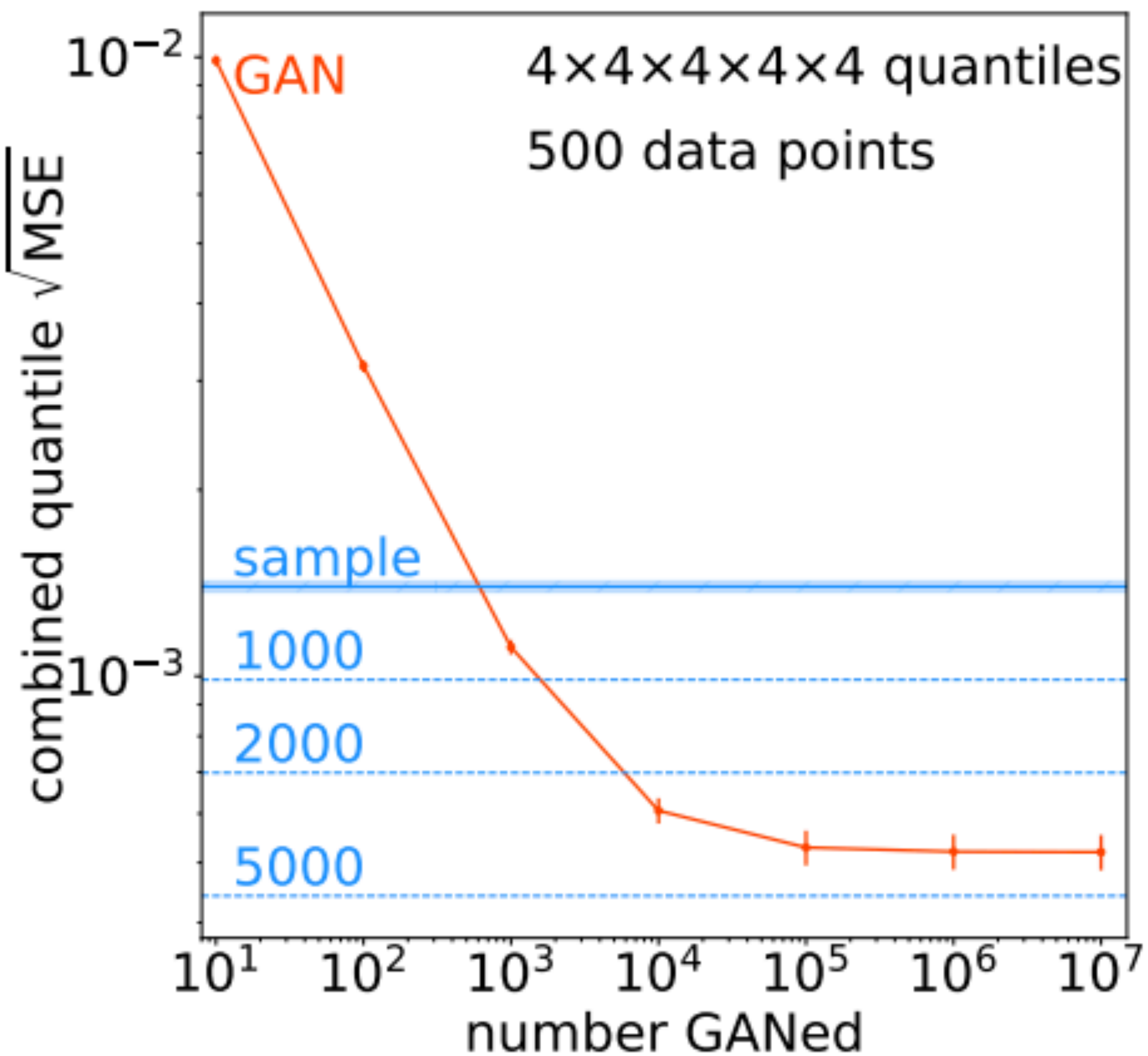


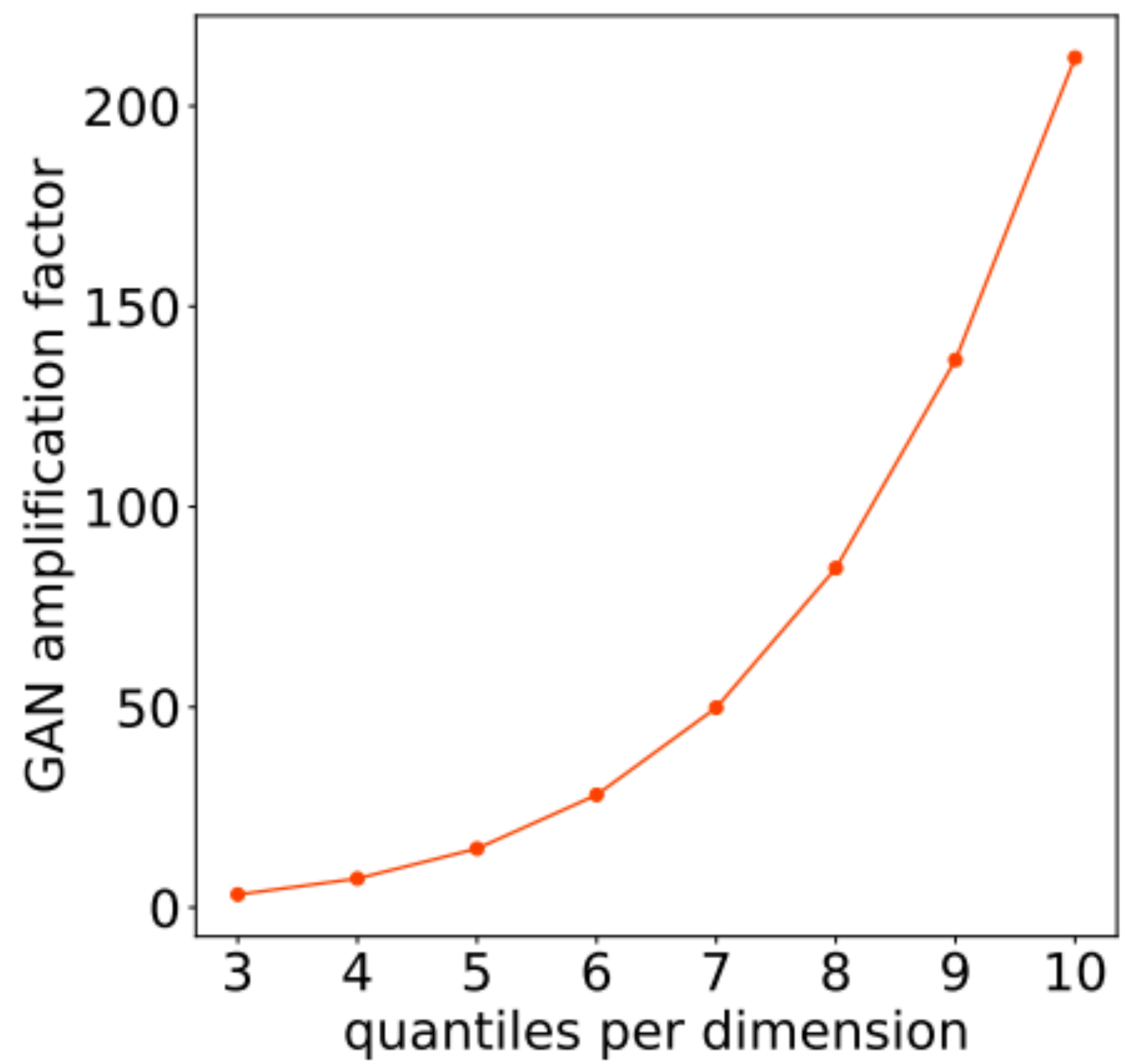
Improve statistics of training sample by interpolation

2008.06545

Amplification 5D

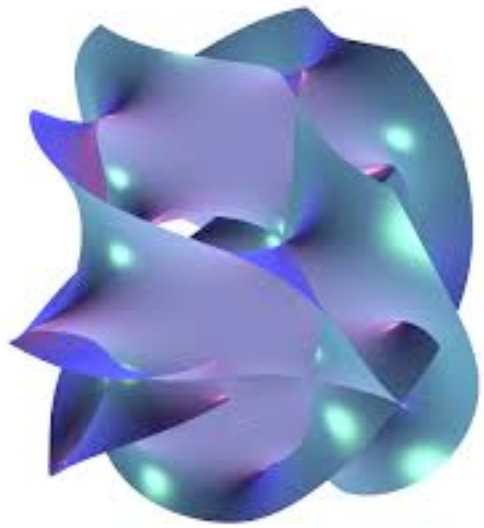
Use spherical shell instead of double Gaussian



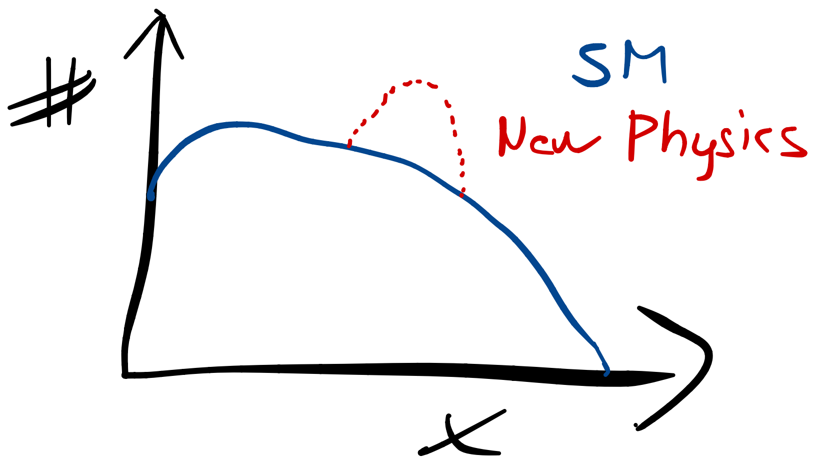


2008.06545

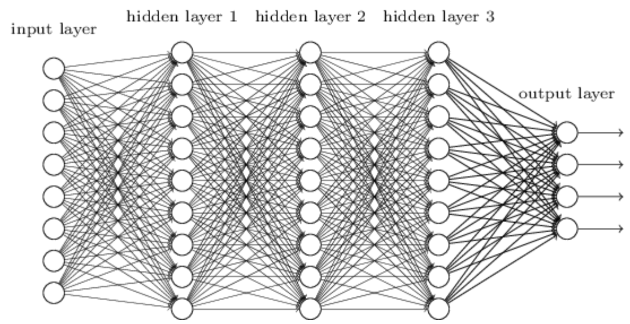
Unsupervised anomaly detection



Decide new physics model to test

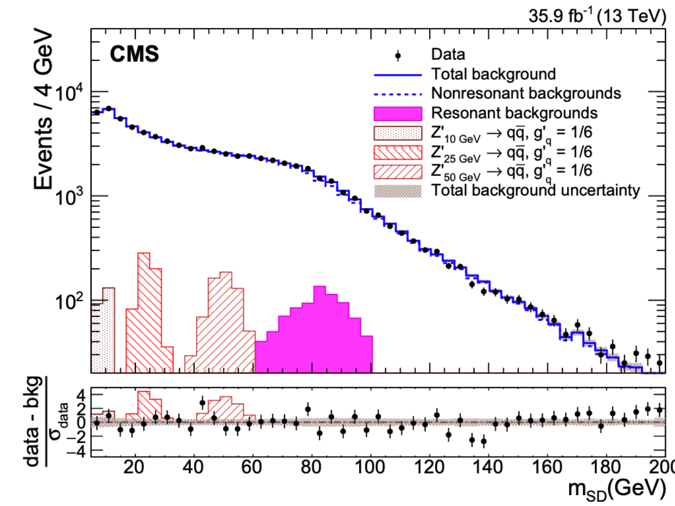


Use Monte Carlo simulation to provide realistic estimate of effect new physics and Standard Model prediction

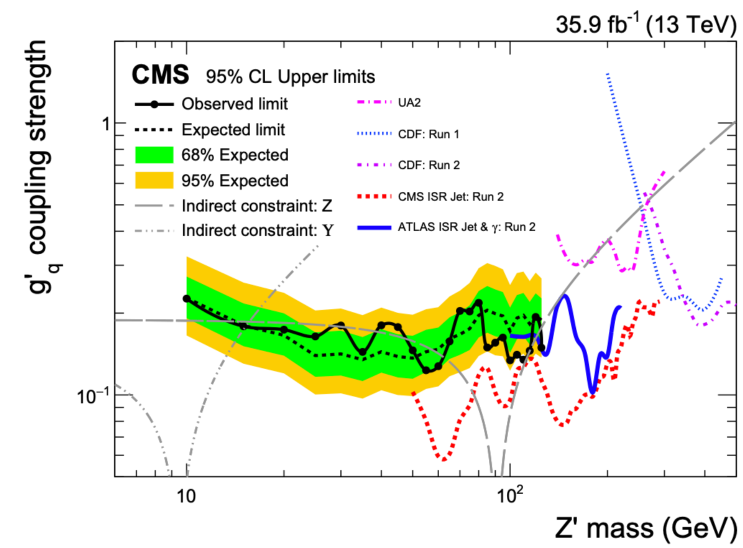


Find a test statistic (e.g. selection criteria and classifier output)

Apply to measured data

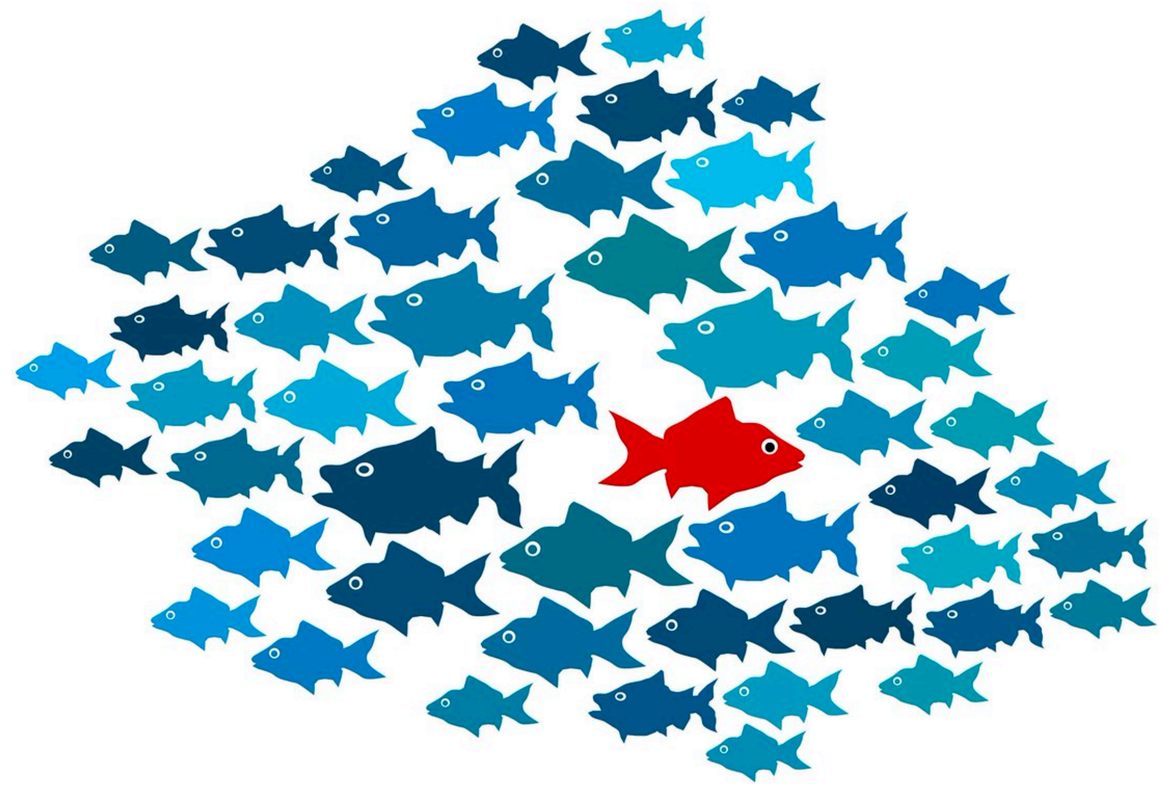


Perform statistical analysis / hypothesis test



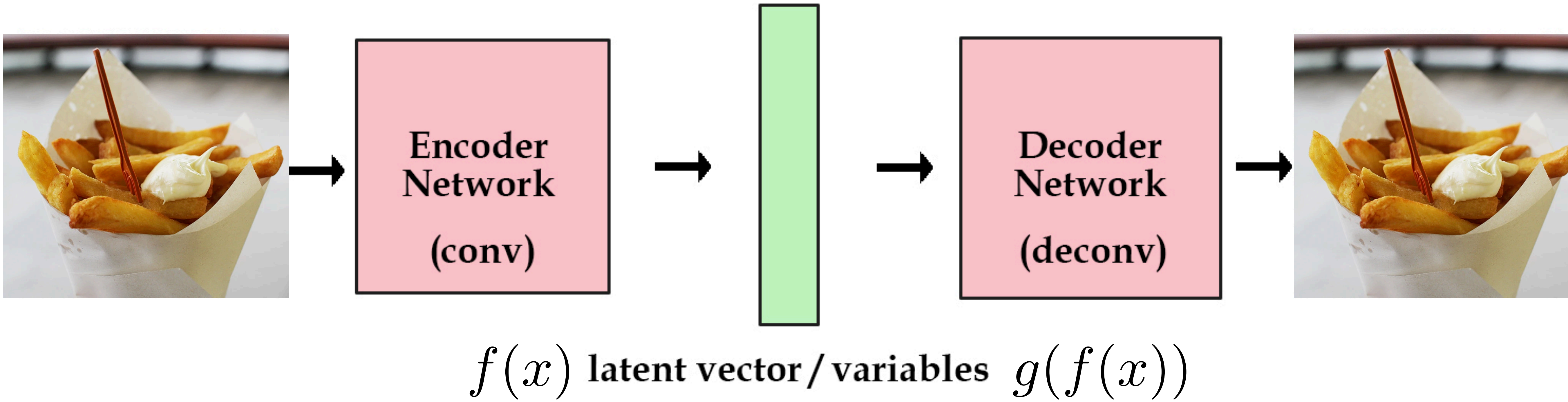
Collect Nobel prize

Problem: The potential space of new theories is **HUGE**. Cannot cover all possible models.



Can we look for new physics, without knowing what to search for?

Autoencoder

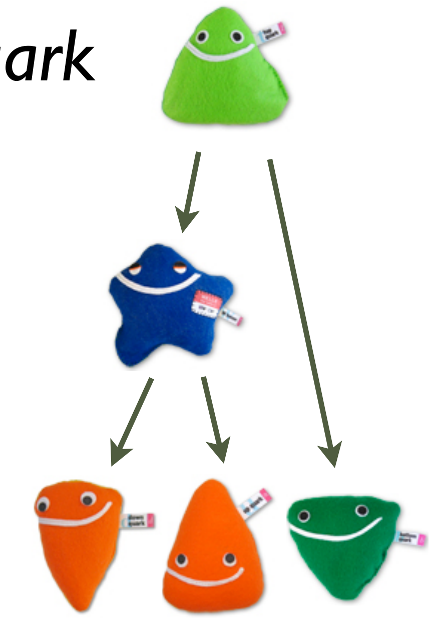


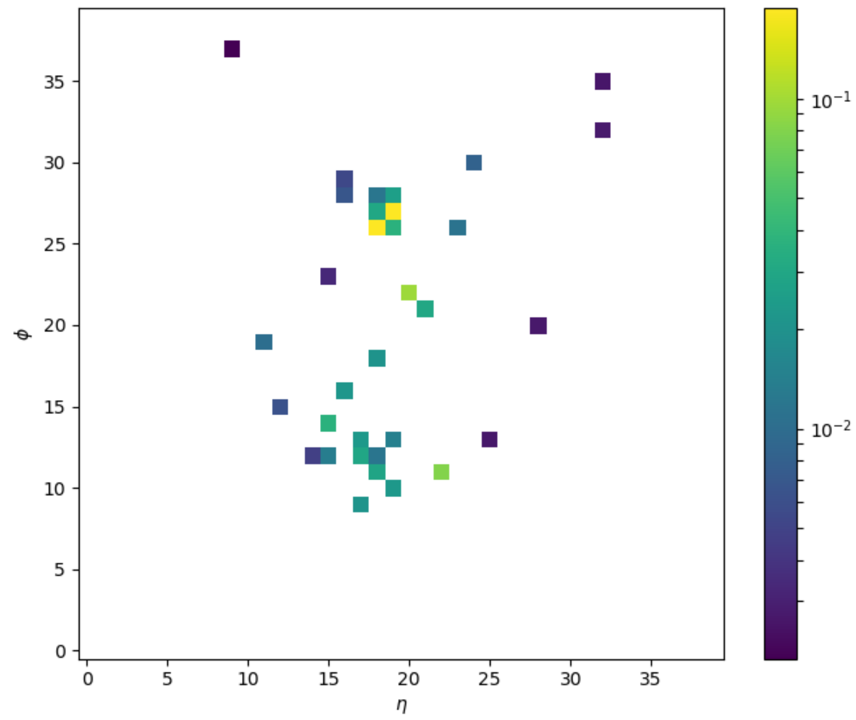
$$L = (\hat{y} - g(f(x)))^2$$

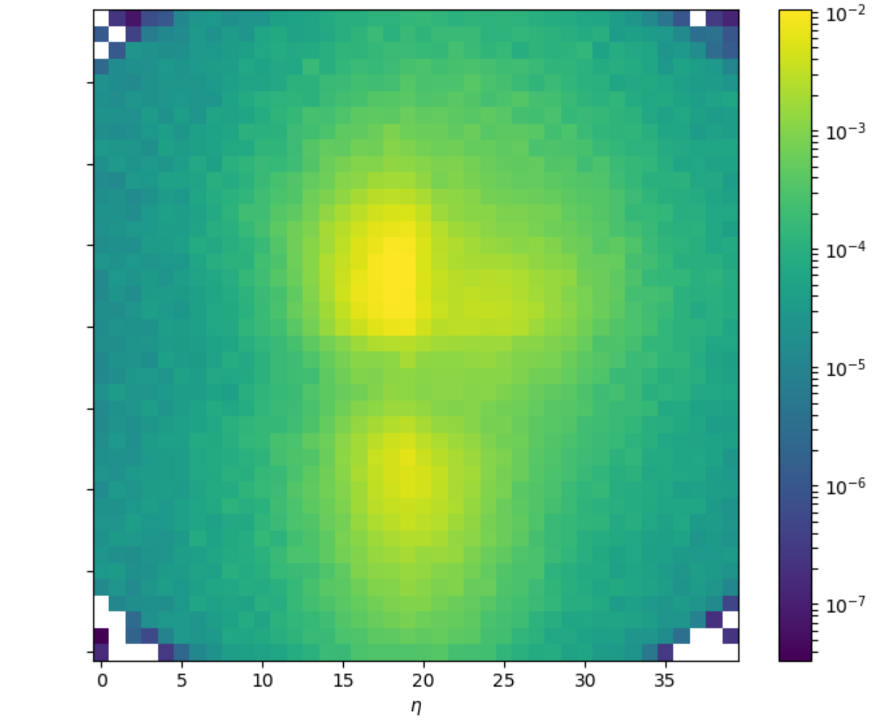
- Weakly supervised learning
- *Latent space/bottleneck* with compressed representation
- Dimension reduction
- Denoising

kvfrans
deeplearningbook.org

Top Quark Jet

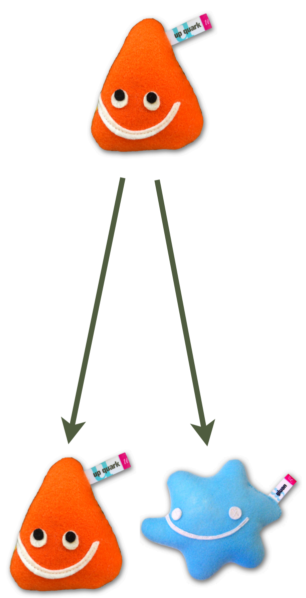


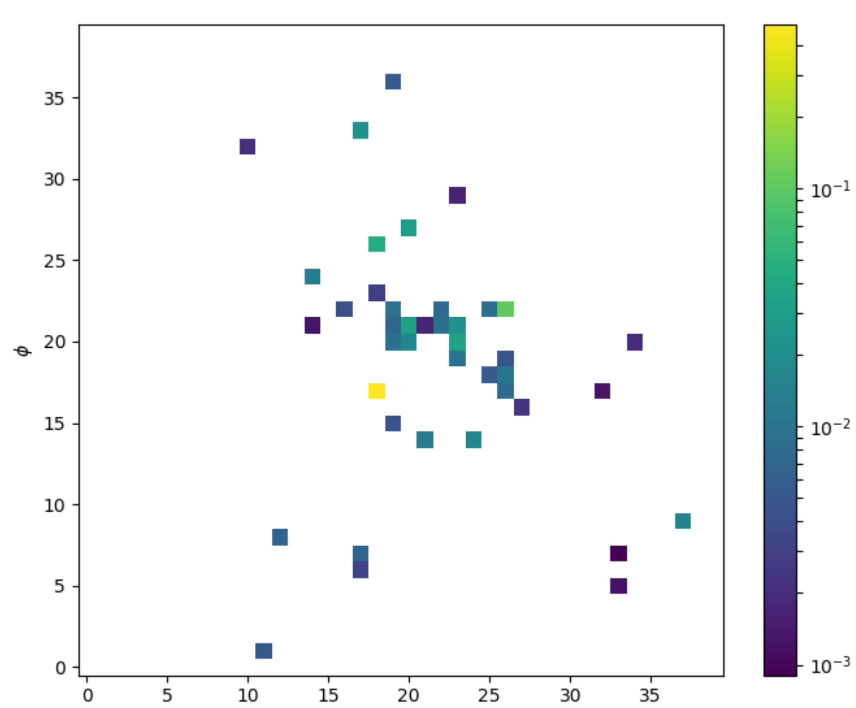


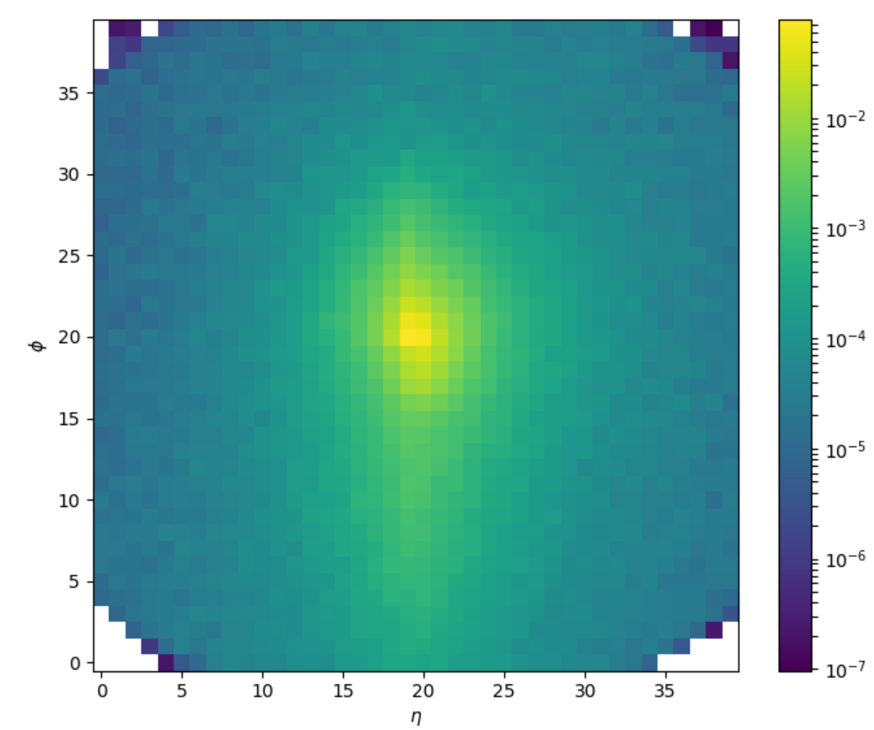


Example: Jet Images

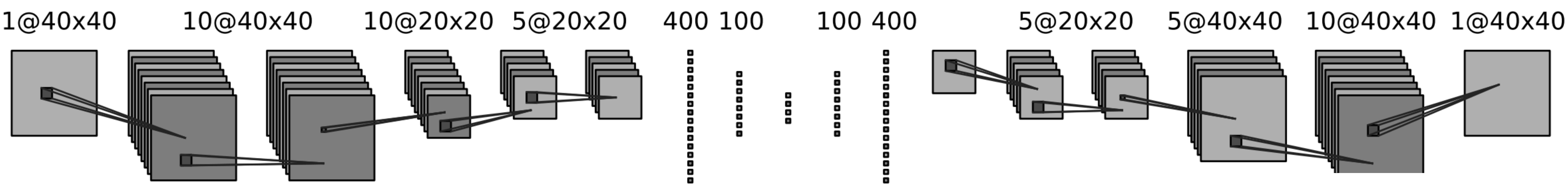
QCD Jet

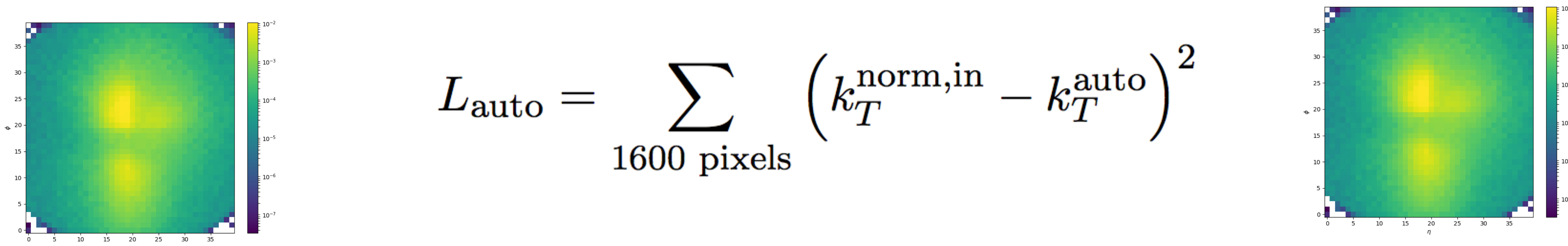






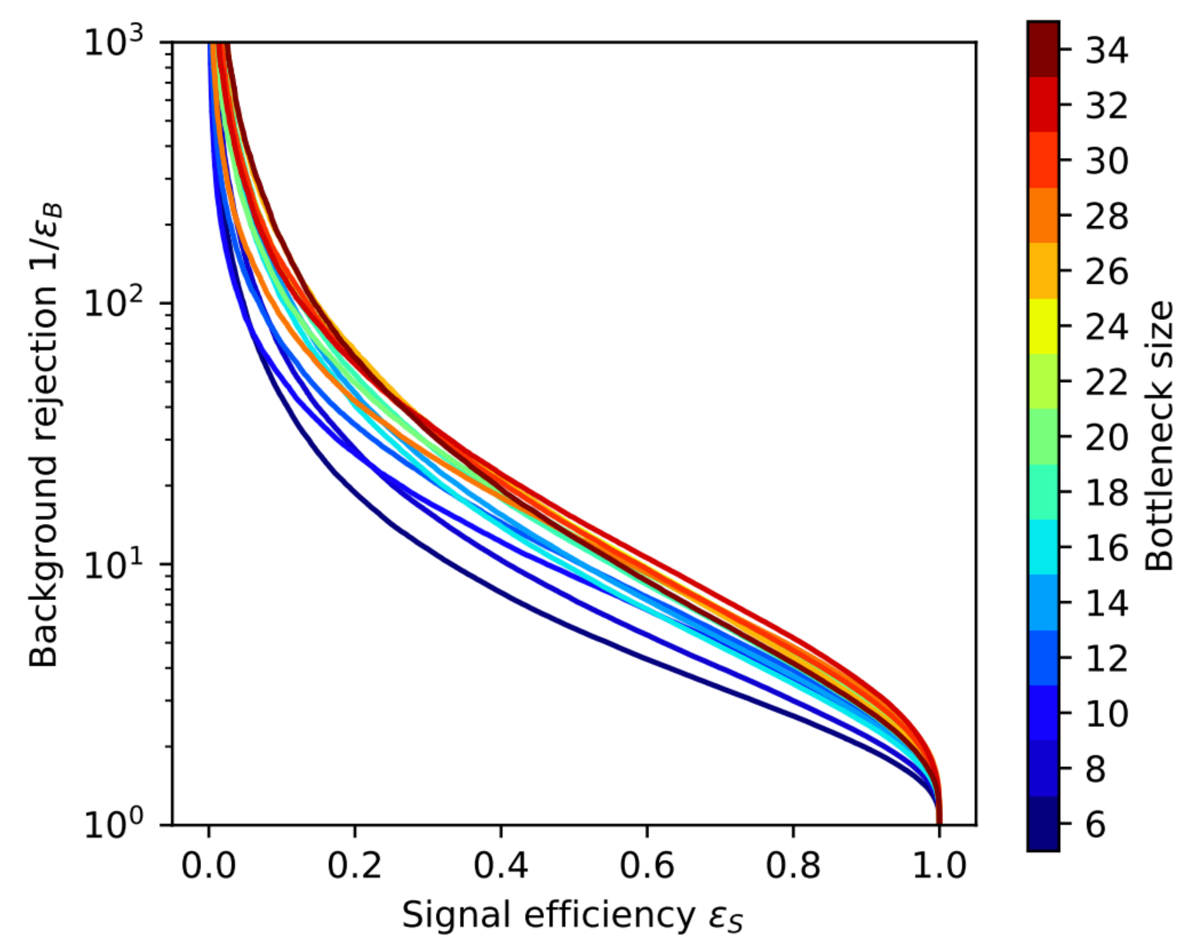
Autoencoder





$$L_{\text{auto}} = \sum_{1600\ \text{pixels}} \left(k_T^{\text{norm,in}} - k_T^{\text{auto}} \right)^2$$

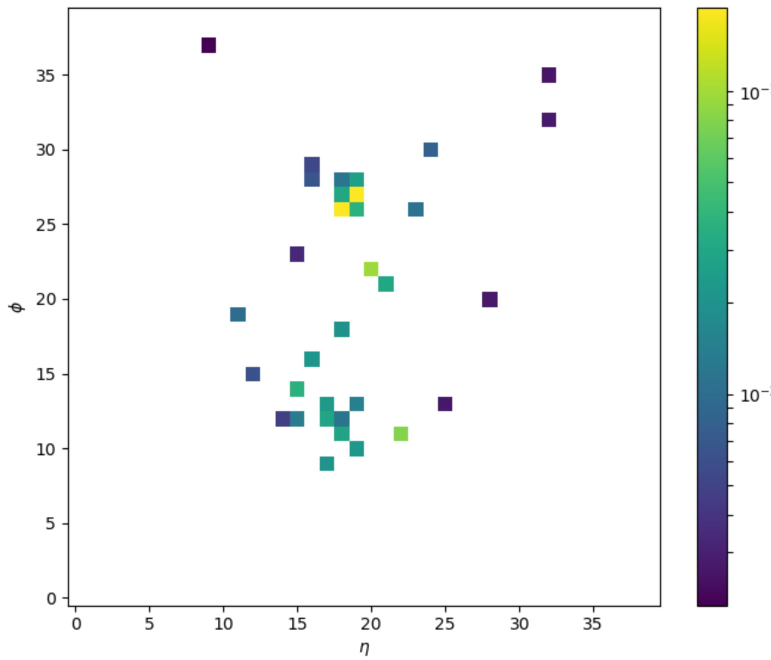
- Train on pure QCD light quark/ gluon jets and apply to top tagging
- Top quarks/ new physics identified as anomaly

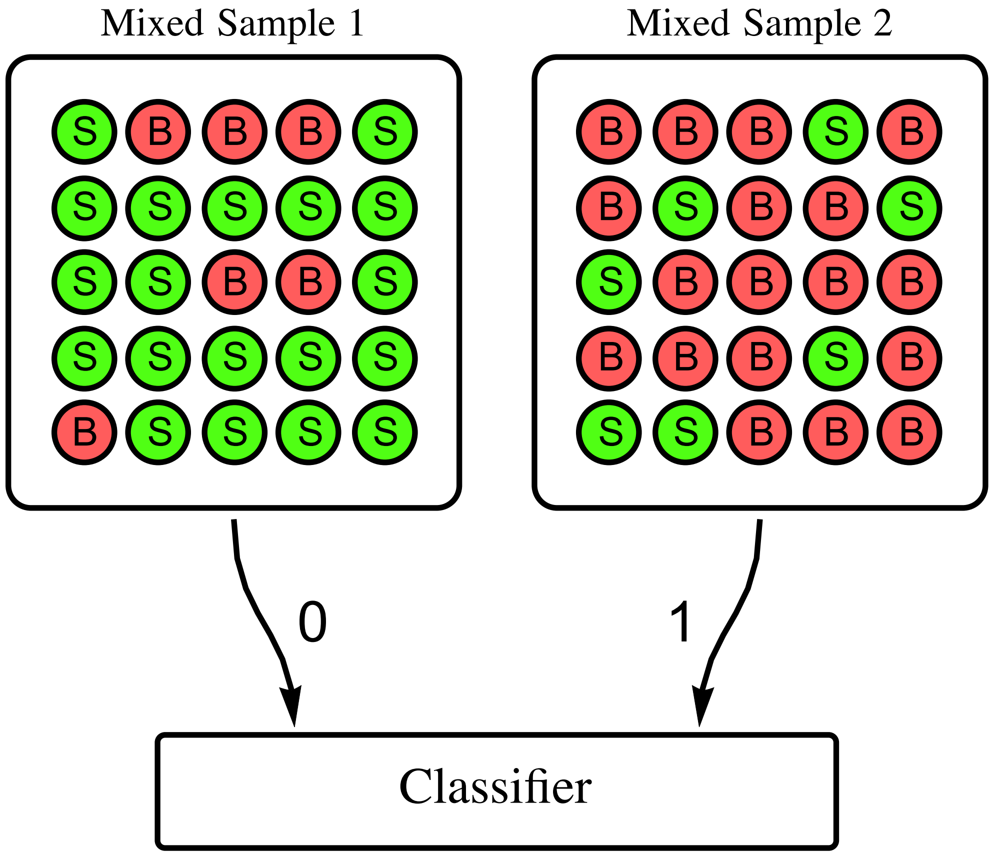


QCD or What?
T Heimel, GK, T Plehn, JM Thompson, 1808.08979
Searching for New Physics with Deep Autoencoders
M Farina, Y Nakai, D Shih, 1808.08992

Caveats

- Anomaly score for a given signature depends on complexity of signal/background in addition to training data
- We are not looking for individual anomalous events but anomalous regions of phase space
- Usual L2 difference not optimal as loss:
  - Different distributions of pixels compatible with same physics
- Potential improvements from Variational Autoencoders

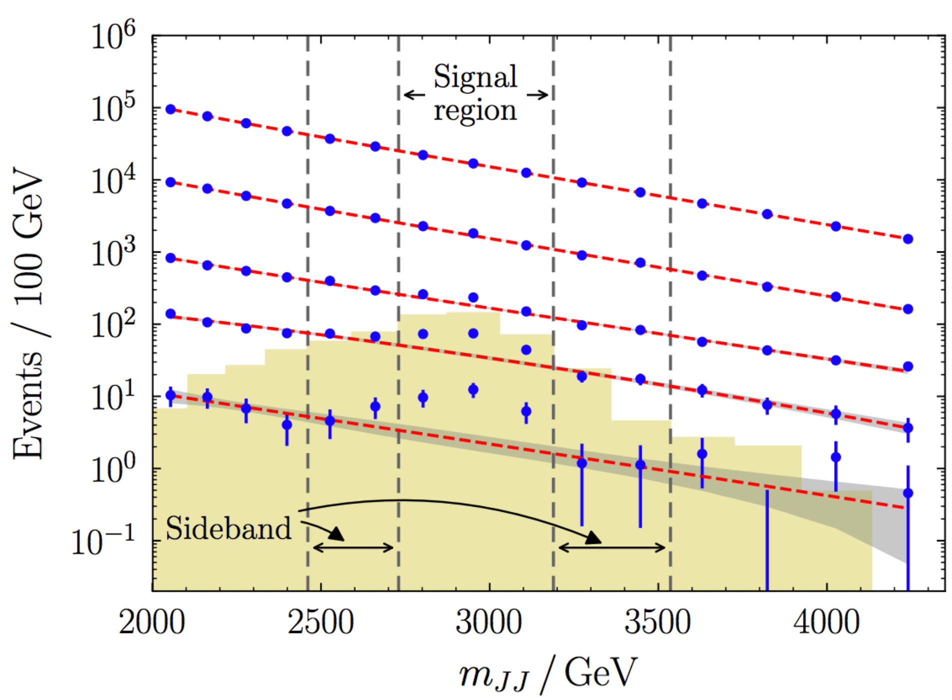


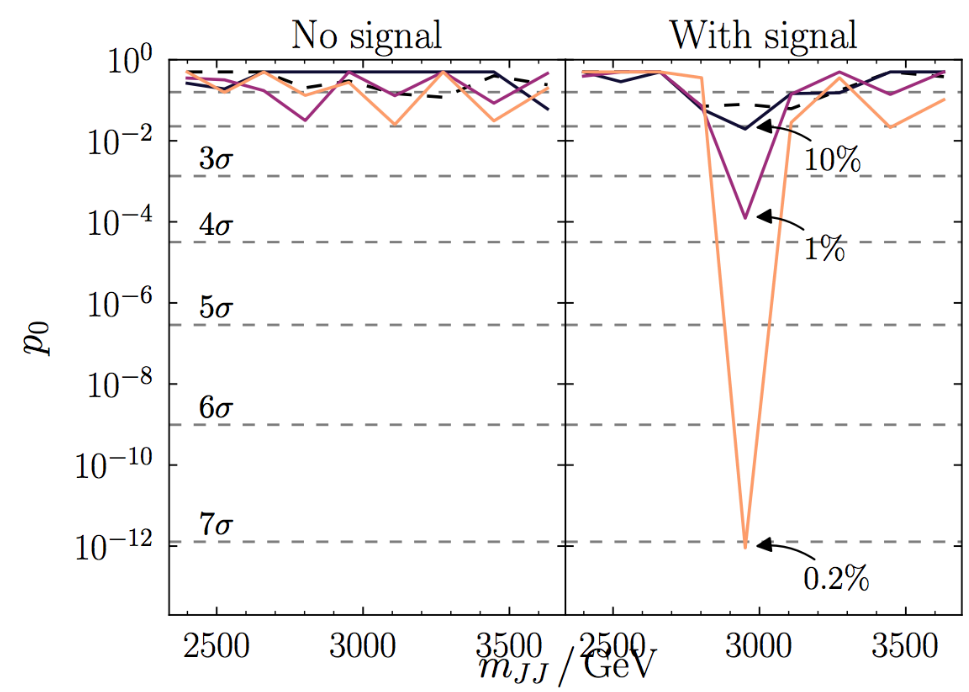


CWola Hunting

$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1 - f_1)\, p_B}{f_2\, p_S + (1 - f_2)\, p_B} = \frac{f_1\, L_{S/B} + (1 - f_1)}{f_2\, L_{S/B} + (1 - f_2)}$$

Distinguishing mixed samples is equivalent to signal/background classification!





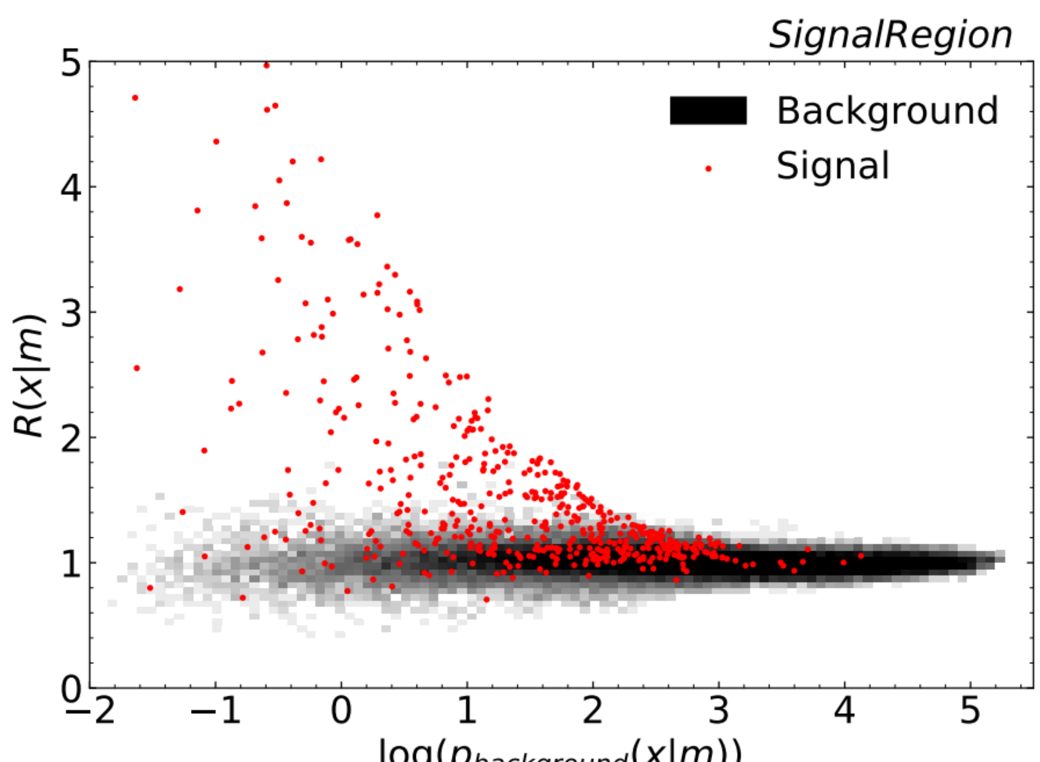
- Assume signal is resonant in one variable
- Define signal region and sidebands
- Train classifier and look for excess

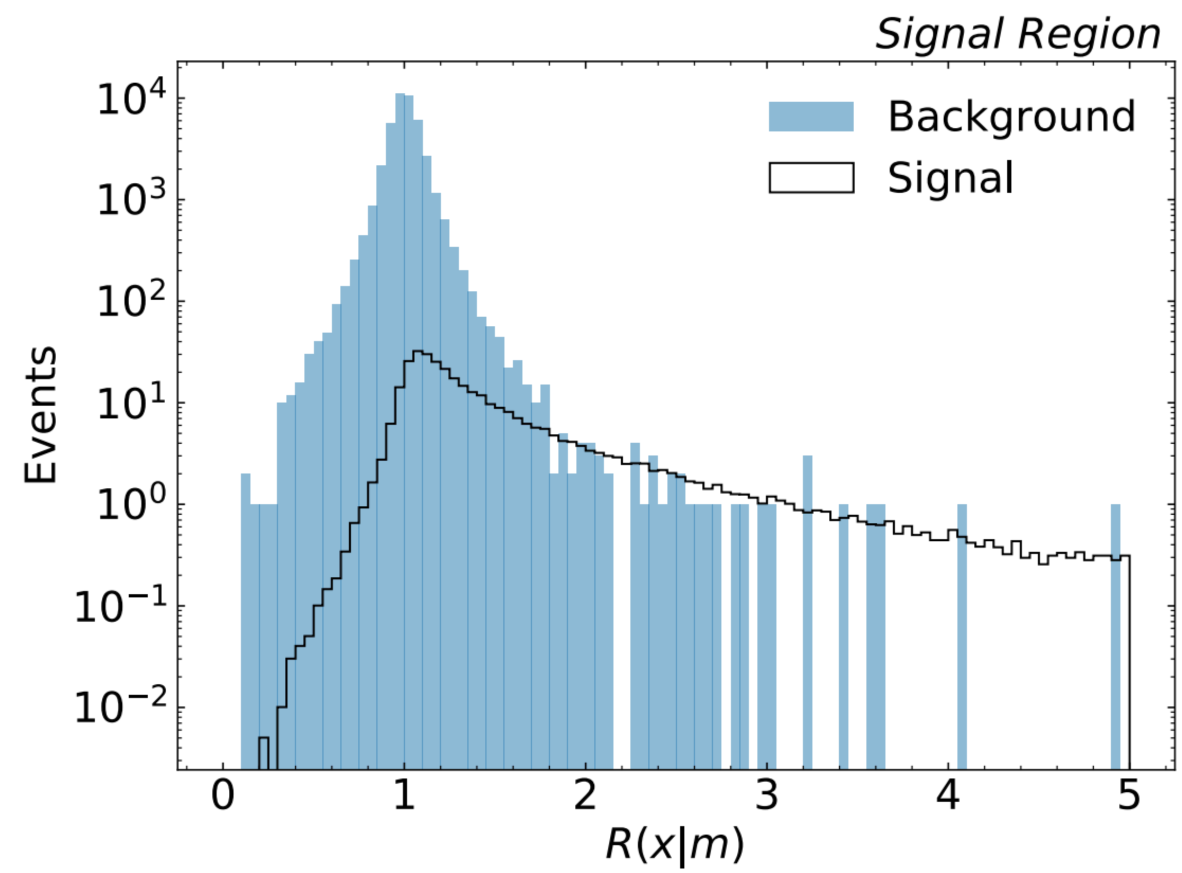
Classification without labels: Learning from mixed samples in high energy physics, EM Metodiev, B Nachman, J Thaler, 1708.02949
Anomaly Detection for Resonant New Physics with Machine Learning
JH Collins, K Howe, B Nachman
1805.02664

ANODE: ANOmaly detection with Density Estimation

An anomaly is a local over density of events

- Build density estimator in sideband region P_{SB}
- Extrapolation to signal region gives background estimate P_{SB} -> P_{BG}
- Build density estimator in signal region P_{SR}
- Likelihood ratio $R = P_{SR}/P_{BG}$
- *Density estimation via MAF (1705.07057) (Masked Autoregressive Flow)*





Anomaly Detection with Density Estimation, B Nachman, D Shhih 2001.04990

LHC Olympics 2020

- For more on anomaly detection see material at the recent workshop: https://indico.desy.de/e/anomaly2020

Conclusions

- Deep Learning for particle physics is rapidly developing solutions to a wide range of problems
  - Object and Event classification
  - Anomaly detection
  - Robustness and uncertainties
  - Fast reconstruction and simulation
- Further reading
  - Basic concepts: http://www.deeplearningbook.org/
  - Overview of ML in HEP papers: https://iml-wg.github.io/HEPML-LivingReview/

Thank you!