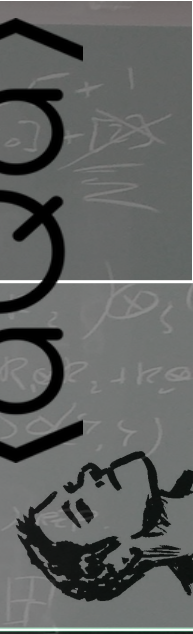
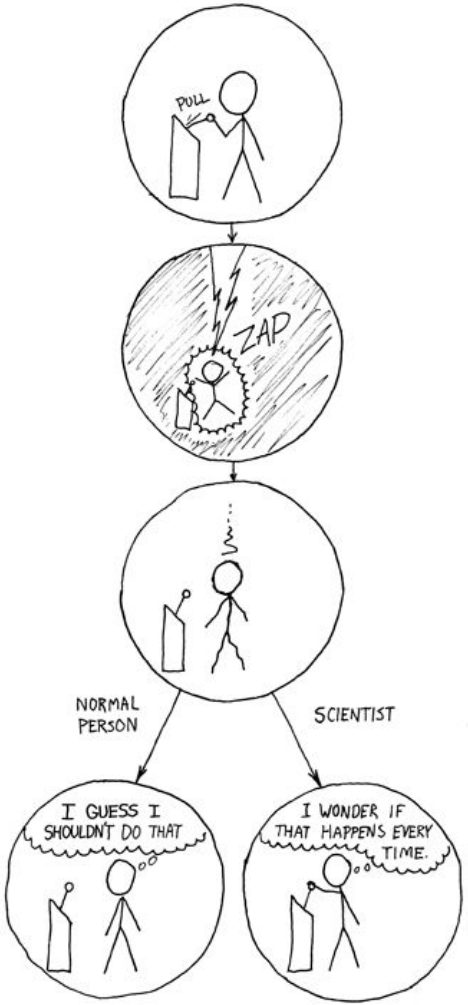
Overview of Classical Data Analysis and ML in HEP

Vince Croft



Monday 12 September 2022



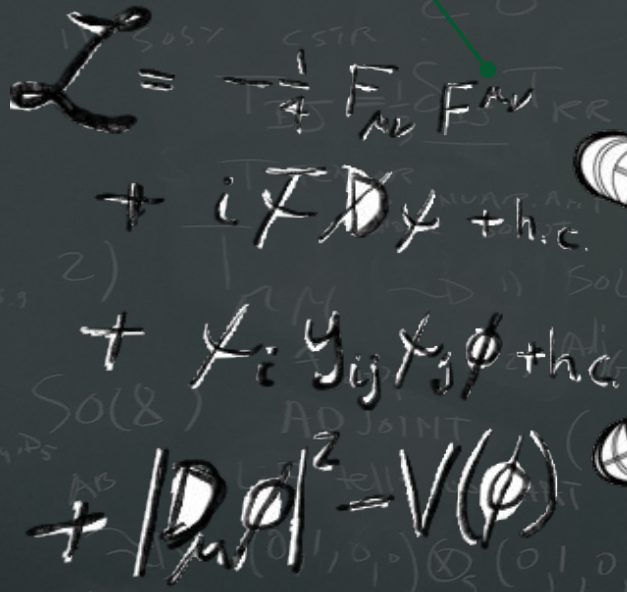
Observation

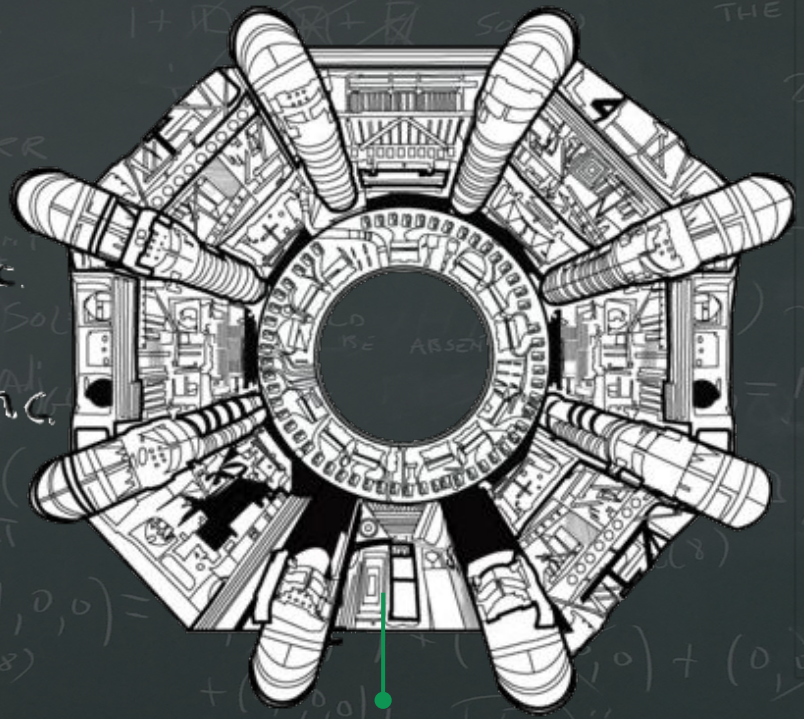
Hypothesis

SCIENCE!

Test

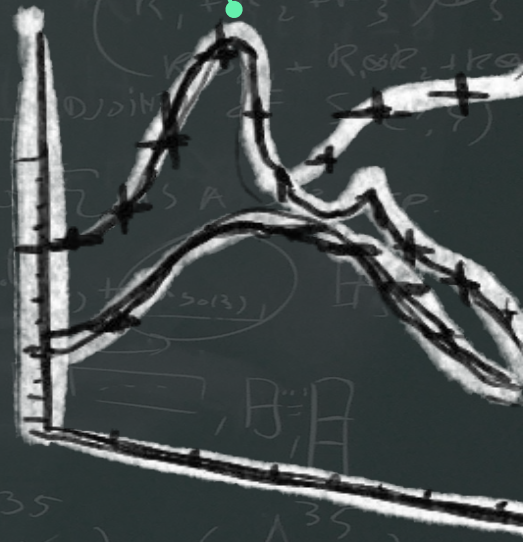
Hypothesis





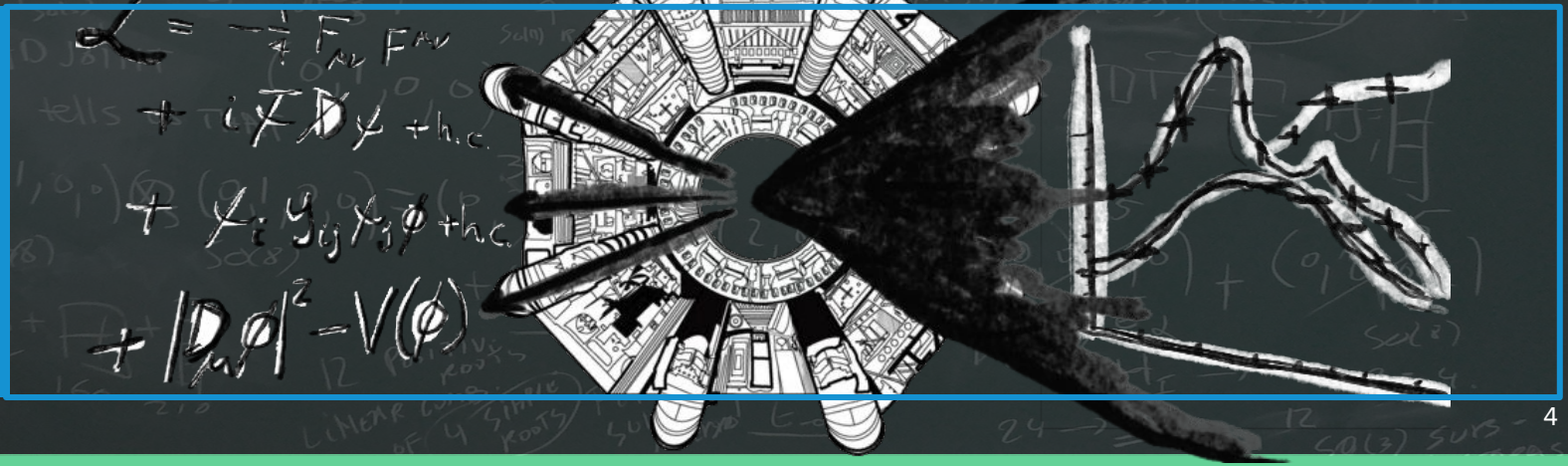
Test

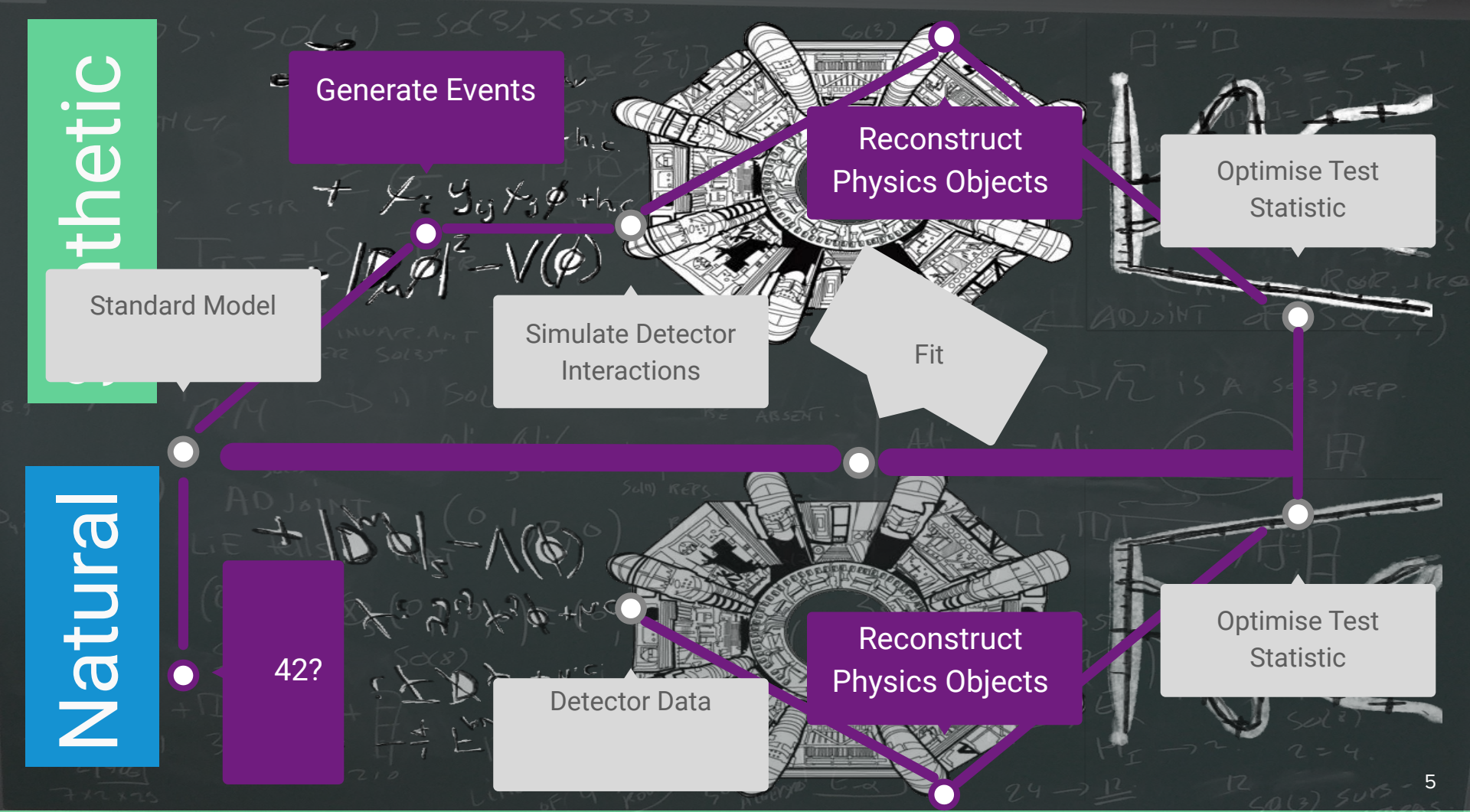
Observation

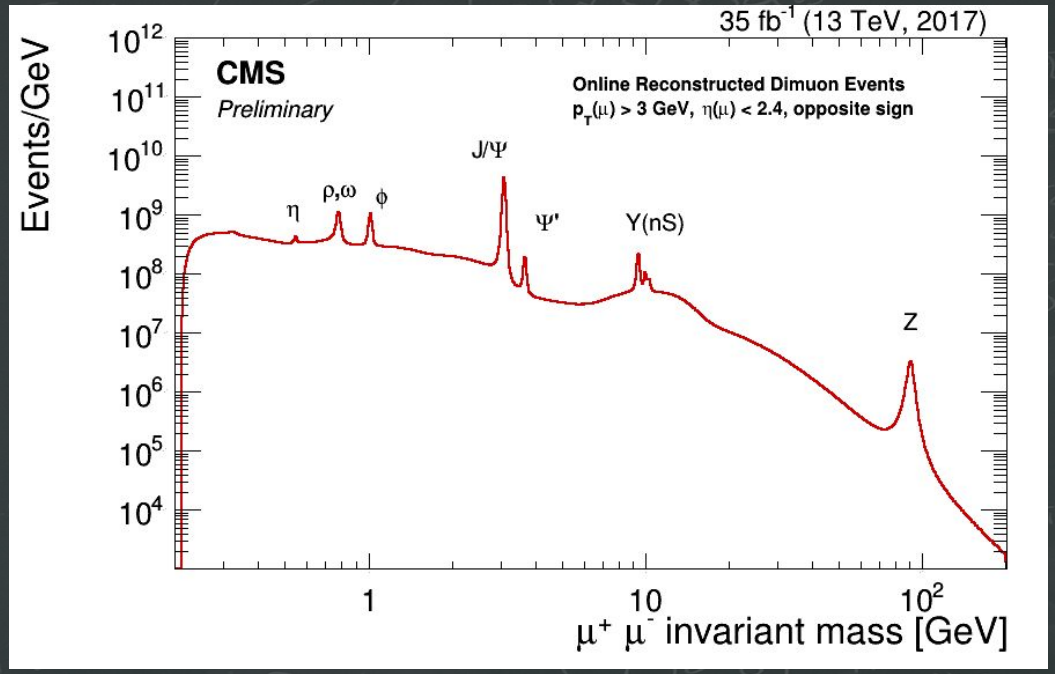


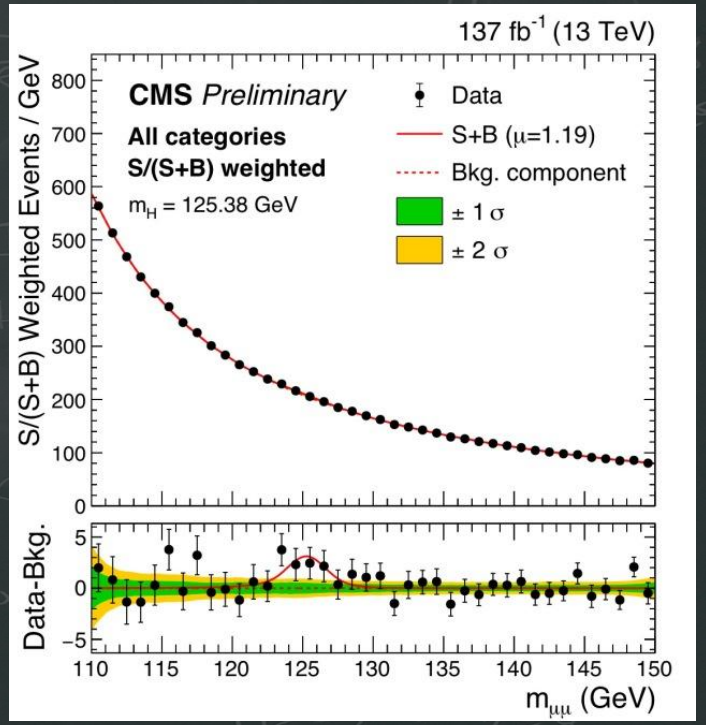
Synthetic

Natural




thetic
Generate Events
Reconstruct Physics Objects
Optimise Test Statistic
Standard Model
Simulate Detector Interactions
Fit
Natural
42?
Detector Data
Reconstruct Physics Objects
Optimise Test Statistic


35 fb-1 (13 TeV, 2017)
CMS
Preliminary
Online Reconstructed Dimuon Events
pT(μ) > 3 GeV, η(μ) < 2.4, opposite sign
Events/GeV
η
ρ,ω
ϕ
J/Ψ
Ψ'
Υ(nS)
Z
μ+ μ- invariant mass [GeV]


137 fb-1 (13 TeV)
CMS Preliminary
All categories
S/(S+B) weighted
mH = 125.38 GeV
Data
S+B (μ=1.19)
Bkg. component
± 1 σ
± 2 σ
S/(S+B) Weighted Events / GeV
Data-Bkg.
mμμ (GeV)

thetic

Generate Events

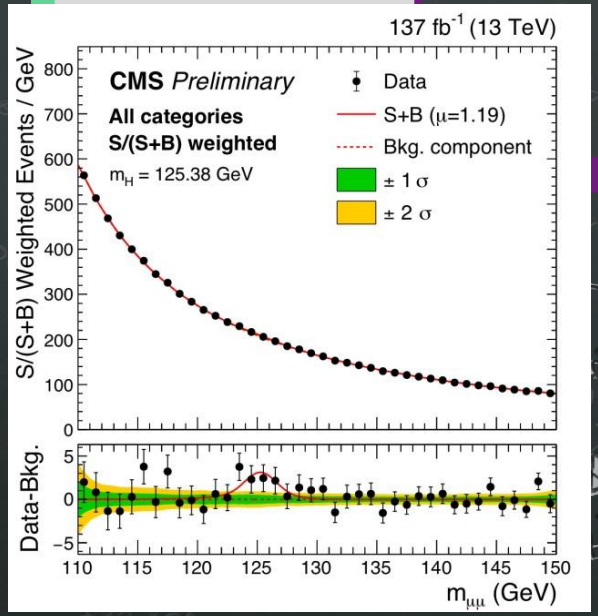
Standard Model

Simulate Detector Interactions

Reconstruct Physics Objects

Fit

Optimise Test Statistic



Detector Data

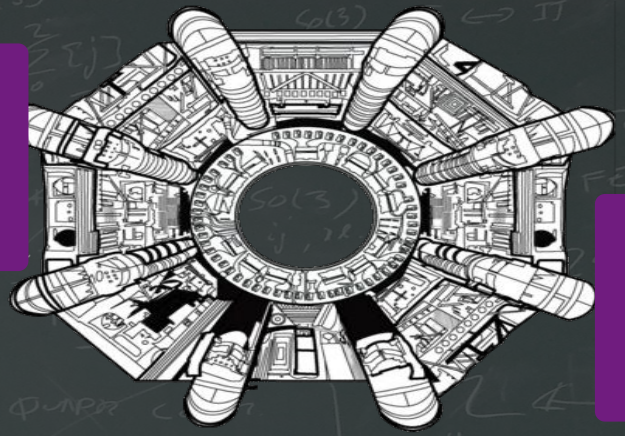
Reconstruct Physics Objects

Optimise Test Statistic

Synthetic

Fast Simulation/Generation

Parameter Estimation



Classification

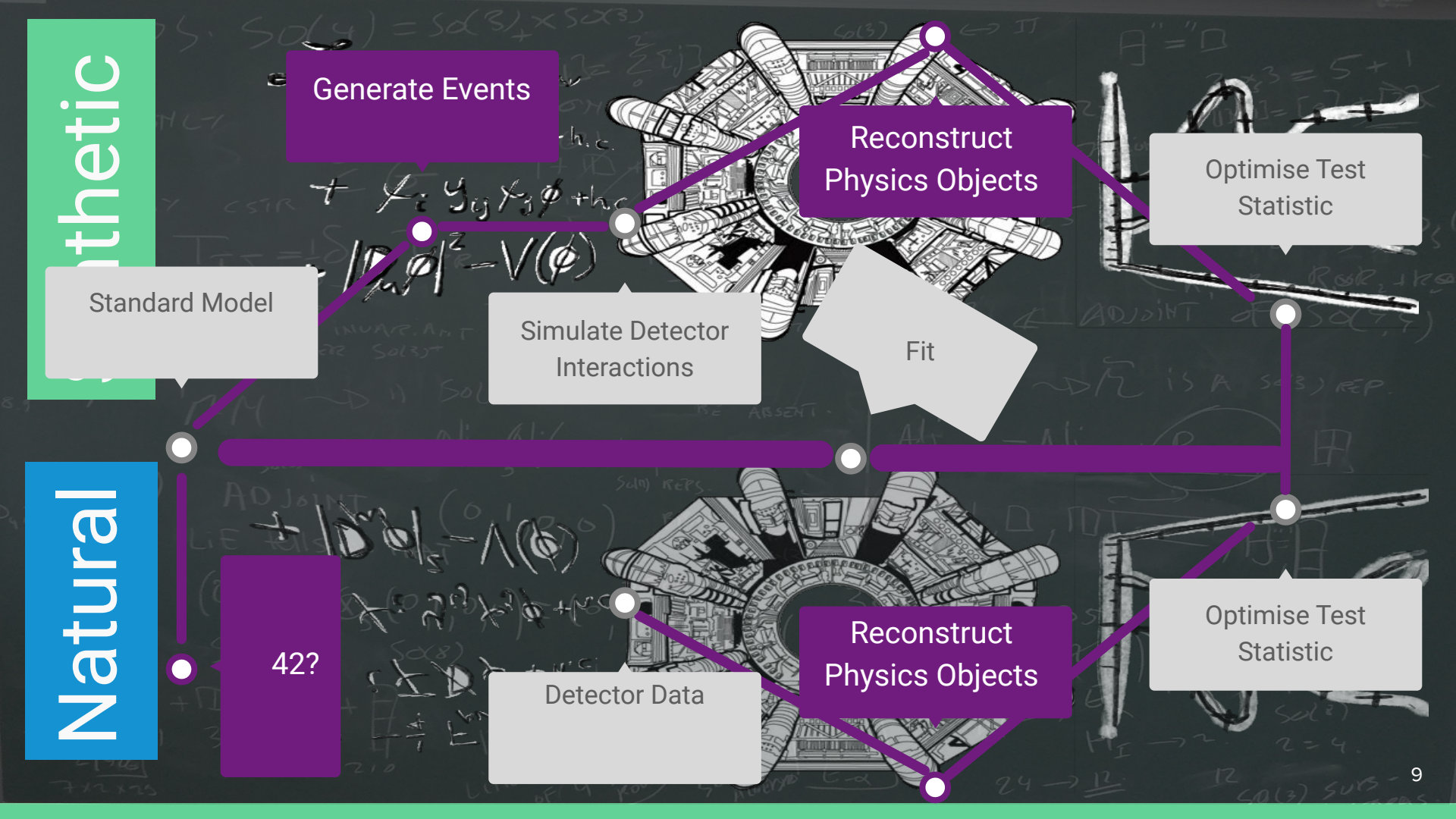
Background Modelling

Natural

Online Processing

Triggers/Quality Control/ Anomalies

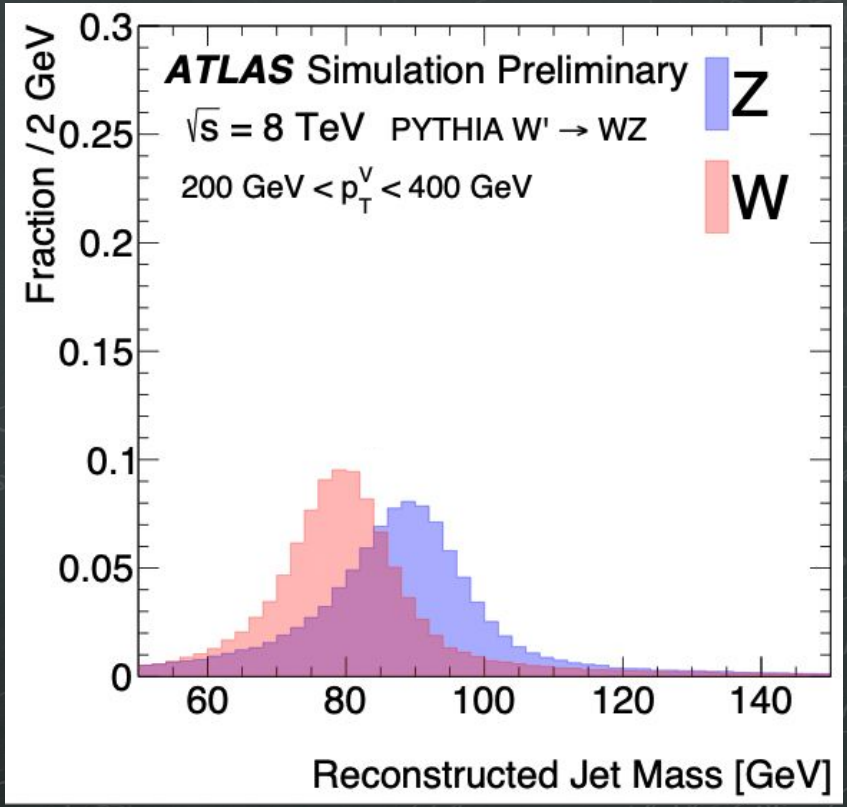
- Calibration
- Clustering
- Tracking
- Noise Mitigation
- Particle ID/Efficiency

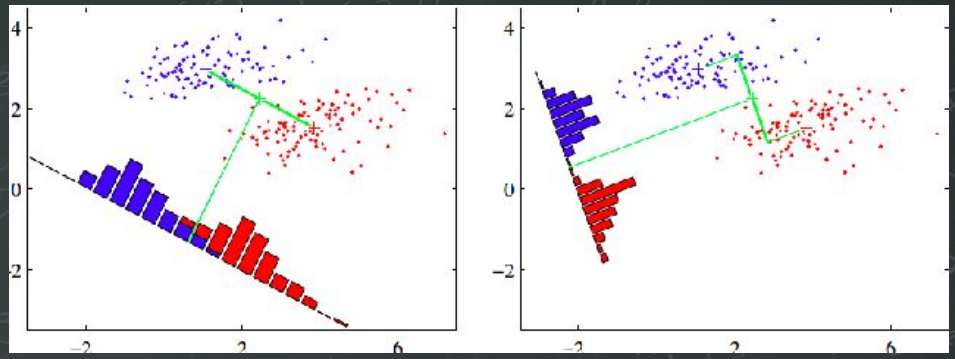

thetic
Generate Events
Reconstruct Physics Objects
Optimise Test Statistic
Standard Model
Simulate Detector Interactions
Fit
Natural
42?
Detector Data
Reconstruct Physics Objects
Optimise Test Statistic

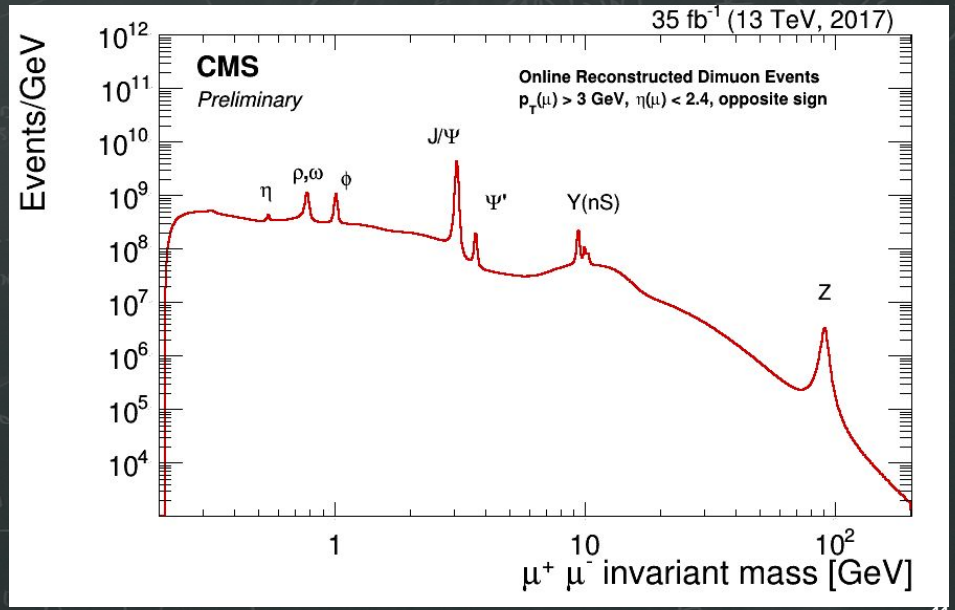
Checklist 1:

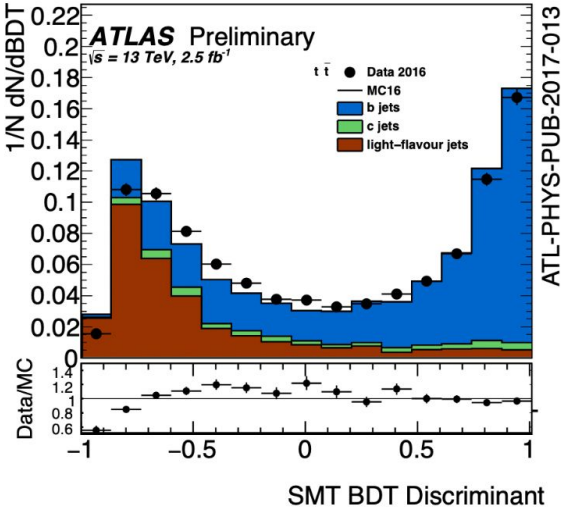
- Science is science.
- Particle physics is all about precision.
- Can simulate huge amounts of high quality, information dense, labeled data.
- ML has clear goals:
  - Performance
  - Accurate generalisation
  - Speed
  - Translatability (controlled biases)
- ML aims to replace physical simulation with functional approximations.

Step -2: Classification





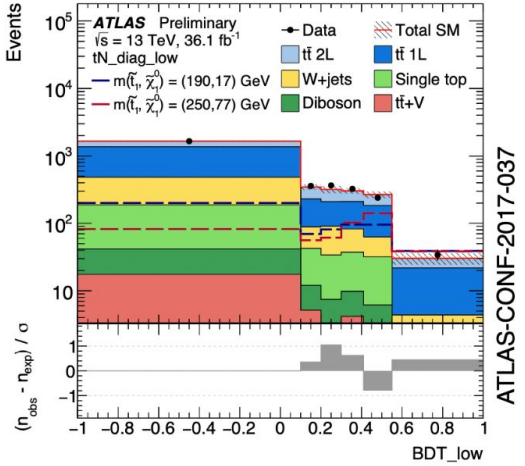


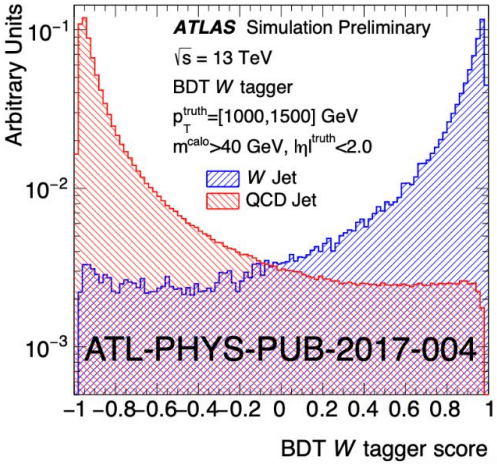


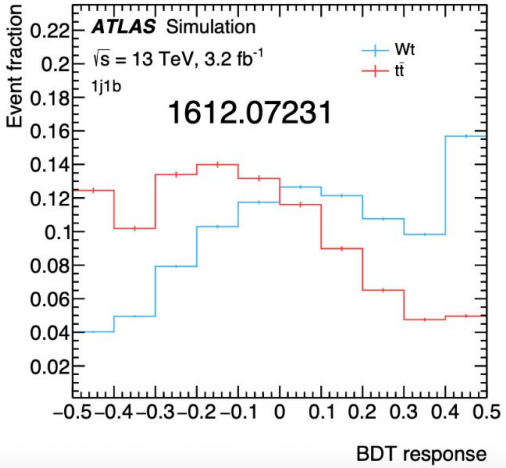
We love BDTs.

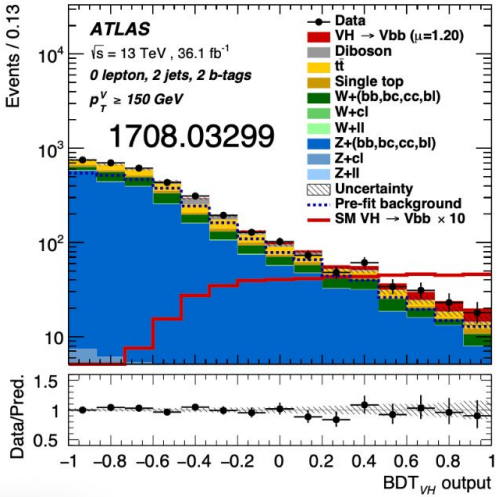
If 3 < dim(feature vector) < O(10)

this is probably right for you!

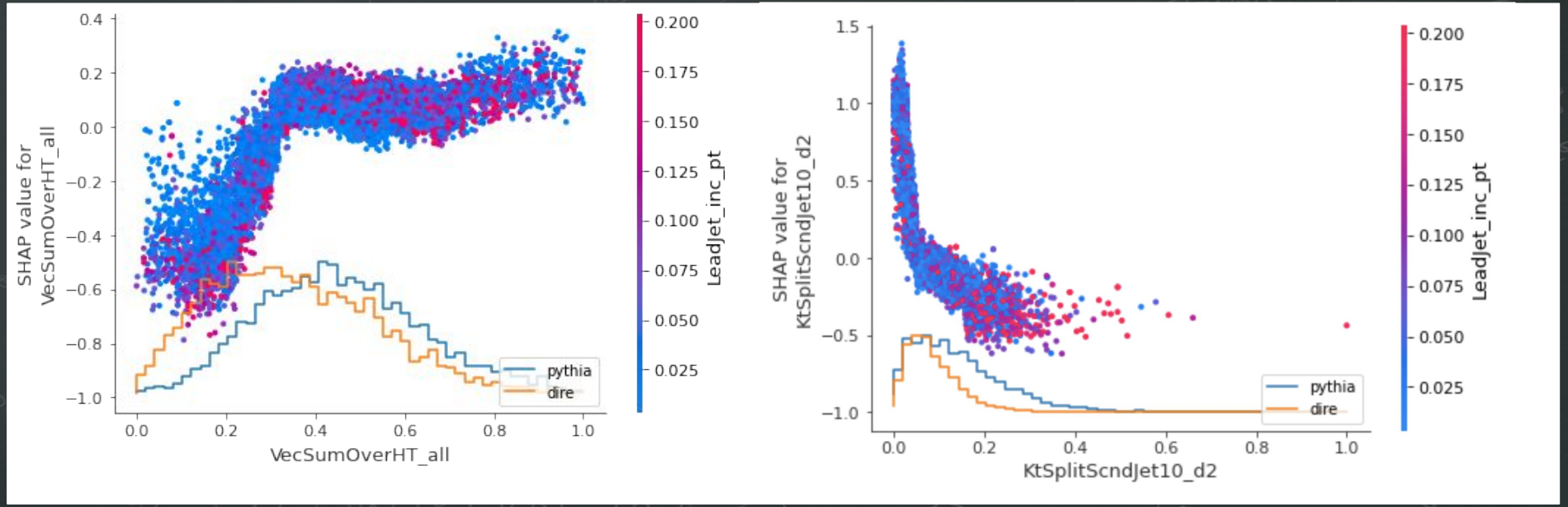








SHAP Distributions

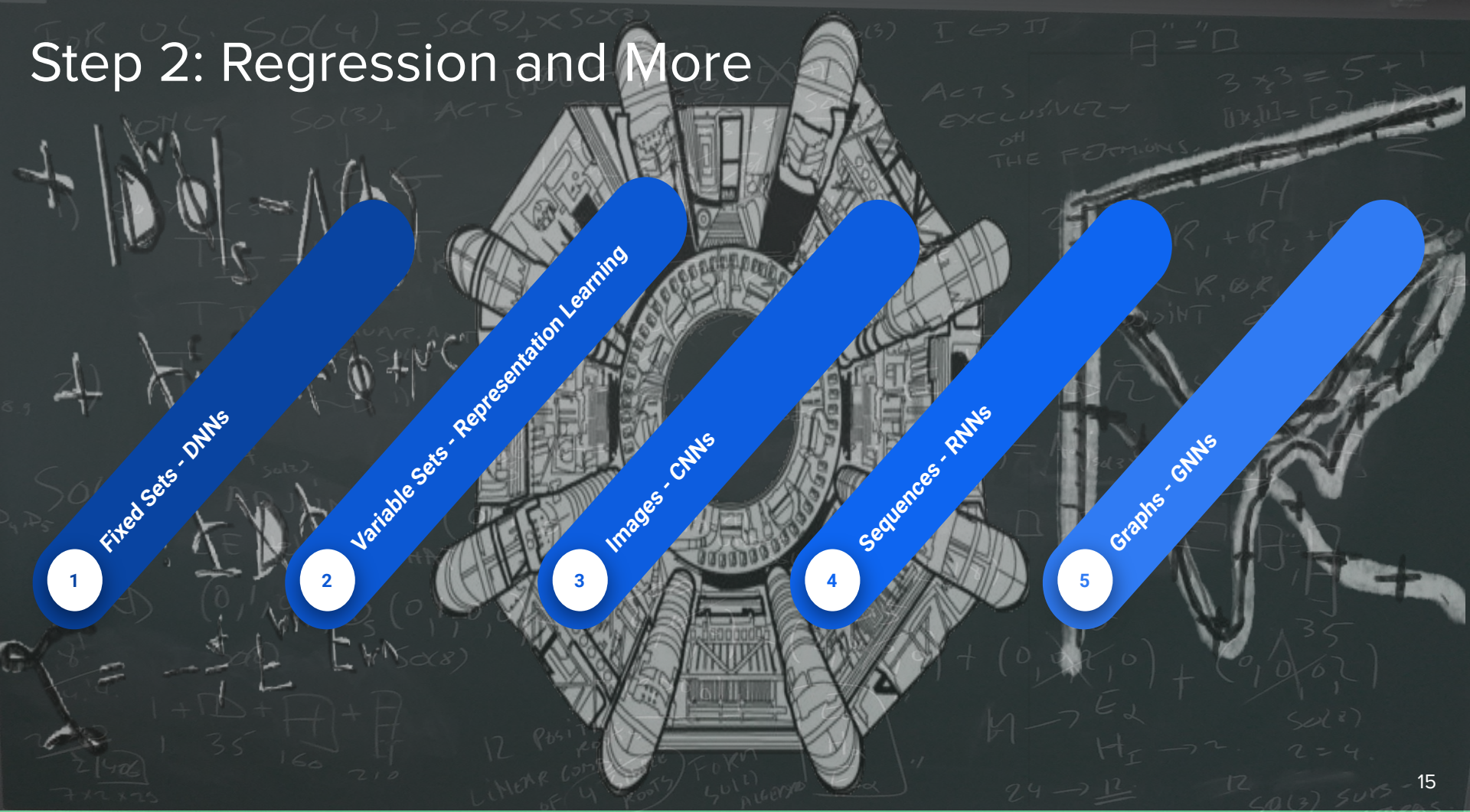


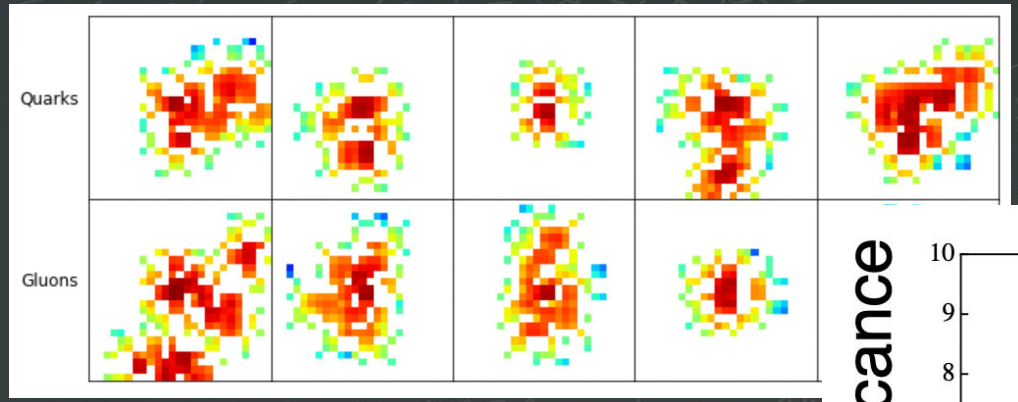
- Difference from 0 is what's important
- Correlations taken into account
- Also useful to explore data and optimize cuts (or debug data selection)
- Anomaly detection

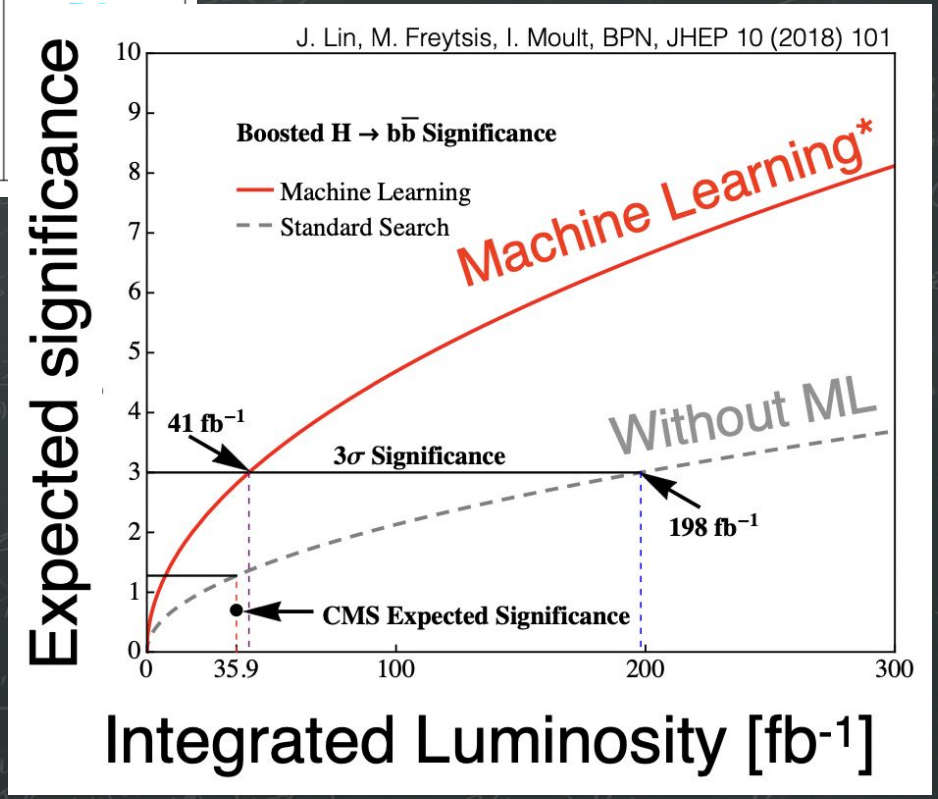
Checklist 2:

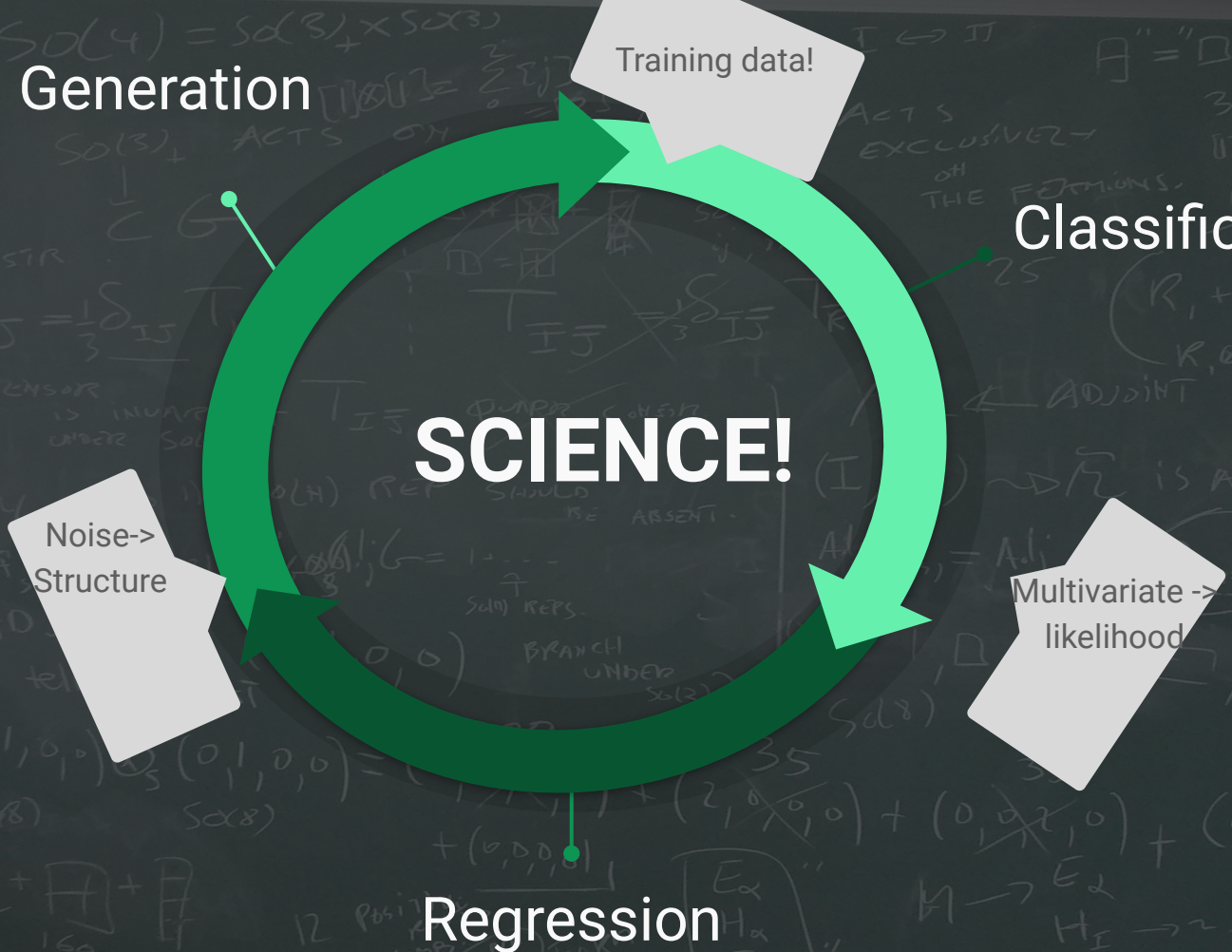
- Signal and Background!
- Big gains in HEP
- Nice processes
- Resonant processes
- Need to reconstruct the properties not just A vs B

Step 2: Regression and More




Quarks
Gluons

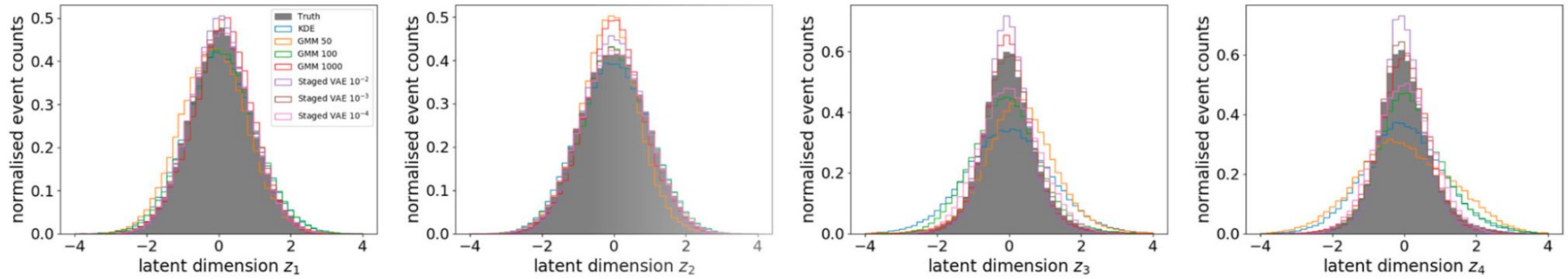

J. Lin, M. Freytsis, I. Moult, BPN, JHEP 10 (2018) 101
Boosted H → bb̄ Significance
Machine Learning
Standard Search
Machine Learning*
Without ML
41 fb⁻¹
3σ Significance
198 fb⁻¹
CMS Expected Significance
Expected significance
Integrated Luminosity [fb⁻¹]
0
1
2
3
4
5
6
7
8
9
10
0
35.9
100
200
300

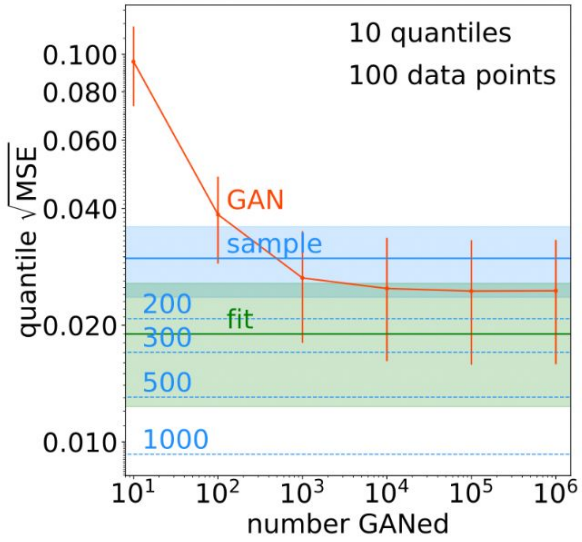

Generation
Training data!
Classification
SCIENCE!
Noise->
Structure
Multivariate ->
likelihood
Regression

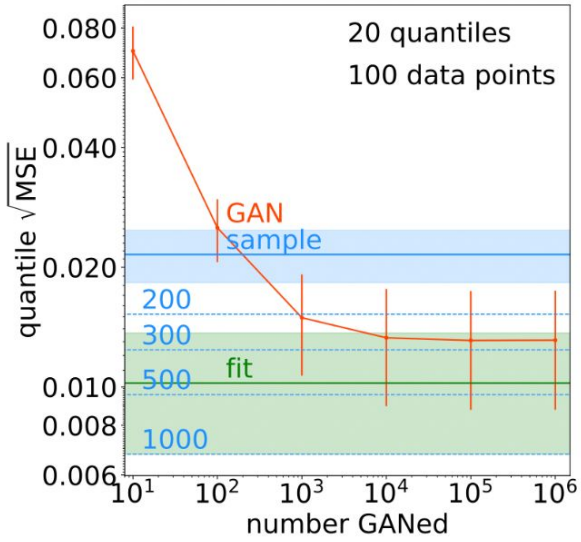
Checklist 3:

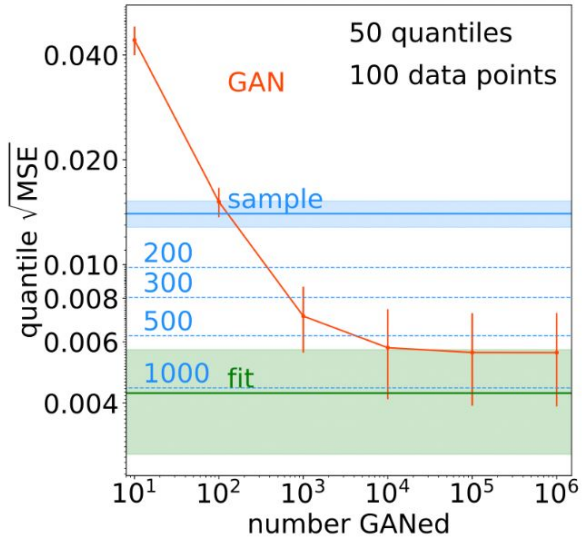
- ML doesn’t replace science.
- Multitude of approaches.
- Many clear benefits of importing computing tools to HEP

Step ±1: Theory and Data

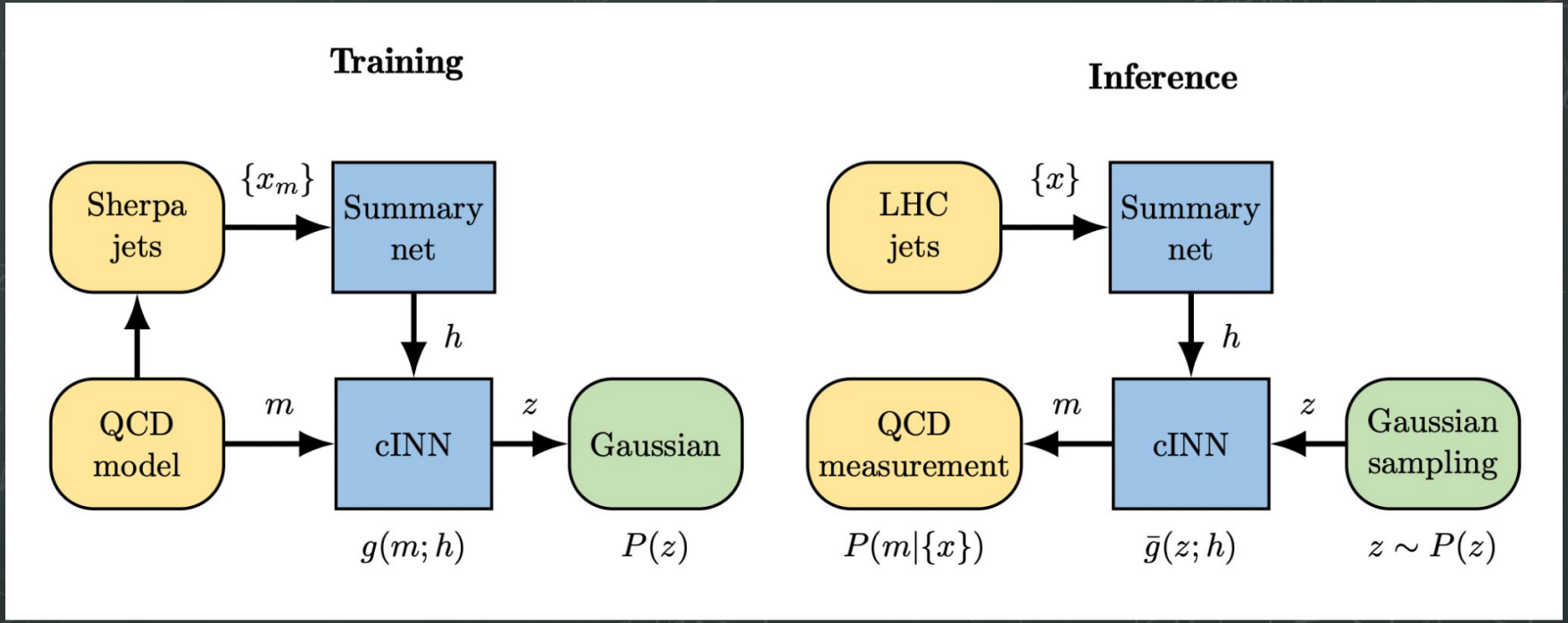








See also 1909.03081, 2002.06307, 2104.04543 (Generative Bayesian NNs), and 2107.08979 (“resampling”)



See also 2006.06685, 2205.01697, 2110.13632

Checklist 4:

- Generative tools change the game.
- Can generative inference replace traditional inference methods?

Summary

- HEP is a wonderful playground for cutting edge and impactful research in data science and ML.
- We need big data to probe fundamental quantum processes.
- The data we have is not perfect. But ML is improving all the time.

Issues

- Care needs to be taken not to lose important physics oversight.
- It’s easy to spend a year making and validating a ML prediction that doubles the speed of a function that takes hours to run.
- Our data is fixed, we’ve already achieved so much more than our detectors were designed for.
- Quantum Doesn’t Scale.