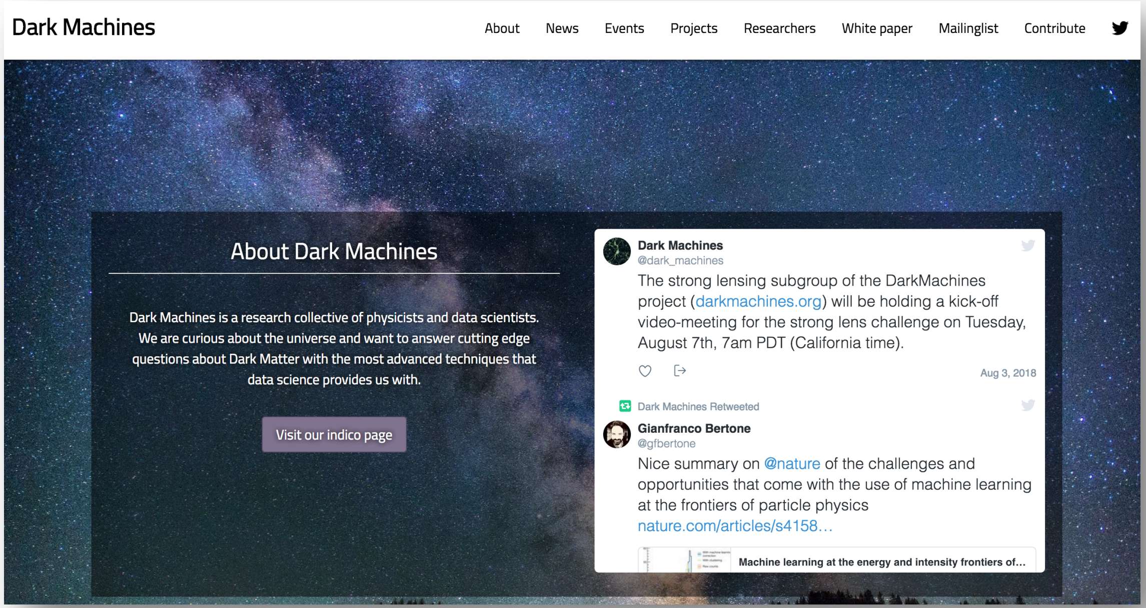
Darkmachines
(and other activities)

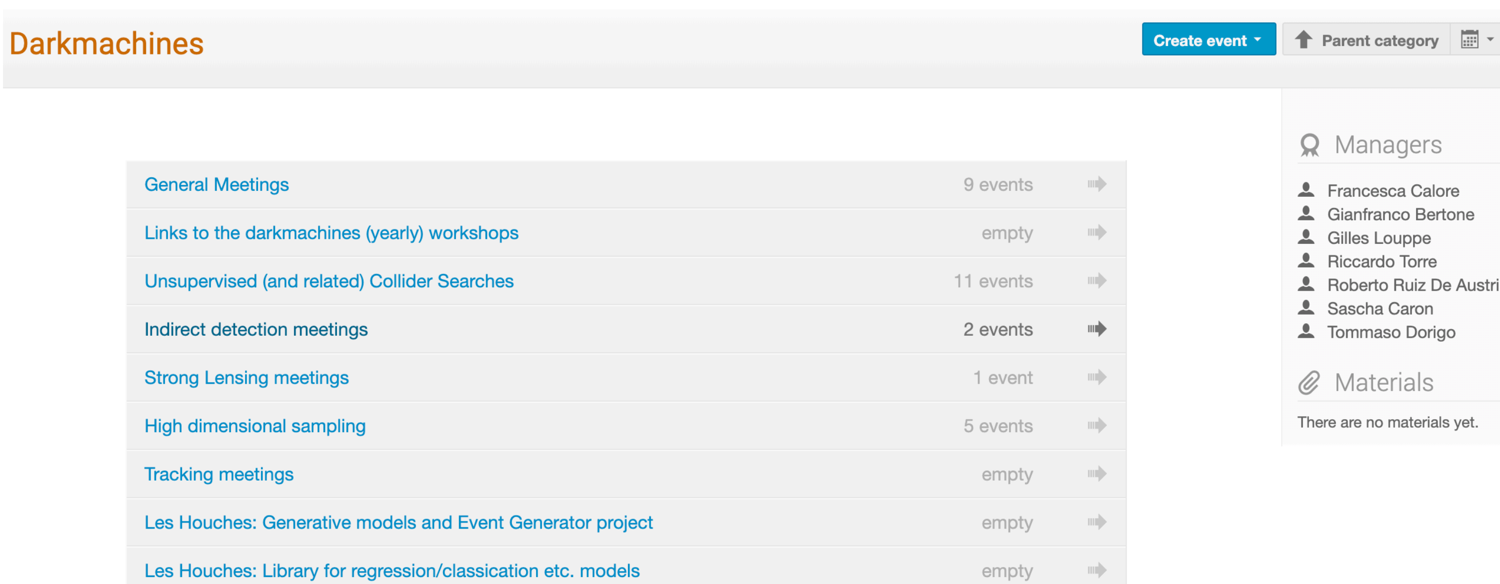
Building machines to understand the Dark Universe



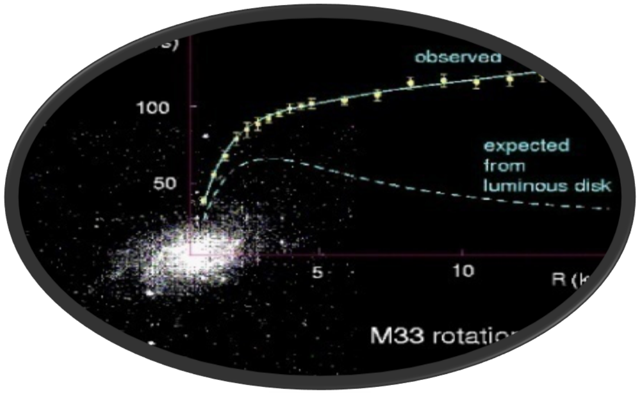
Sascha Caron
(Radboud University and Nikhef)

darkmachines

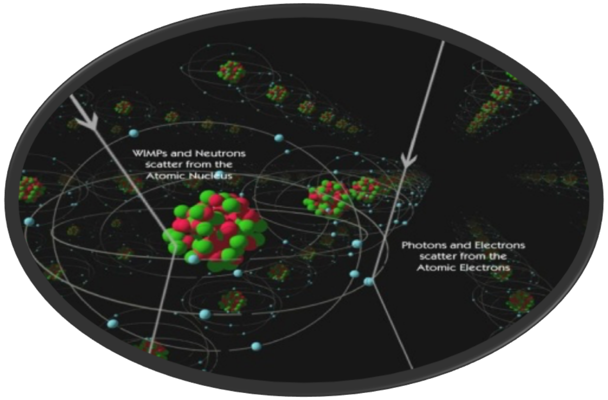
- Yearly meeting with about 80 scientists
(2018 Leiden, 2019 Trieste, 2020 CERN, 2021 ?, …)
- Video talks on ML
- Collaboration of ML experts, Astronomy and Physics
- Work by defining “challenges” → 10 challenges ongoing
- >250 scientists signed up at www.darkmachines.org



Dark Matter data gathering pillars

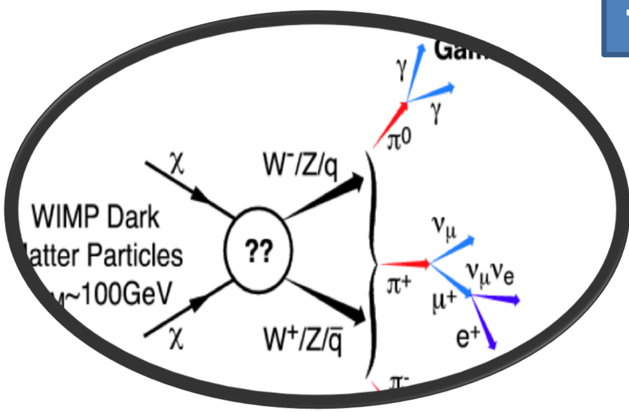


Gravitational interactions

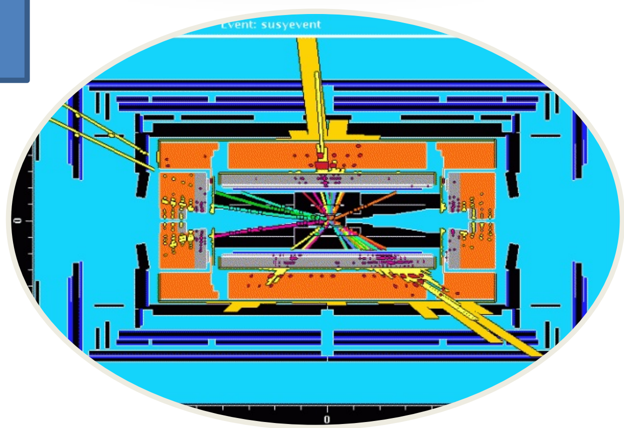


Direct Detection

Message of this slide:
All this is connected
→ So connect all this



Indirect Detection



Production

Machine Learning summary Les Houches 2019

Collaborate on 3 projects:

- Event generation with Generative models
- Database of Networks (regression, classification)
- Anomaly detection

Melissa v. Beekveld, Wolfgang Woltenberger, Richard Ruiz, Sydney Otten, Andrea Coccaro, Roberto Ruiz, Riccardo Torre, Sascha Caron, Sezen Sekmen, Sanmay Ganguly, Giovanni Zevi, Bob Stienen, Maurizio Pierini, Sabine Kraml, Jan Heisig, Luca Silvestrini, Seung Lee, …

The Deep Learning revolution !?

- 2014 First applications in HEP

→ Better classification of LHC events with DL on 4-vectors compared to traditional methods

- 2017/18 ? First application of deep generative models

Expect radical shift in next years from traditional analysis techniques to advanced ML / DL

→Radically new avenues for discovery

Why ML ?

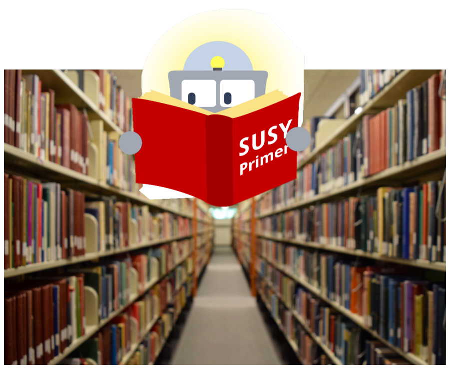
Traditional pipeline:

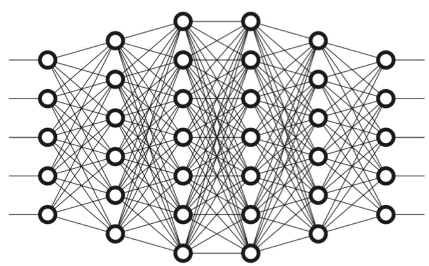
Model → Simulation → >days → Experiment

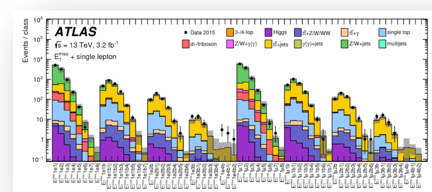
10

ML pipeline:

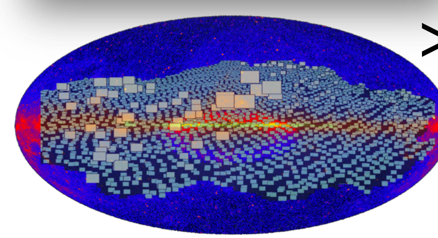
< 1 msec







>100000



Construct
fast and reversable pipeline
→ Optimization of pipeline

Higgs: You got a new toy, it's a playmobil castle with a size between 0.1-100 cm. Can you find it ?

Today: I have a new toy for you, I put it somewhere in your room. The size is 0.1-100 cm. Can you find it ?

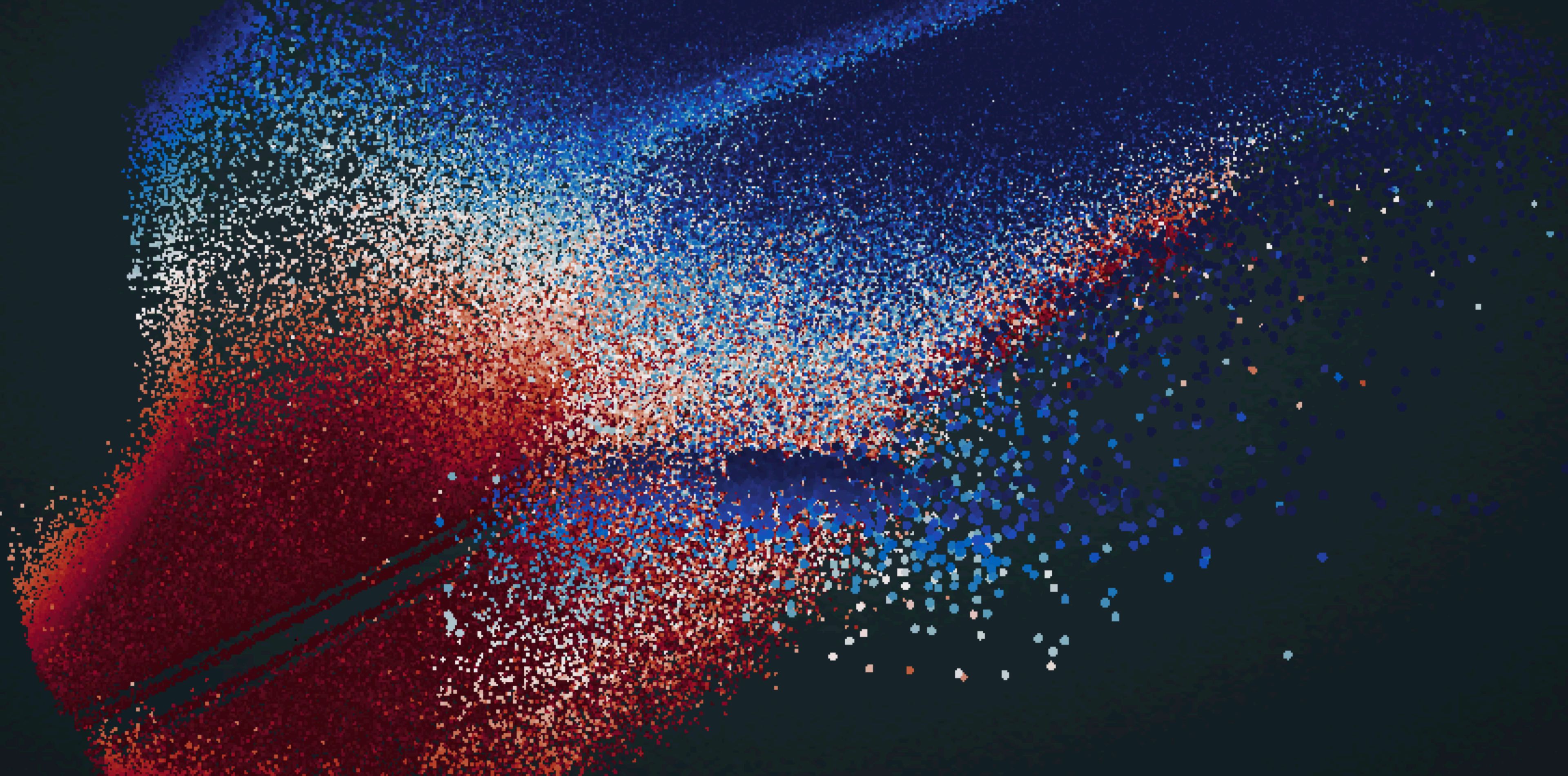
We work to implement more “automatization” to „scan“ the full room for something interesting…

→This can help LHC, but might also work for astrophysics

→We can embed into this “scan“ our prejudice how new physics looks like, e.g. in this case it would be „toy“ detection software trained on all known toys…

Models (EFT, SUSY) are in reality very very complicated
We humans simplify them

Can we broaden the search strategy ?
Can we also fine the model outside of the box ?



Bob Stienen

Accelerating BSM model predictions

Inputs → ***Long simulations + many programs*** → Output

Train classification / regression tool to improve ***this*** by ML

Advantages:

- **Speed !**
- **Generality !**

Coupling Theory and Machine Learning part 1

"Learning a function" from datasets with known labels sounds boring and old-fashioned.

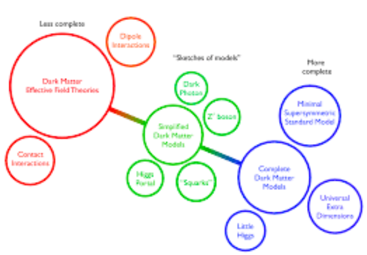
However we can couple it to simulators+ experiments + phenomenology ….

Coupling Theory and Machine Learning part 1

"Learning a function" from datasets with known labels sounds boring and old-fashioned.

However we can couple it to simulators+ experiments + phenomenology ….

Supersymmetry → Experimental Test → Machine Learning to generalize over Supersymmetry



Coupling Theory and Machine Learning part 1

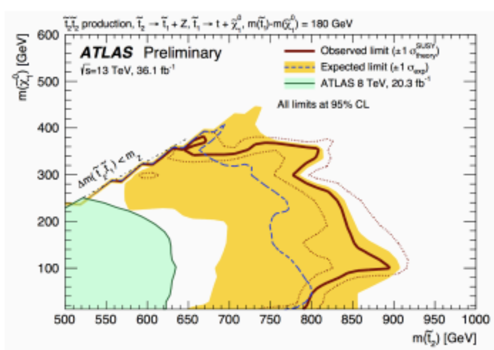
"Learning a function" from datasets with known labels sounds boring and old-fashioned.

However we can couple it to simulators+ experiments + phenomenology ….

Training data provided by ATLAS

SUSY-AI

- Exclusion determination in 19d pMSSM
- 310,324 model points with known exclusion as data input
- Algorithm: a collection of decision trees (Random Forest)
- **Idea: going from 2d slices to N-dim representations**



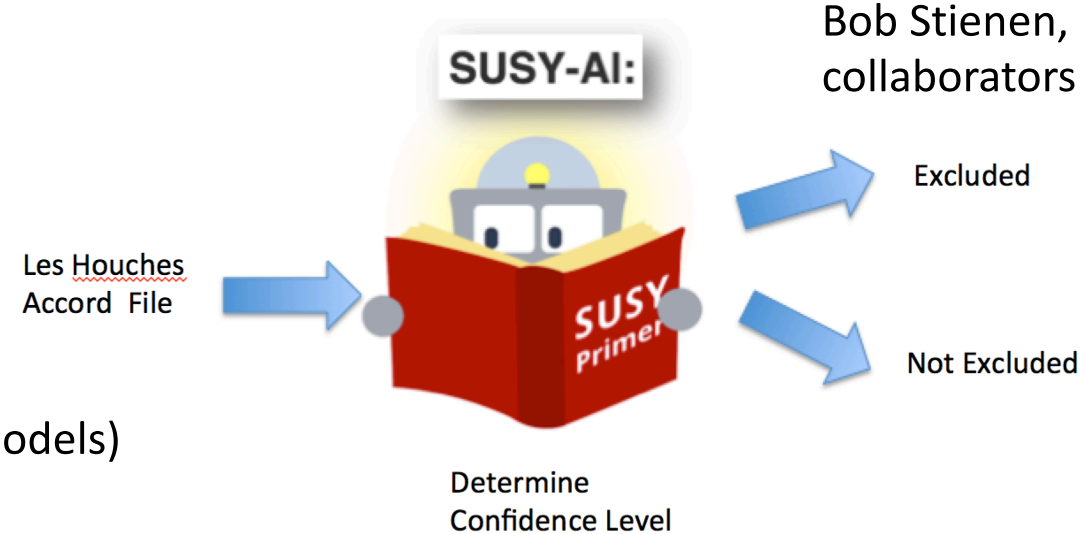
Prevent overfitting: Boosting: many trees + not subset of all features for each tree
*Bagging: random picking training data -> each tree of the forest sees only 0.68*data (see extra slides)*

SUSY-AI

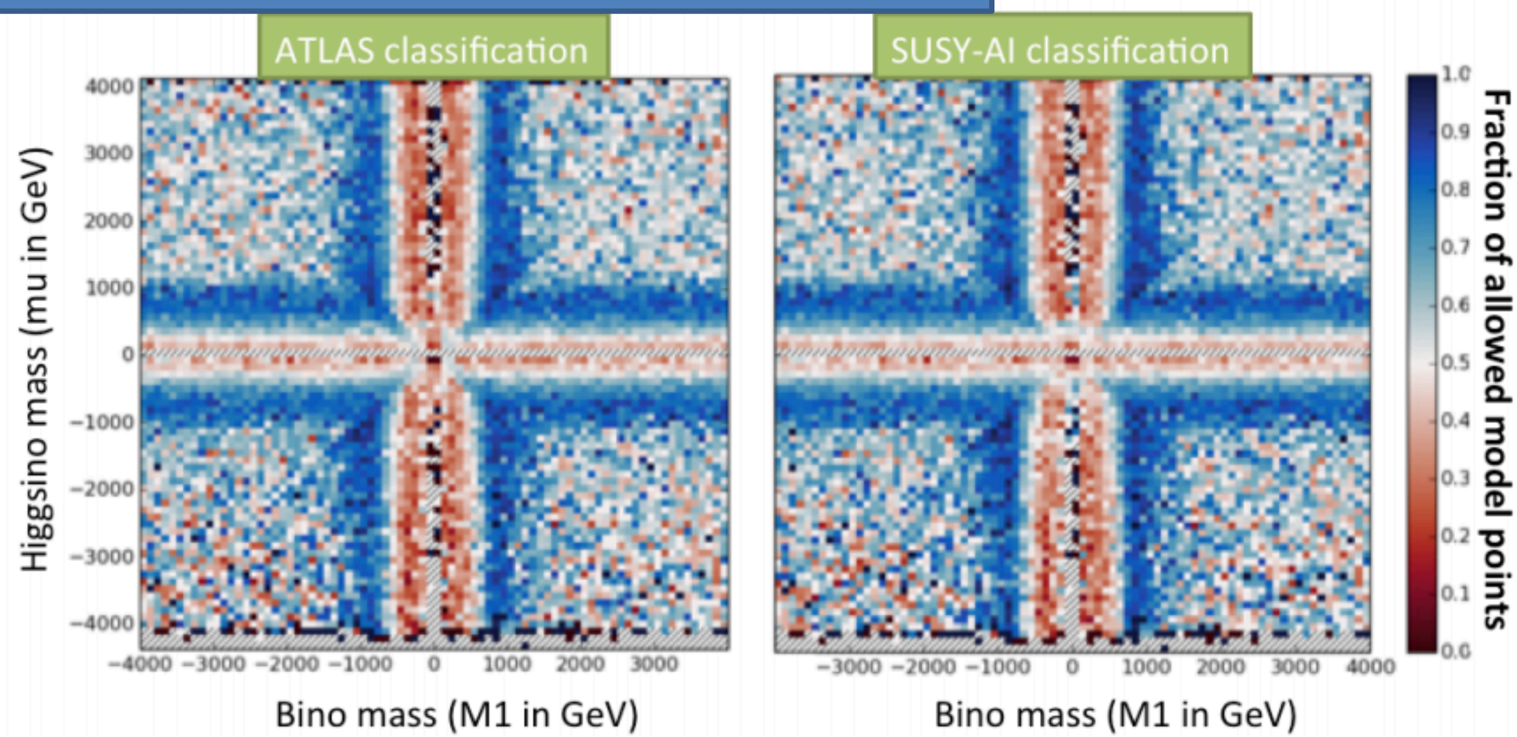
Encoding of model constraints with Machine Learning

Aim: Generic framework (**all** models)

Bob Stienen, >10 collaborators



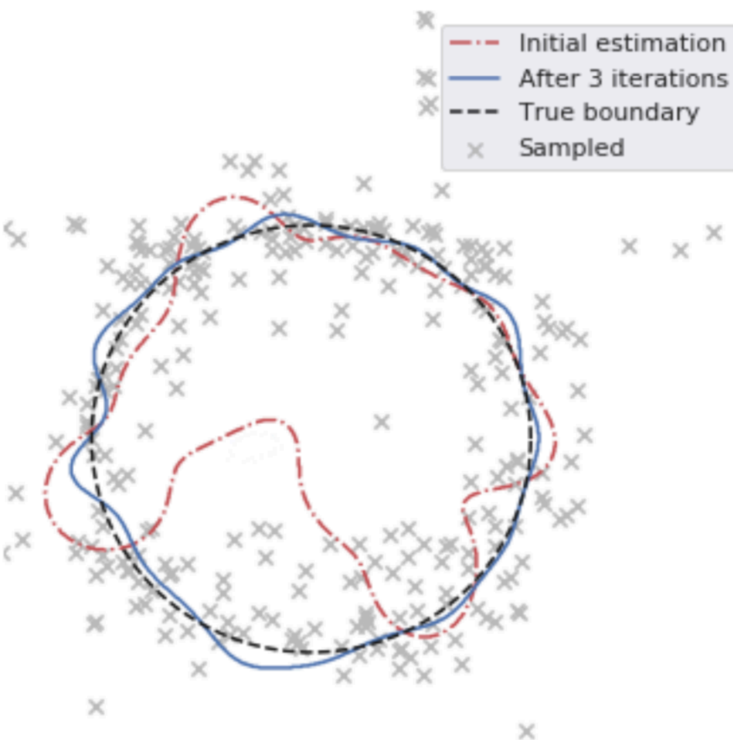
Testing with out-of-bag estimation (remember 0.68!)

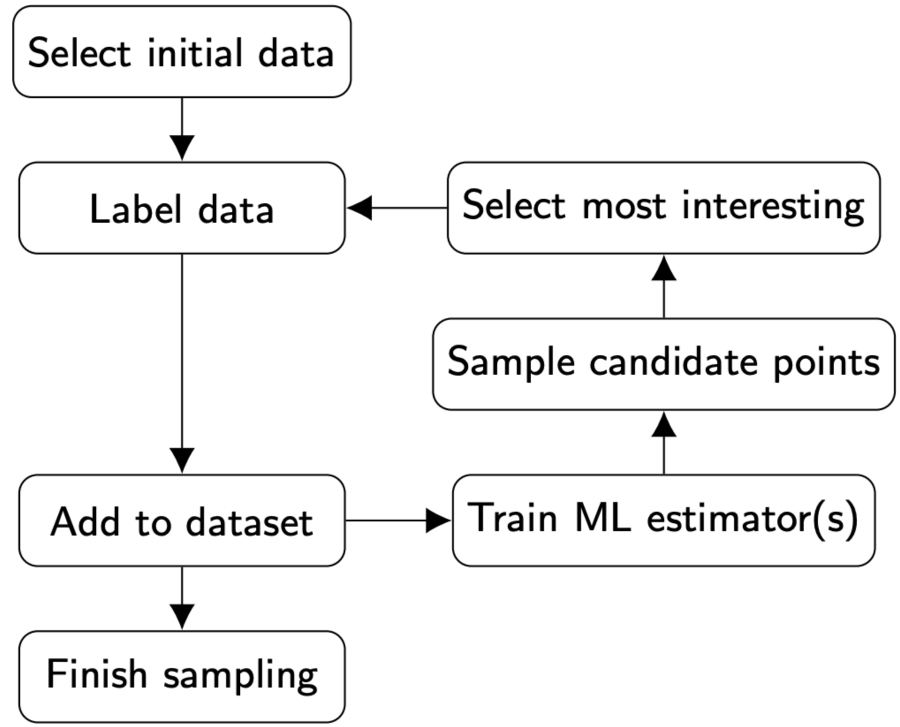


How to sample points ?

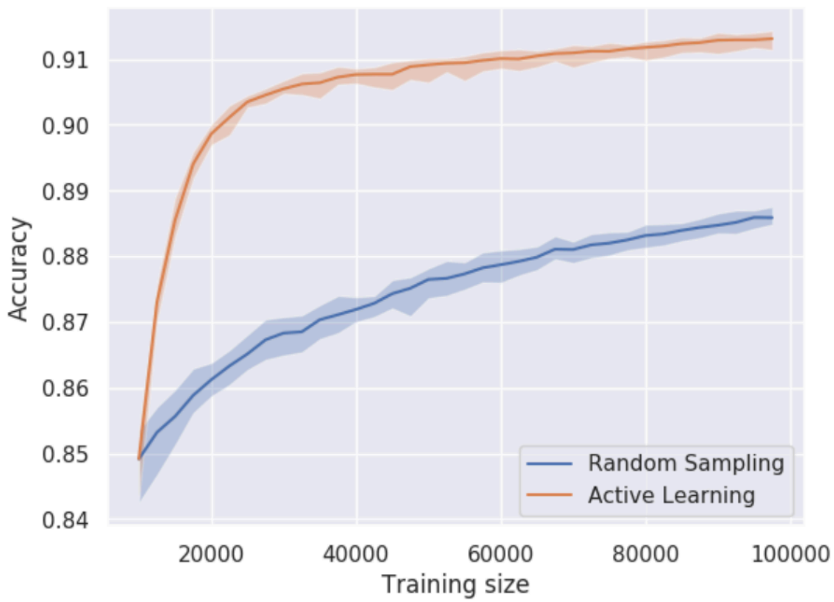
We try "active learning":

Arxiv 1905.08628





Active learning by "committee"
Random forest and "MC dropout" networks



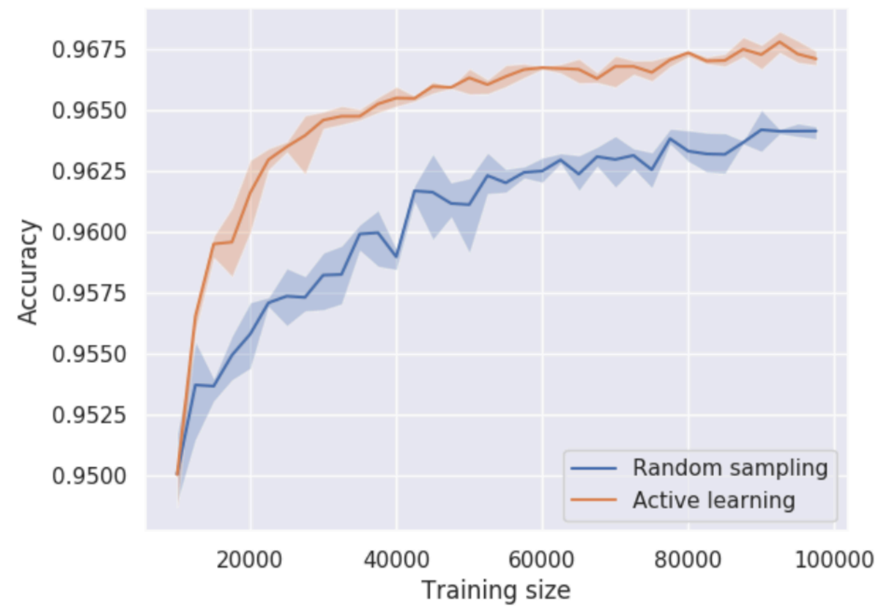


FIG. 4. Accuracy development on model exclusion of the 19-dimensional model for new physics (pMSSM) for random sampling and active learning using a random forest as algorithm and an infinite pool. True labeling was provided by a machine learning algorithm trained on model points and labels provided by ATLAS [1]. Here active learning is vastly superior over random sampling, yielding a gain in computational time of a factor of 5 to 6. The bands around the curves show the range in which all curves of that colour lie when the experiment was repeated 7 times.

FIG. 5. Accuracy development on model exclusion of the 19-dimensional model for new physics (pMSSM) for random sampling and active learning using a dropout neural network with infinite pool. True labeling was provided by a machine learning algorithm trained on model points and labels provided by ATLAS [1]. The gain of active learning with respect to random sampling (as described by Equation 2) is 3 to 4. The bands show the range in which all curves of that colour lay when the experiment was repeated 7 times.

”Sampling” depends on the task
→ Wanna learn a likelihood ?

DarkMachines

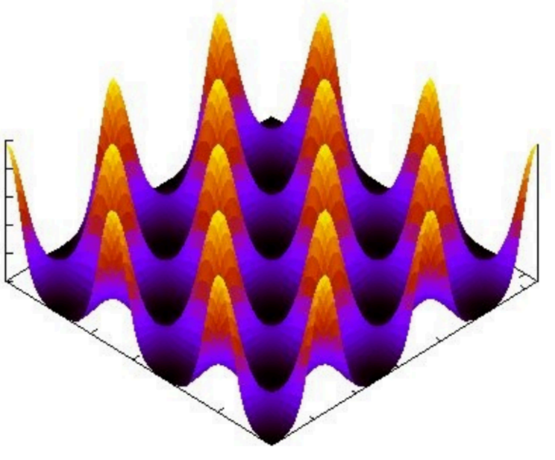
High dimensional sampling project

Martin White, Joaquin Vanschoren

Various approaches
→ compare + develop

Basic idea

- Repeat a ScannerBit style study with a wider variety of techniques, and a series of toy functions + physics cases
- Have initially settled on the MultiNest "eggbox" likelihood for testing



Regression: Predicting real numbers → Cross sections

Andy Buckley:

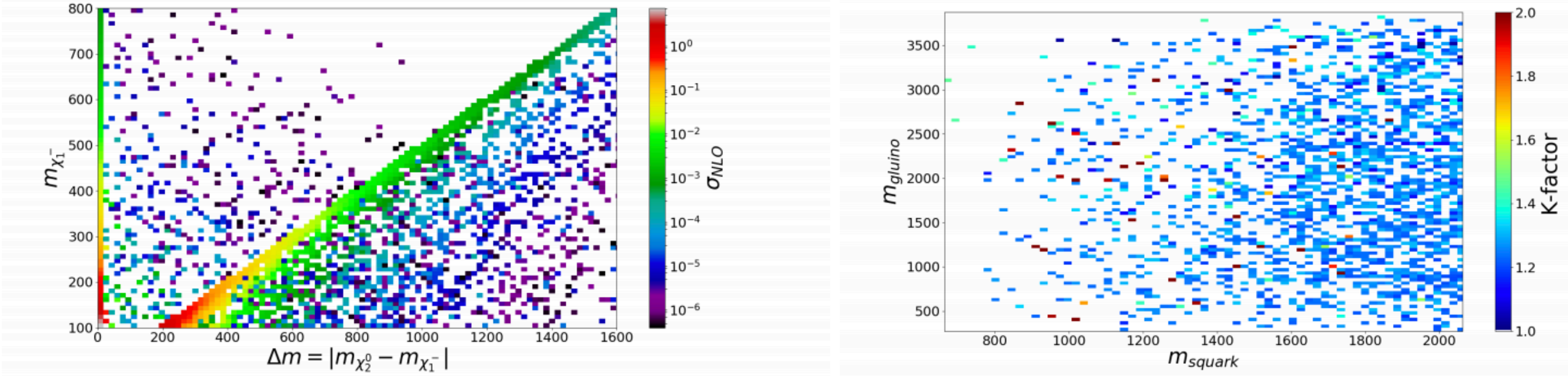
use of fast BSM cross-section prediction, cf. FastLim, EWKFast, and there is prelim. work for Gambit

DeepXs: SUSY EW Cross sections

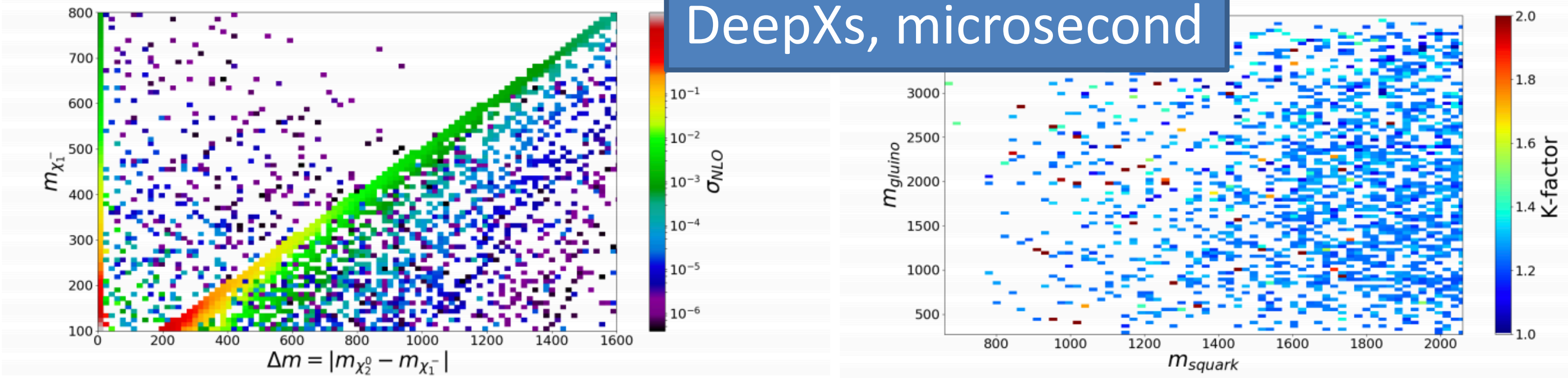
https://arxiv.org/abs/1810.08312

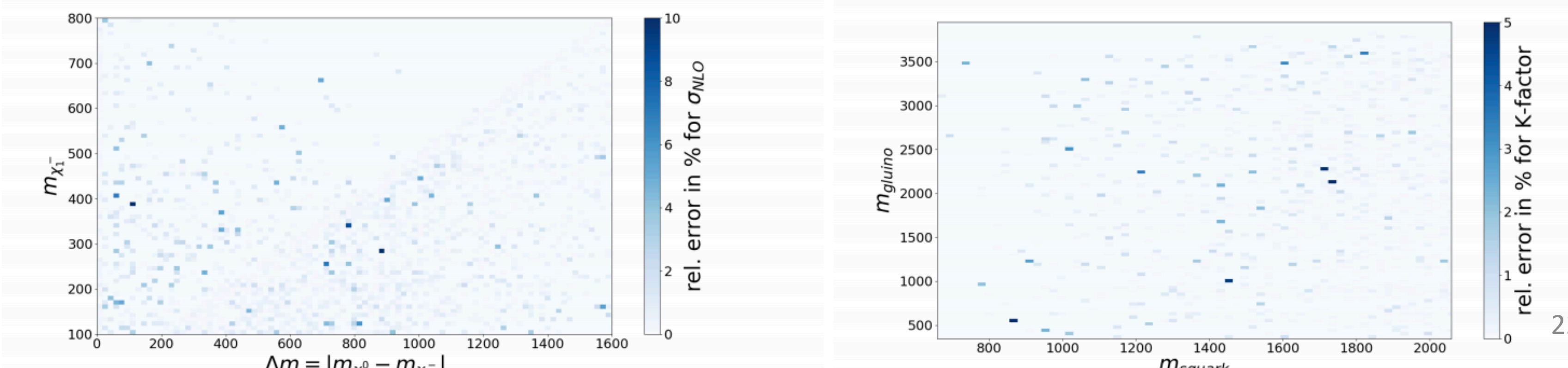
- Running NLO code to derive SUSY cross sections can take up to 10 minutes
- Can we "learn the cross sections" and derive in a microsecond for *any* model parameter set? ➔

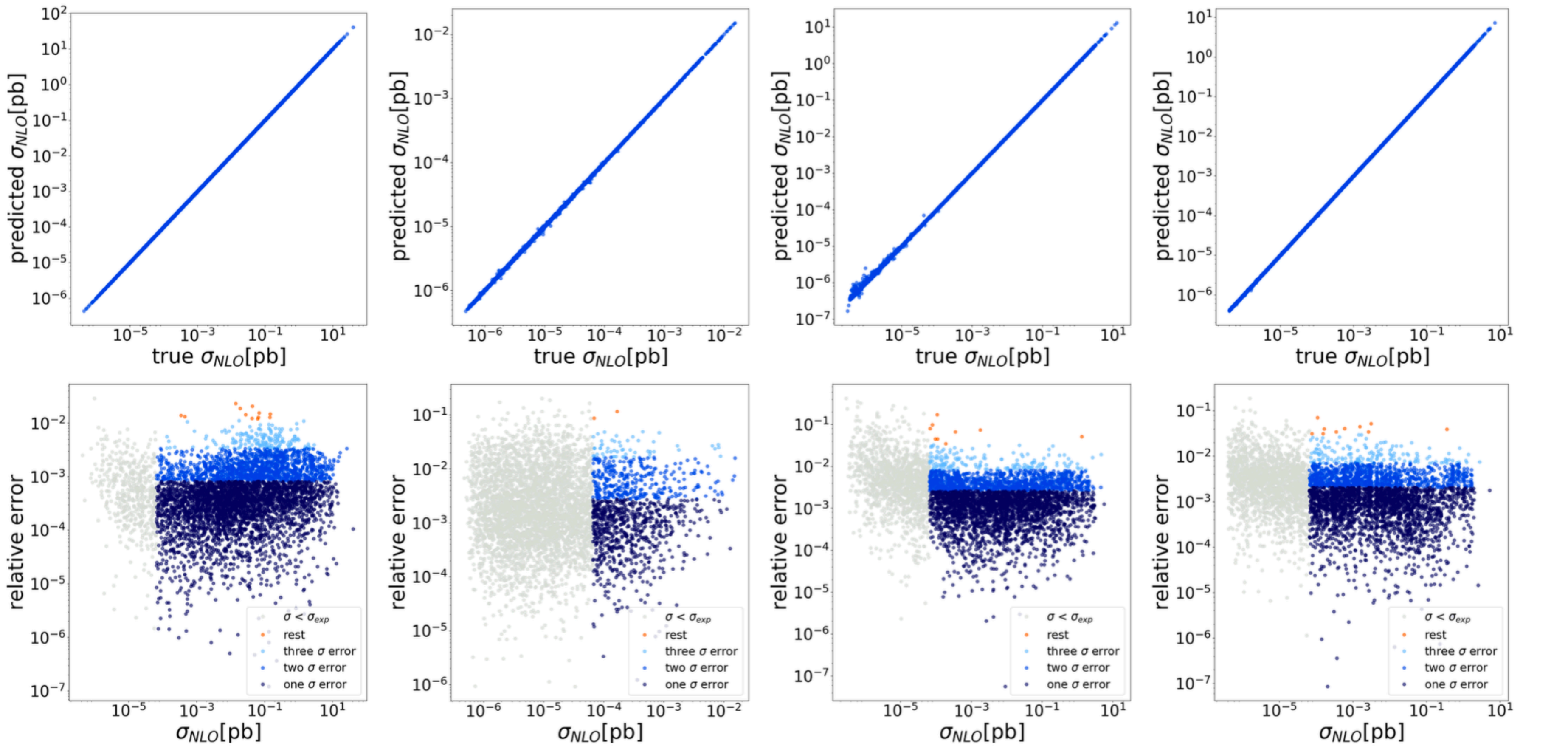
Prospino, minutes



DeepXs, microsecond





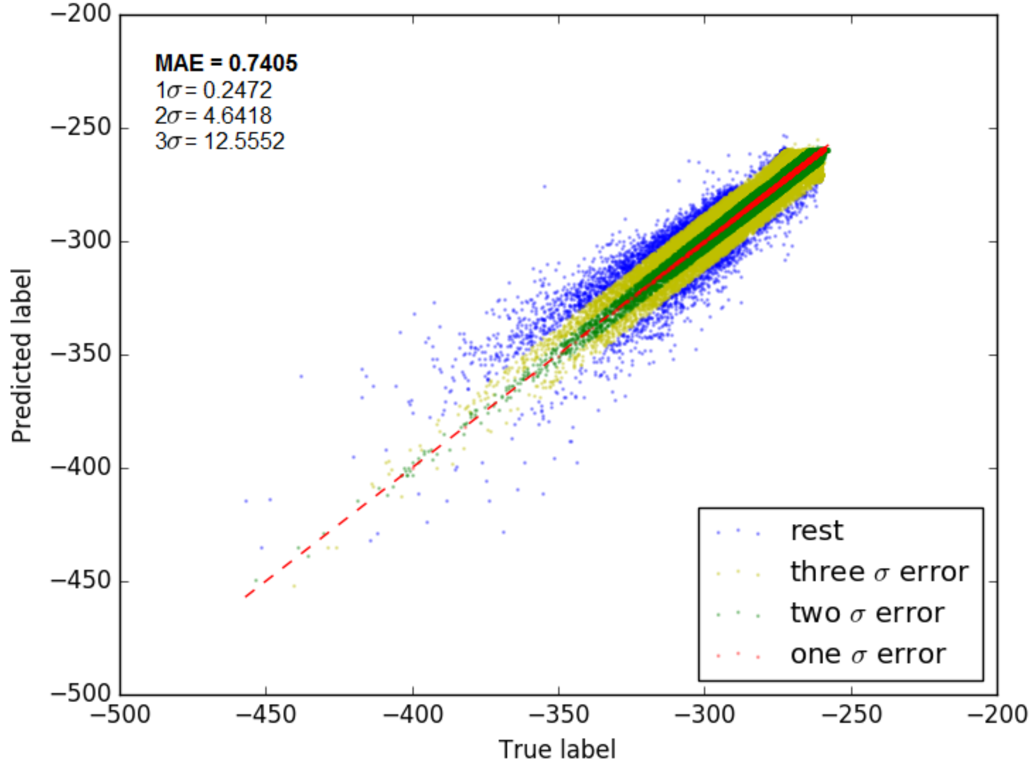


a) $\tilde{\chi}_1^+\tilde{\chi}_1^-$ **b)** $\tilde{\chi}_2^0\tilde{\chi}_2^0$ **c)** $\tilde{\chi}_2^0\tilde{\chi}_1^+$ **d)** $\tilde{\chi}_2^0\tilde{\chi}_1^-$

FIG. 1: The true vs. predicted NLO cross-sections (top) and the relative error vs. true NLO cross-section with confidence intervals (bottom) for the same 10^4 samples in both plots

inference at NLO with inference times that improve the Monte Carlo integration procedures that have been available so far by a factor of ≈ 6.9 million from ≈ 3 minutes to $\approx 26\mu s$ per evaluation.

BSM-AI regression example… Learning GAMBIT likelihoods



MSSM - 7

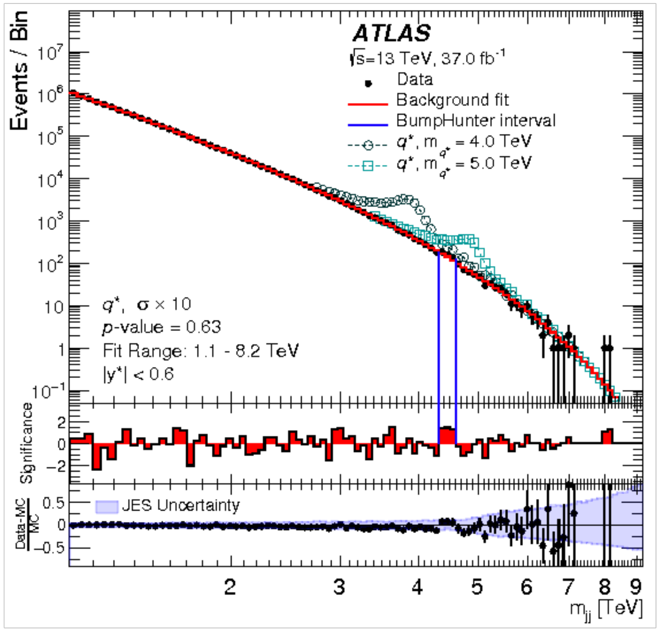
https://arxiv.org/abs/1705.07917

Plot by Sydney Otten

Regression: Background modelling

- Currently: Guess background distribution
- Try: Use NN to fit all kinds of mass distributions

→NN may get general feeling how mass distributions can be predicted given input final state and observable



PhenoAI ?

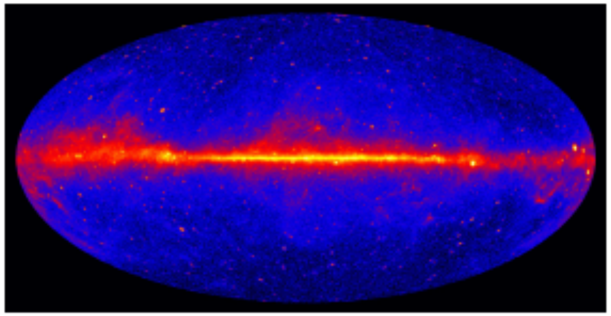
- Encoding regression/classification for everybody

”Les Houches” project + darkmachines project
→ You can join / help !

Astroparticle DM searches with Machines

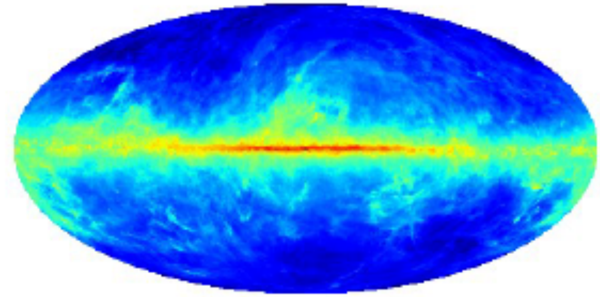
DM searches in the inner Galactic region with Fermi LAT

Fermi LAT; > 1 GeV

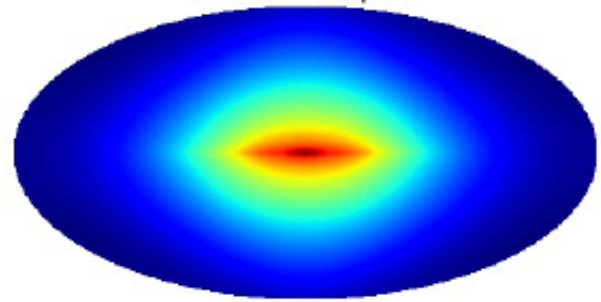


Subtract
1) Known point sources
2) Diffuse foregrounds

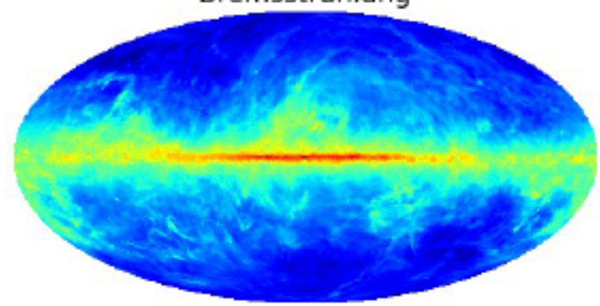
Pion decay



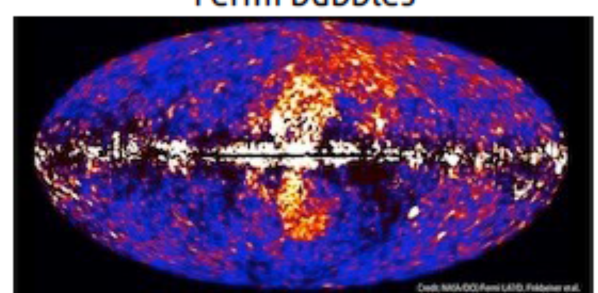
Inverse Compton



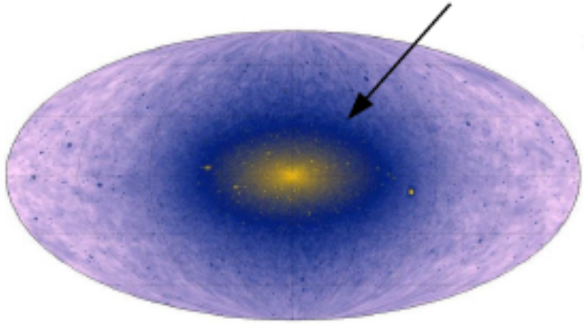
Bremsstrahlung



Fermi bubbles

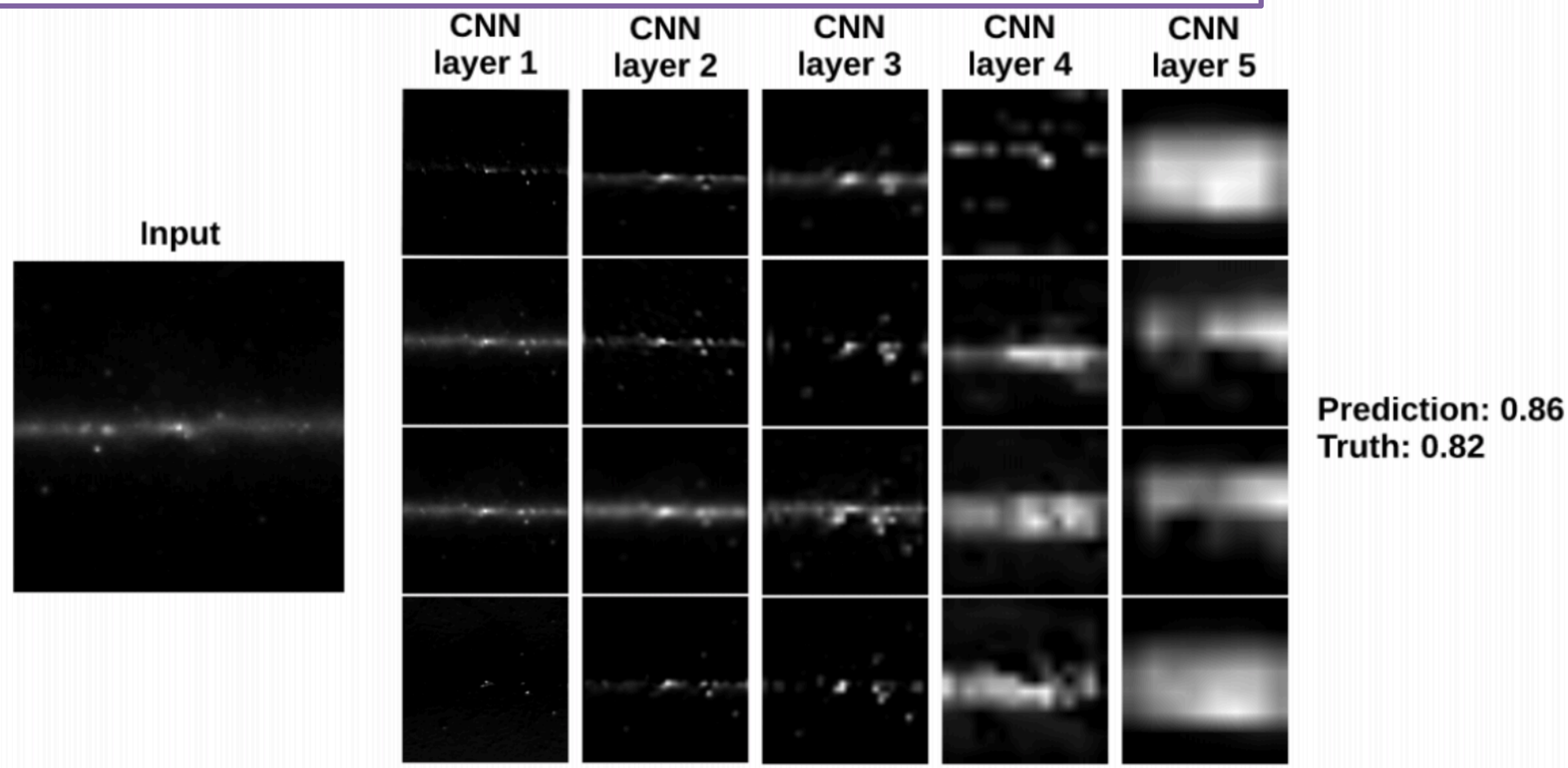


Do residuals look like this?



Isotropic or point sources: A Deep Convolutional Network approach

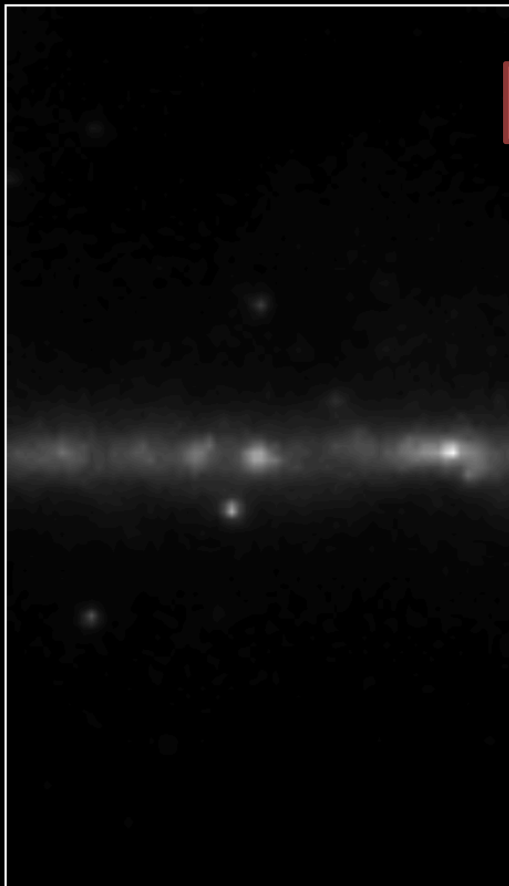
Output of the 5 convolutional layers can be "visualized" per event.



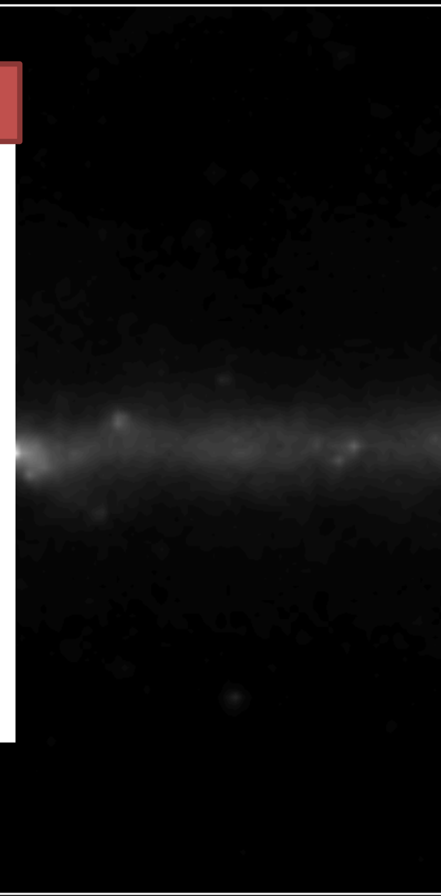
Activations of the network. Only four filters per layers are shown for clarity, between 256 and 65 filters are used for the different layers

arXiv:1708.06706

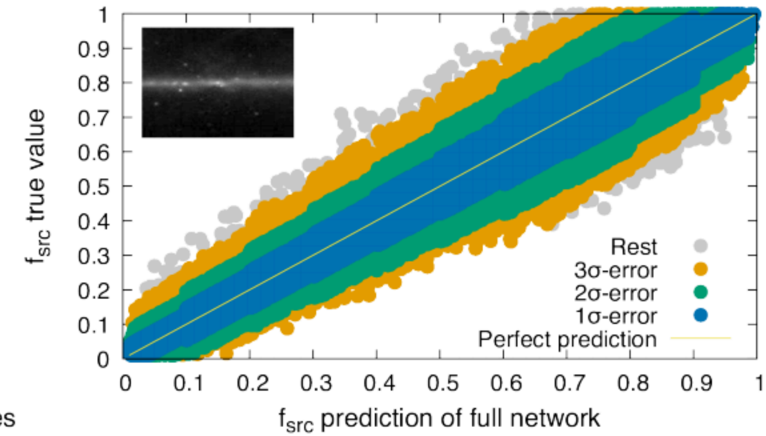
What is this fraction?



This is 0.5



Network can generalize over randomness



(b) Prediction of the full network versus true values.

Your prediction:
Invert image:

Truth: 0.052
Network: 0.1230
Your guess: 0.5
Who is better? The network

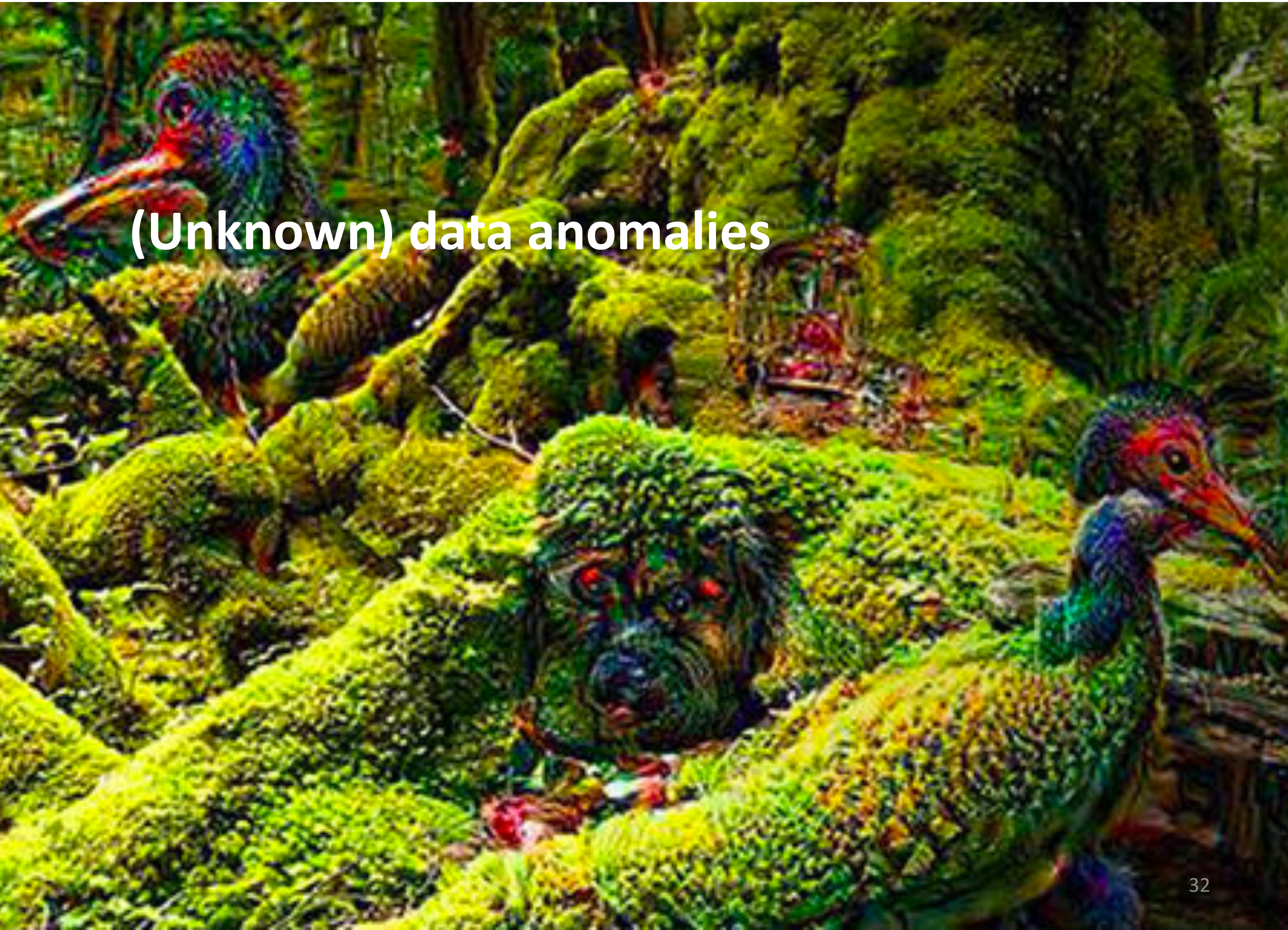
Interpretation here is frequentists and relies on the model to be correct (uncertainties from toy experiments, no p-value yet)

Next steps

- Categorize objects on the gamma-ray sky

Also point source detection now → see e.g. recent paper called "deepsource" https://arxiv.org/abs/1807.02701

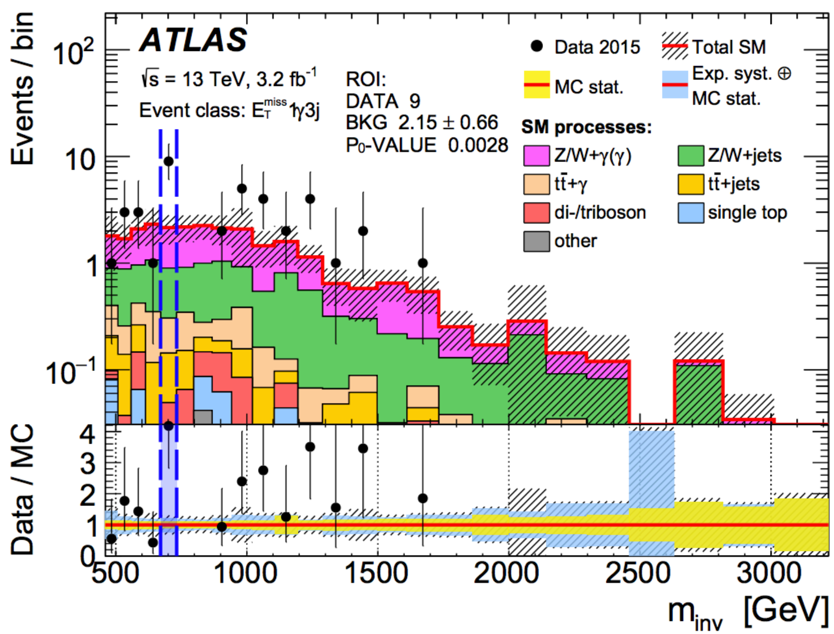
(Unknown) data anomalies



Our recent ATLAS approach

- Look everywhere for new overdensities
- Compare data to the SM using a test statistics and a scan algorithm

➔ e.g. General Search (on arxiv now: https://arxiv.org/abs/1807.07447)



Automatize:
>1600 distributions
>800 channels
>10^5 regions

Which quantity is optimal ?
How to determine background ?
How many hypothesis tests are optimal?

New ideas for searches with unknown signal -> Selection of recent developments in 2017/2018 !

- Fit a ML based background model to be less sensitive on MC prediction (gaussian processes in **arXiv:1709.05681** **)**
- Autoencoders as “filters” for SM events 1808.08992
- Unsupervised techniques (clustering as hypothesis test…)

K- Nearest Neighbour to estimate the point density of two samples, KL-test statistics to compare the samples

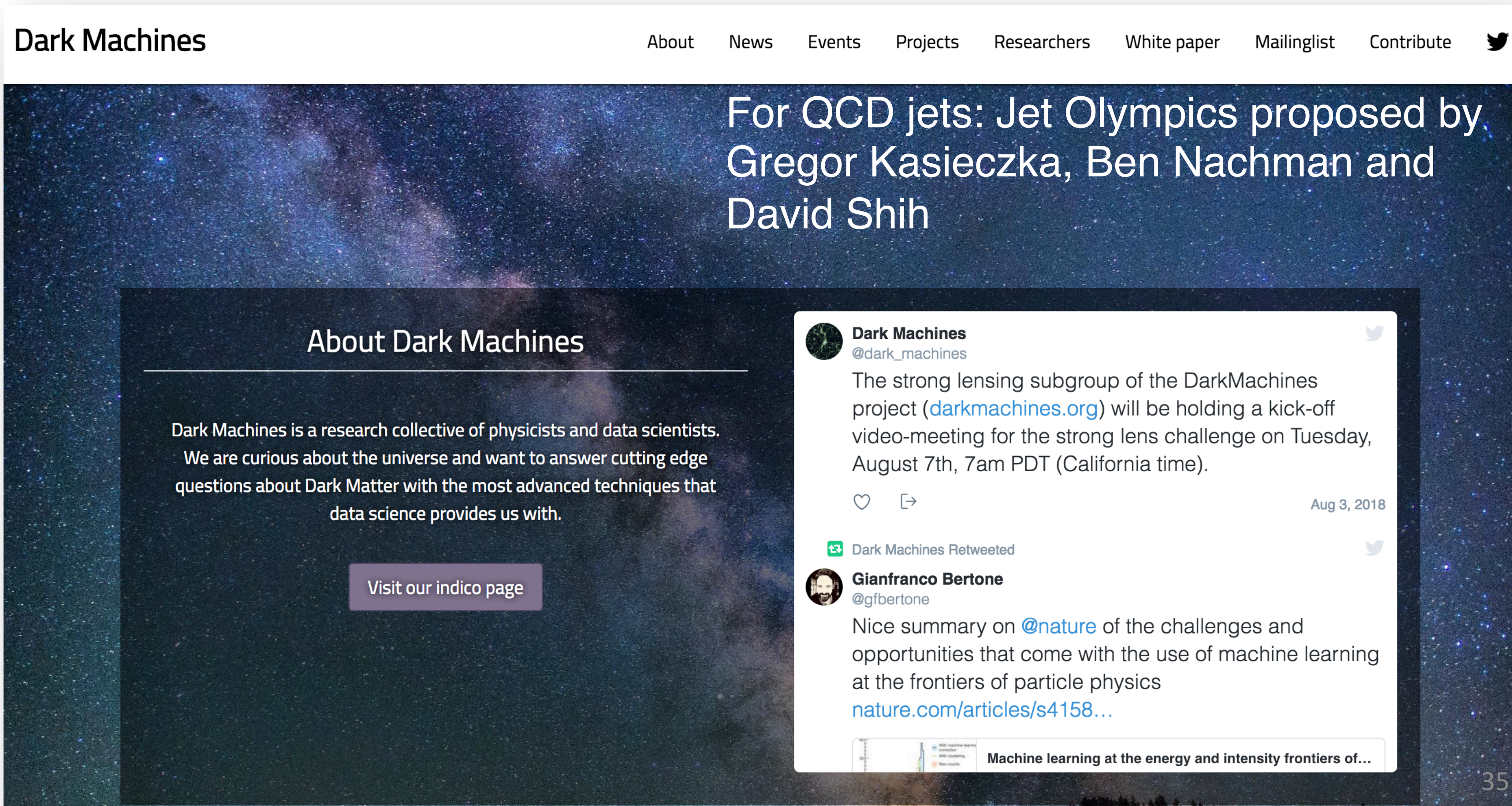
- Classification without Labels (CWOLA) **arXiv:1805.02664:**

Here the idea is to train a NN to seperate signal region + sideband region (as two samples) --> this can be possible due to a signal in the signal region …

- “Novelty detection algorithm“ **arXiv:1807.10261 ,**
- unsupervised KL divergence **arXiv:1807.06038**
- Self-organzing maps…
- **Optimal „distance measures“ between events** https://arxiv.org/abs/1902.02346
- … various more !!! (can‘t catch up anymore, can you ?)
- **Which one is good ? Which one to use ? Need comparison !!!**

Next steps: Compare / Optimize different approaches

e.g. in „unsupervised searches“ group of darkmachines
(Amir Farbin, Erzebet Merenyi, Andrea di Simone, Maurizio Pierini)
e.g. in ATLAS with General Search as prototype data ?



Generating known (and unknown) physics

Question: Can we make physical (collider, astroparticle, etc) events ?

Crazy Idea:
„Maybe the optimal generator is the best tool to search for new physics“

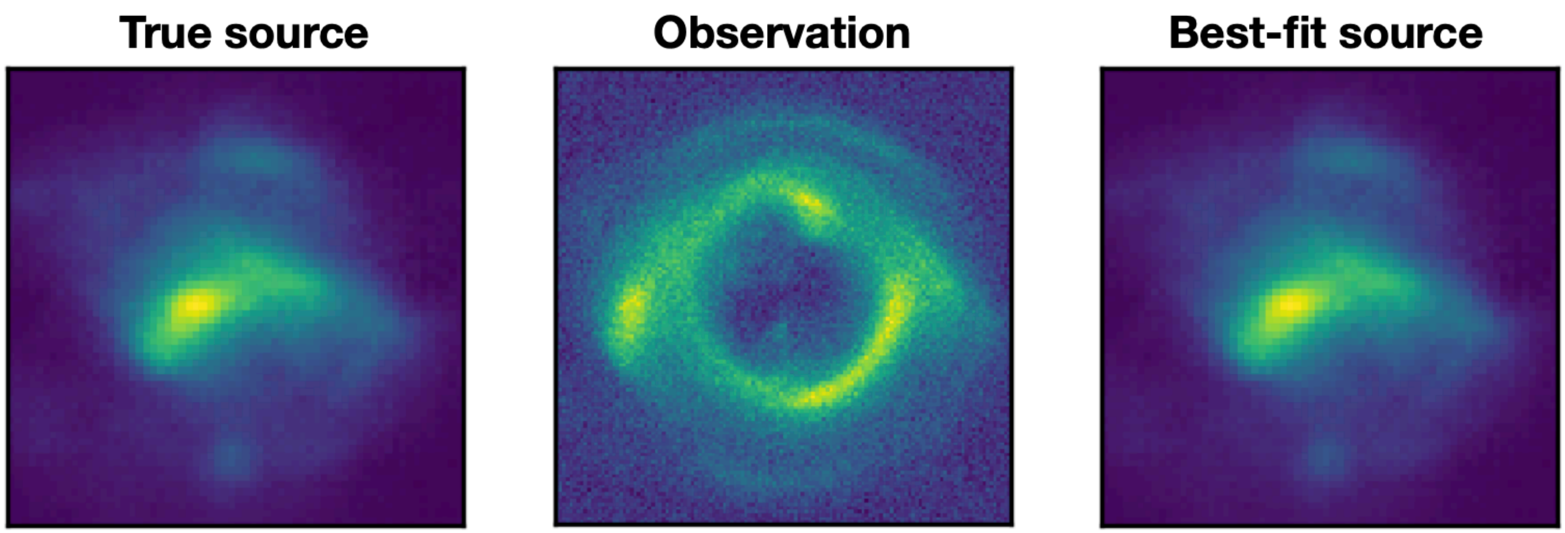
Strong Gravitational Lensing and ML: generative models for galaxies

Adam Coogan

Dark Machines workshop
ICTP, 8-12 April 2019

VAEs + Inverse autoregressive flows to sample latent space variables → see slides on darkmachines Trieste workshop

Lensing galaxies



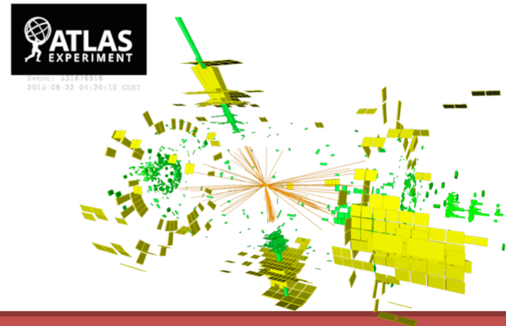
True Einstein radius: 2.3
Best-fit value: 2.29

***Very preliminary, simplified analysis**

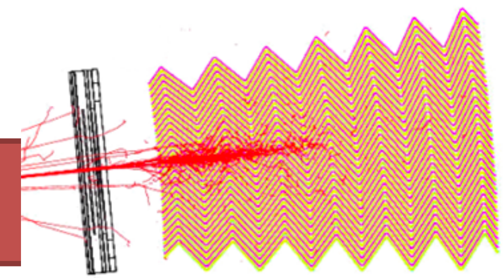
- Let’s have a look at a HEP example

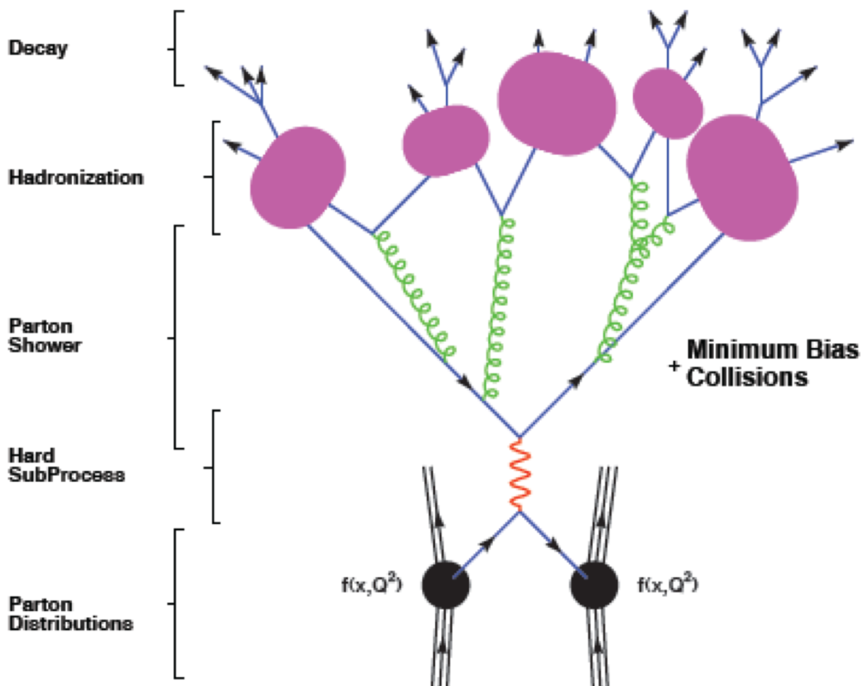
Simulation: Traditional

Energy and angles of reconstructed particles



Detector Simulator

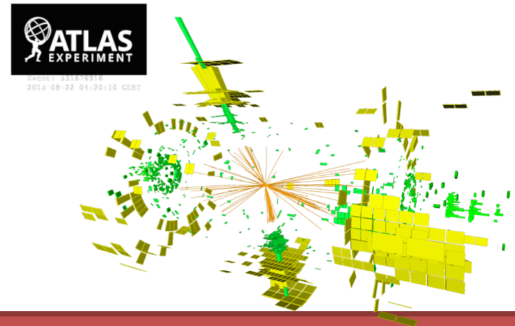




Input:
Random numbers

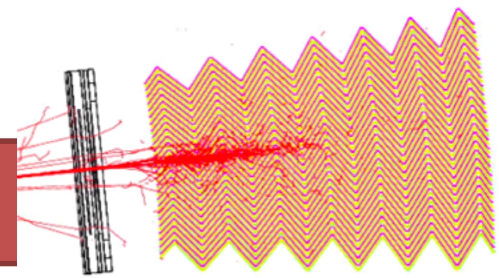
Simulation: Traditional

Energy and angles of reconstructed particles

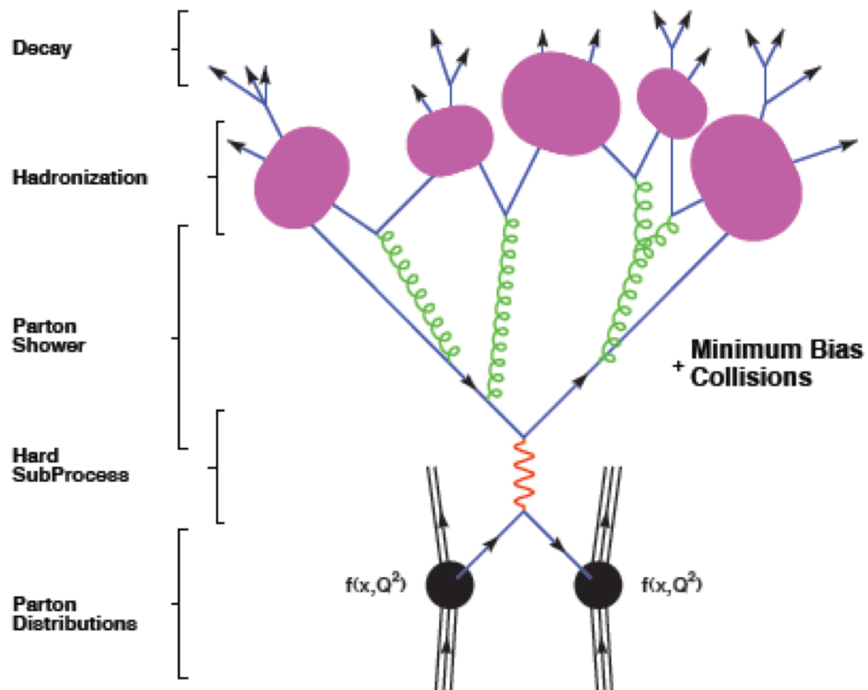


Delphes-AI ?

Detector Simulator



CaloGAN project

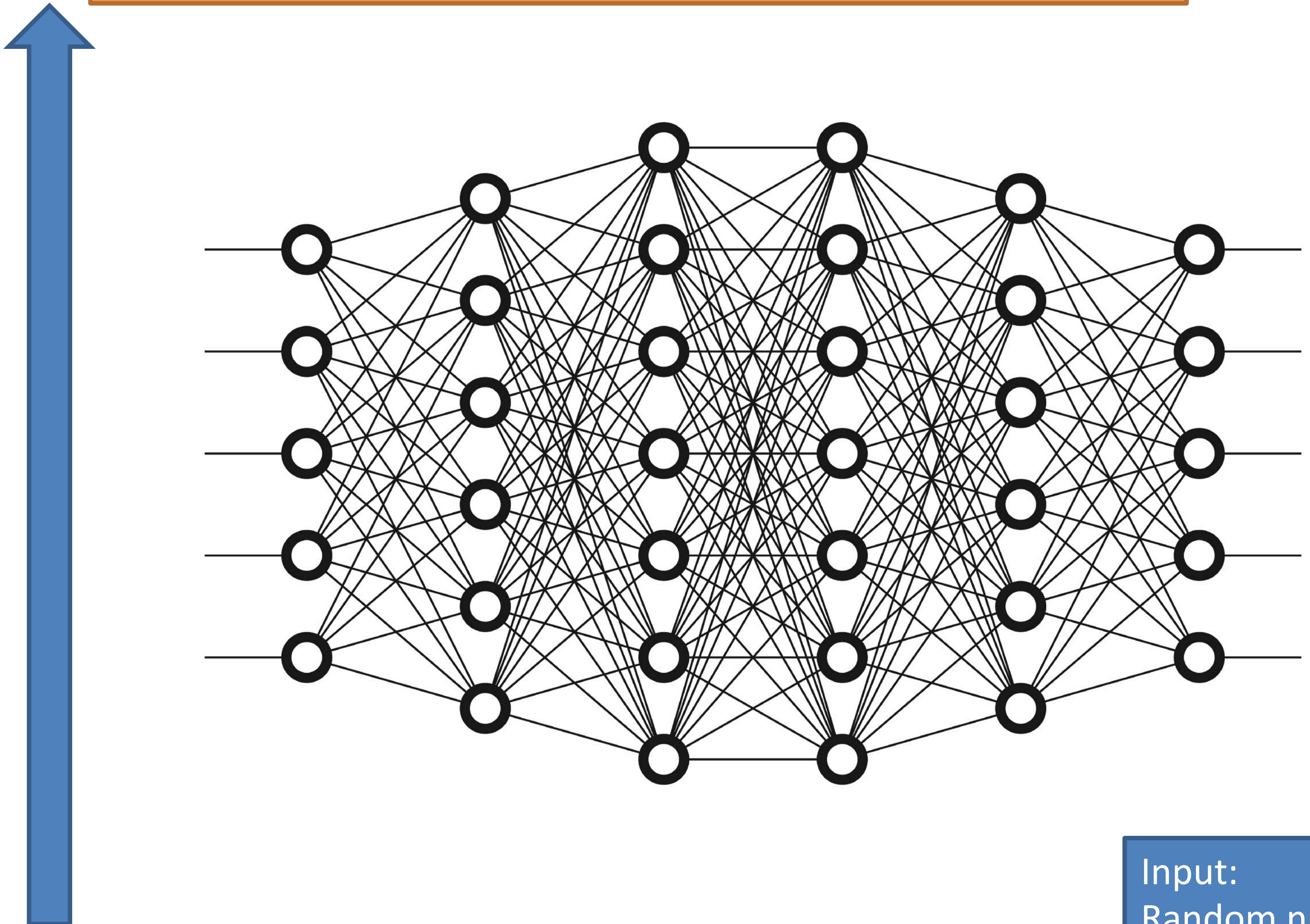


NN Cross sections

Input:
Random numbers

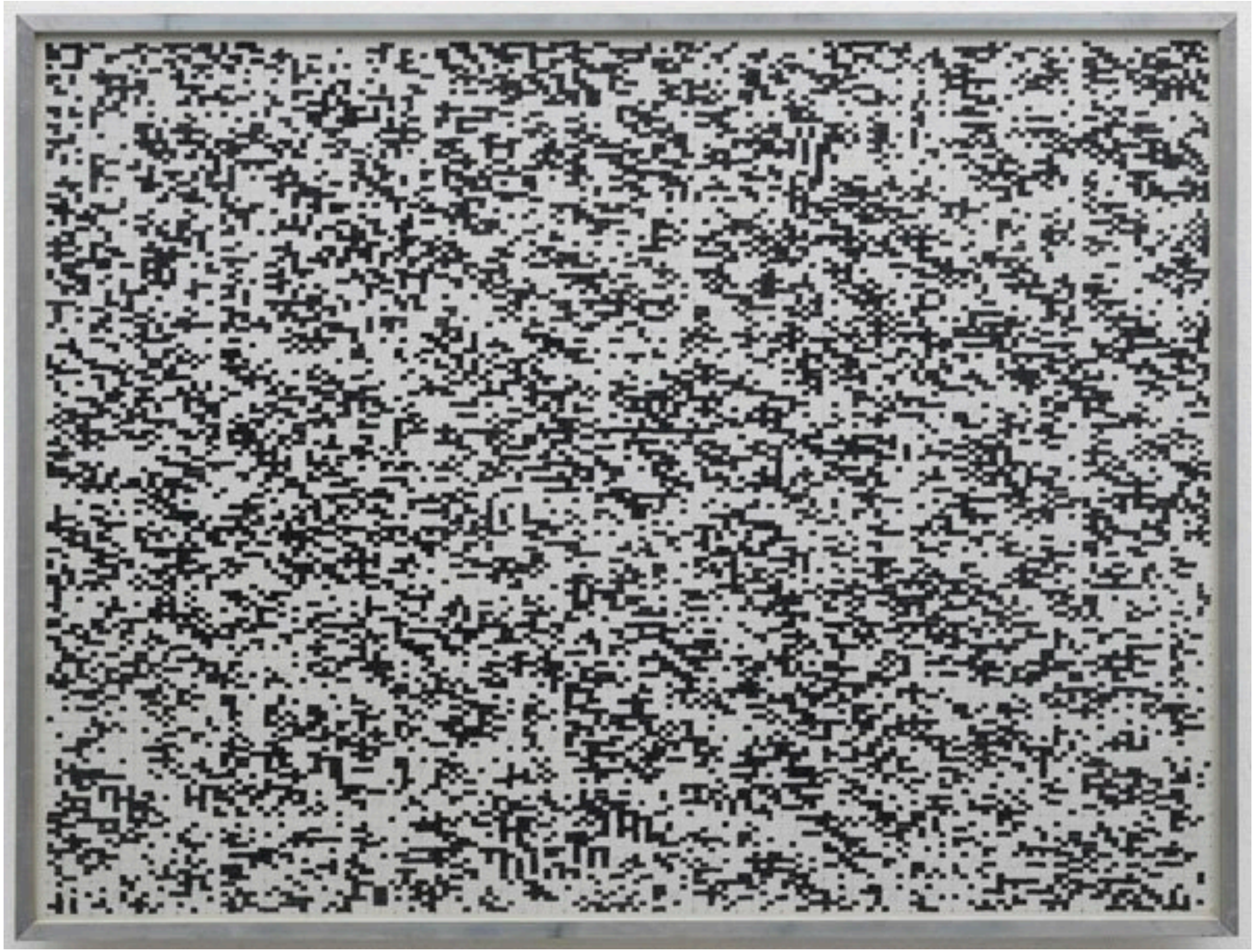
Simulation: Radical

Energy and angles of reconstructed particles



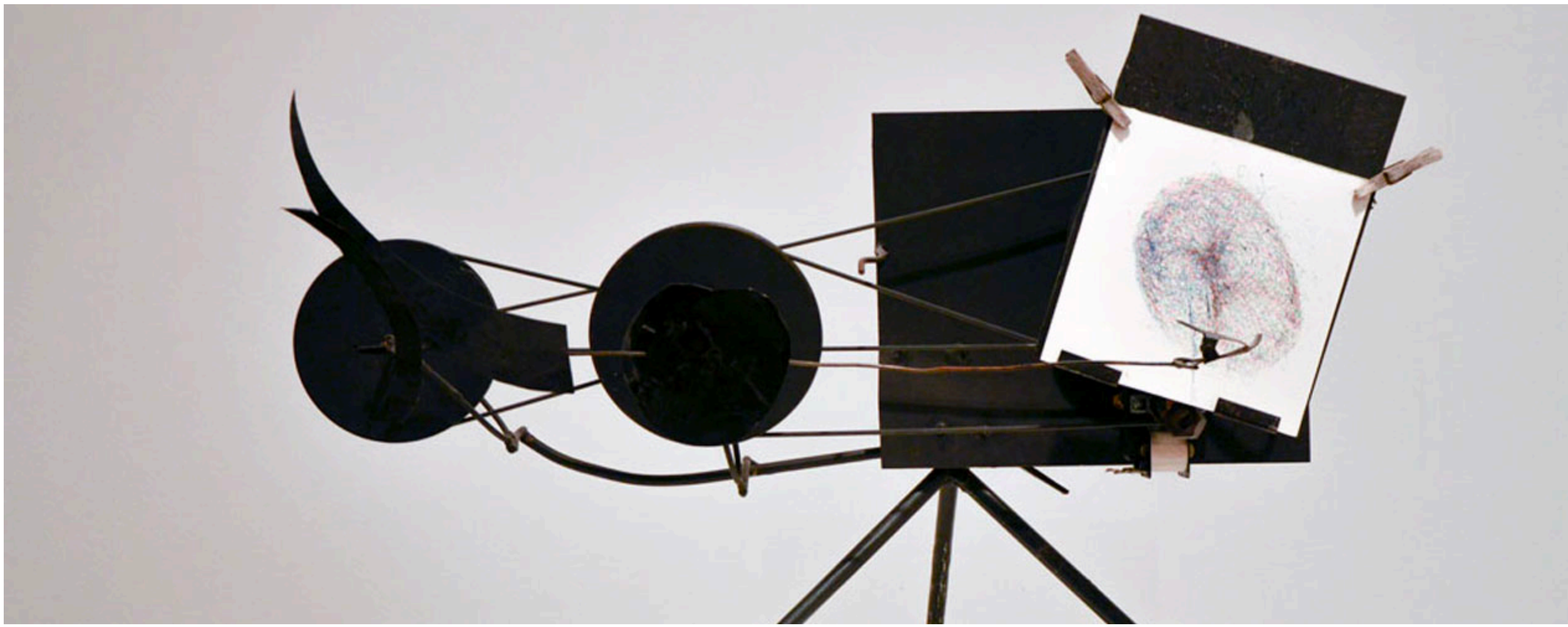
Input:
Random numbers

Random numbers…



Götz, Karl Otto: Statistisch-metrischer
Versuch 4:2:2:1, Entwurf Sommer 1959

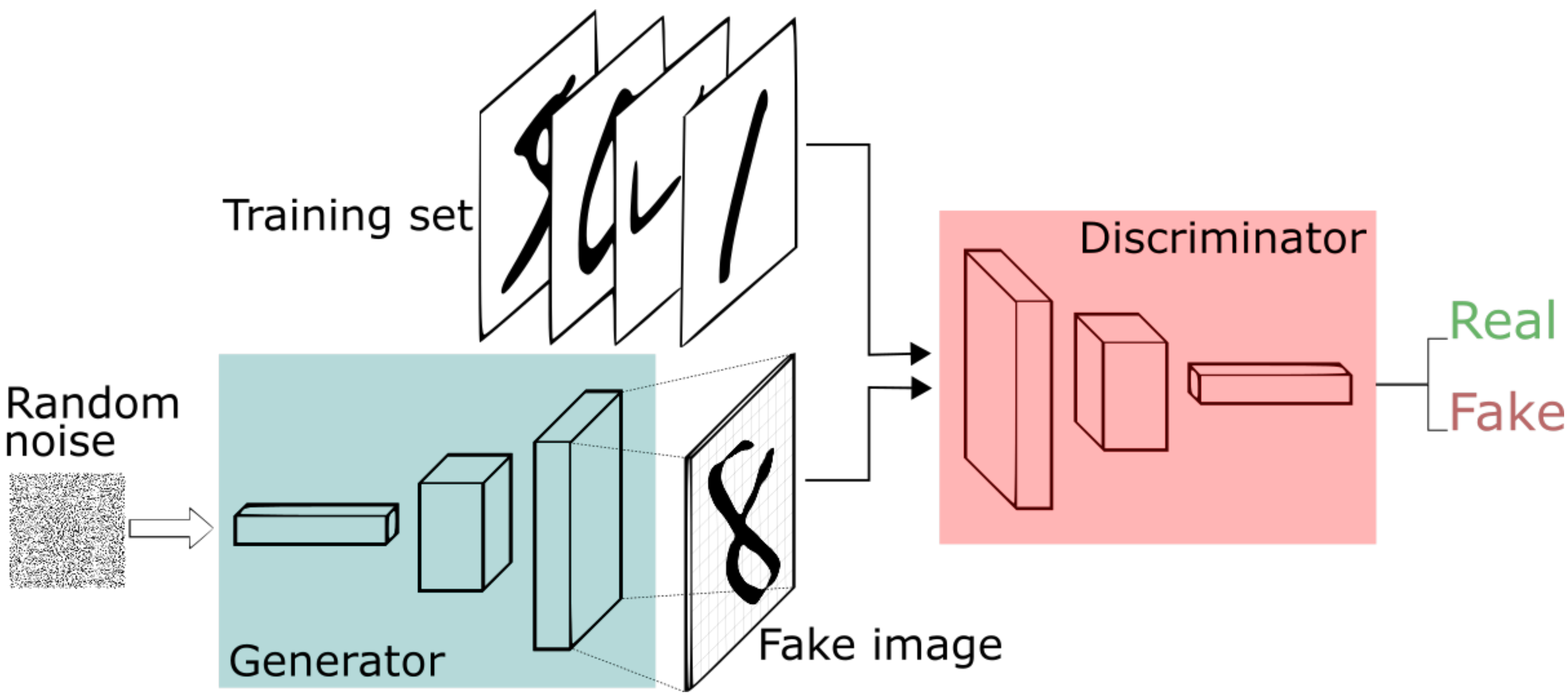
Random input → Art



Tinguely, Meta Matics

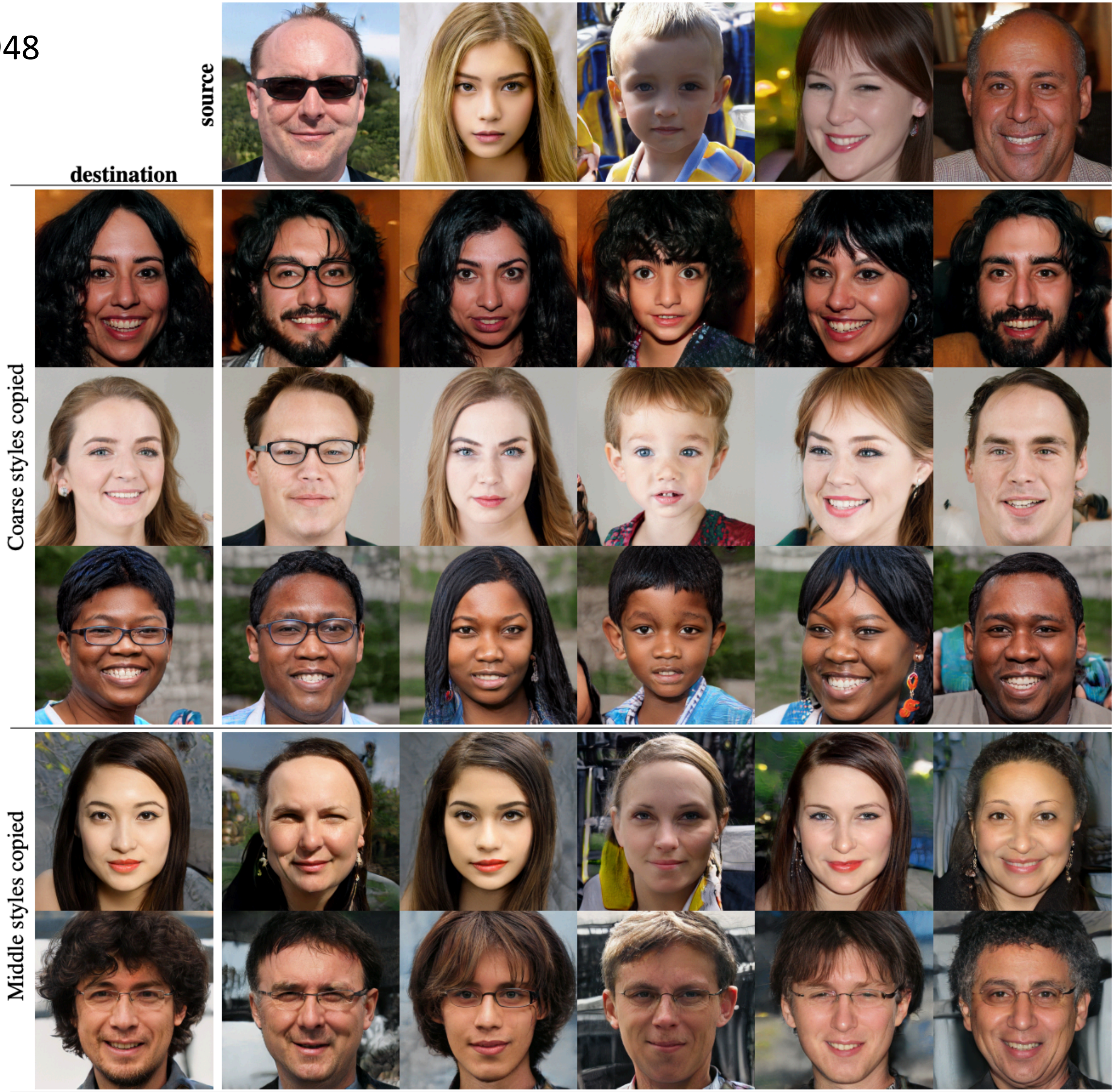
Network simulations ?

Generative Adversarial Networks state of the art:

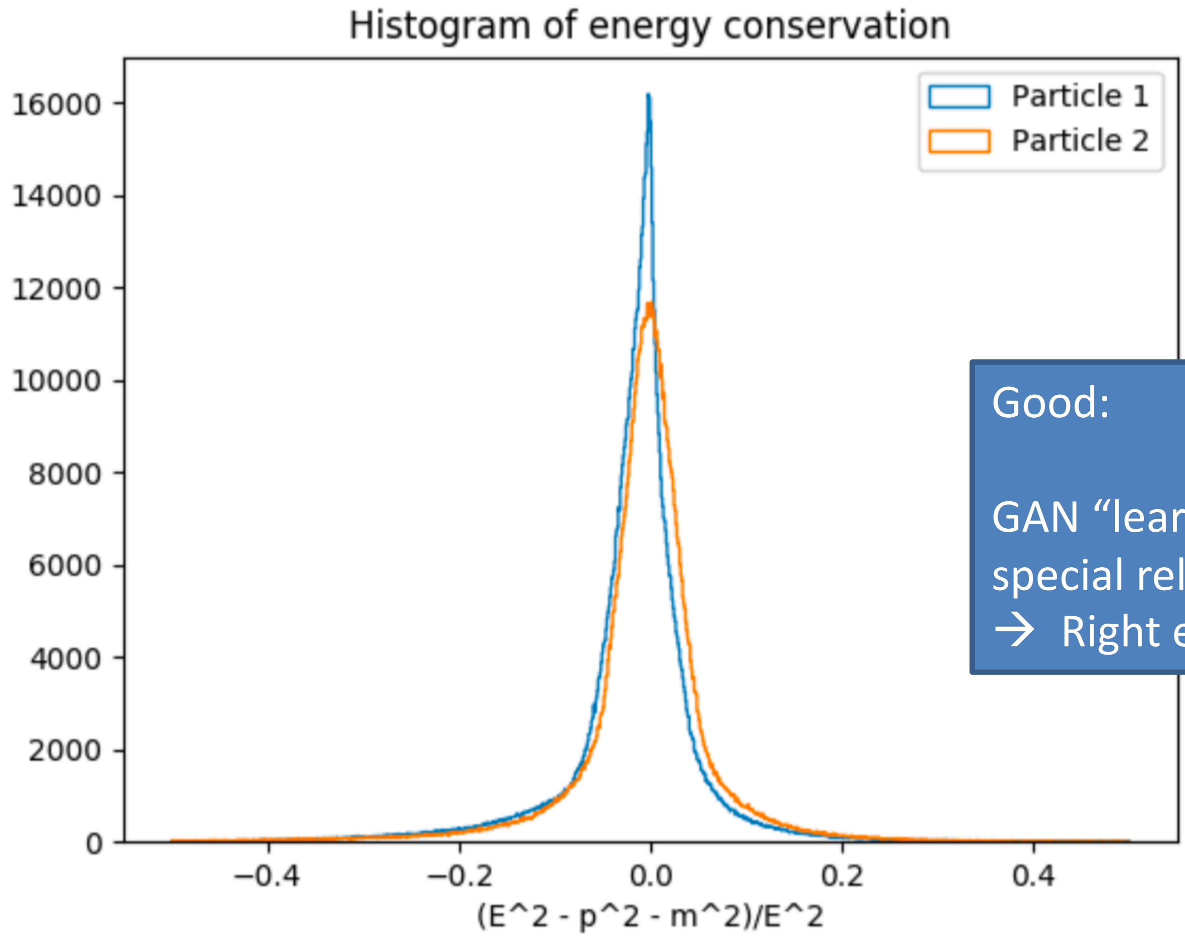


1812.04948

(


source
destination
Coarse styles copied
Middle styles copied

Distributions of Particle Collision “Events“ with GANs

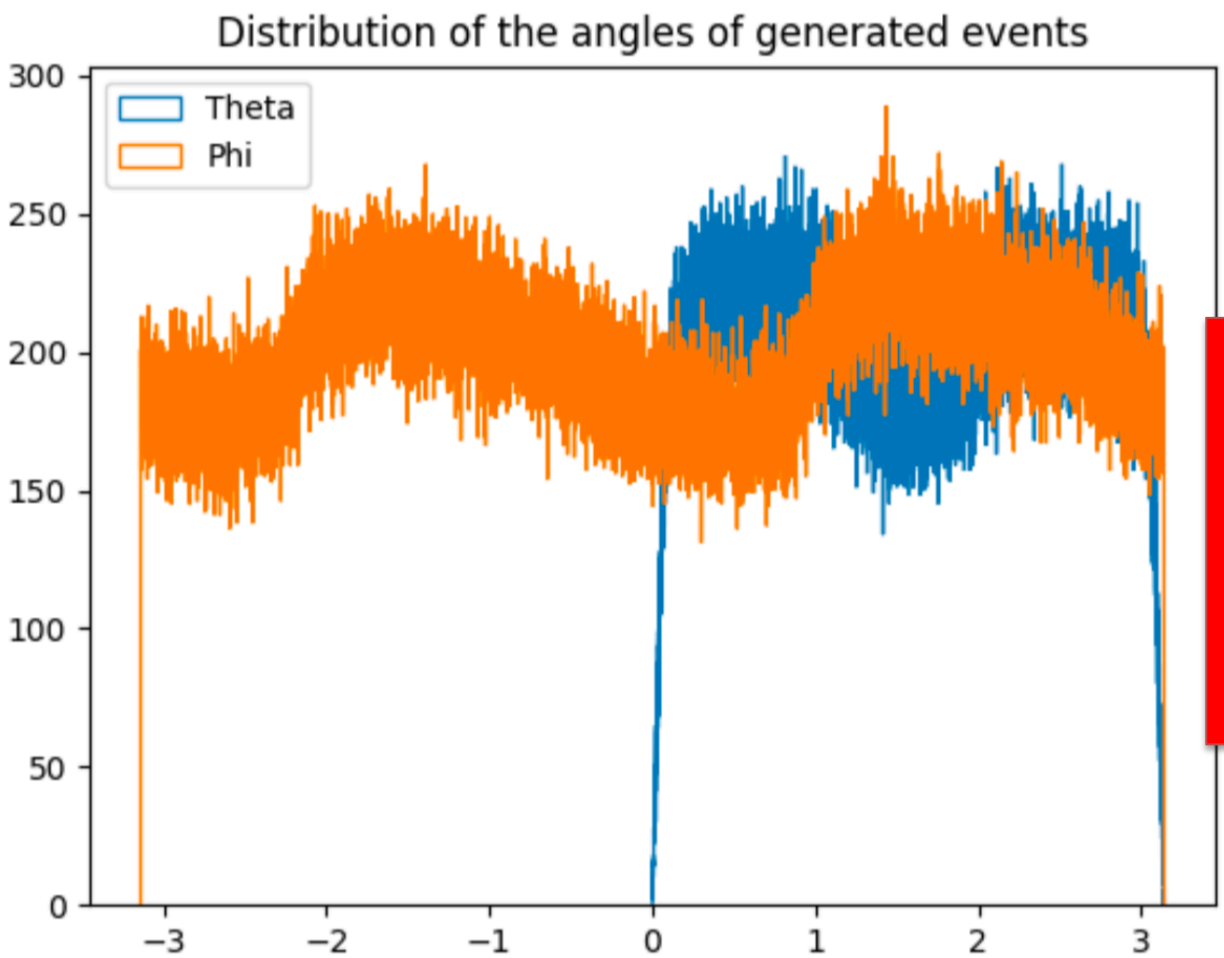


Good:

GAN “learns“ special relativity !
→ Right events

Various other GANs on arxiv, e.g. a[rXiv:1901.05282]() **and various jet GANs**

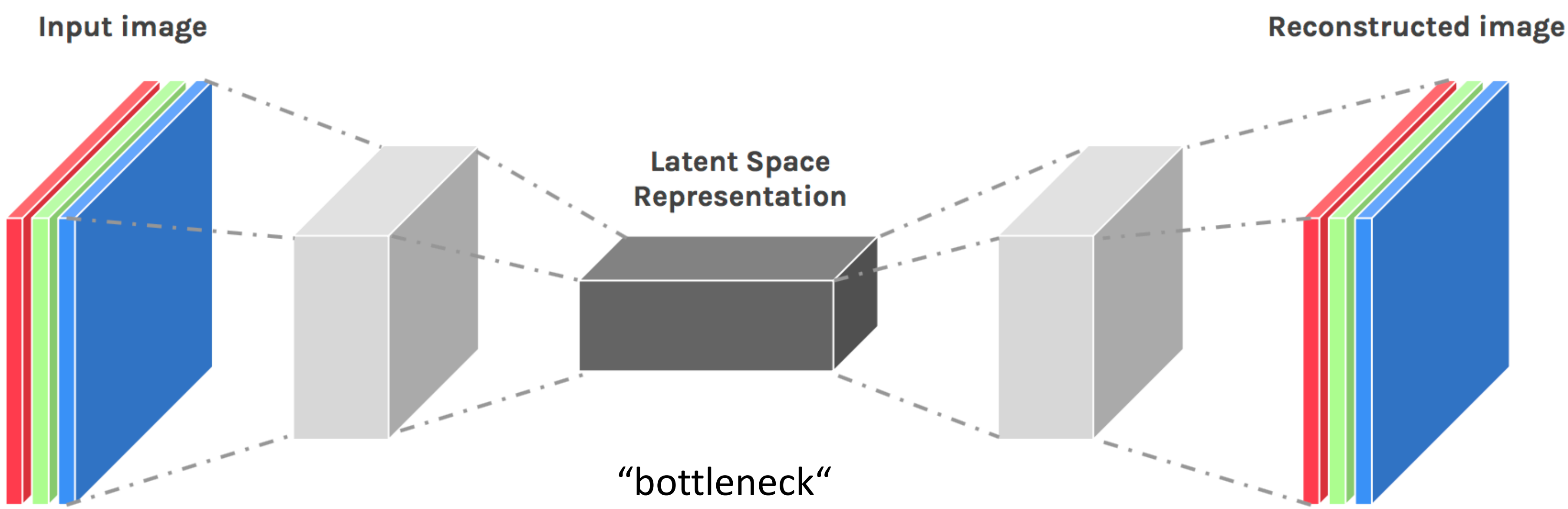
Distributions of Particle Collision "Events" with GANs



BAD:

GAN does not make
events of different types with right frequencies !

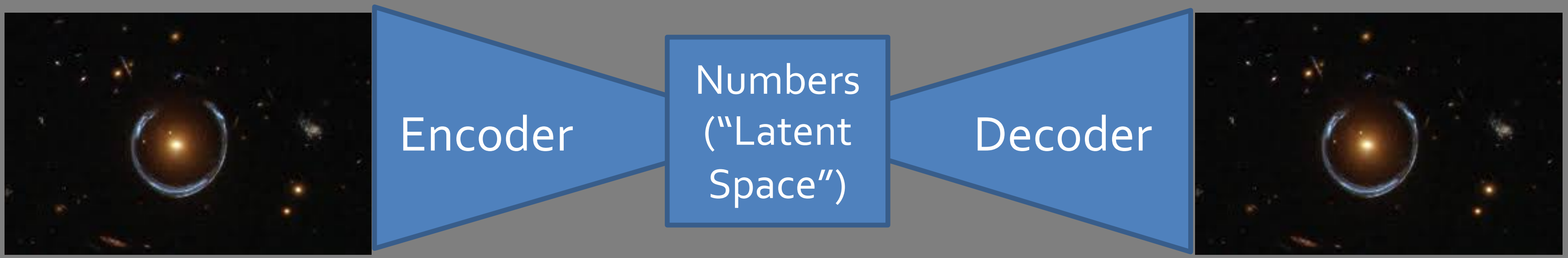
Autoencoders



Can find new physics
with reconstruction Loss

We actually use a better version:
„Dutch“ Autoencoder
(Variational Autoencoder by Dederik Kingma and Max Welling)

Example: Autoencoders

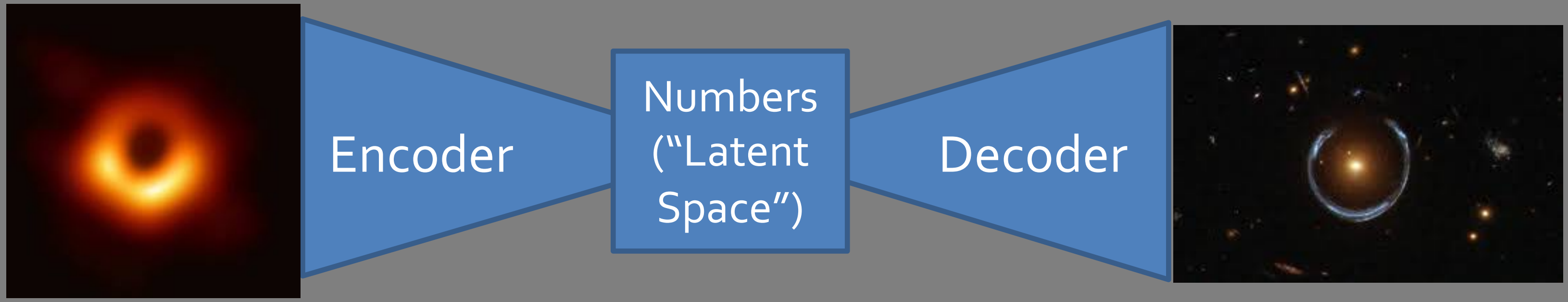


You can encode physics here !

„Dutch" Autoencoder
(Variational Autoencoder by Dederik Kingma and Max Welling)

Example: Autoencoders

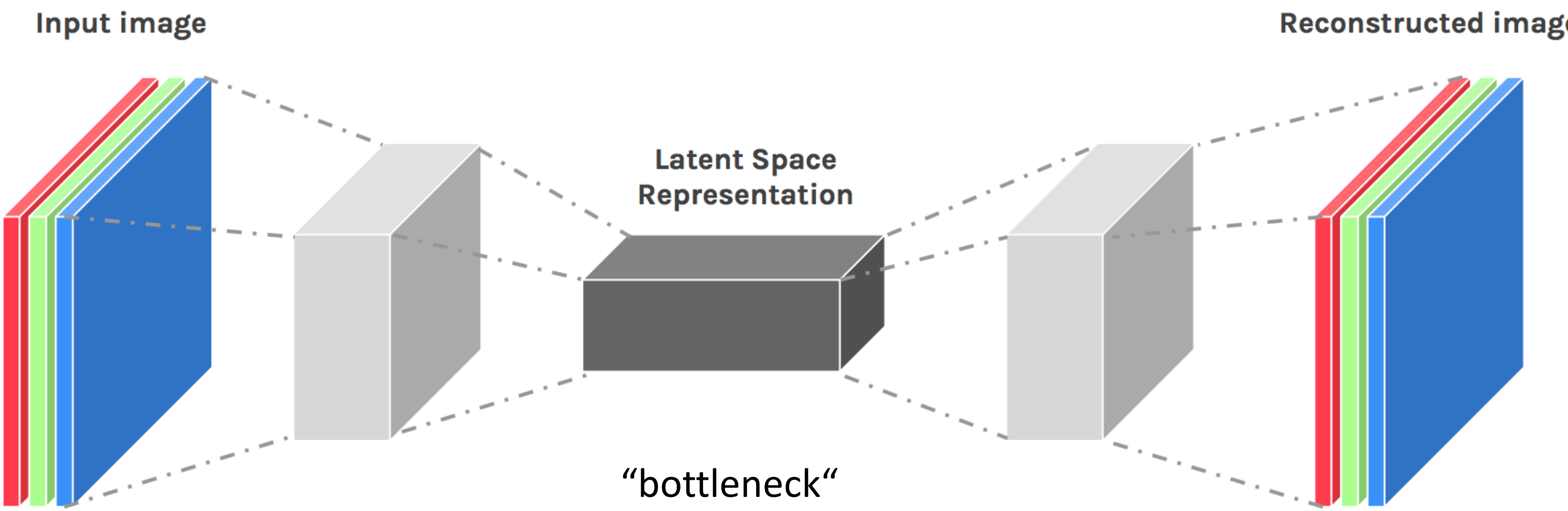
Anomaly Detected !



"bottleneck"

Allows to search
for new physics
(badly reconstructed,
or low density in latent
space)

Variational Autoencoders

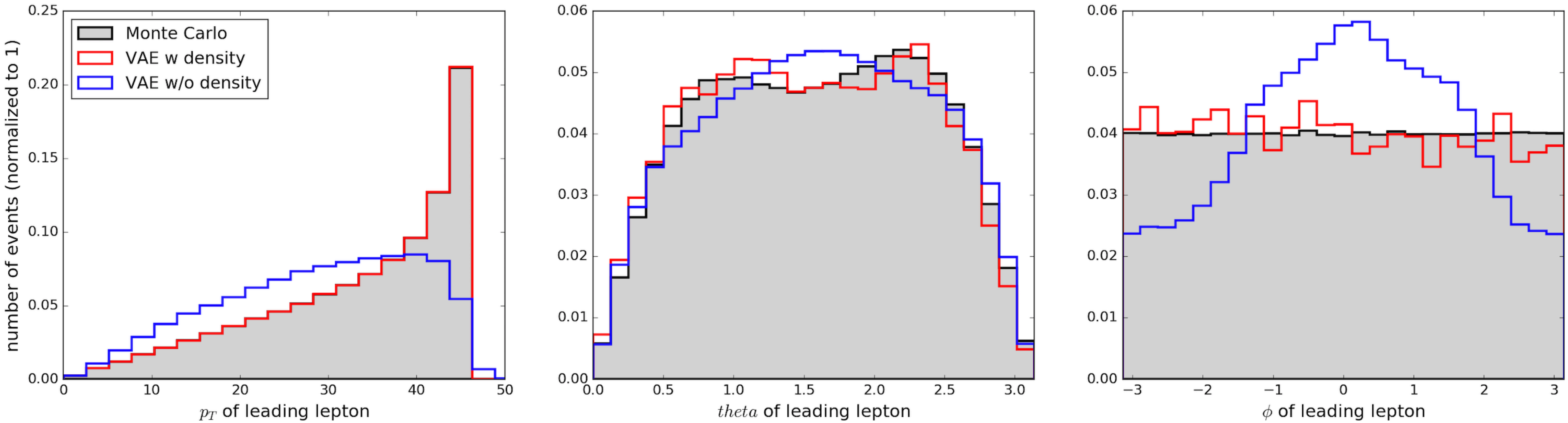


“bottleneck“

Encoder output is the mean and the variance of d Gaussians
Decoder input is z : a sample drawn from these d Gaussians

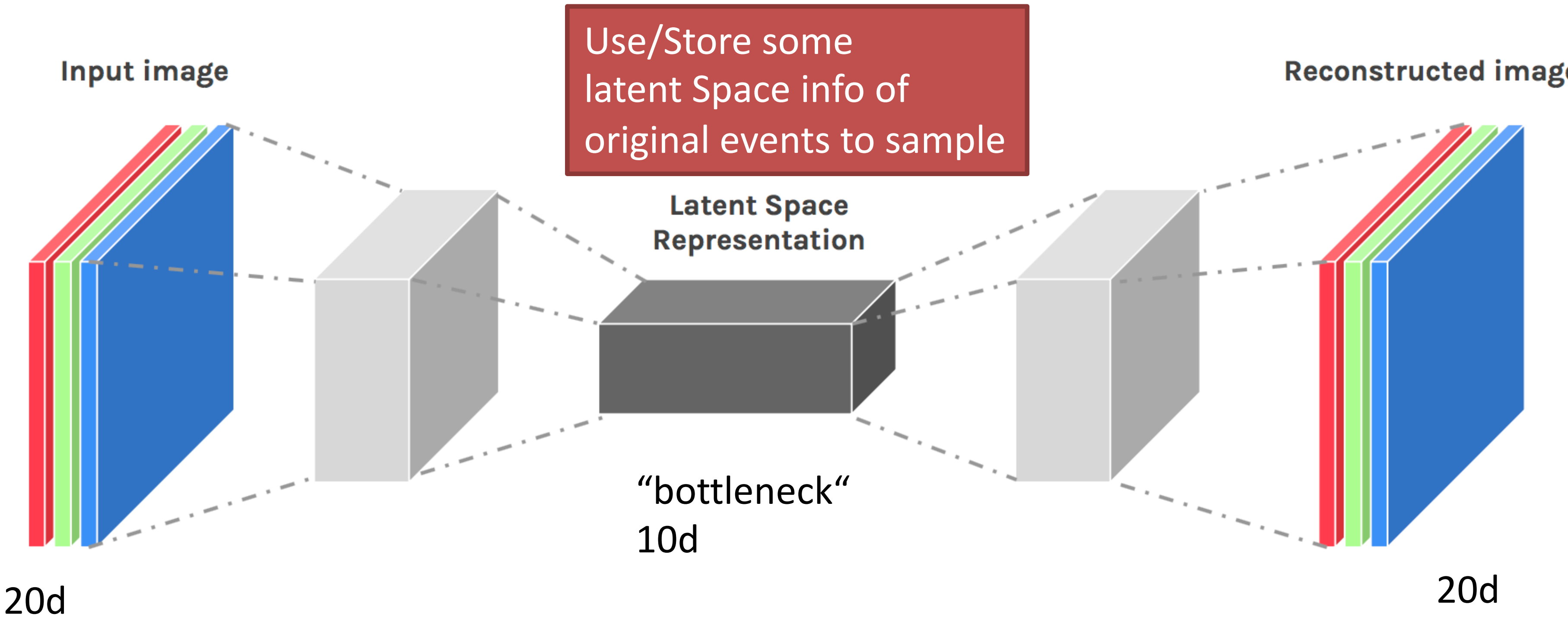
”Dutch Autoencoder” Kingma and Welling…

Distributions of Particle Collision “Events“ with variational autoencoders



BAD:
Autoencoder typically does not
make events of different types with right
frequencies !

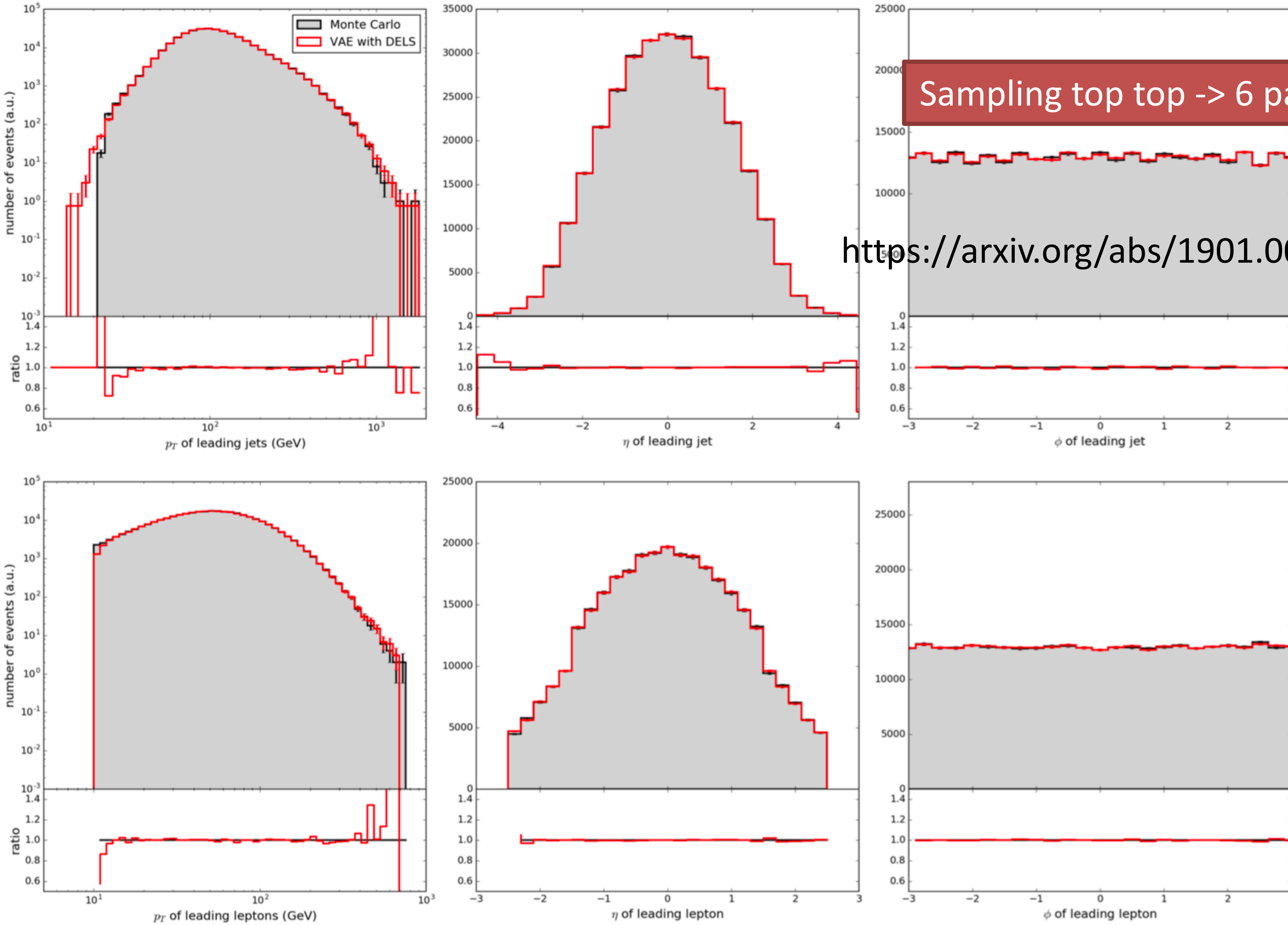
Autoencoder
(+ event info in latent space)



VAE: p(z) is typically from d dimensional gauss,
VAE with buffer: $$p(z) = \sum_{i=1}^{N} p(z|x_i)p(x_i)$$

https://arxiv.org/abs/1901.00875

Distributions of Particle Collision “Events“ with „latent space buffering“ of variational autoencoders



Sampling top top -> 6 particles

https://arxiv.org/abs/1901.00875

Increasing the gaussians in latent space → smudge factor effect

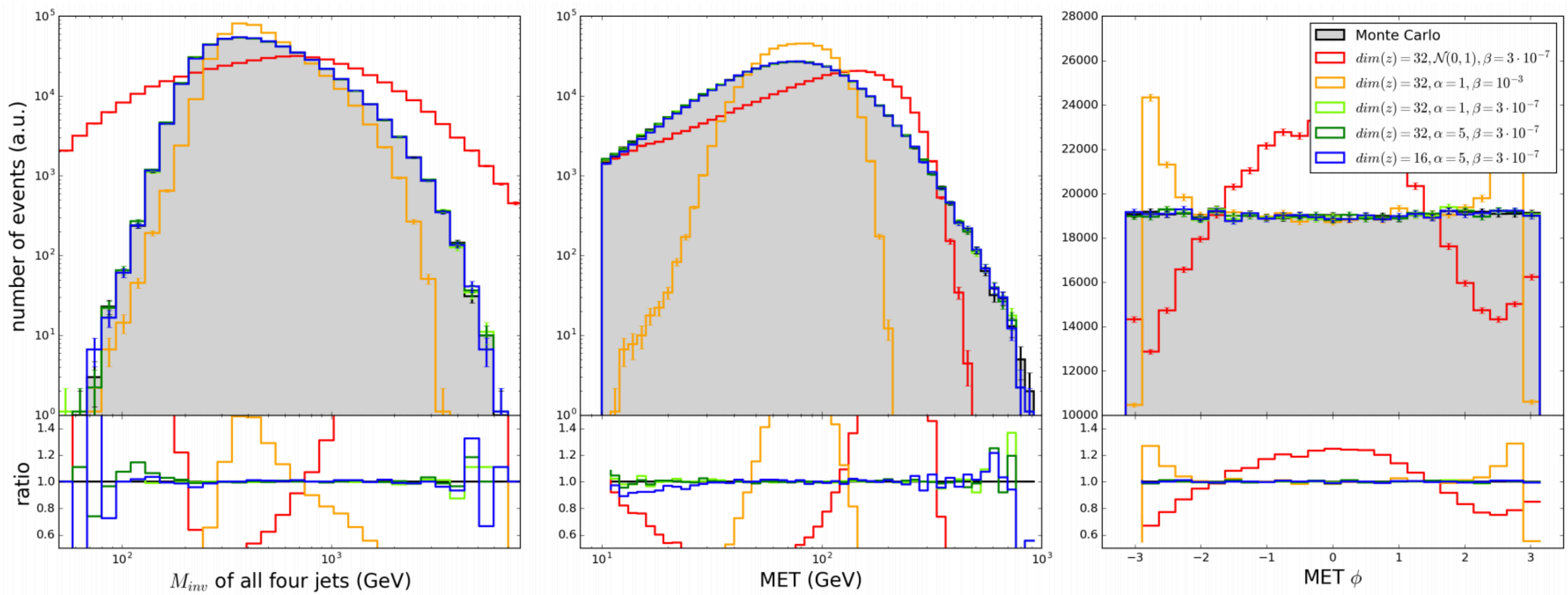


FIG. 6: Events that are generated by the Monte Carlo generator for the $pp \to t\bar{t}$ process (gray), by the standard VAE (red line) and by the B-VAE (rest) for different values of α, β and $\dim(z)$. Shown from left to right is the

Monte Carlo made from data !

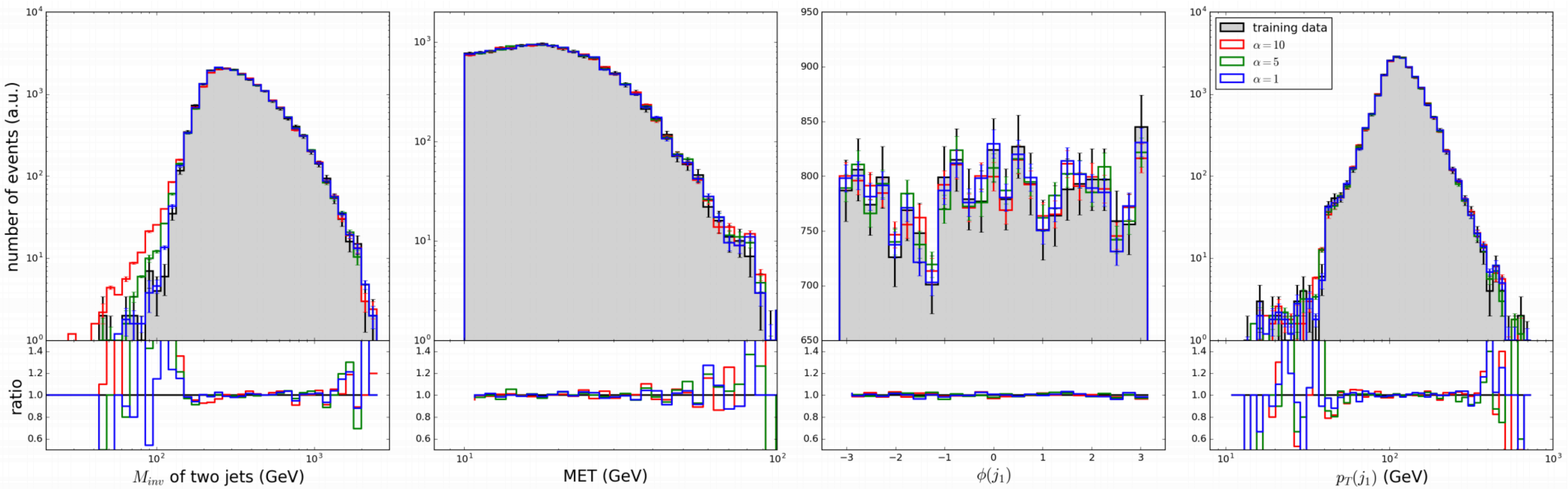
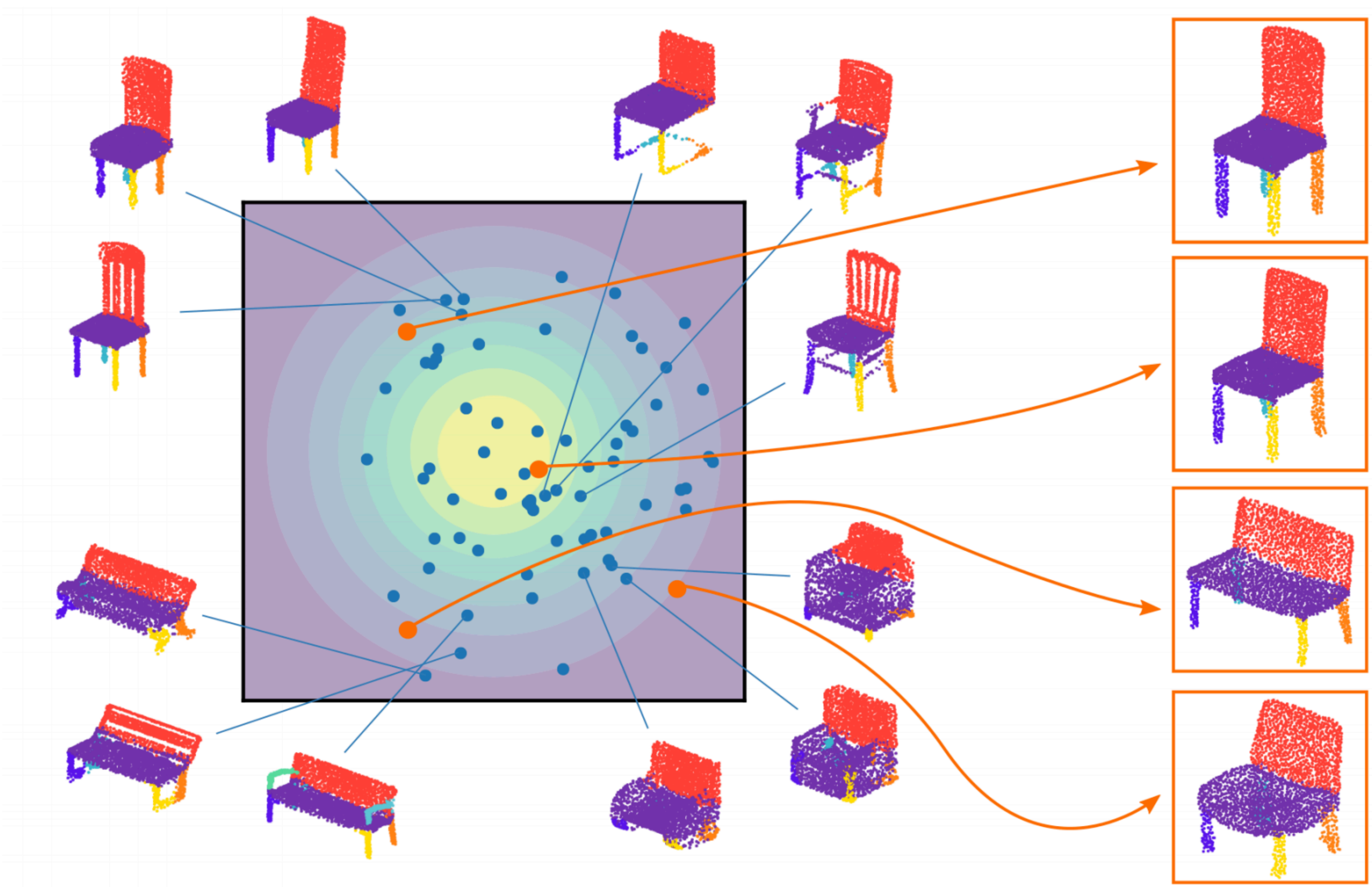


FIG. 10: Experimental events that are taken from the MultiJet primary dataset from CMS open data [40] (gray) and three B-VAE configurations with $\alpha = 1$ (blue), $\alpha = 5$ (green) and $\alpha = 10$ (red). Shown are the invariant mass distribution for the leading and subleading jet, the missing transverse energy, as well as ϕ and p_T of the leading jet.

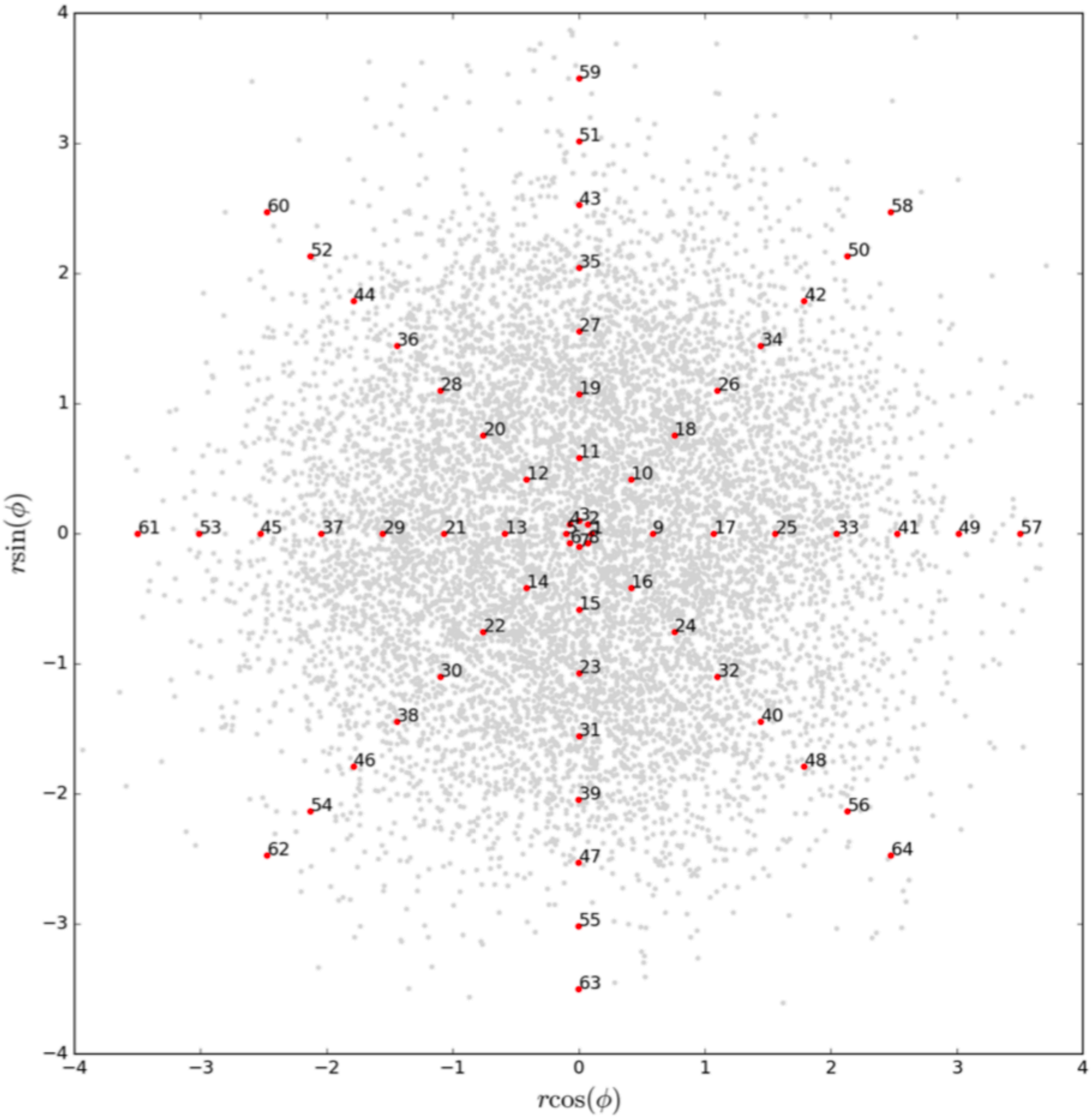
Data preservation in form of MCs ?

Concept of a latent space of sofas and chairs



arxiv 1610.07584

Top top Latent space PCA1 vs PCA2



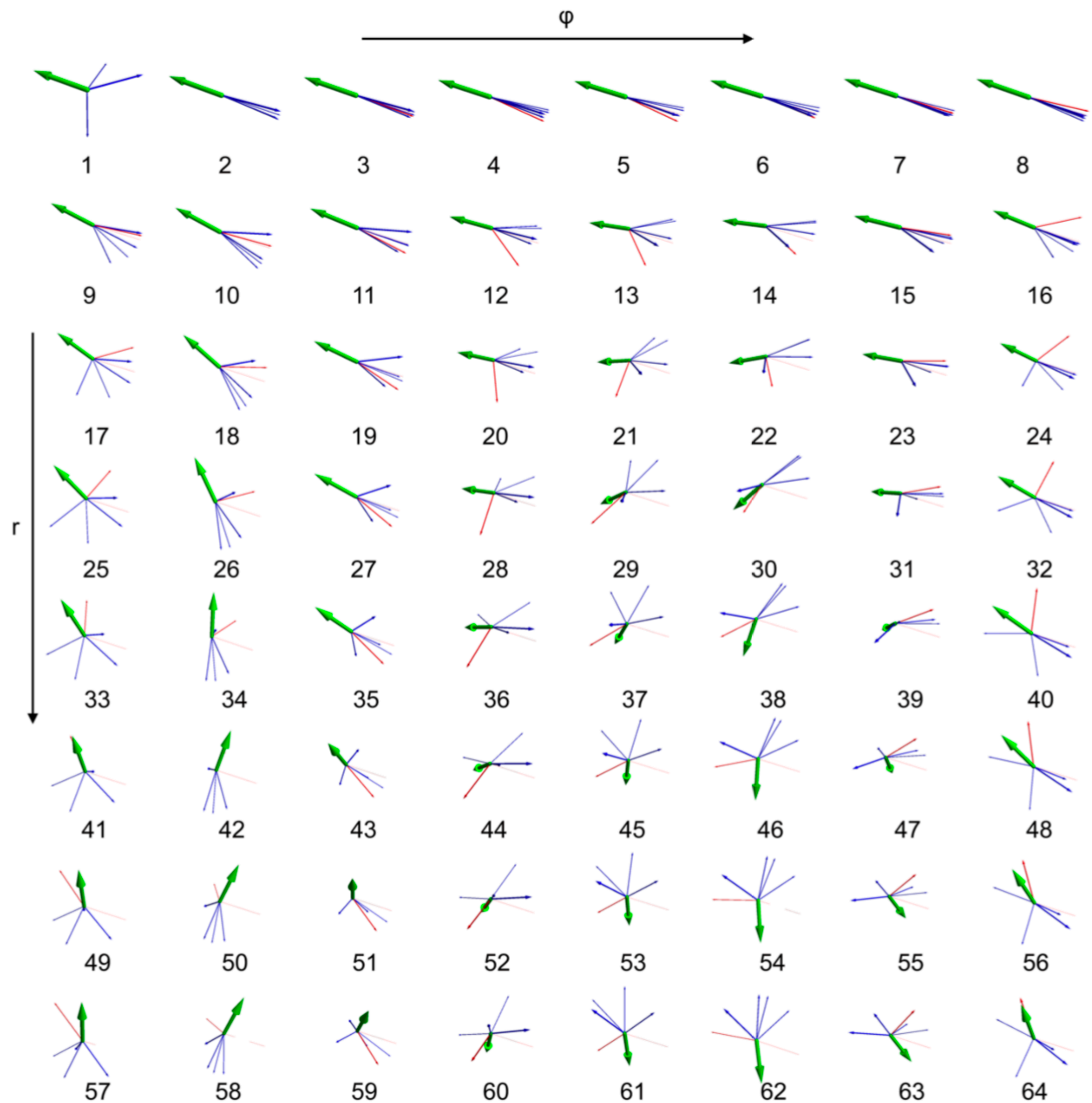


FIG. 7: Visualization of the first two components of a principal component analysis of encoded Monte Carlo events in latent space. This shows an 8×8 grid of event displays following the red dots in Figure 6. These 64 points chosen

Discussion: Follow up via Les Houches

- What can we do with this ? → Let‘s discuss

- For which cases does it sample better than sqrt (N_local) ?

- Inverse of generators ?

Summary

- Help us to accelerate (friendly) science with machine learning

Extra Slides