Machine Learning in High-Energy Physics

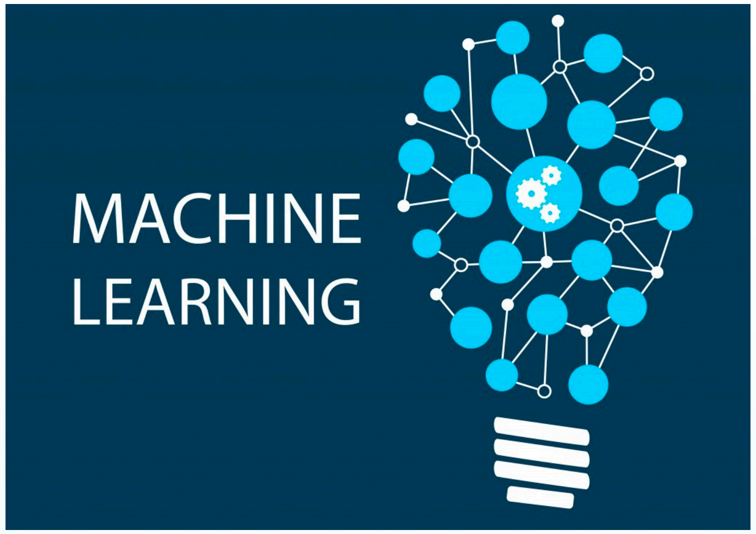
Juan Rojo
VU Amsterdam & Theory Group, Nikhef

Nikhef Topical Lectures on Machine Learning
Nikhef, Amsterdam, 6/6/2018

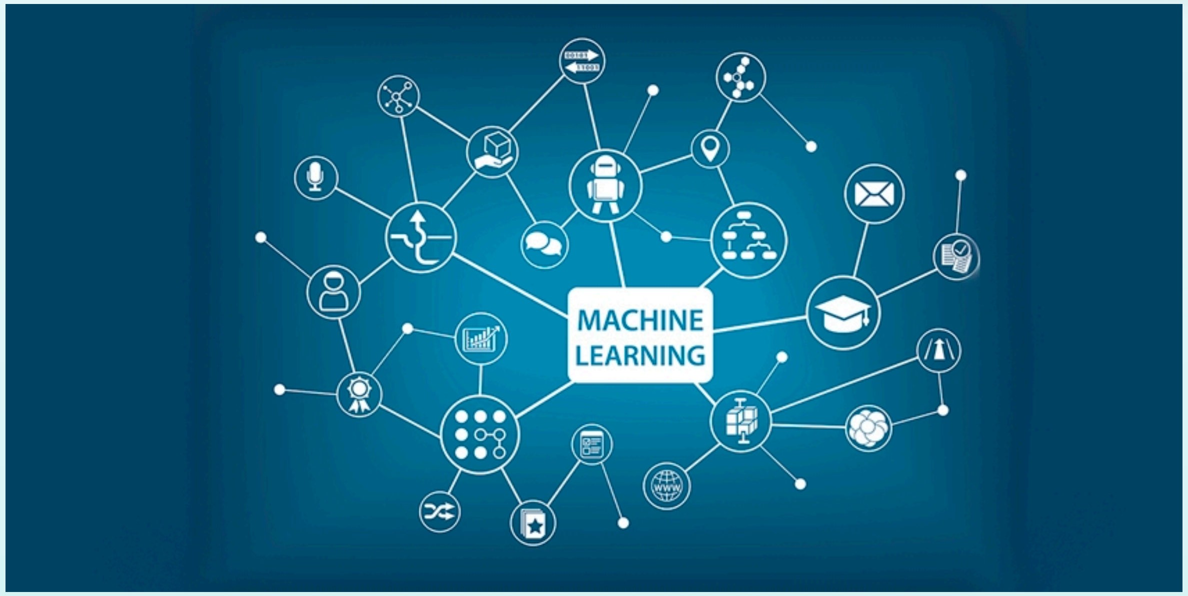
Machine Learning in HEP

- **Huge, fast growing field**, with new applications being proposed every day
- Here restrict ourselves to a **few representative examples**: if you want to learn more about other applications, don't hesitate to ask!
- For further **overviews of ML applications to HEP** and related fields please see *e.g.*:

- ☑ ***Big data tools in Physics and Astronomy*** (Amsterdam, https://indico.cern.ch/event/622093/)
- ☑ ***Machine learning for Phenomenology*** (Durham, https://conference.ippp.dur.ac.uk/event/660/)
- ☑ ***Inter-Experimental LHC Machine Learning WG*** (https://iml.web.cern.ch/)
- ☑ ***Accelerating searches for Dark Matter with Machine Learning*** (https://indico.cern.ch/event/664842/)
- ☑ ***CERN Data Science seminars*** (https://indico.cern.ch/category/9320/)



Why we need ML in HEP?



Outstanding questions in Particle Physics

The Higgs boson

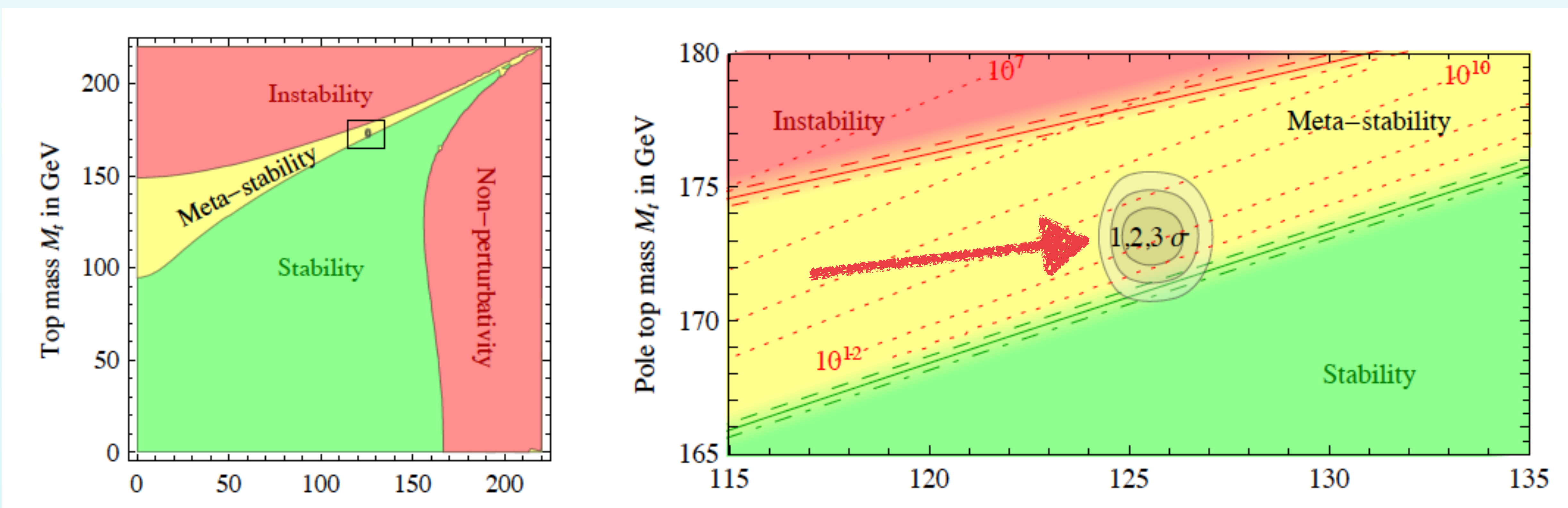
- Huge gap, 10^{17}, between **Higgs and Plank scales**
- **Elementary or composite**? Additional Higgs bosons?
- Coupling to **Dark Matter**? Role in cosmological phase transitions?
- Is the **vacuum state of the Universe** stable?

10^{17} GeV (Planck scale)

With radiative corrections, the *natural* value of the Higgs mass is Planck scale

125 GeV (Higgs mass)

1 GeV (Proton mass)



Degrassi et al 12

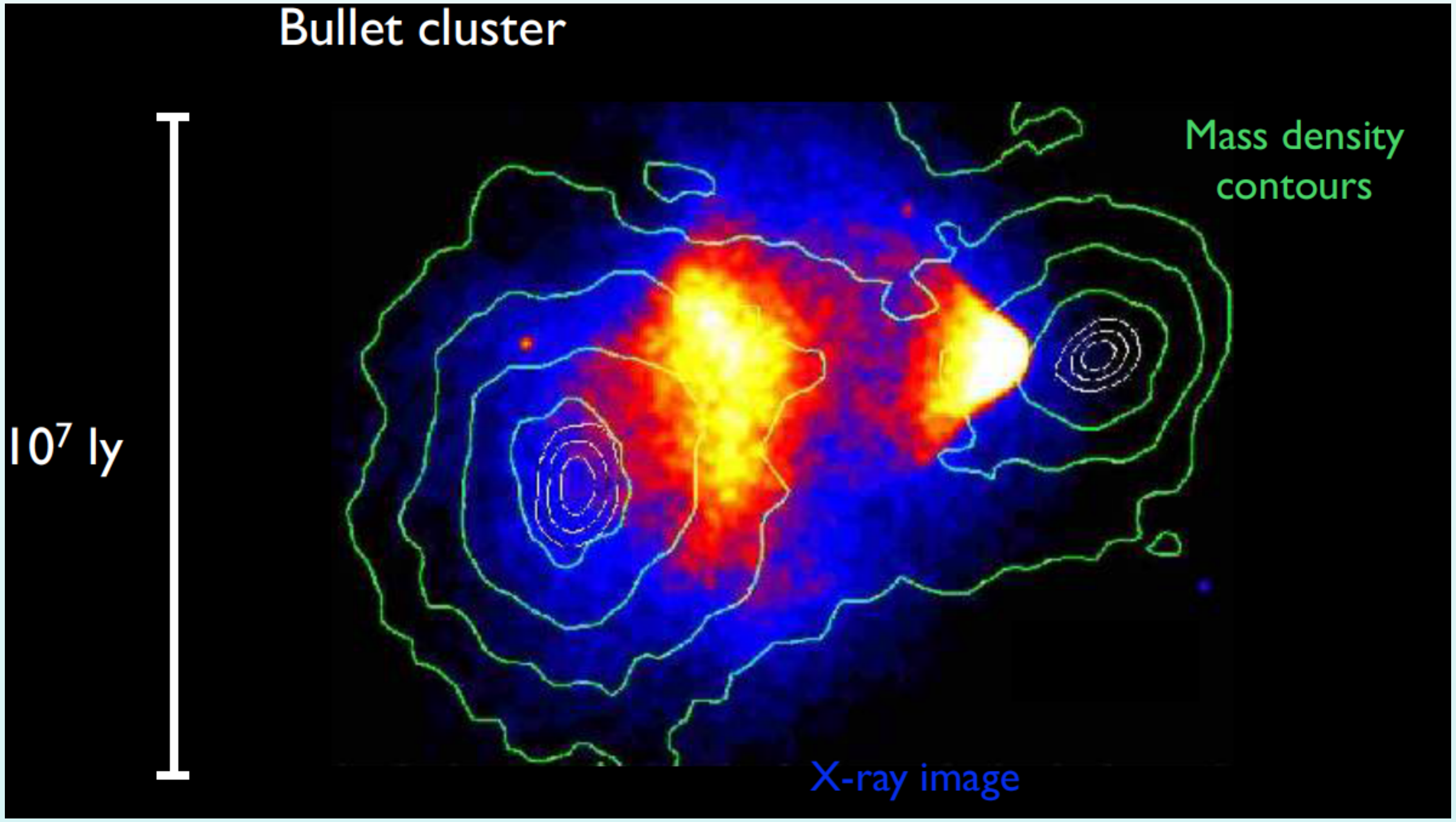
Outstanding questions in Particle Physics

The Higgs boson

- Huge gap, 10^{17}, between **Higgs and Plank scales**
- **Elementary or composite**? Additional Higgs bosons?
- Coupling to **Dark Matter**? Role in cosmological phase transitions?
- Is the **vacuum state of the Universe** stable?

Dark Matter

- **Weakly interacting massive particles**? Sterile neutrinos? **Extremely light particles** (axions)?
- **Interactions** with Standard Model particles?
- What is the **structure of the Dark Sector**? Is Dark Matter self-interacting?



Outstanding questions in Particle Physics

The Higgs boson

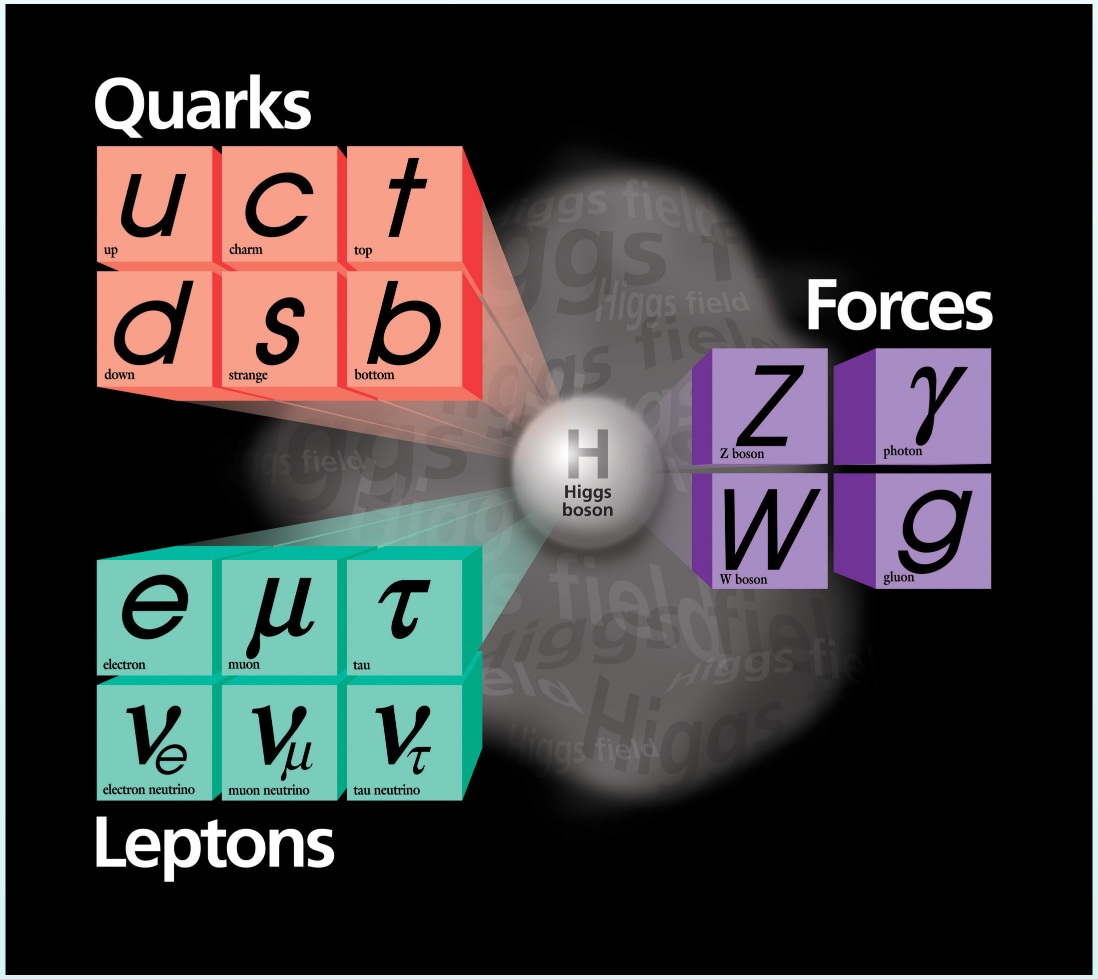
- Huge gap, 10^{17}, between **Higgs and Plank scales**
- **Elementary or composite**? Additional Higgs bosons?
- Coupling to **Dark Matter**? Role in cosmological phase transitions?
- Is the **vacuum state of the Universe** stable?

Quarks and leptons

- **Why three families?** Can we explain **masses and mixings**?
- Origin of **Matter-Antimatter asymmetry** in the Universe?
- Are **neutrinos Majorana or Dirac**? CP violation in the lepton sector?

Dark Matter

- **Weakly interacting massive particles**? Sterile neutrinos? **Extremely light particles** (axions)?
- **Interactions** with Standard Model particles?
- What is the **structure of the Dark Sector**? Is Dark Matter self-interacting?



Outstanding questions in Particle Physics

Huge ga

Elemen
bosons?

Couplin
phase tra

Is the **va**

Why t
and mi

Origin
the Uni

Are r
violatio

- The LHC will provide **crucial inputs to these open puzzles** …

- …however we may need to search for **subtle signals** (*e.g.* deviation with respect SM) in the **very messy environment of hadron collisions**

- Requires not only state-of-the-art theory calculations but also exploiting **recent developments in Data Science and Machine Learning** techniques

- The LHC is an amazing machine: let's make sure we extract **the best possible physics output from it**!!

es? Sterile
(axions)?

articles?

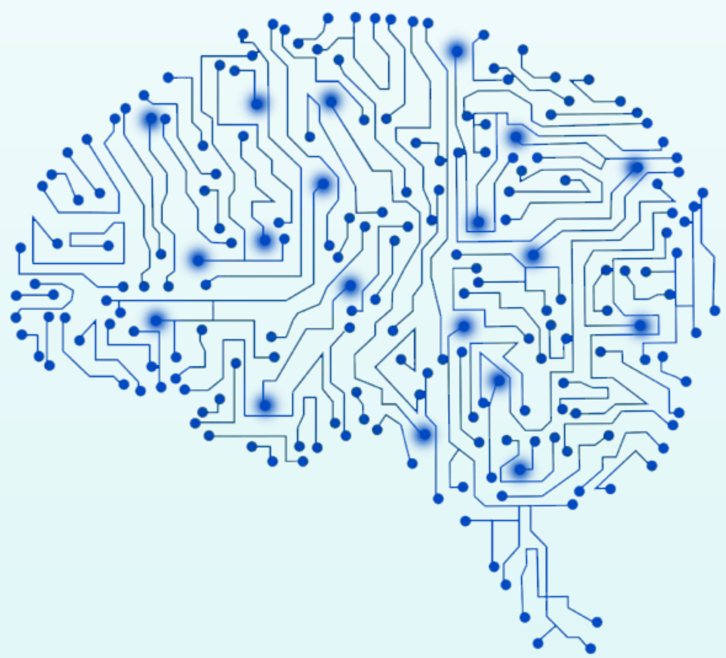
Sector? Is

es

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions



QCD-aware NNs for jet physics

Automated bSM exclusion limits

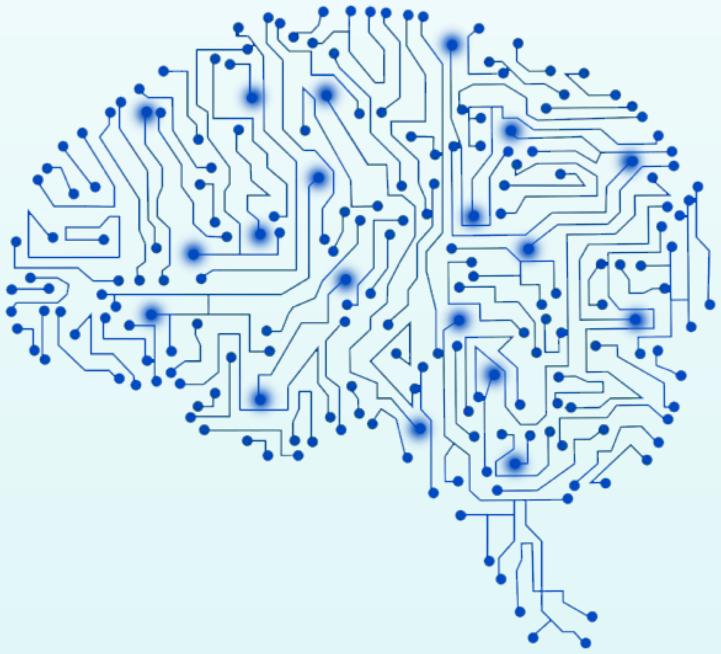
Boosting bSM searches

HEP detector simulation

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

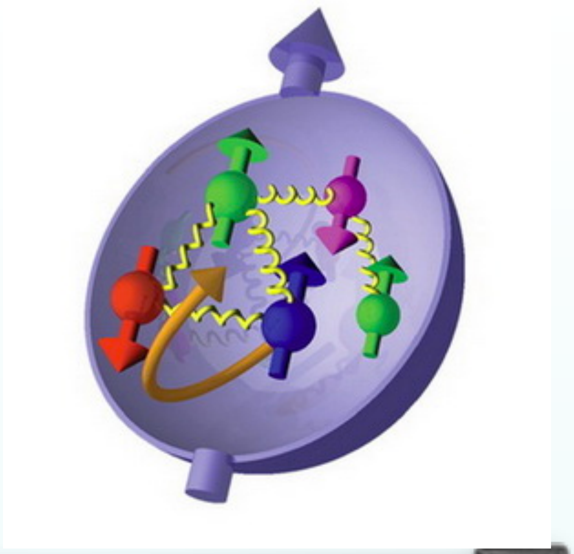


QCD-aware NNs For jet physics

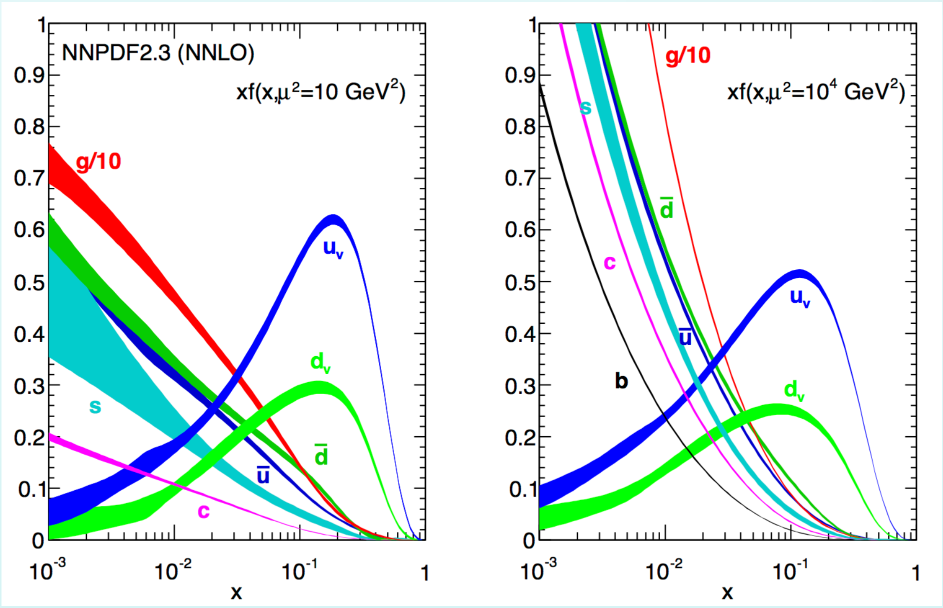
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation



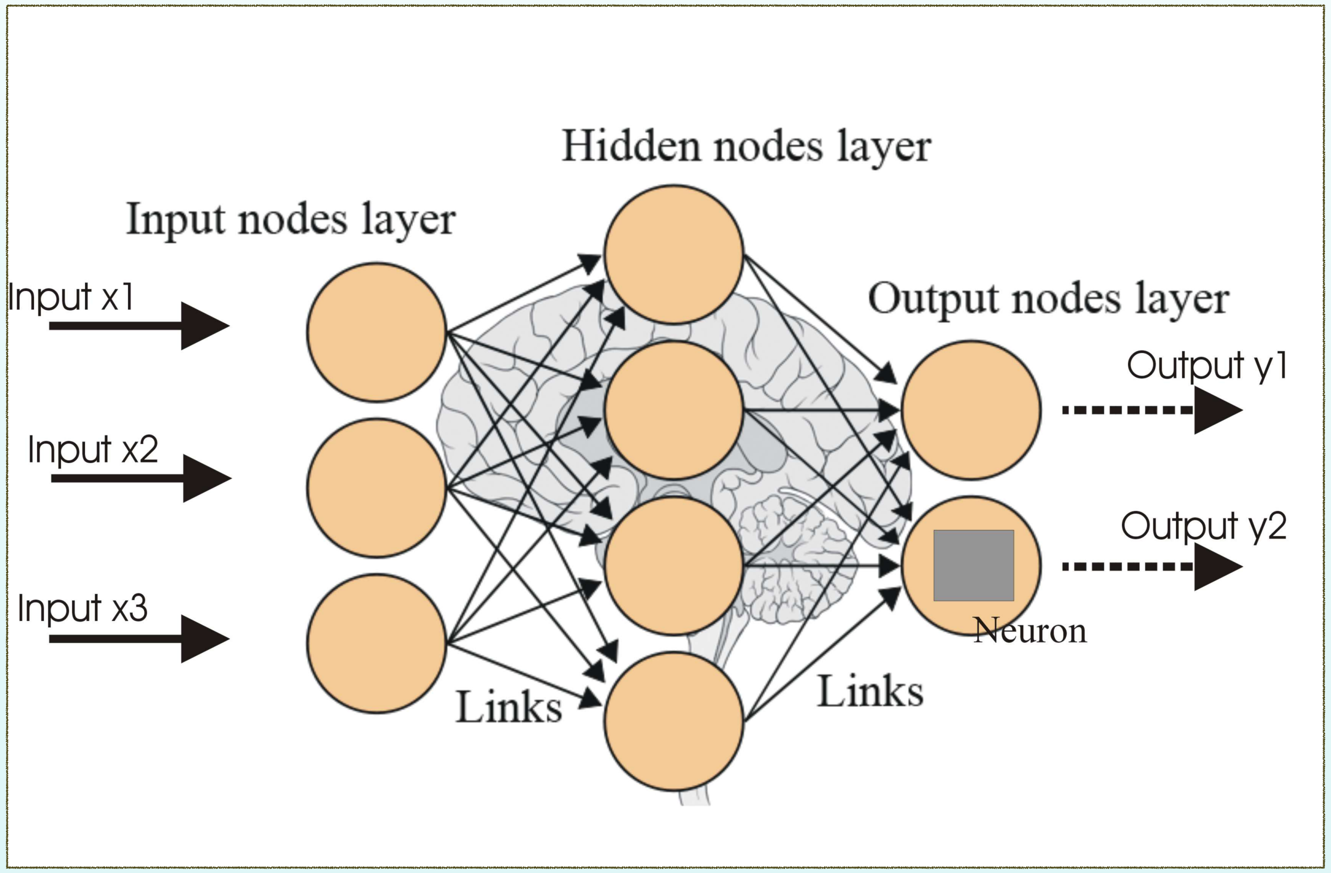
The inner life of protons with artificial neural networks



Artificial Neural Networks

Inspired by **biological brain models**, **Artificial Neural Networks** (ANNs) are **mathematical algorithms** widely used in a wide range of applications, from **HEP** to **targeted marketing** and **finance forecasting**

From biological to artificial neural networks



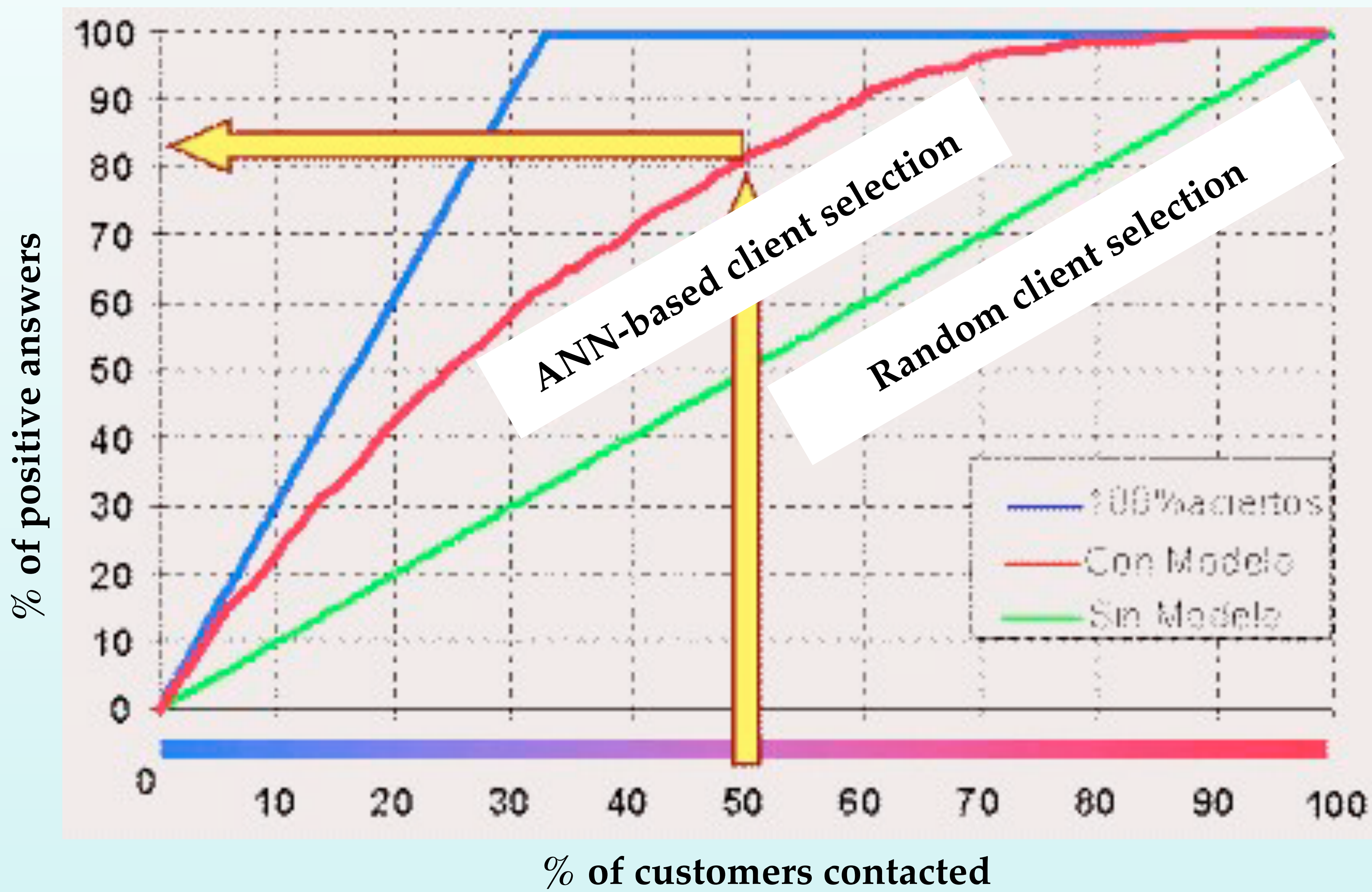
Artificial neural networks aim to excel where domains as their **evolution-driven counterparts outperforms traditional algorithms in tasks such as pattern recognition, forecasting, classification**, …

ANNs - a marketing example

A bank wants to offer a new credit card to their clients. Two possible strategies:

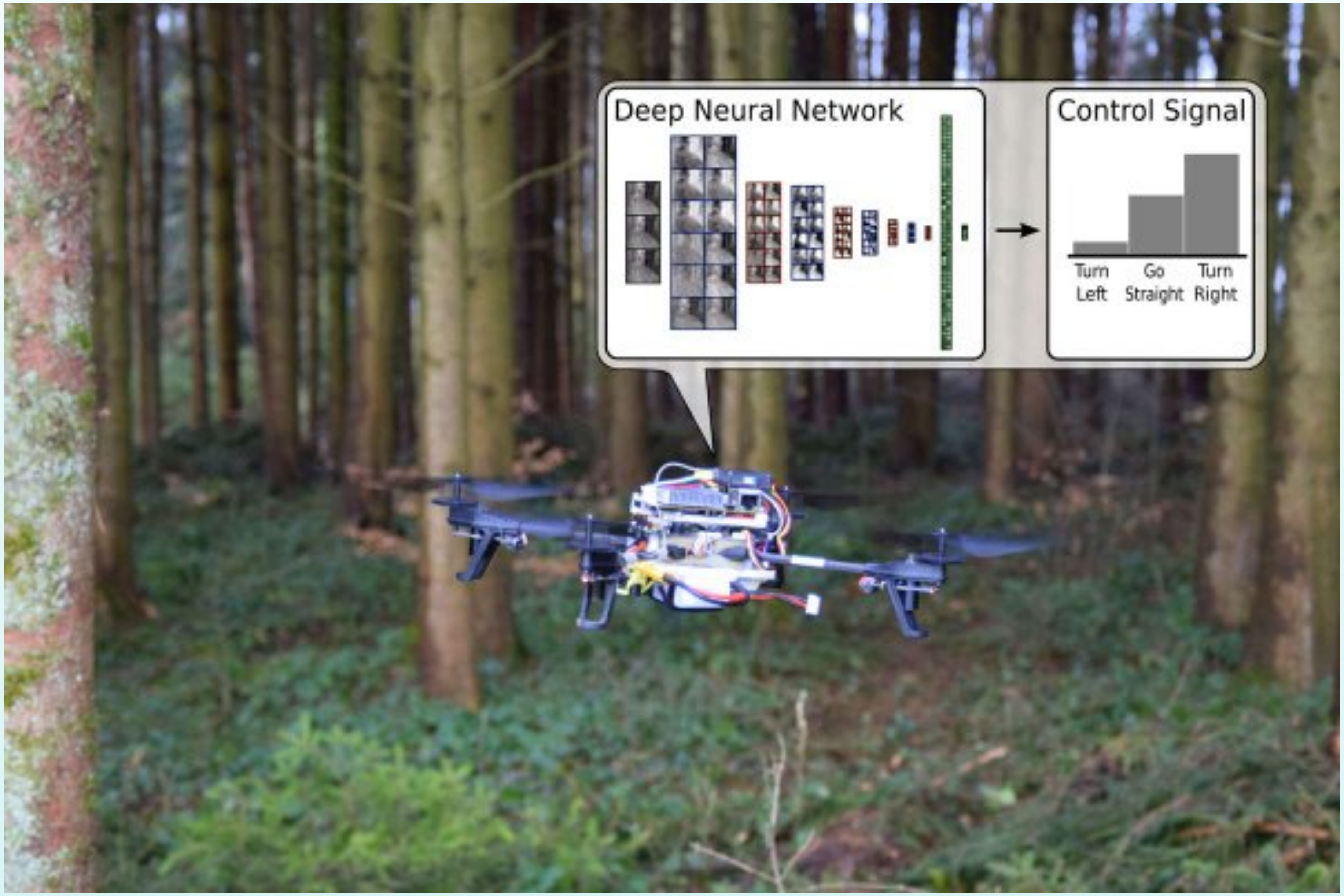
- **Contact all customers**: slow and costly
- Contact **5%** of the customers, **train a ANN with their input** (gender, income, loans) and **their output** (yes/no) and use the information to c**ontact only clients likely to accept the product**

Cost-effective method to improve marketing performance!



ANNs and pattern recognition

- ANNs can enable an **autonomous vision-control drone** to recognise and follow forest trails
- Image classifier operates directly on **pixel-level image intensities**
- If a trail is visible, the **software steers the drone** in the corresponding direction

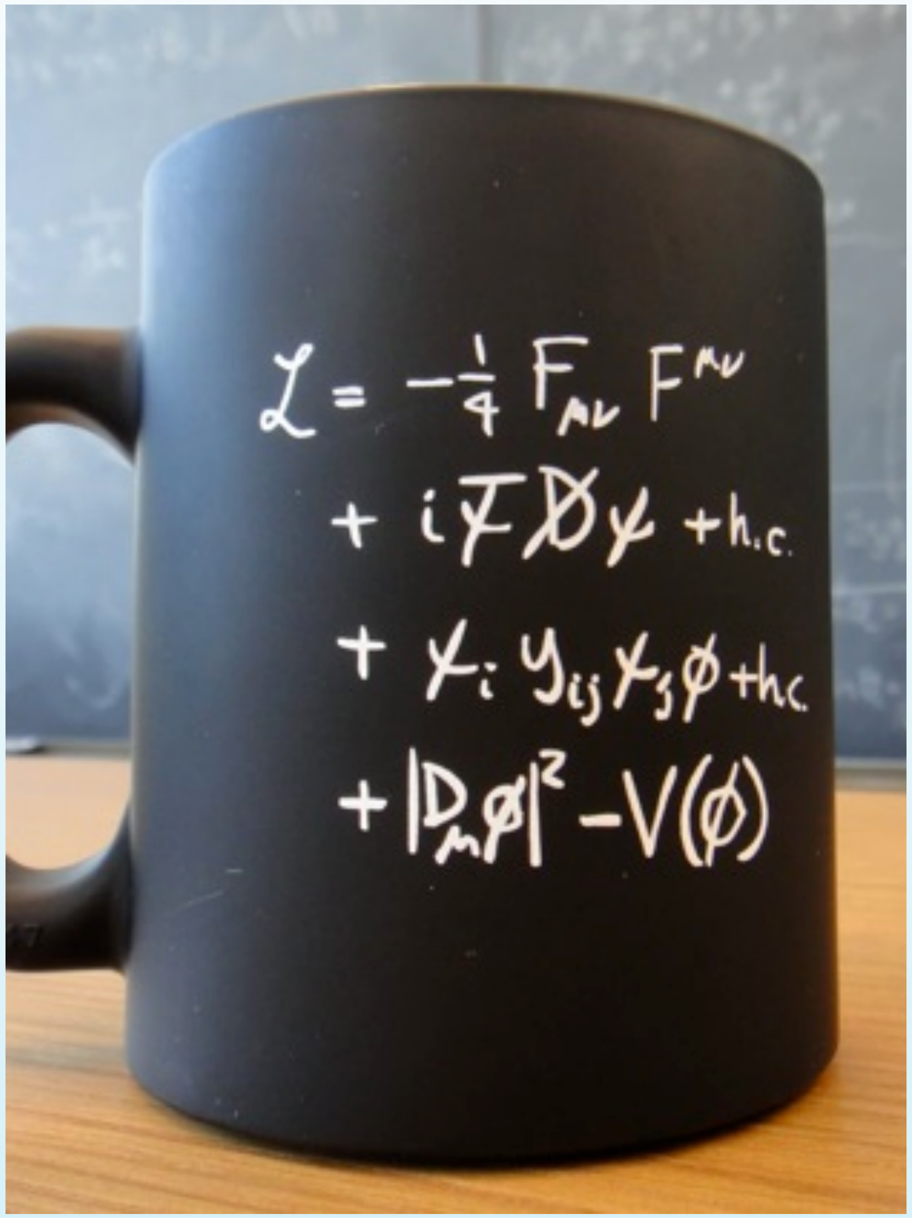


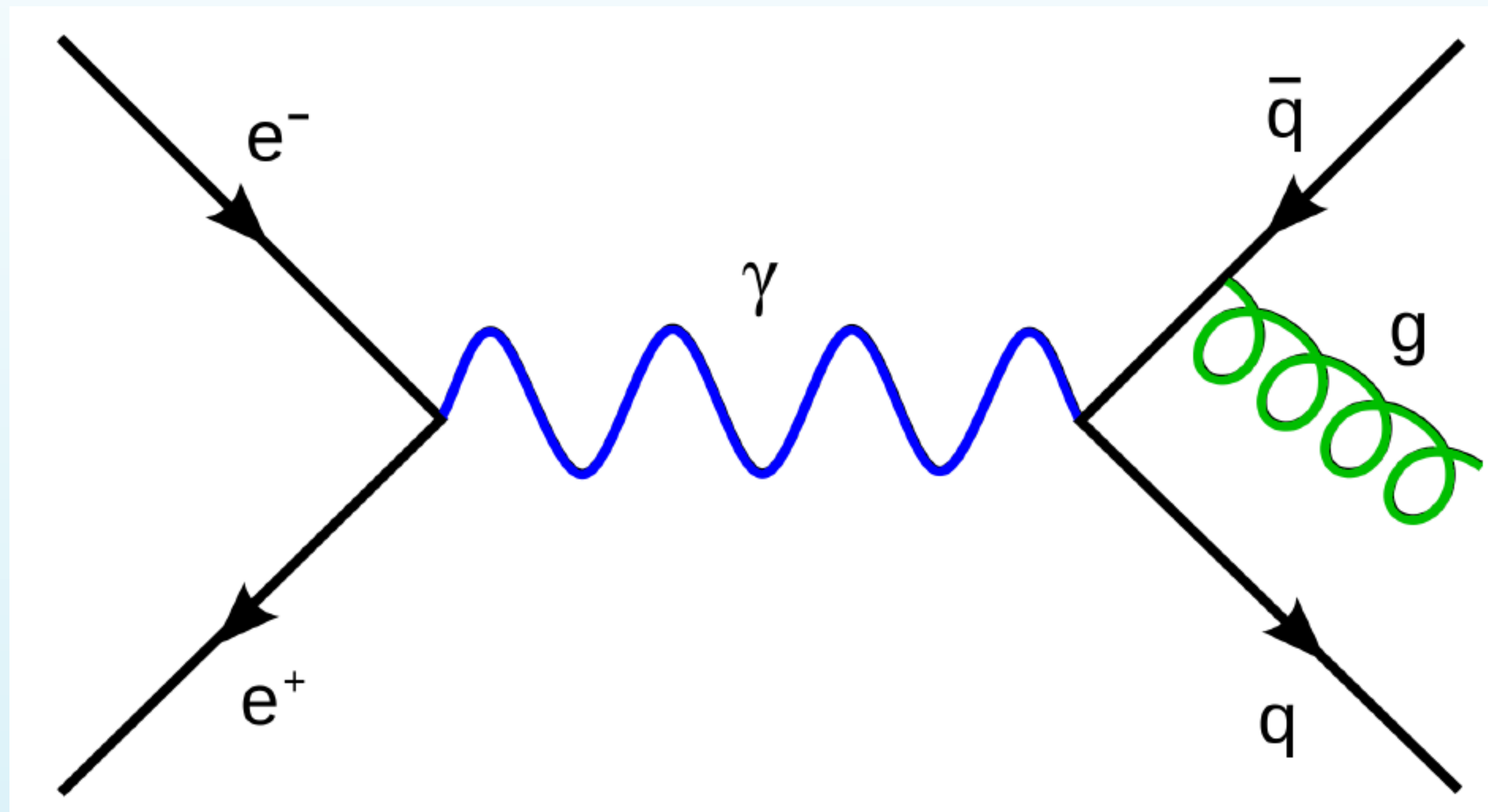
Giusti et al, IEEE Robotics and Automation Letters, 2016

Similar algorithms at work in self-driving cars!

Lepton vs Hadron Colliders

In high-energy **lepton colliders**, such as the **Large Electron-Positron Collider** (LEP) at CERN, the collisions involve **elementary particles** without substructure

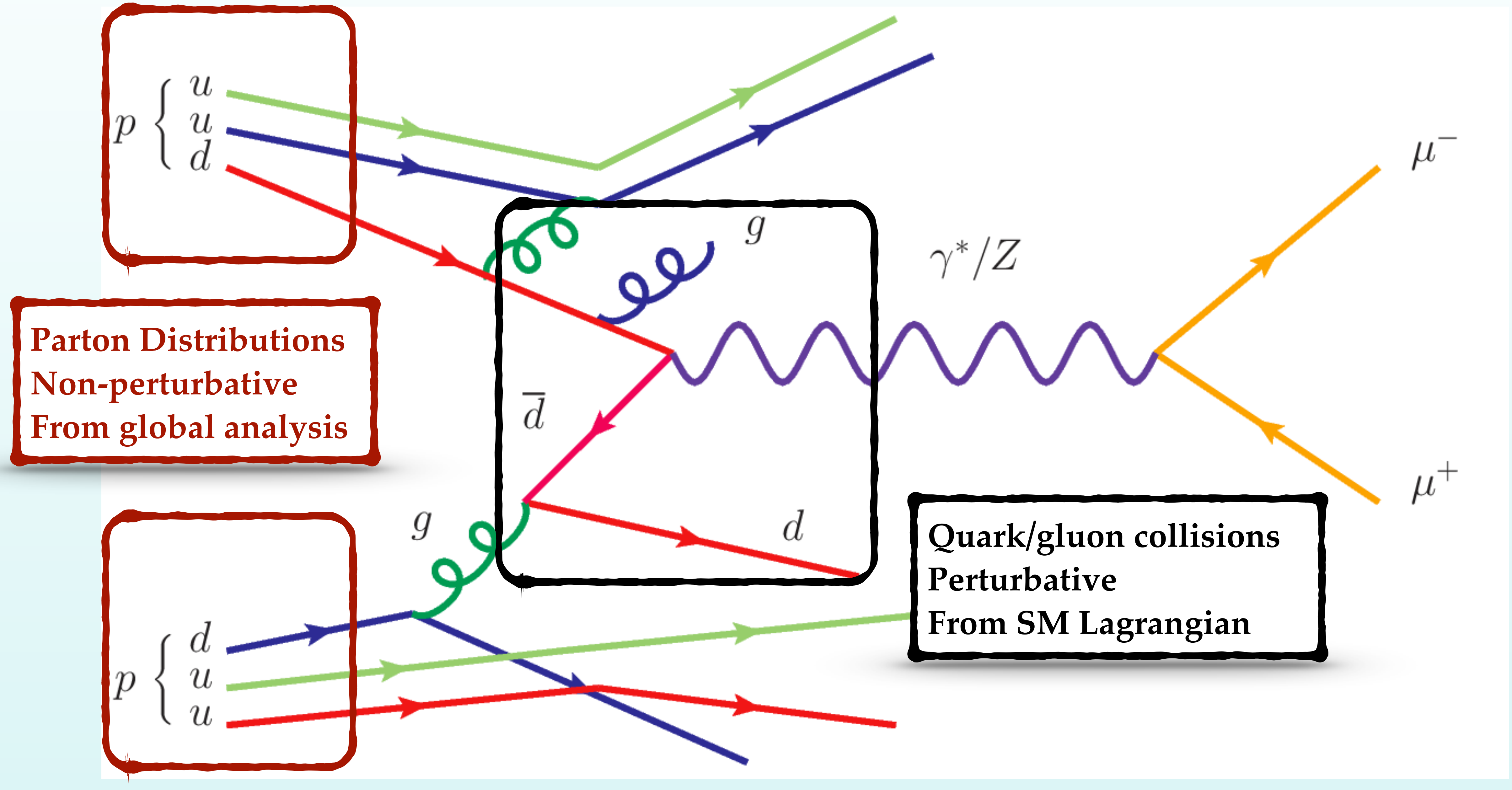




Cross-sections in lepton colliders can be computed in perturbation theory using the **Feynman rules of the Standard Model Lagrangian**

Anatomy of a proton-proton collision

In high-energy **hadron colliders** the collisions involve **composite particles** (protons) with internal substructure (quarks and gluons): the LHC is actually a quark/gluon collider!



Calculations of **cross-sections** in hadron collisions require the combination of **perturbative cross-sections** with **non-perturbative parton distribution functions (PDFs)**

Anatomy of hadronic collisions

In high-energy **hadron colliders** the collisions involve **composite particles** (protons) with internal substructure (quarks and gluons): the LHC is actually a quark/gluon collider!

DGLAP evolution kernel (perturbative)

$$\sigma^{(\mathrm{th})}\left(\{a_i\}\right) = \widehat{\sigma}_{ij}(Q^2) \otimes \Gamma_{ij,kl}(Q^2, Q_0^2) \otimes q_k\left(Q_0, \{a_i\}\right) \otimes q_l\left(Q_0, \{a_i\}\right)$$

Hard-scattering partonic cross-sections (perturbative)

PDFs at input parametrisation scale (non-perturbative)

Calculations of **cross-sections** in hadron collisions require the combination of **perturbative cross-sections** with **non-perturbative parton distribution functions (PDFs)**

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x, Q)$$

Q: **Energy of the quark/gluon collision**
Inverse of the resolution length

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction *x* of the proton momentum, when probed with energy *Q*

x: **Fraction of the proton's momentum**

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x, Q)$$

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction *x* of the proton momentum, when probed with energy *Q*

Q: **Energy of the quark/gluon collision**
Inverse of the resolution length

x: **Fraction of the proton's momentum**

PDFs are determined by **non-perturbative QCD dynamics:** cannot be computed from first principles, and need to be **extracted from experimental data** with a **global analysis**

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x, Q)$$

Q: **Energy of the quark/gluon collision**
Inverse of the resolution length

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction *x* of the proton momentum, when probed with energy *Q*

x: Fraction of the proton's momentum

PDFs are determined by **non-perturbative QCD dynamics:** cannot be computed from first principles, and need to be **extracted from experimental data** with a **global analysis**

- **Energy conservation** $$\int_0^1 dx \left(g(x,Q) + \sum_q q(x,Q) \right) = 1$$

- **Dependence with quark/gluon collision energy Q** determined in perturbation theory

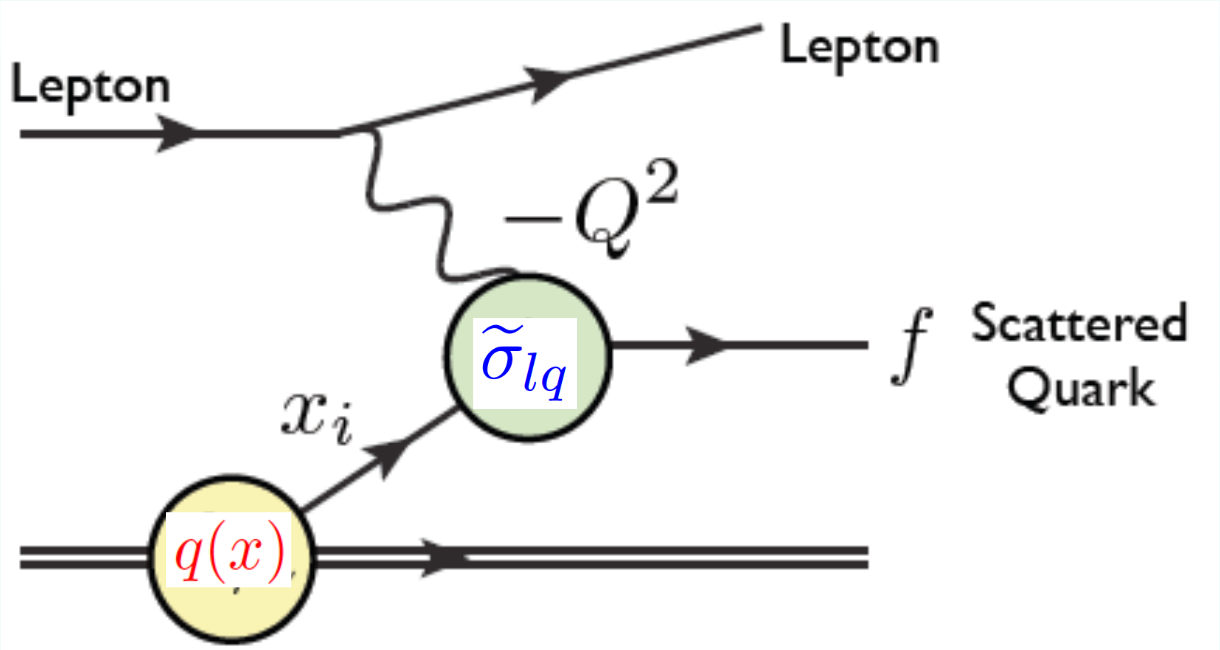
$$\frac{\partial g(x,Q)}{\partial \ln Q} = P_g\left(\alpha_s\right) \otimes g(x,Q) + P_q\left(\alpha_s\right) \otimes q(x,Q)$$

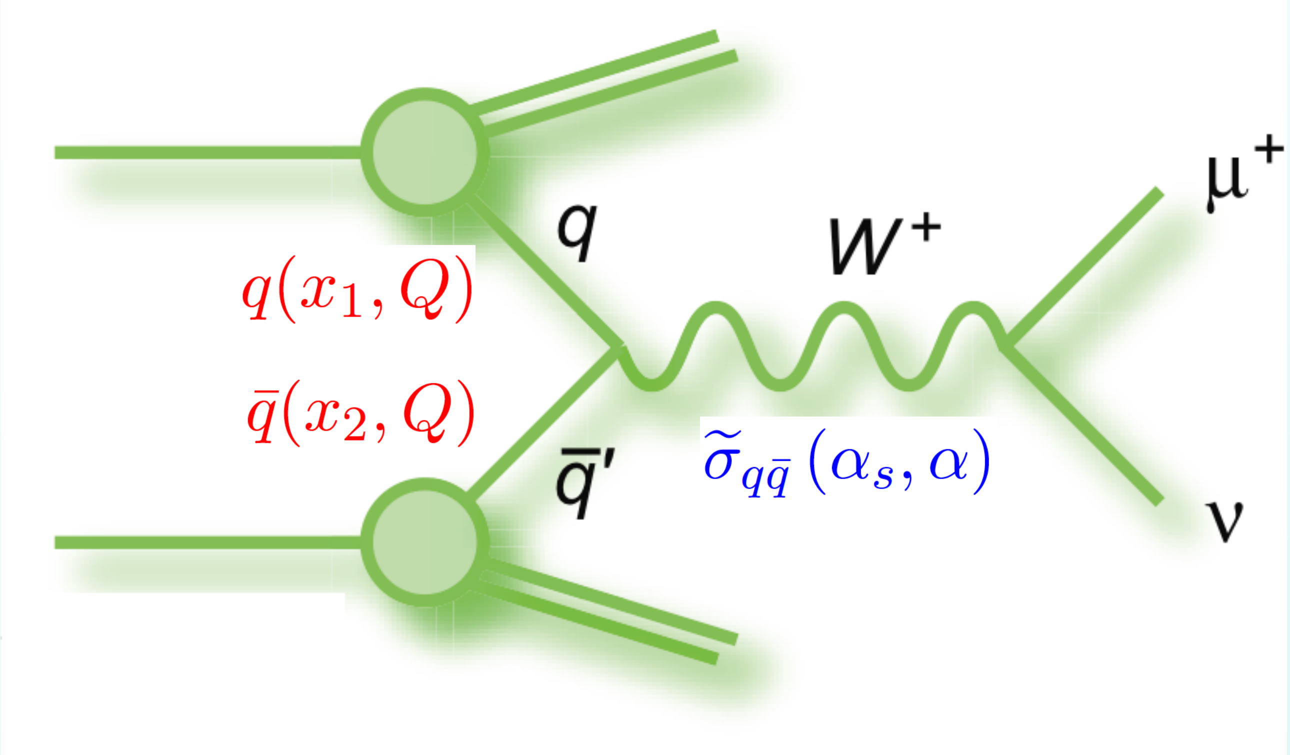
The Factorization Theorem

The **QCD Factorization Theorem** guarantees **PDF universality: extract them from a subset of** process and use them to provide pure predictions for new processes

$$\sigma_{lp} \simeq \widetilde{\sigma}_{lq}\left(\alpha_s, \alpha\right) \otimes q(x, Q)$$

$$\sigma_{pp} \simeq \widetilde{\sigma}_{q\bar{q}}\left(\alpha_s, \alpha\right) \otimes q(x_1, Q) \otimes \bar{q}(x_2, Q)$$





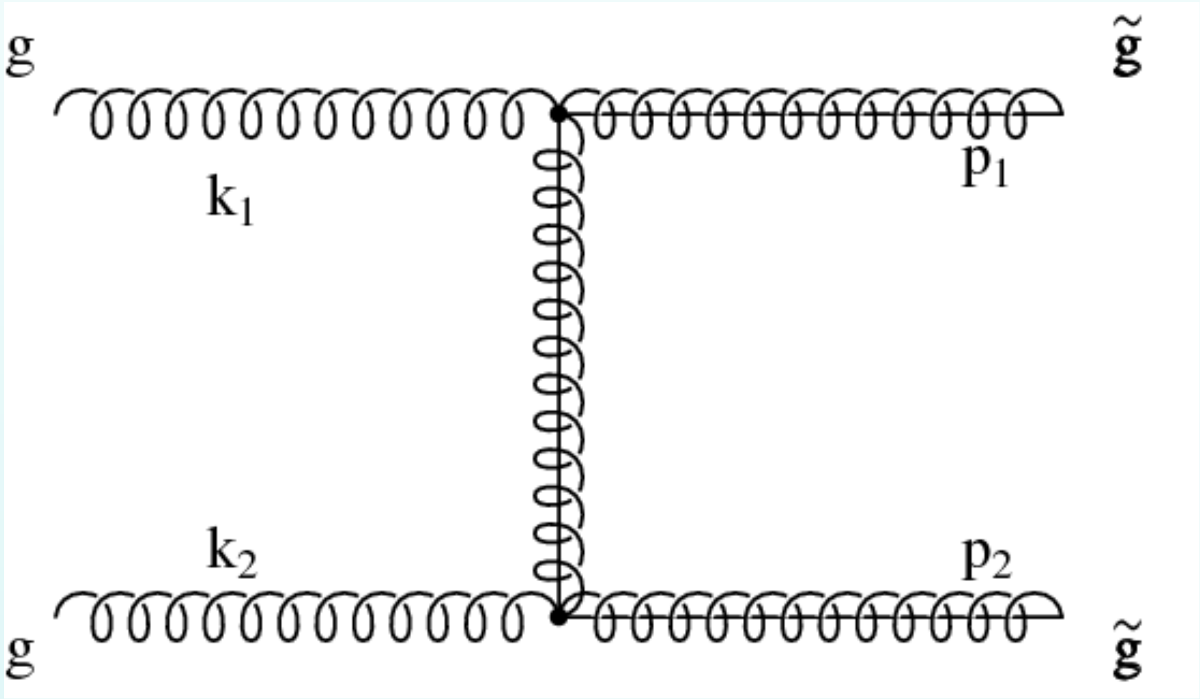
Determine PDFs in lepton-proton collisions ....

And use them to compute cross-sections in proton-proton collisions at the LHC

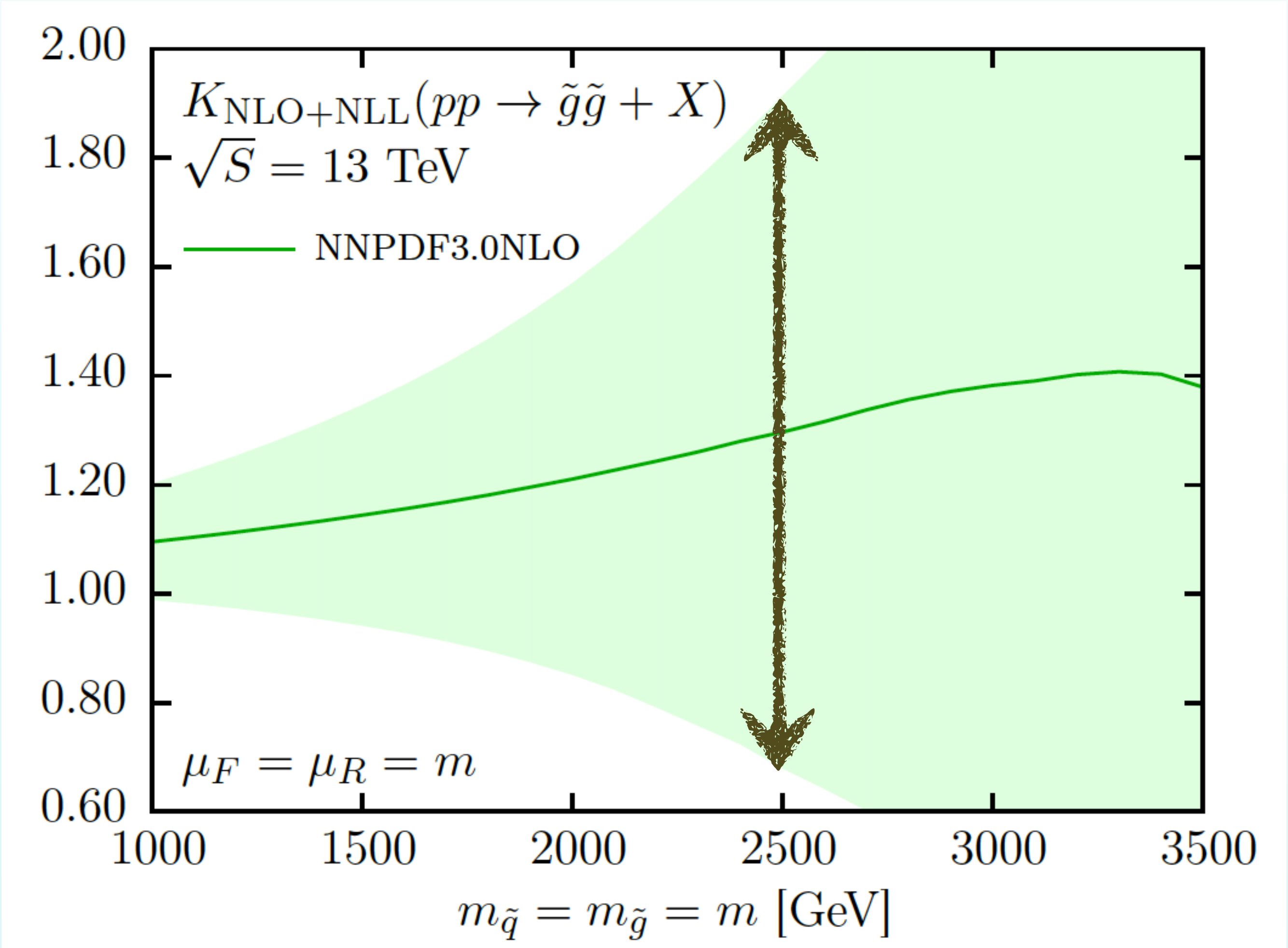
Beyond BSM discovery

PDF uncertainties in the production of **New Physics heavy resonances** can be al large as **100%!**

Crucial *i.e.* in searches for ***supersymmetry*** and any BSM scenario that predicts new heavy particles within the reach of the LHC



Gluino pair production at the LHC



Beenakker, Borchensky, Kramer, Kulesza, Laenen, Marzani, Rojo 15

*Unless we **improve PDF uncertainties**, even if we discover New Physics, it will be extremely **difficult to characterise the underlying BSM scenario***

ANNs as universal unbiased interpolants

ANNs provide **universal unbiased interpolants** to parametrize the non-perturbative dynamics that determines the **size and shape of the PDFs** from experimental data

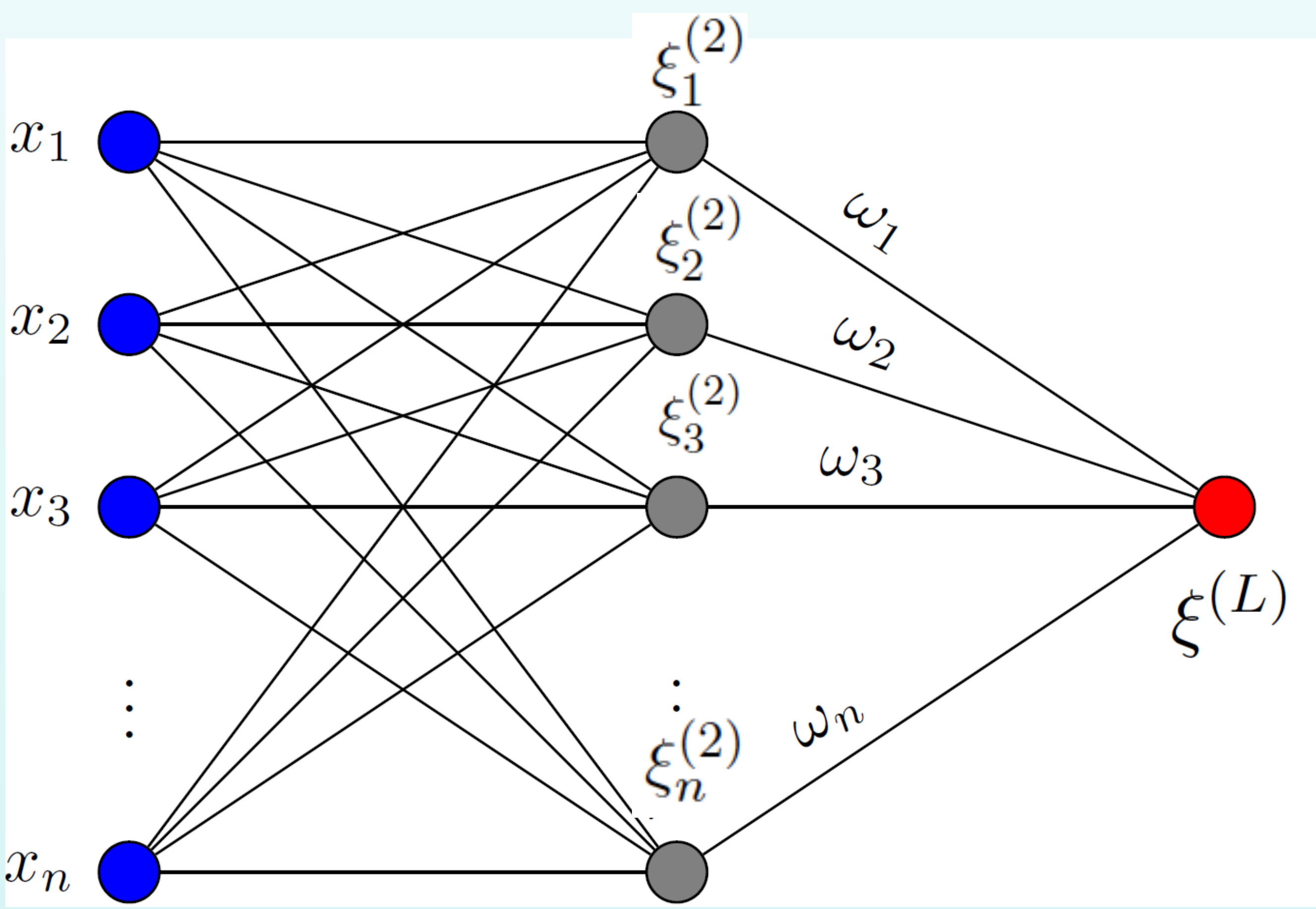
not from QCD!

Traditional approach

$$g(x,Q_0) = A_g(1-x)^{a_g}x^{-b_g}\left(1 + c_g\sqrt{s} + d_g x + \ldots\right)$$

NNPDF approach

$$g(x,Q_0) = A_g \mathrm{ANN}_g(x)$$



$$\mathrm{ANN}_g(x) = \xi^{(L)} = \mathcal{F}\left[\xi^{(1)}, \{\omega_{ij}^{(l)}\}, \{\theta_i^{(l)}\}\right]$$

$$\xi_i^{(l)} = g\left(\sum_{j=1}^{n_{l-1}} \omega_{ij}^{(l-1)}\xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

- ANNs eliminate **theory bias** introduced in PDF fits from choice of *ad-hoc* functional forms
- NNPDF fits used **O(400) free parameters**, to be compared with O(10-20) in traditional PDFs. Results stable if **O(4000) parameters used!**

Machine Learning for PDF fits

NNPDF

Experimental data

Fixed-target & collider DIS
Tevatron and LHC measurements
Jets, DY, top, Z pT, ...

Statistical framework

PDF parametrisation,
PDF uncertainties and propagation
Model and theory uncertainties

Theory calculations

The global QCD fit

Minimise figure of merit (*) and
determine PDF parameters

fit validation, statistical estimators, diagnosis tools

APFEL WEB

http://apfel.mi.infn.it/

on-line plotting toolbox

LHAPDF

lhapdf.hepforge.org

standard interface for public PDF delivery

(*)

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right)\left({\rm cov}_{\rm exp} + {\rm cov}_{\rm th}\right)^{-1}_{mn}\left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$

APPLgrid, FastNLO, aMCfast....

NNLO DGLAP evolution
DIS structure functions

APFEL, HOPPET, QCDNUM, ...

Fast NLO grids
NNLO QCD &
NLO EW K-factors

External (N)NLO codes

MCFM, NLOjet++, FEWZ,
DYNNLO, private codes...

Machine Learning for PDF fits

Experimental data

Fixed-target & collider DIS
Tevatron and LHC measurements
Jets, DY, top, Z pT, ...

Statistical framework

PDF parametrisation,
PDF uncertainties and propagation
Model and theory uncertainties

Theory calculations

NNPDF

Artificial Neural Networks for the PDF parametrisation

fit validation, statistical estimators, diagnosis tools

APFEL WEB

http://apfel.mi.infn.it/

on-line plotting toolbox

LHAPDF

lhapdf.hepforge.org

standard interface for public PDF delivery

(*)

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right)\left(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th}\right)^{-1}_{mn}\left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$

APPLgrid, FastNLO, aMCfast....

NNLO DGLAP evolution
DIS structure functions

APFEL, HOPPET, QCDNUM, ...

Fast NLO grids
NNLO QCD &
NLO EW K-factors

External (N)NLO codes

MCFM, NLOjet++, FEWZ, DYNNLO, private codes...

Machine Learning for PDF fits

Experimental data

Fixed-target & collider DIS
Tevatron and LHC measurements
Jets, DY, top, Z pT, ...

Monte Carlo method for robust error estimate and propagation

fit validation, statistical estimators, diagnosis tools

Statistical framework

PDF parametrisation,
PDF uncertainties and propagation
Model and theory uncertainties

The global QCD fit

Minimise figure of merit (*) and determine PDF parameters

APFEL WEB

http://apfel.mi.infn.it/

on-line plotting toolbox

LHAPDF

lhapdf.hepforge.org

standard interface for public PDF delivery

Theory calculations

(*)

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right)\left(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th}\right)^{-1}_{mn}\left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$

APPLgrid, FastNLO, aMCfast....

NNLO DGLAP evolution
DIS structure functions

APFEL, HOPPET, QCDNUM, ...

Fast NLO grids
NNLO QCD &
NLO EW K-factors

External (N)NLO codes

MCFM, NLOjet++, FEWZ,
DYNNLO, private codes...

Machine Learning for PDF fits

Experimental data

Fixed-target & collider DIS
Tevatron and LHC measurements
Jets, DY, top, Z pT, ...

Minimisers/Optimisers: Genetic Algorithms, Stochastic Gradient Descent, Covariance Matrix Adaptation ...

fit validation, statistical estimators, diagnosis tools

Statistical framework

PDF parametrisation,
PDF uncertainties and propagation
Model and theory uncertainties

The global QCD fit

Minimise figure of merit (*) and determine PDF parameters

APFEL WEB

http://apfel.mi.infn.it/

on-line plotting toolbox

LHAPDF

lhapdf.hepforge.org

standard interface for public PDF delivery

Theory calculations

(*)

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right) \left(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th}\right)^{-1}_{mn} \left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$

APPLgrid, FastNLO, aMCfast....

NNLO DGLAP evolution
DIS structure functions

APFEL, HOPPET, QCDNUM, ...

Fast NLO grids
NNLO QCD &
NLO EW K-factors

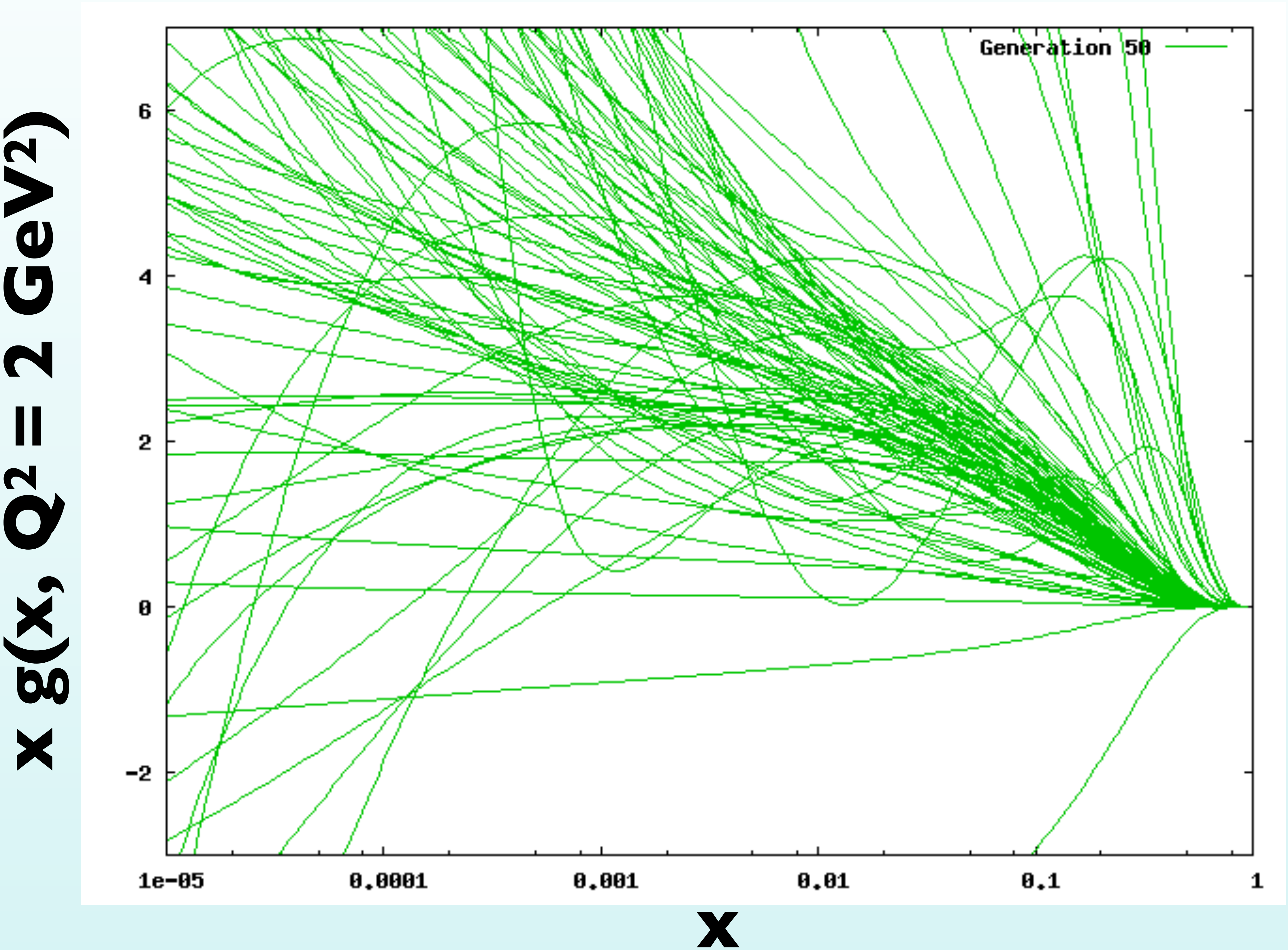
External (N)NLO codes

MCFM, NLOjet++, FEWZ,
DYNNLO, private codes...

PDF Replica Neural Network Learning

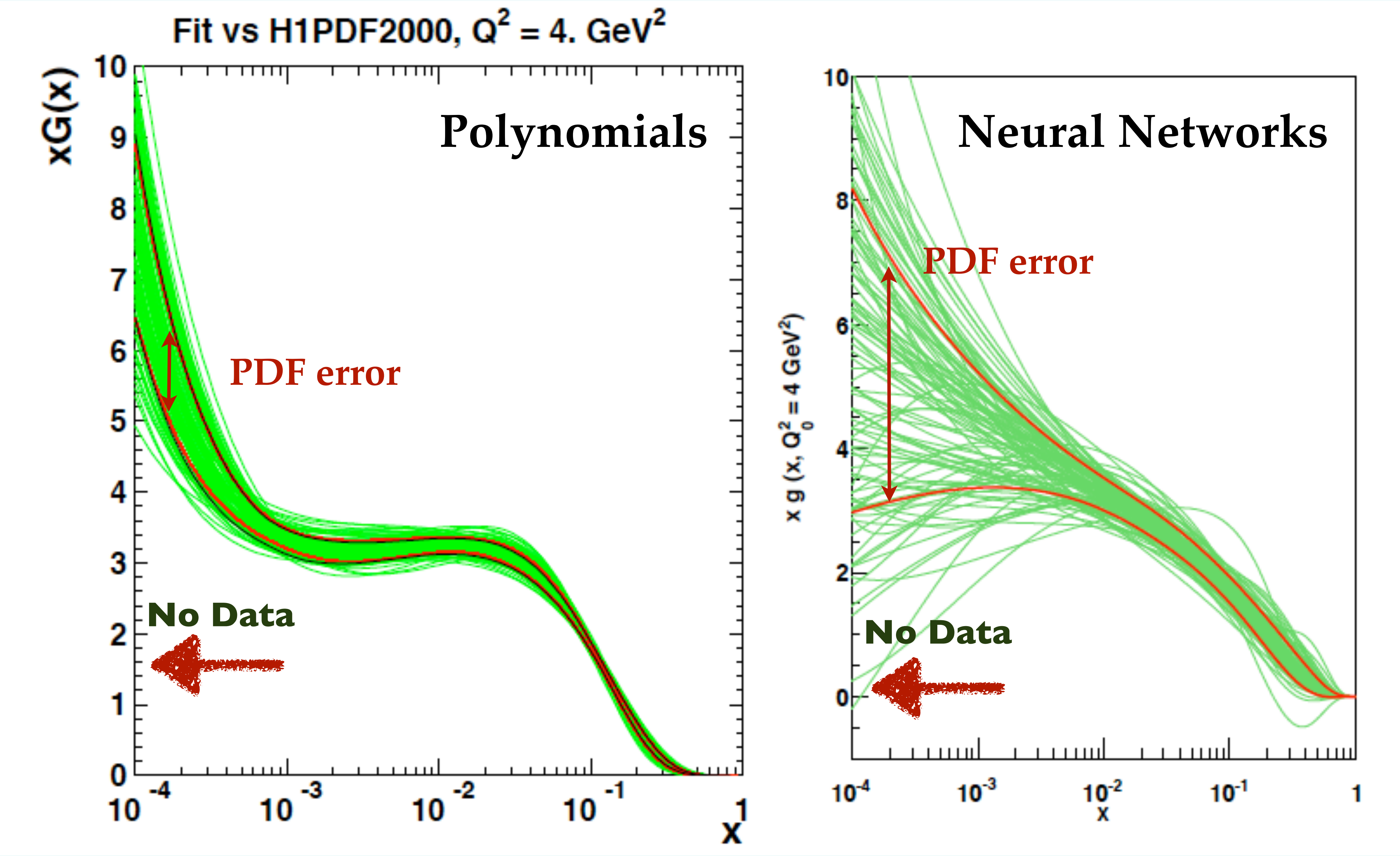
The minimisation of the **data vs theory χ^2** is performed using **Genetic Algorithms**

Each **green curve** corresponds to a **gluon PDF Monte Carlo** replica



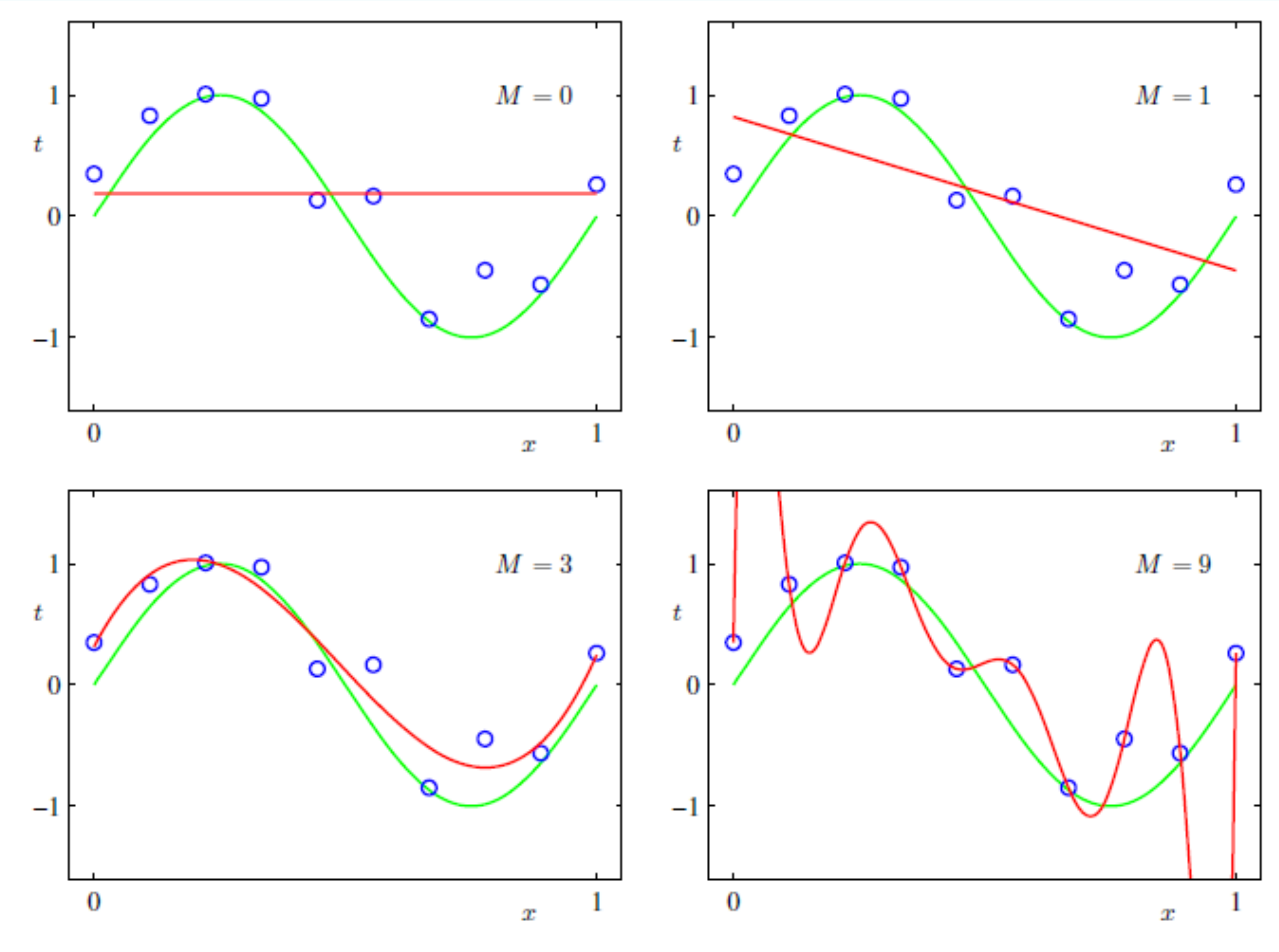
Artificial Neural Networks vs Polynomials

- Compare a **benchmark PDF analysis** where **the same dataset** is fitted with **Artificial Neural Networks** and with **standard polynomials,** other settings identical)
- **ANNs** avoid biasing the PDFs, **faithful extrapolation at small-x** (very few data, thus error blow up)



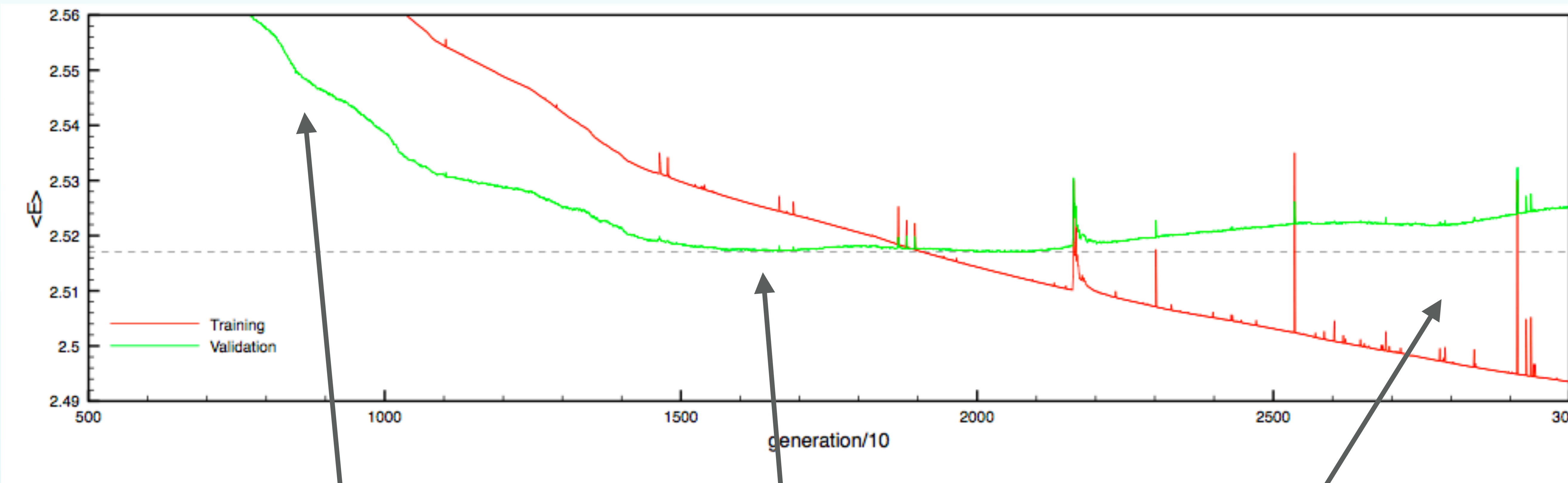
Avoiding overfitting

For a **flexible enough input functional form for the PDF**, one might end up **fitting statistical fluctuations** rather than the underlying physical law!



Avoiding overfitting

- Separate the input measurements into a **training** and a **validation** sample
- The validation sample is never trained, only used to monitor the quality of the fit to the training sample
- The optimal stopping point is at the **global minimum of the validation χ^2**



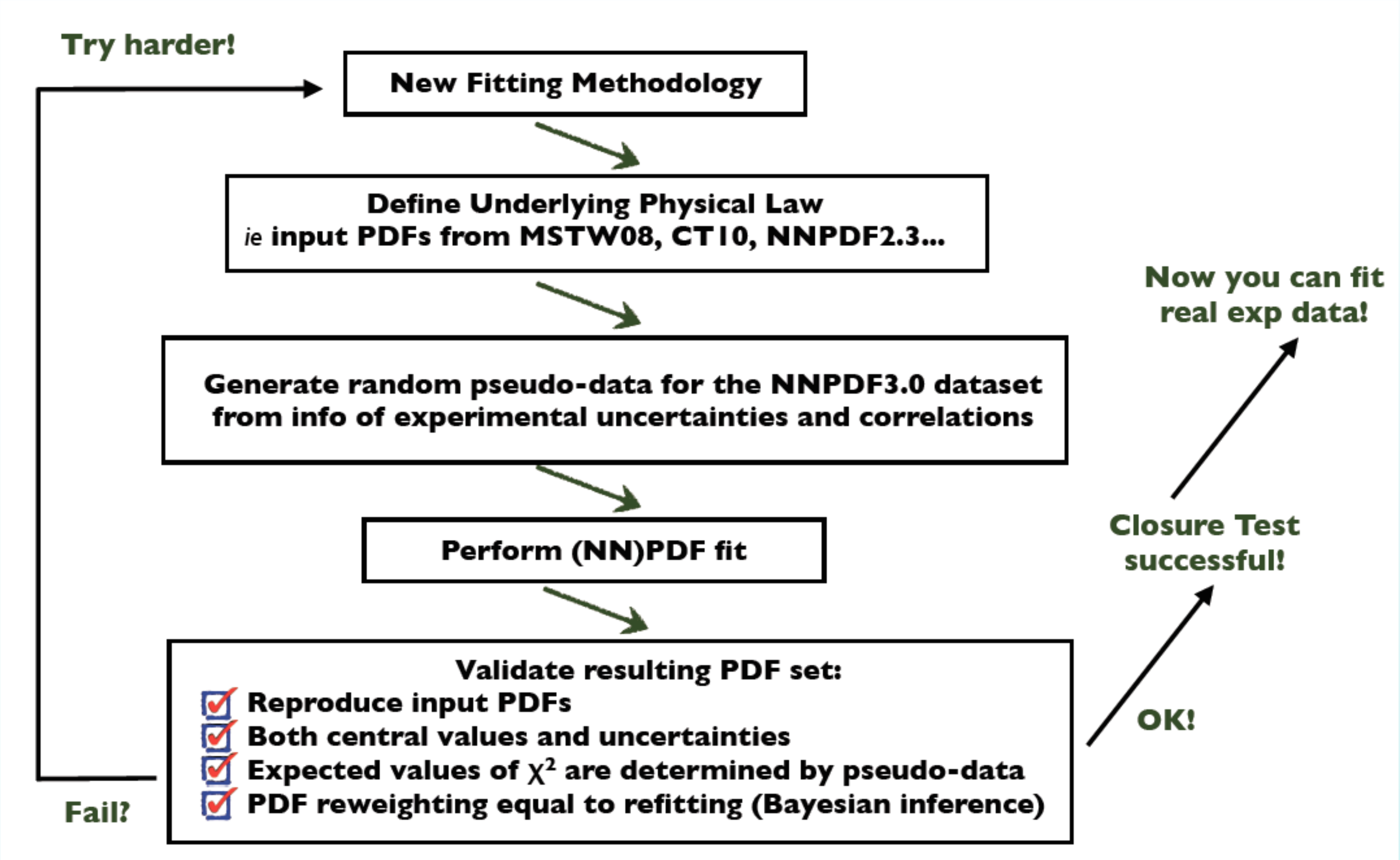
Undertrained

Optimal stopping point

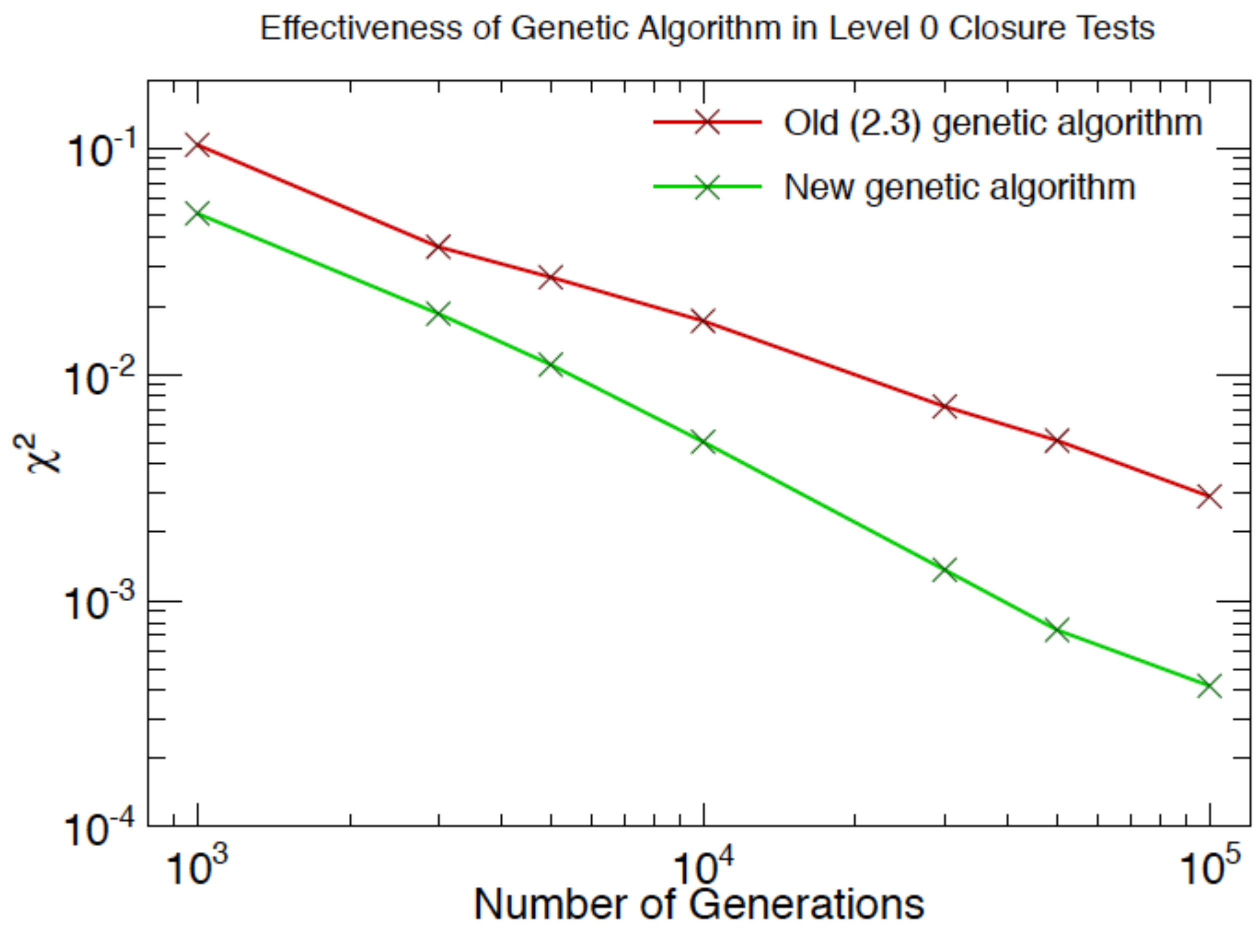
Overtrained

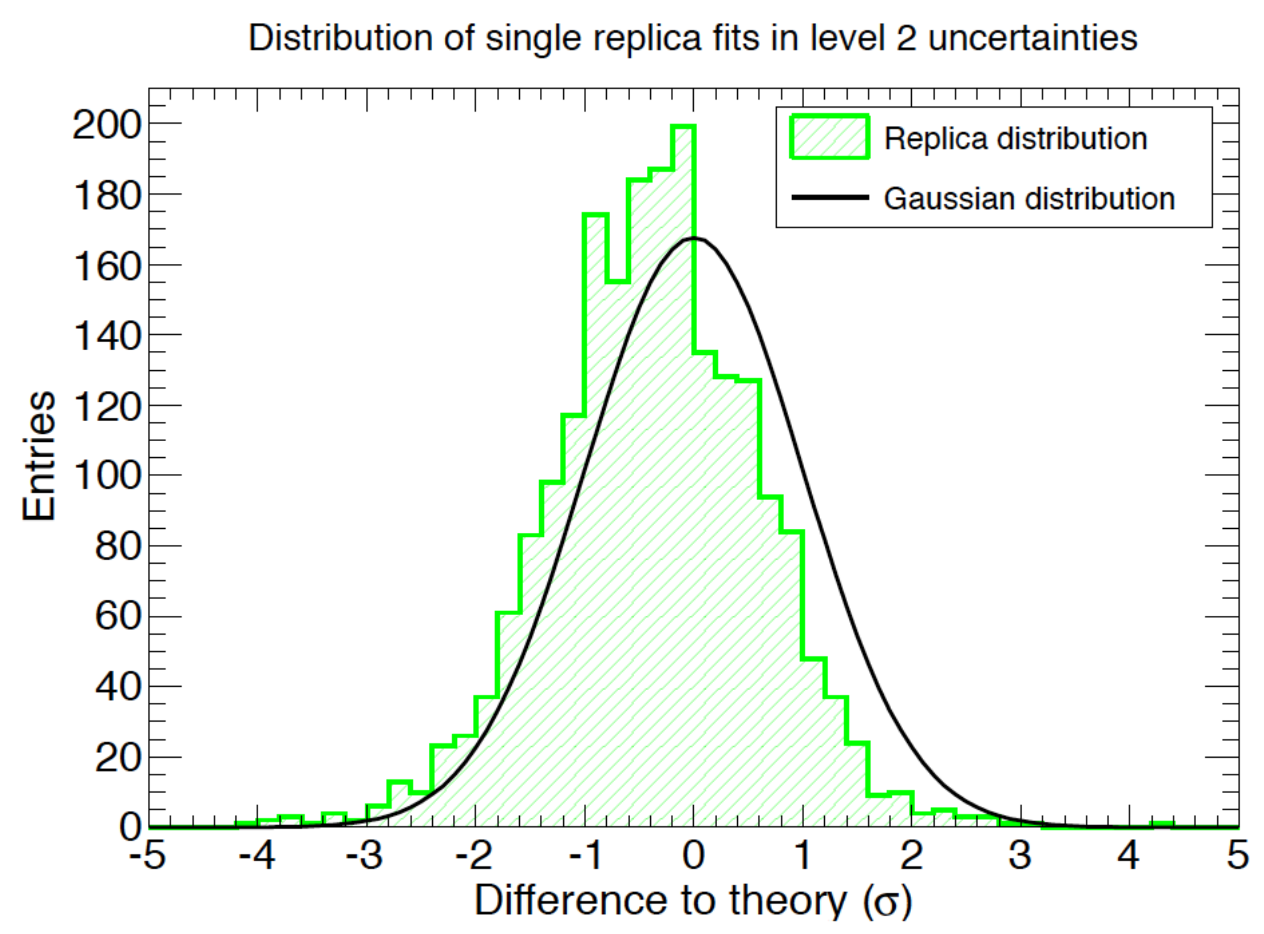
Closure testing the fitting methodology

Generate **pseudo-data based on a known theory** and **test fitting methodology in this fully controlled environment**, free of the noise and other complications (imperfect theory, data inconsistencies) of real world



Closure testing the fitting methodology





Carefully benchmark which training strategy is more efficient

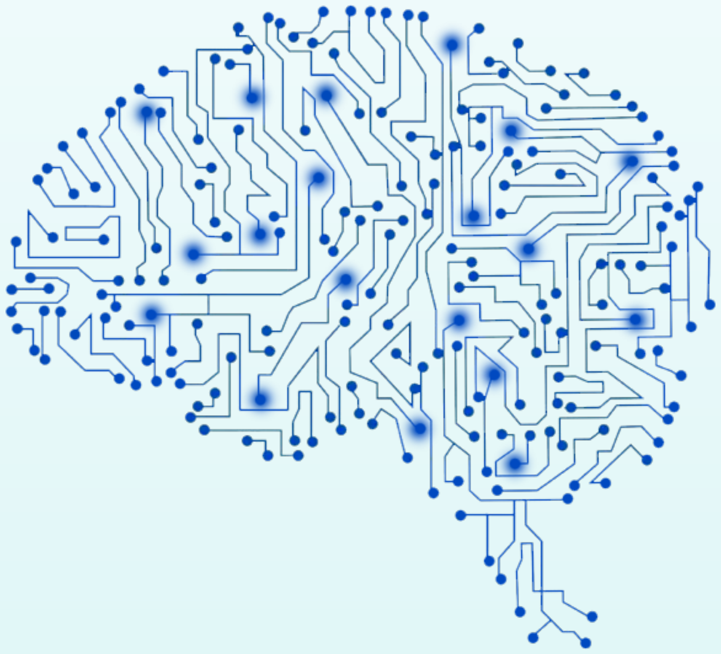
Robust statistical interpretation of PDF uncertainties (from repeating ``runs of the world!")

Generate **pseudo-data based on a known theory** and **test fitting methodology in this fully controlled environment**, free of the noise and other complications (imperfect theory, data inconsistencies) of real world

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

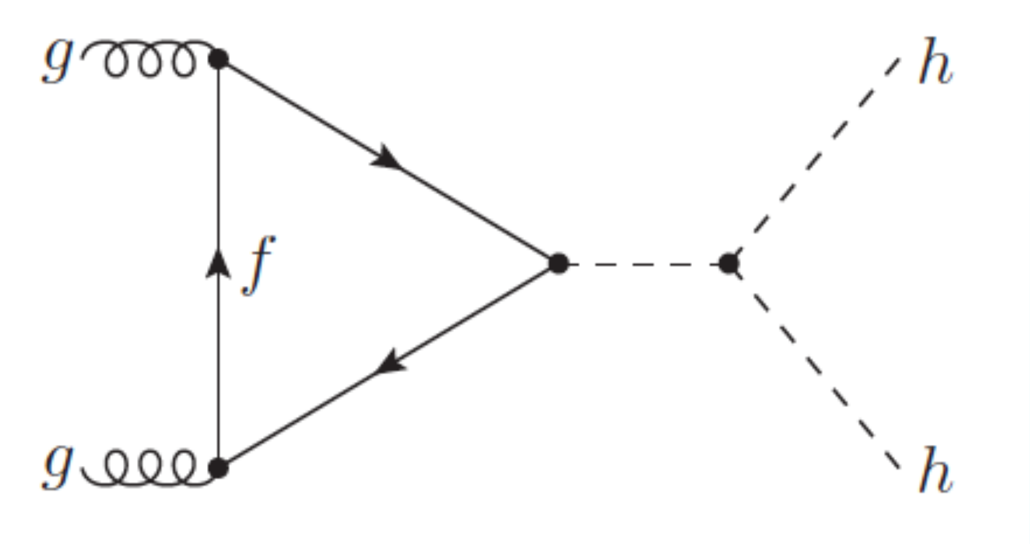


QCD-aware NNs for jet physics

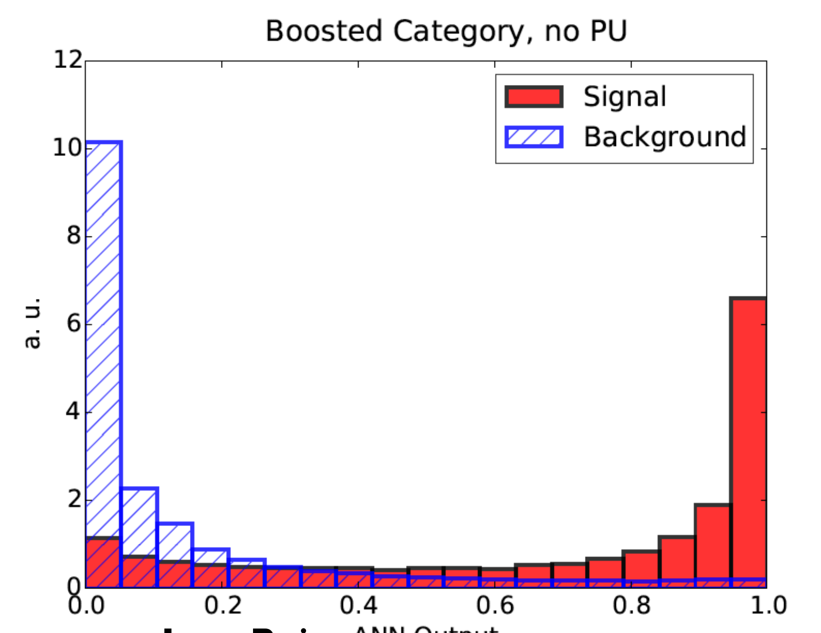
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation

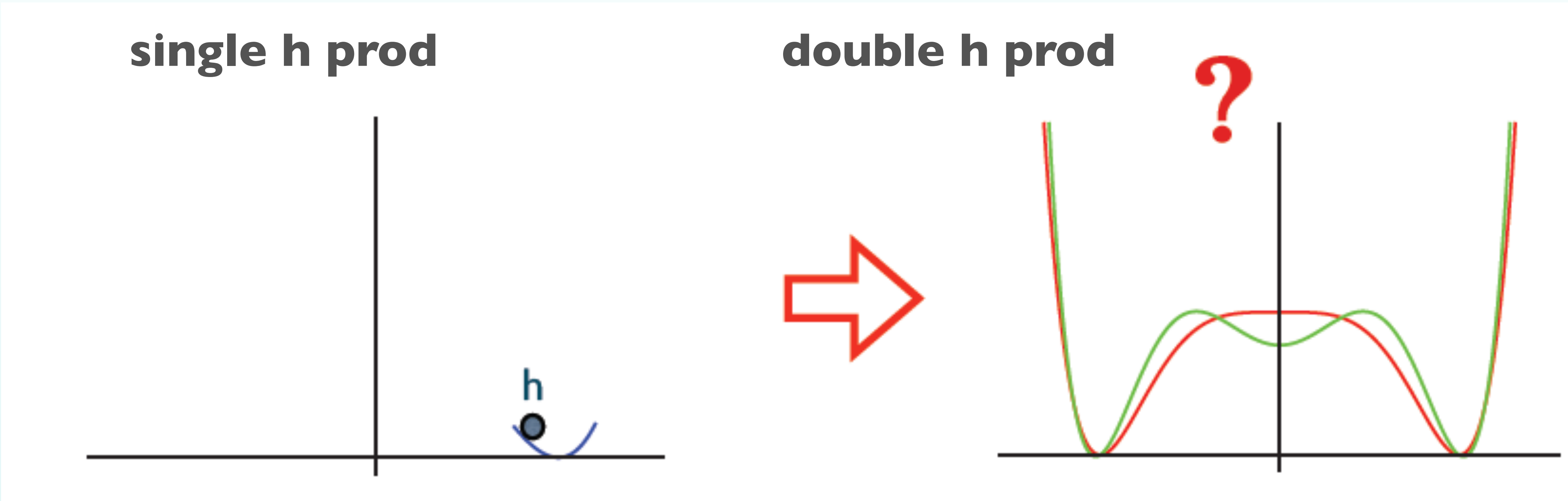


Unravelling the Higgs Self-Coupling



Probing Electroweak Symmetry breaking

- **Current measurements** (couplings in single Higgs production) probe **Higgs potential close to minimum**
- Double Higgs production essential to **reconstruct the full Higgs potential** and clarify EWSB mechanism
- Higgs SM potential is *ad-hoc*: not fixed by the SM symmetries, **many other EWSB mechanisms conceivable**



Higgs mechanism

$$V(h) = m_h^2 h^\dagger h + \frac{1}{2}\lambda(h^\dagger h)^2$$

Coleman-Weinberg mechanism

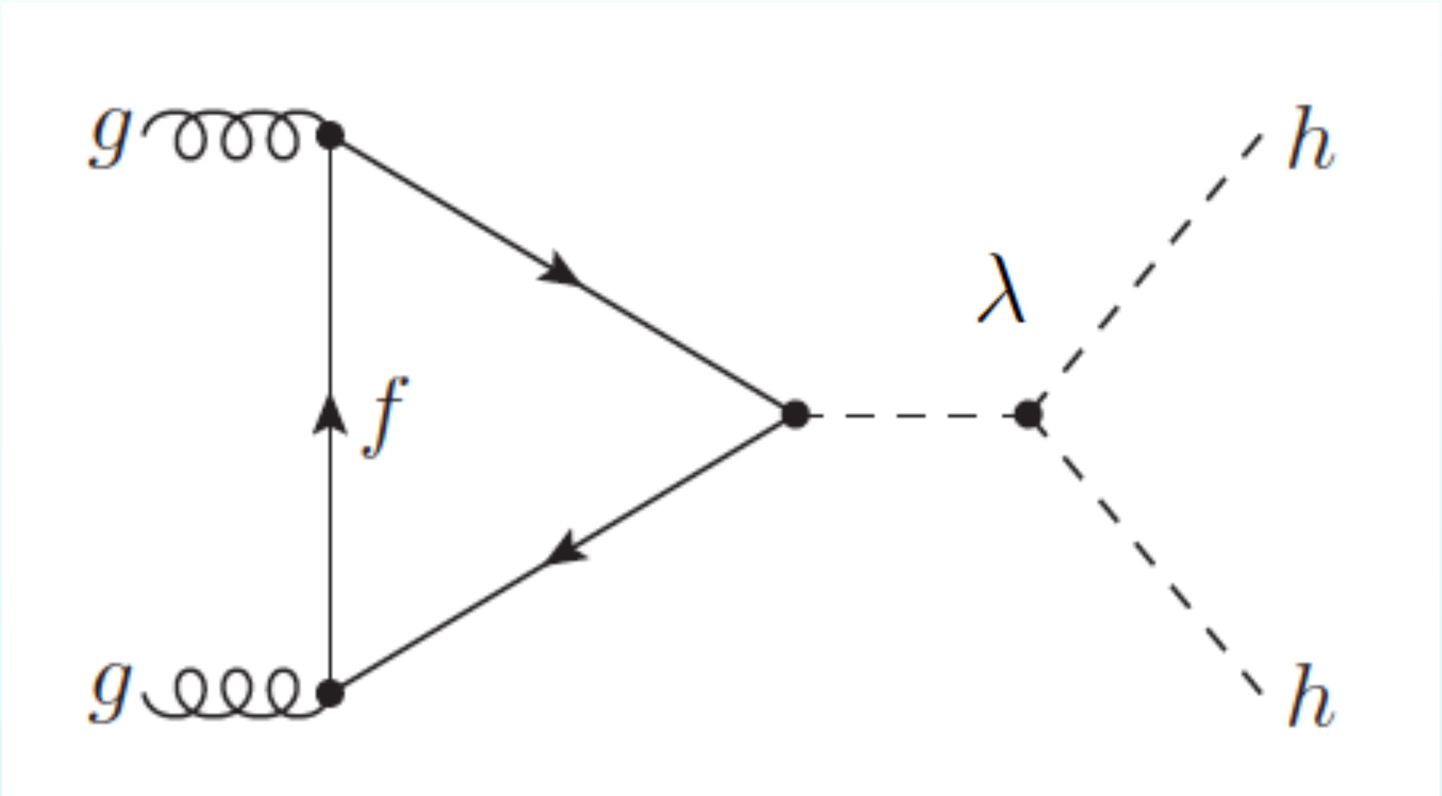
$$V(h) \to \frac{1}{2}\lambda(h^\dagger h)^2 \log\left[\frac{(h^\dagger h)}{m^2}\right]$$

Each possibility associated to **completely different EWSB mechanism**, with crucial implications for the **hierarchy problem**, the structure of quantum field theory, and **New Physics at the EW scale**

Arkani-Hamed, Han, Mangano, Wang, arxiv:1511.06495

Probing Electroweak Symmetry breaking

- **Current measurements** (couplings in single Higgs production) probe **Higgs potential close to minimum**
- Double Higgs production essential to **reconstruct the full Higgs potential** and clarify EWSB mechanism
- The Higgs potential is *ad-hoc*: **many other EWSB mechanisms conceivable**



Higgs mechanism

$$V(h) = m_h^2 h^\dagger h + \frac{1}{2}\lambda(h^\dagger h)^2$$

Coleman-Weinberg mechanism

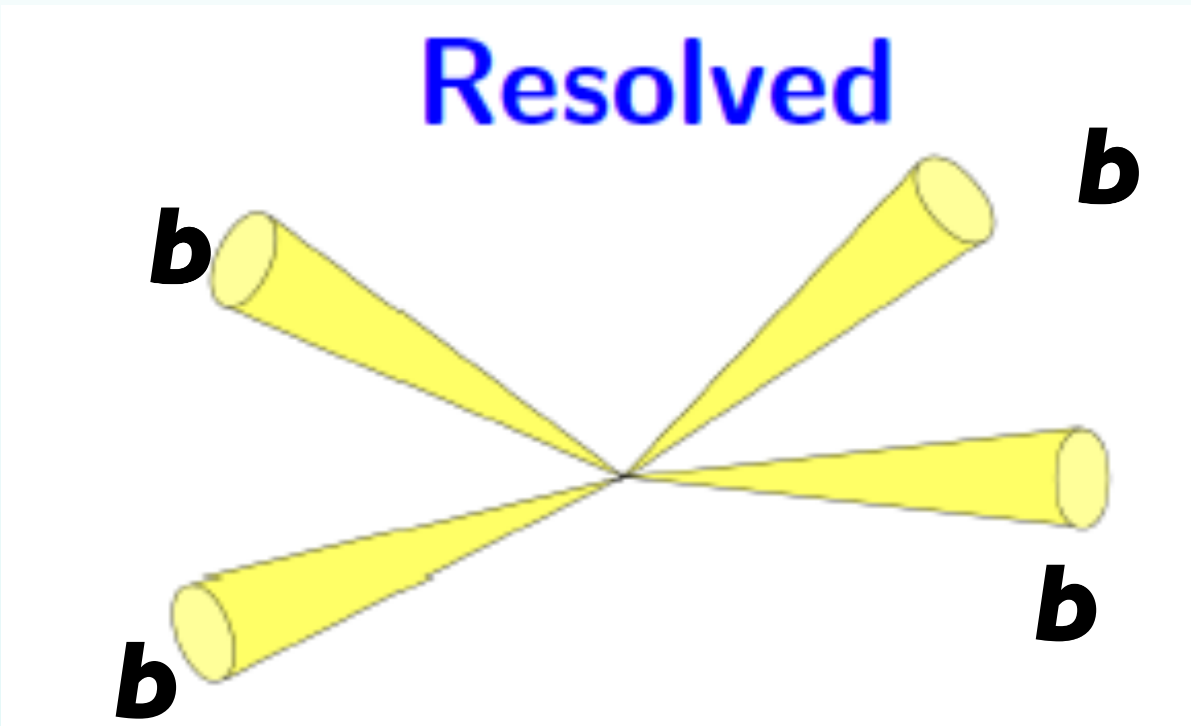
$$V(h) \to \frac{1}{2}\lambda(h^\dagger h)^2 \log\left[\frac{(h^\dagger h)}{m^2}\right]$$

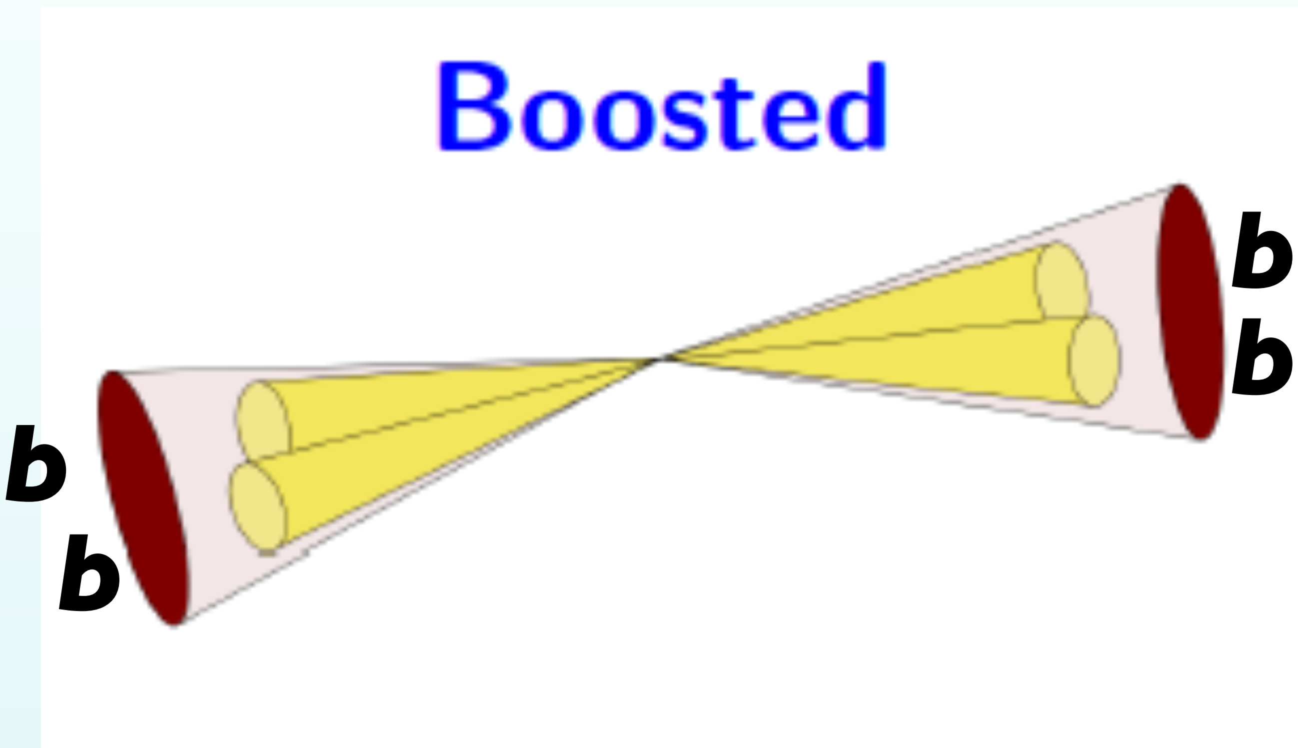
Each possibility associated to **completely different EWSB mechanism**, with crucial implications for the **hierarchy problem**, the structure of quantum field theory, and **New Physics at the EW scale**

Arkani-Hamed, Han, Mangano, Wang, arxiv:1511.06495

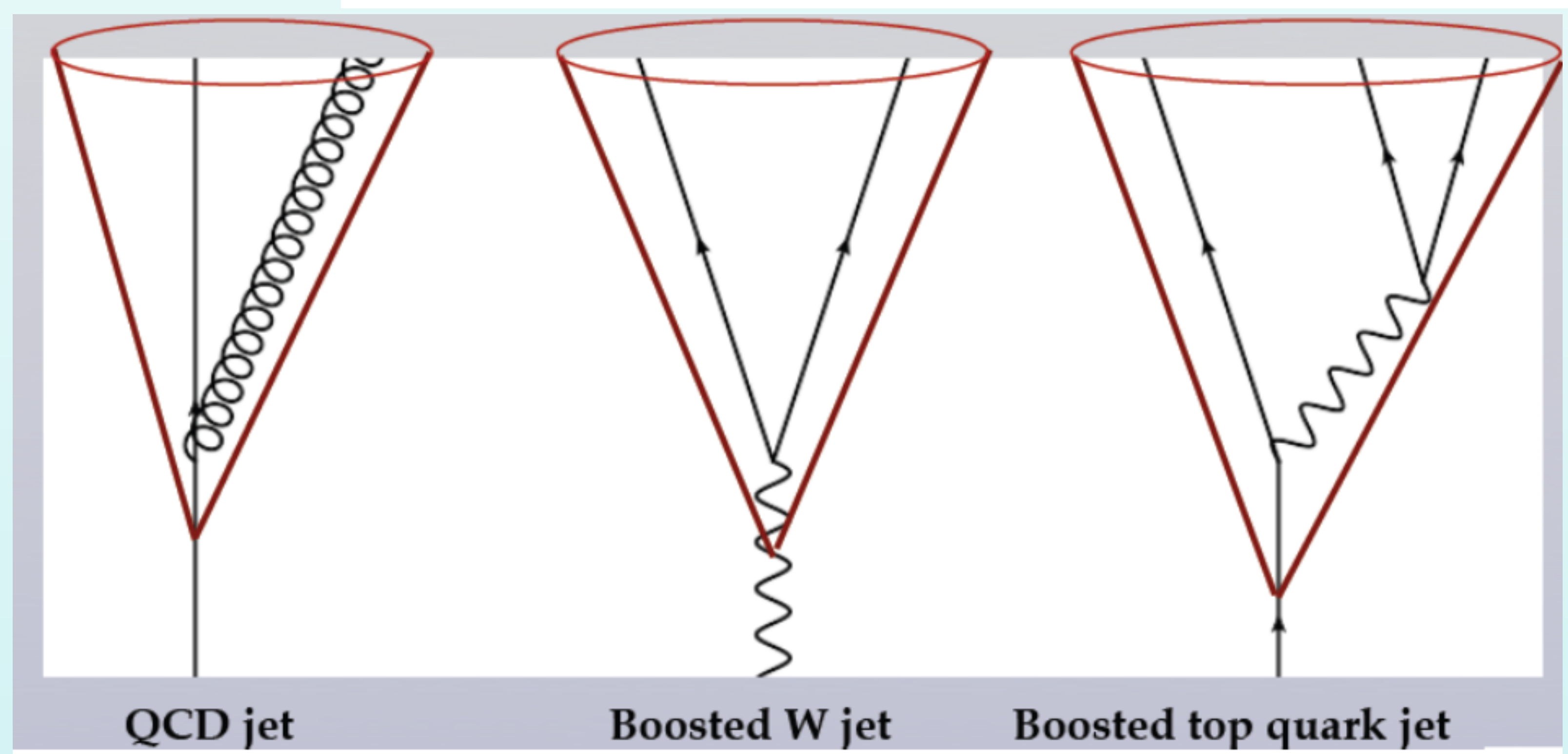
hh->bbbb: selection strategy

- Exploit **4b final state**: highest signal yields, but **overwhelming QCD background** (by orders of magnitude!)
- Carefully chosen selection strategies ensure that **all relevant event topologies can be reconstructed**





Recent progress in **jet substructure** techniques important to reduced QCD background in the **boosted regime**



Juan Rojo

Jet substructure

- The **rich substructure of jets** offers a powerful discriminant between QCD and BSM production dynamics
- Several variables have been introduced to **maximise the discrimination potential**
- Recent progress also from the **analytical point of view** has improved our understanding of substructure

Example: N-subjetiness

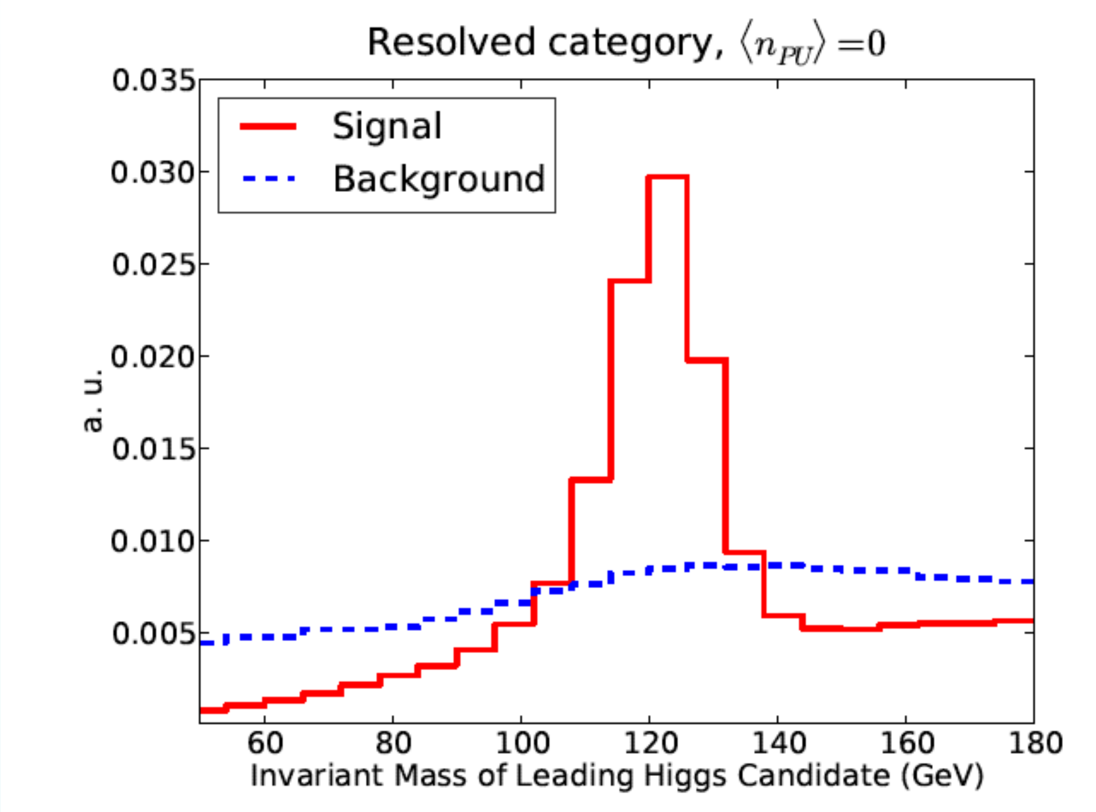
$$\tau_N \equiv \frac{1}{d_0} \sum_k p_{T,k} \cdot \min\left(\delta R_{1k}, \ldots, \delta R_{Nk}\right) \,, \qquad d_0 \equiv \sum_k p_{T,k} R \,,$$

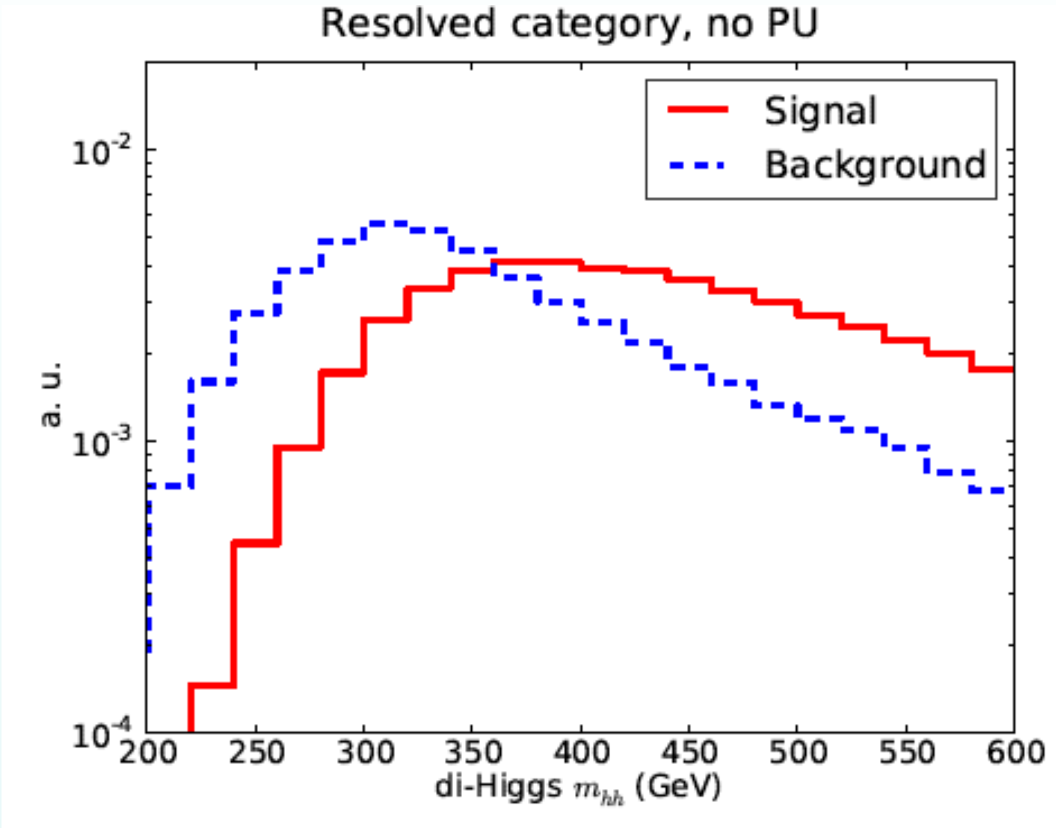
Distance from subjet i to constituent k

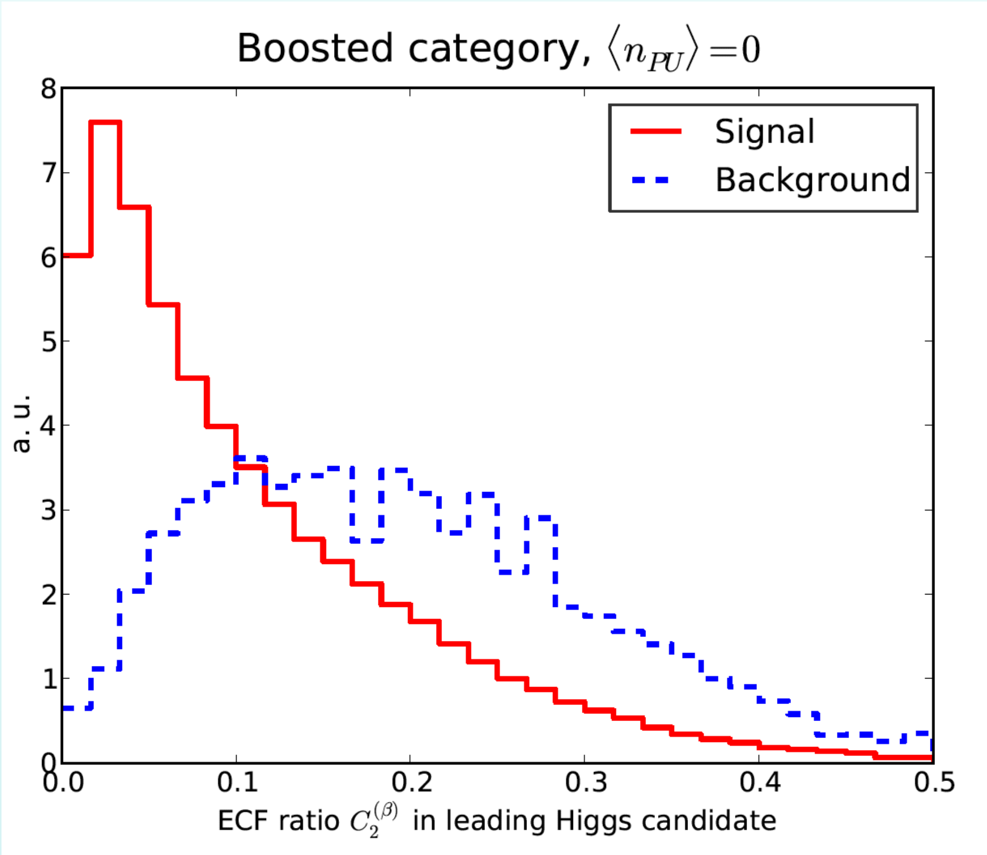
pT of jet constituent k

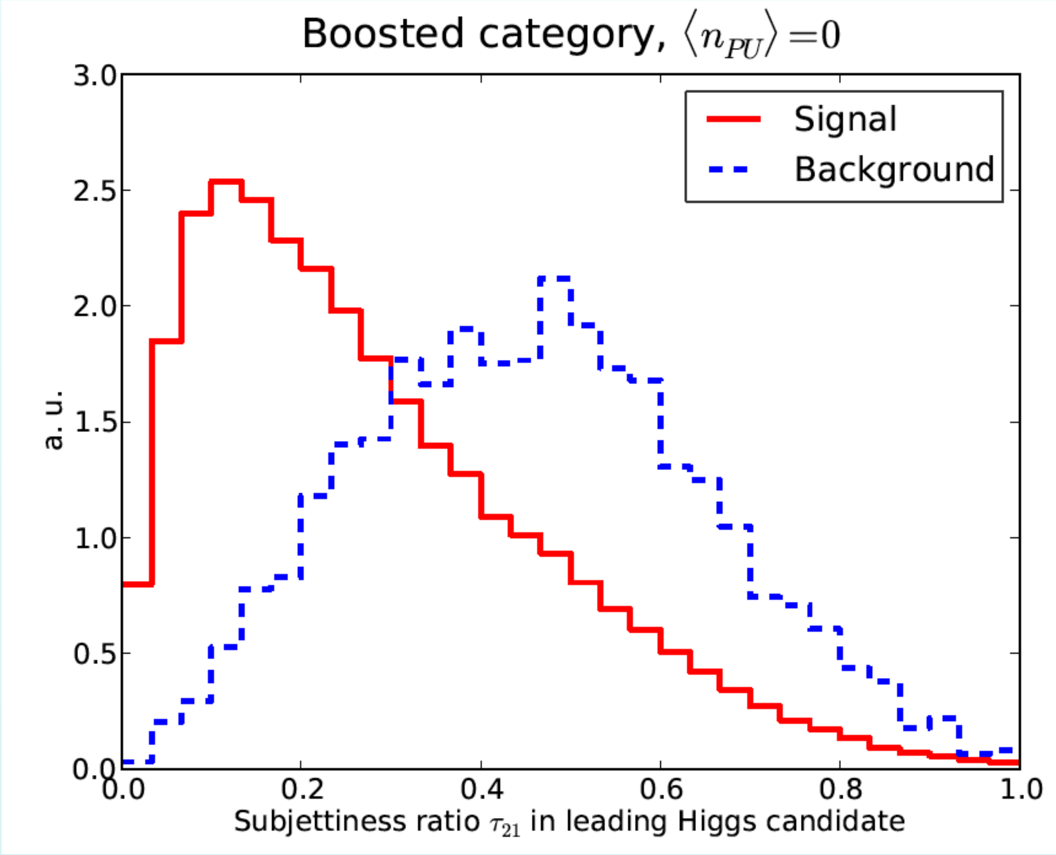
Jet radius

di-Higgs kinematic distributions

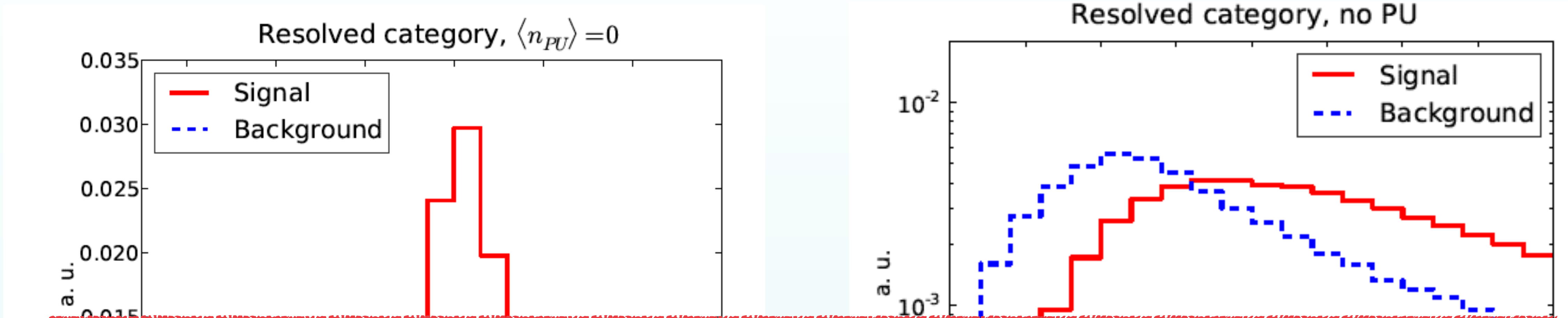








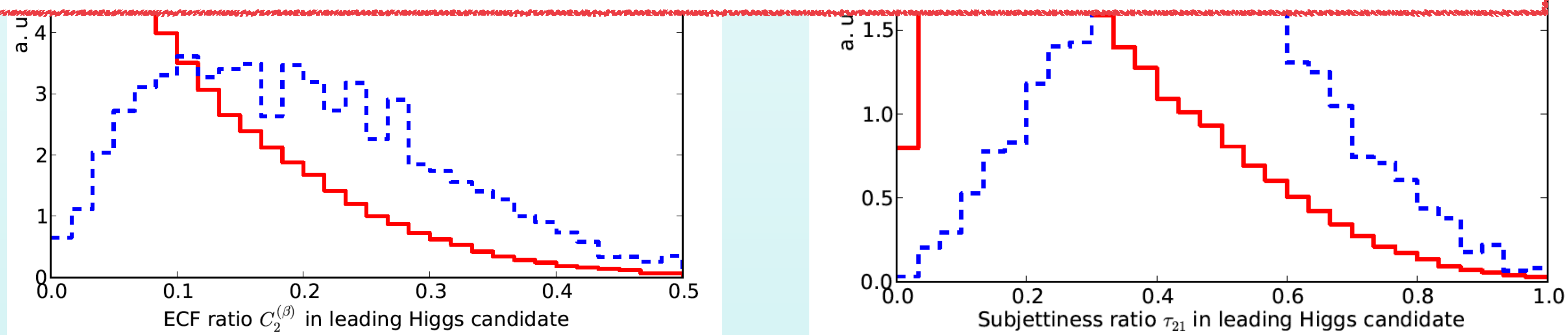
di-Higgs kinematic distributions



Many kinematic variables can be used to **disentangle signal and background**

How do we select which ones to use? And the optical cuts? And the cross-correlations among variables?

We don't need to! Use **ML methods to identify automatically** the combination of kinematical variables with the highest discrimination power!



Multivariate techniques

Input

Higgs pT

Higgs m

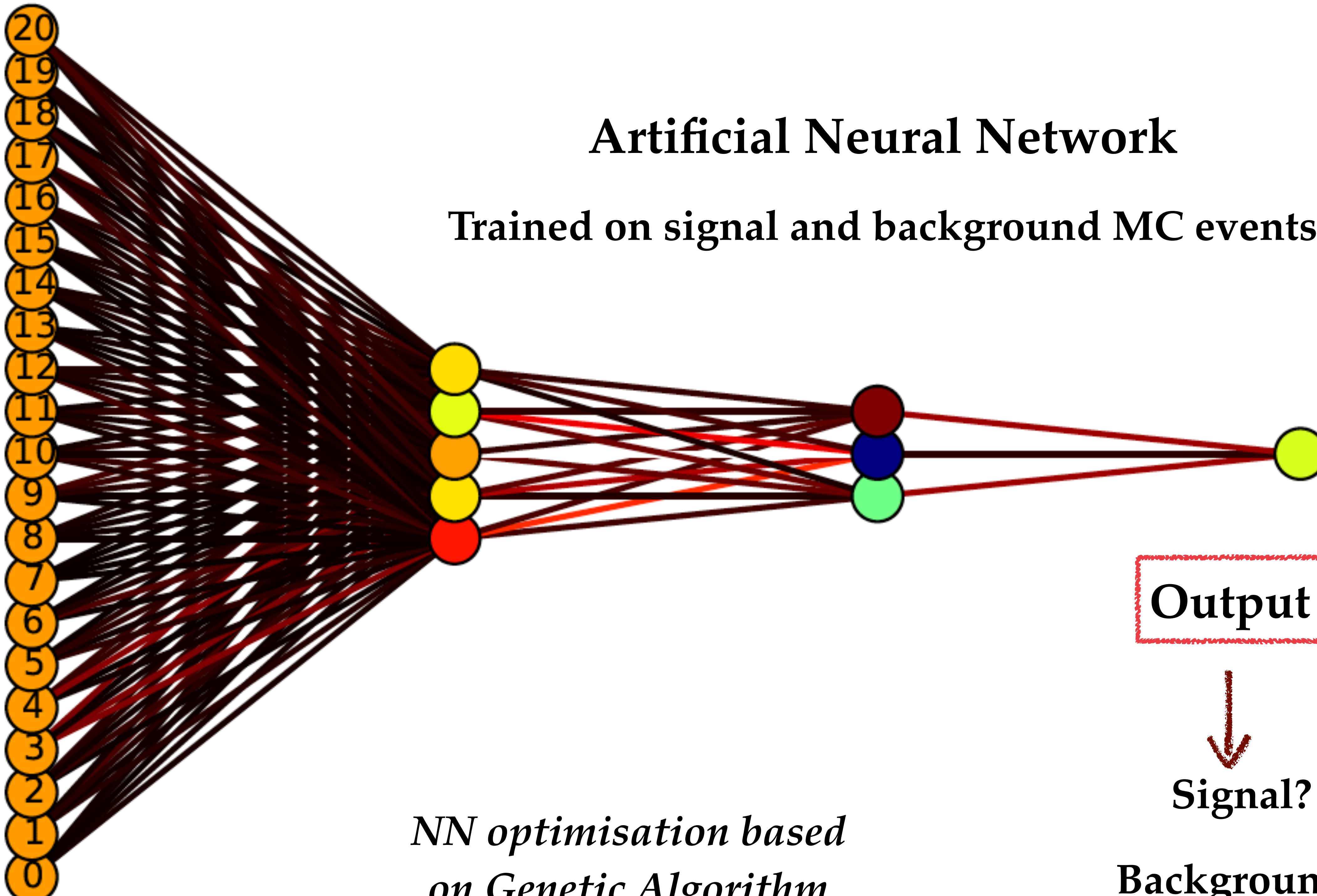
di-Higgs m

ECF

τ_{12}

Subjet pT

…..



Artificial Neural Network

Trained on signal and background MC events

NN optimisation based on Genetic Algorithm

Output

Signal?

Background?

Caveat: in a measurement, training of classifier should be done on ***real data based on control regions***

Multivariate techniques

The optimisation of the classifier is based on the minimisation of the **cross-entropy function**

Number of MC events used for the training

$$E(\{\omega\}) \equiv -\log\left(\prod_i^{N_{\rm ev}} P(y_i'|\{k\}_i, \{\omega\})\right)$$

$$= \sum_i^{N_{\rm ev}} \left[y_i' \log y_i + (1-y_i') \log\left(1-y_i\right)\right]$$

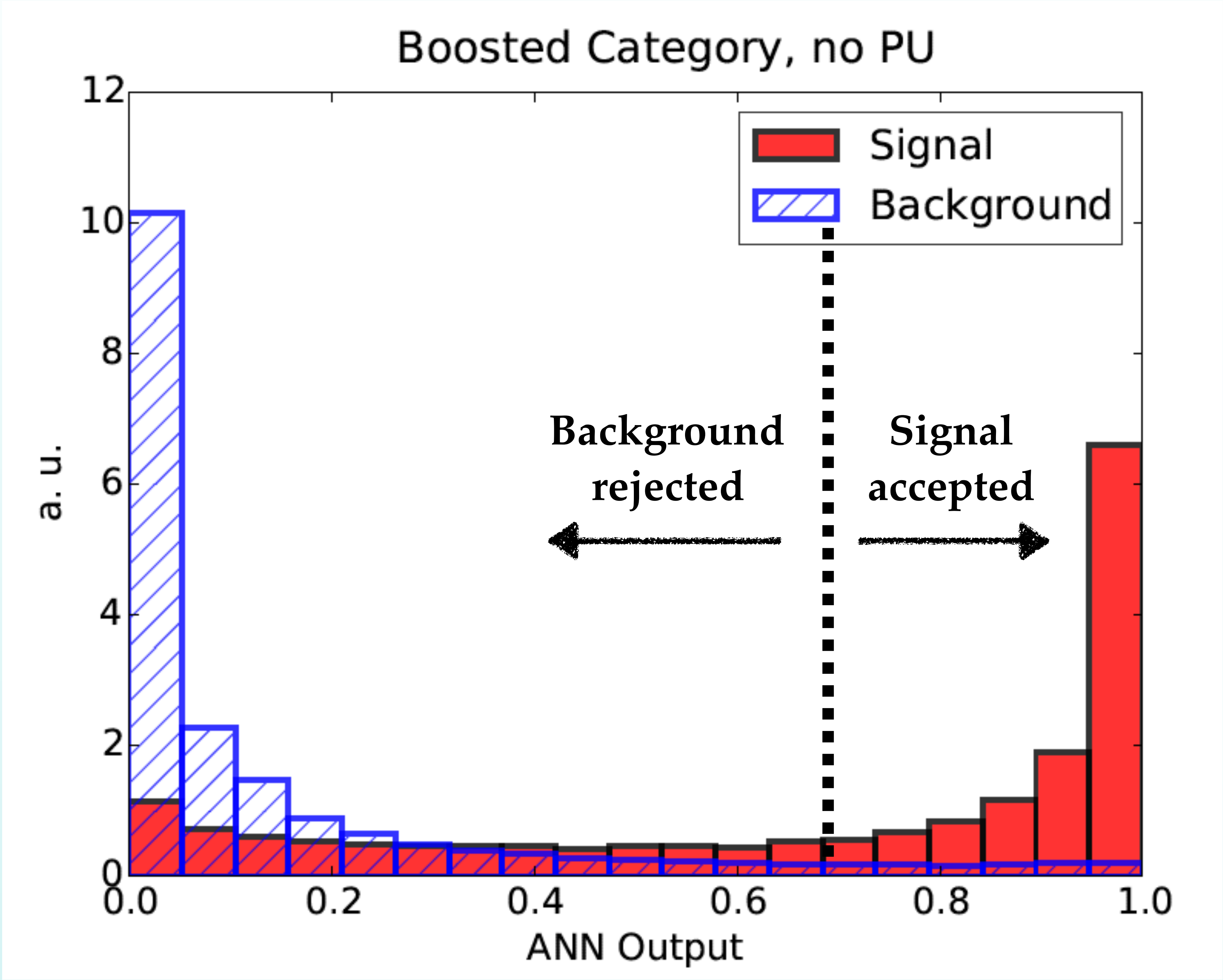
True classification of event: y'_i=0 for background, y'_i=1 for signal

Probability that the event i originates from signal process

aims to achieve the **best possible separation** between signal and background events

Multivariate techniques

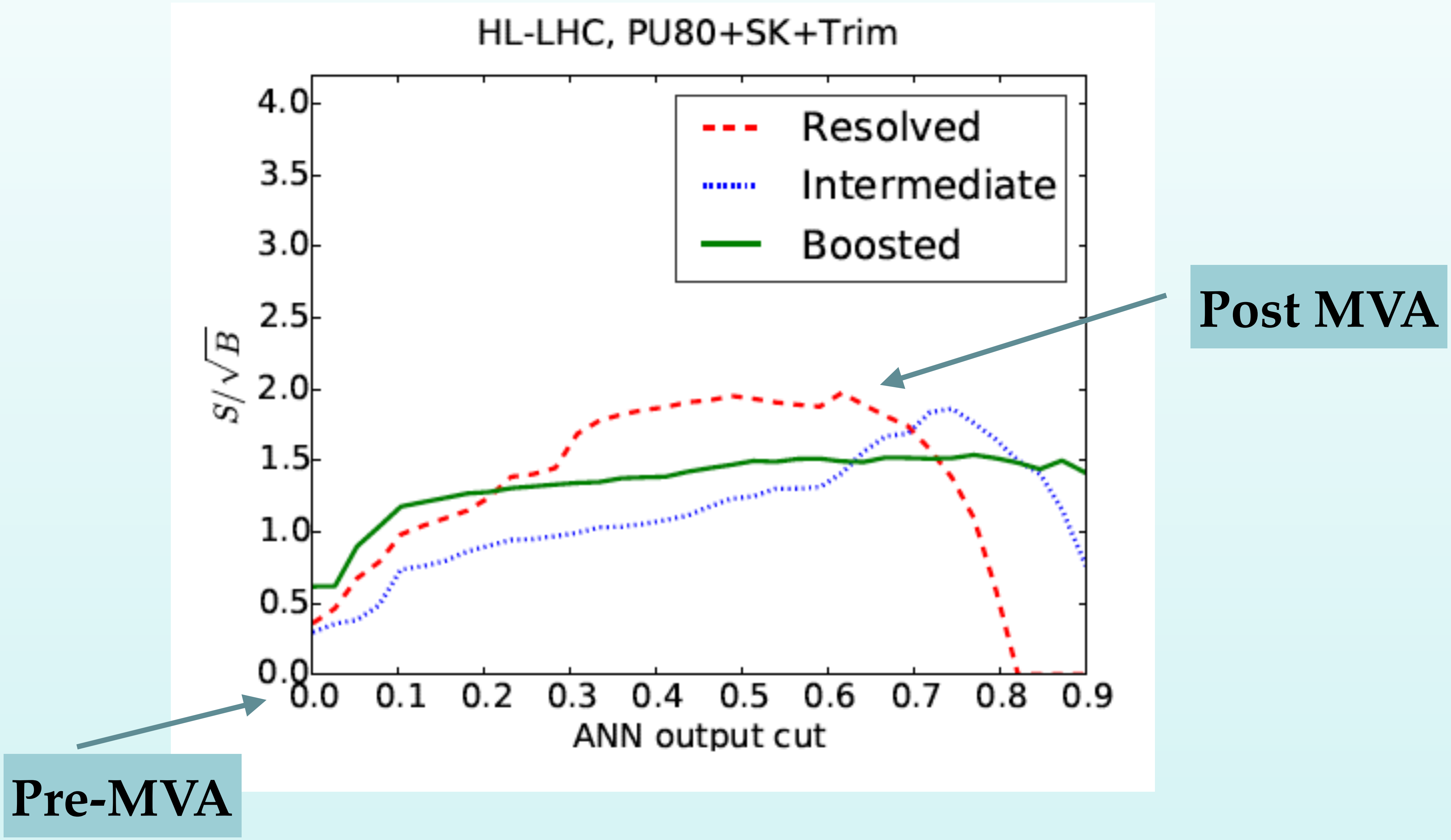
Combining information from all kinematic variables in MVA: **excellent signal/background discrimination**



Discovering Higgs self-interactions

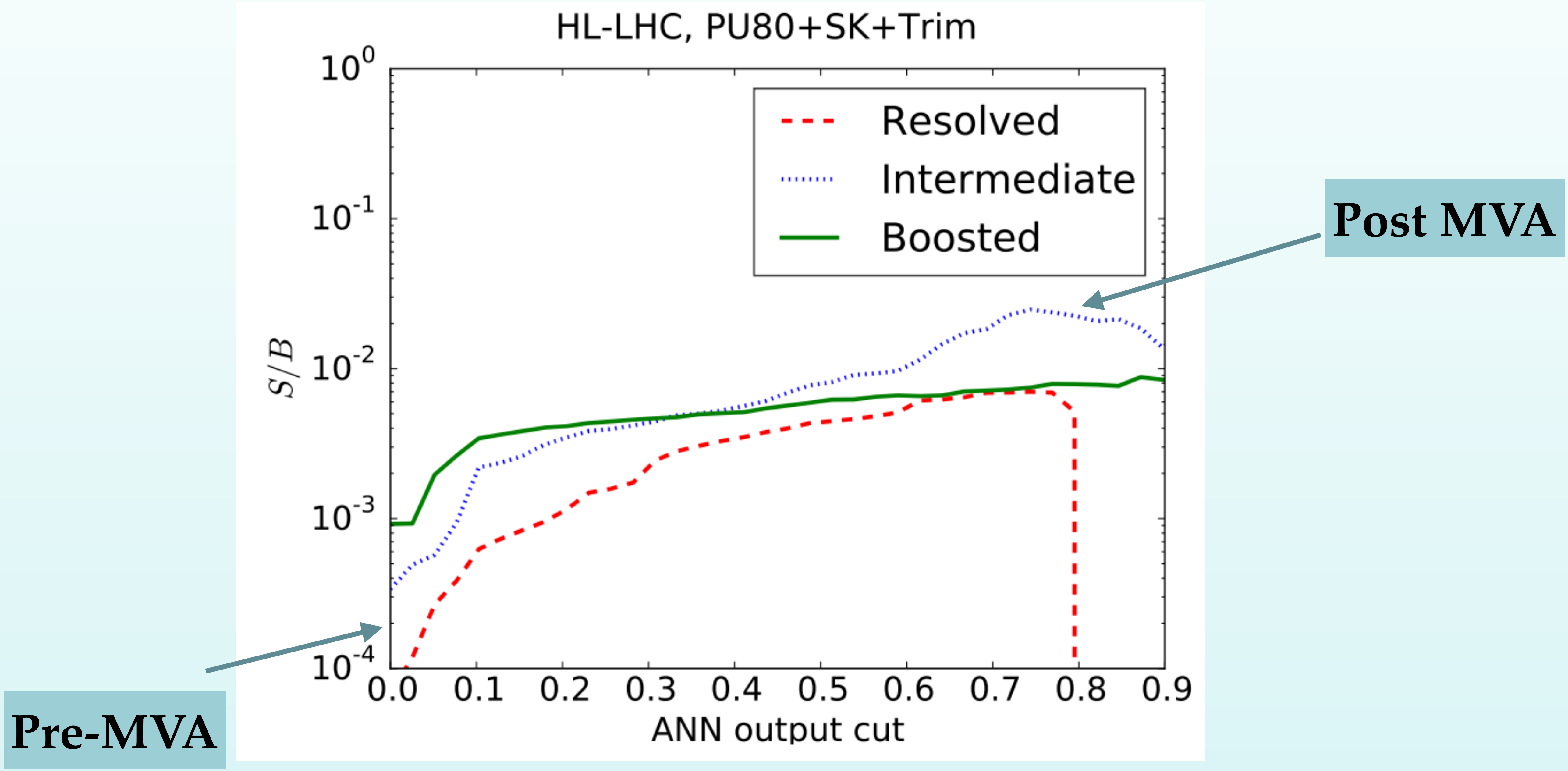
ML techniques allow to **substantially improve the signal significance** for this process **observe Higgs pair production in the 4b final state** at the HL-LHC. Observation (maybe discovery) within reach!

$$\left(\frac{S}{\sqrt{B_{4b}}}\right)_{\rm tot} \simeq 4.7\ (1.5)\,, \quad \mathcal{L} = 3000\ (300)\ {\rm fb}^{-1}$$



Discovering Higgs self-interactions

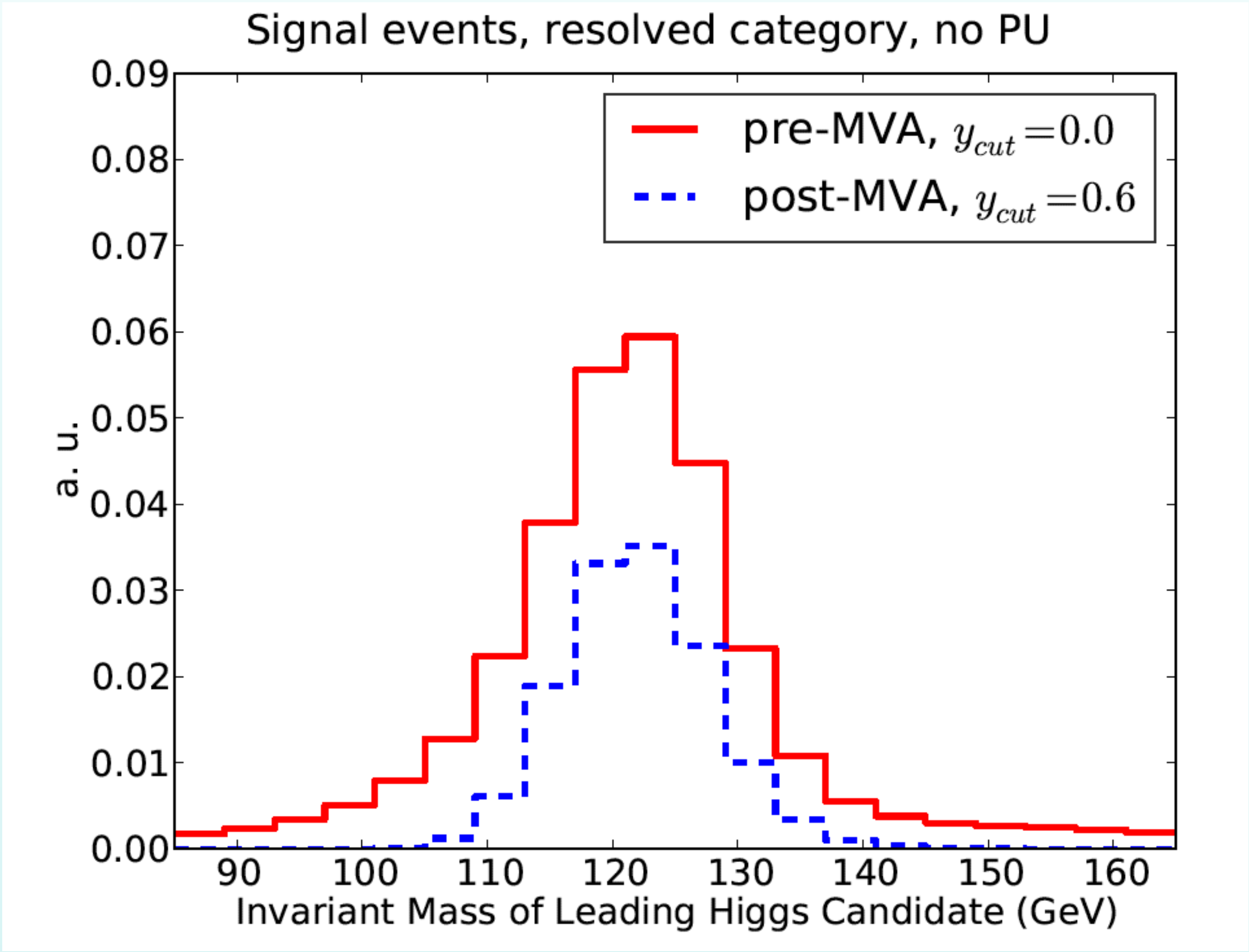
ML techniques allow to **substantially improve the signal significance** for this process **observe Higgs pair production in the 4b final state** at the HL-LHC. Observation (maybe discovery) within reach!



Need to ensure also a high enough signal/background ratio, else experimental systematic errors would kill the signal significance

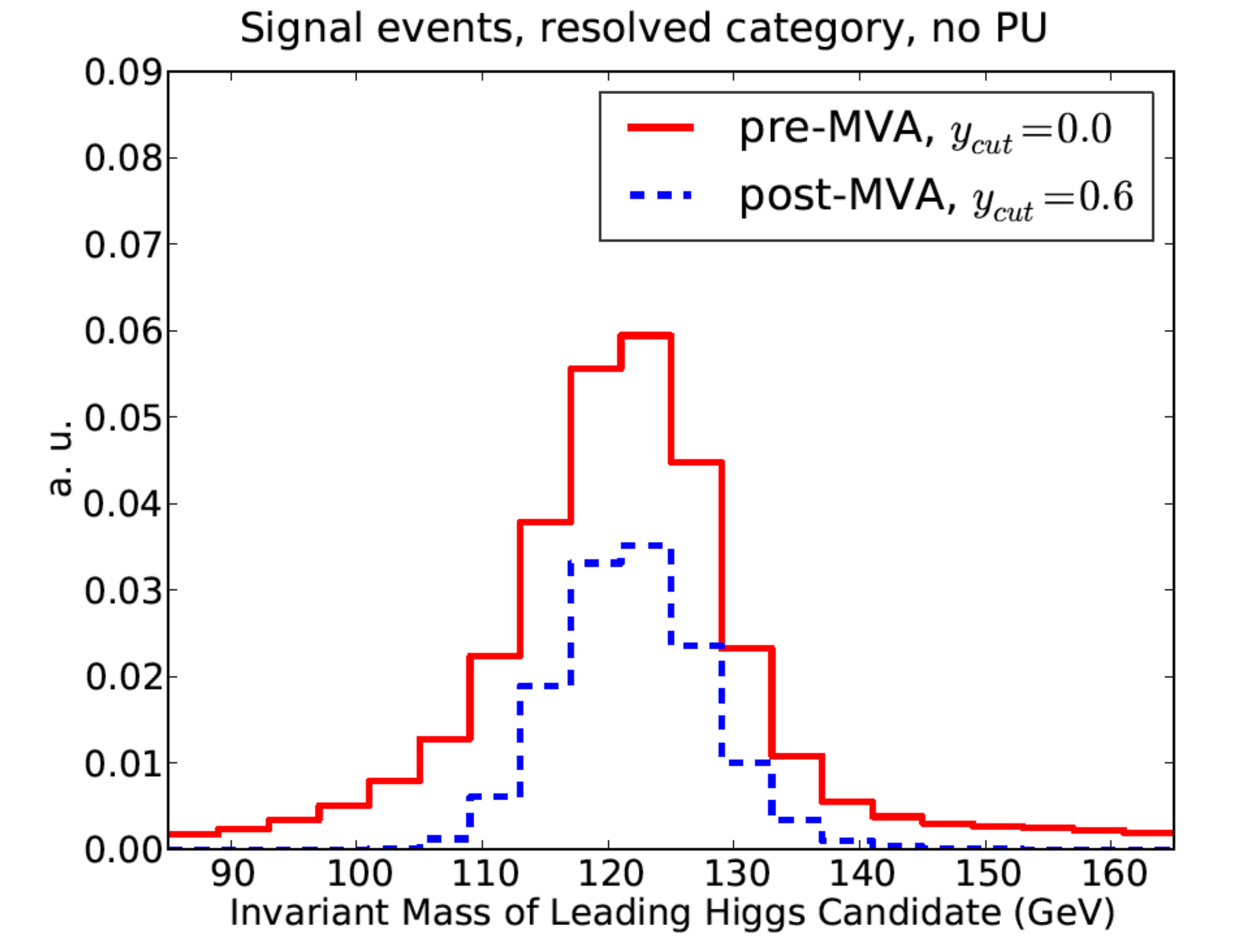
Opening the Black Box

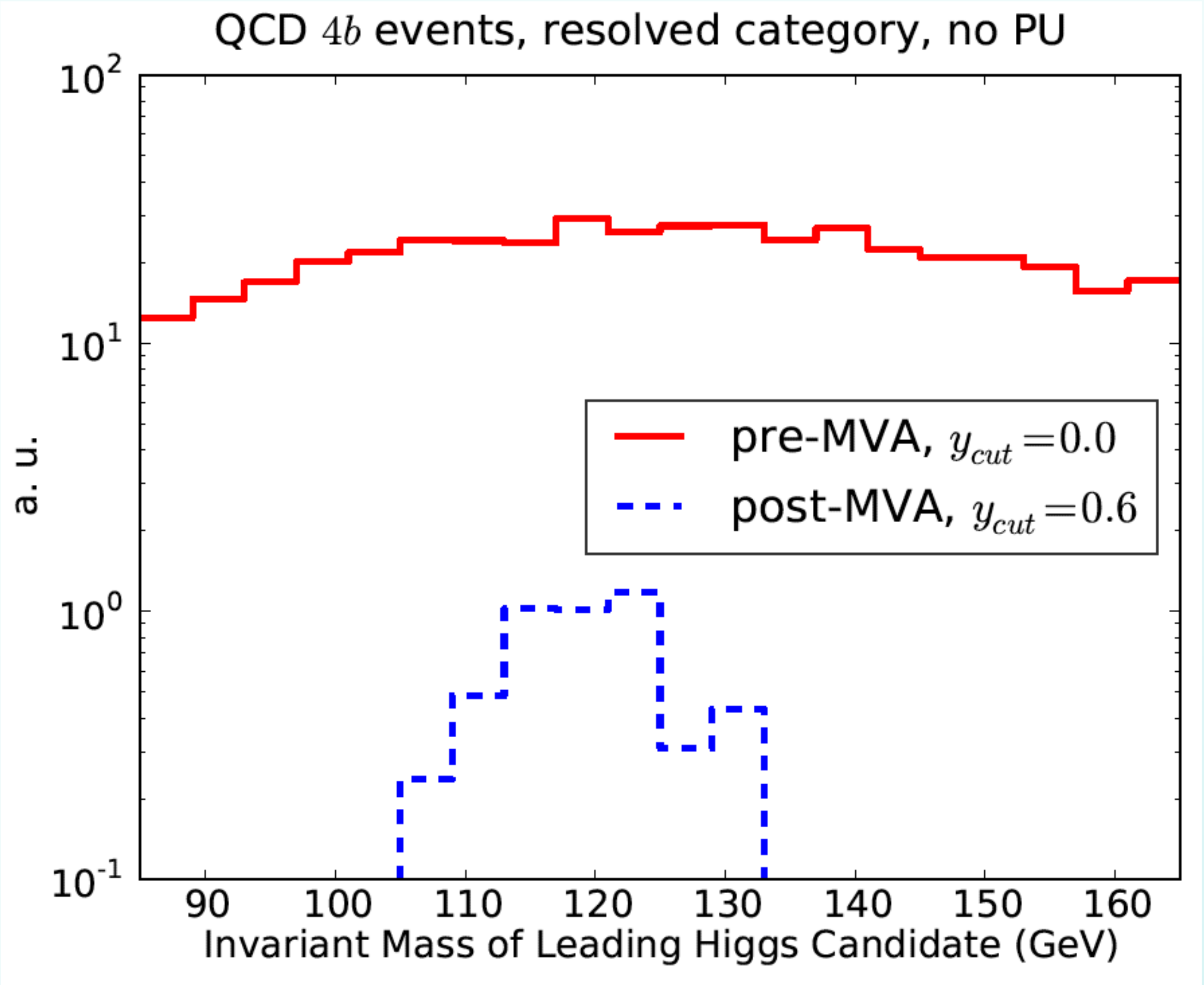
- ANNs are sometimes criticised by being **black boxes**, with little understanding of what happens inside them
- But ANNs are simply a **set of combined kinematical cuts**, nothing mysterious in them!
- Kin distributions **after and before the ANN cut** allow determining the **effective kinematic cuts** being optimised by the MVA, which would allow a cut-based analysis



Opening the Black Box

- ANNs are sometimes criticised by being **black boxes**, with little understanding of what happens inside them
- But ANNs are simply a **set of combined kinematical cuts**, nothing mysterious in them!
- Kin distributions **after and before the ANN cut** allow determining the **effective kinematic cuts** being optimised by the MVA, which would allow a cut-based analysis



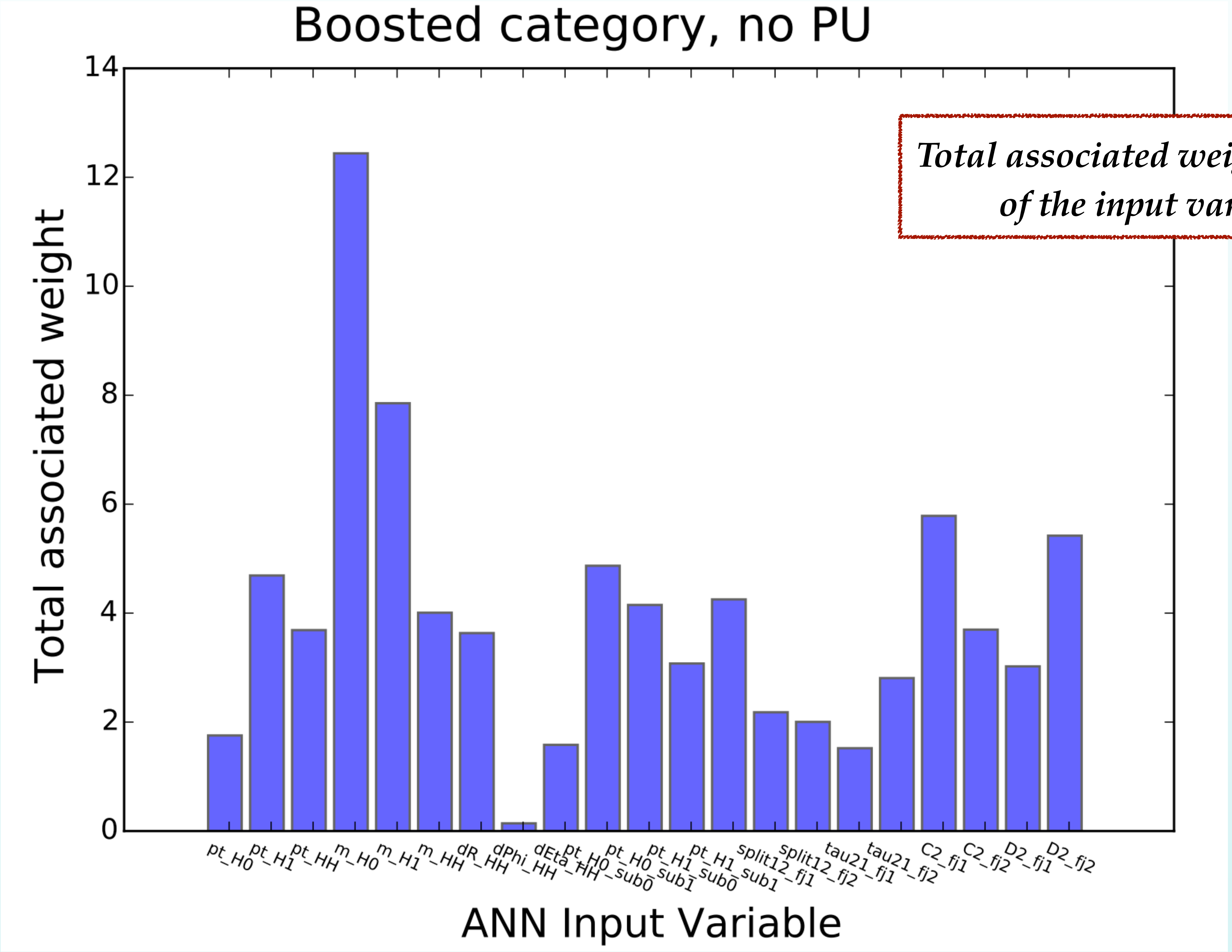


The MVA sculpts a Higgs peak in the QCD background!

Juan Rojo

Some physical insight!

A useful feature of these kind of classifiers is that they made possible **verifying our physical intuition** about which **variables are more important for the discrimination** and which ones are irrelevant

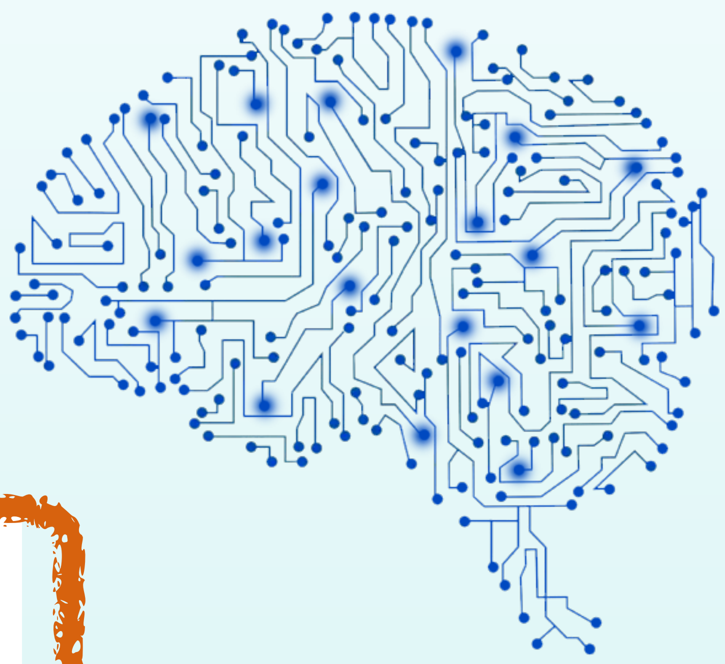


Total associated weight for each of the input variables

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

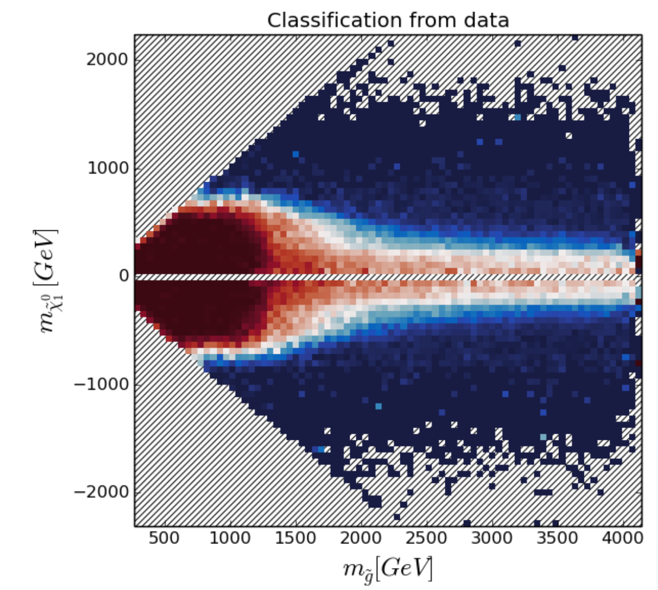


QCD-aware NNs For jet physics

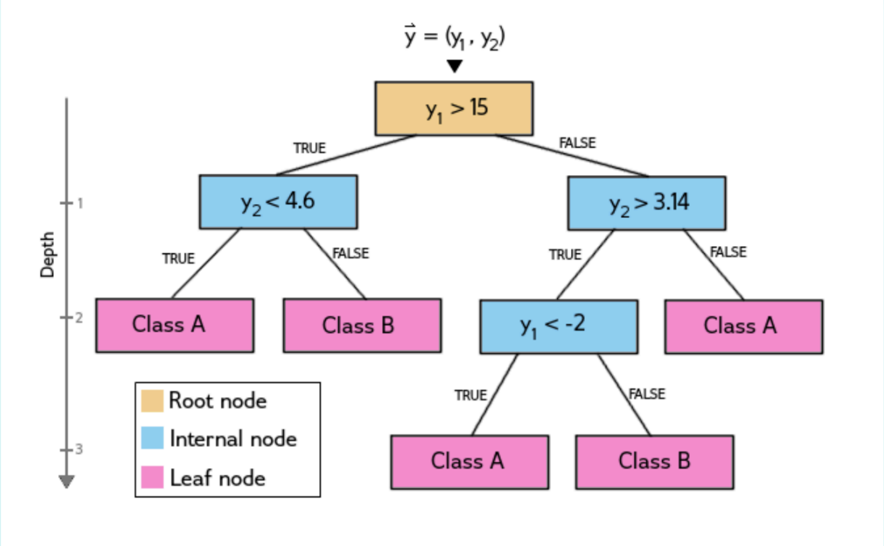
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation



Automated BSM limits with machine learning



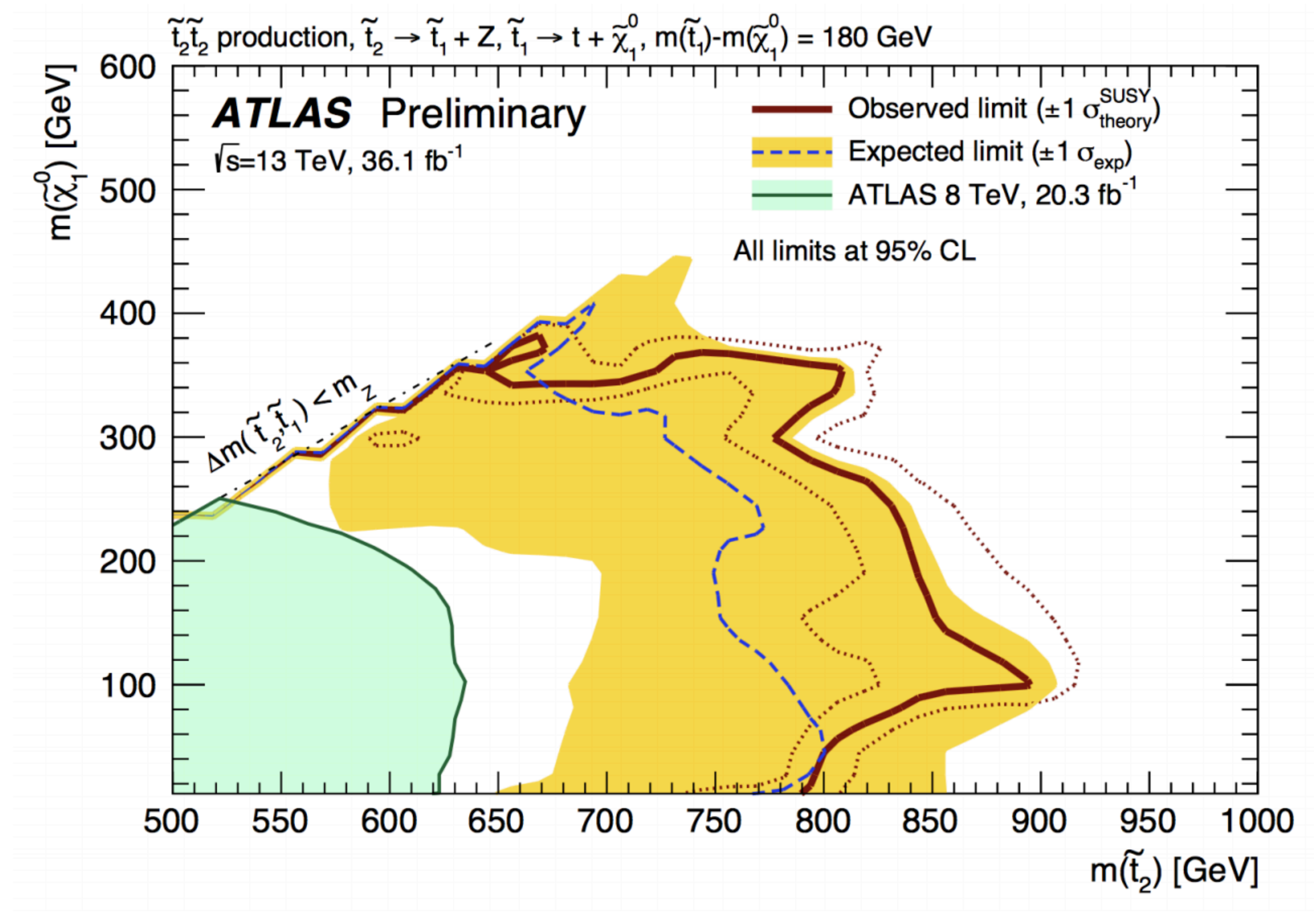
Sascha Caron, Jong Soo Kim, Krzysztof Rolbiecki, Roberto Ruiz de Austri, Bob Stienen

arXiv:1605.02797

Harvesting the LHC data for BSM signals

- In the absence of new particles and/or interactions, LHC searches for BSM physics are used to **derive exclusion ranges for specific scenarios**

- Results presented typically as excluded ranges in a **subset of the full parameter space of the bSM theory**

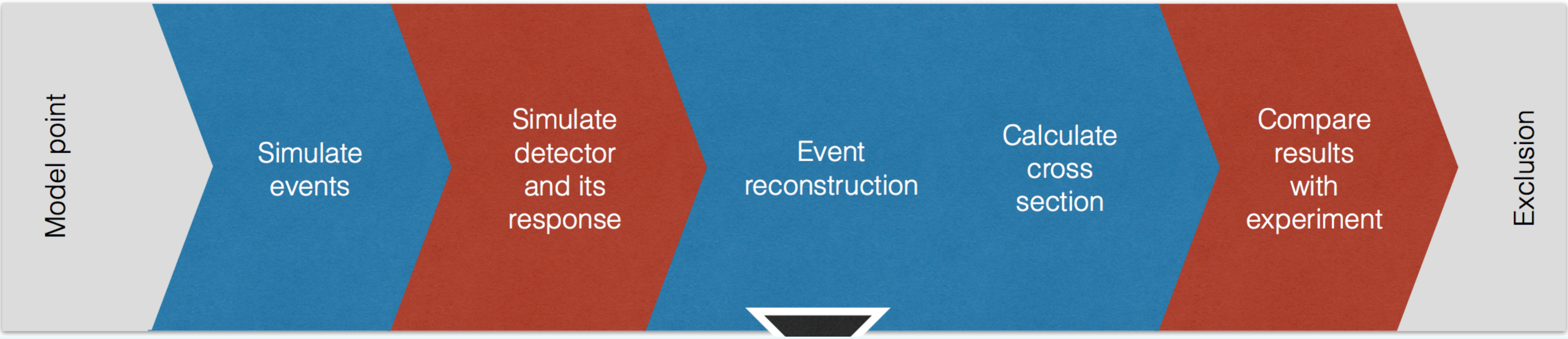


However this is only a **small part of the information contained by the LHC measurements**, ideally we would like the exclusion ranges in the full parameter space of the theory: *e.g.* **19 params in the pMSSM**

Harvesting the LHC data for BSM signals

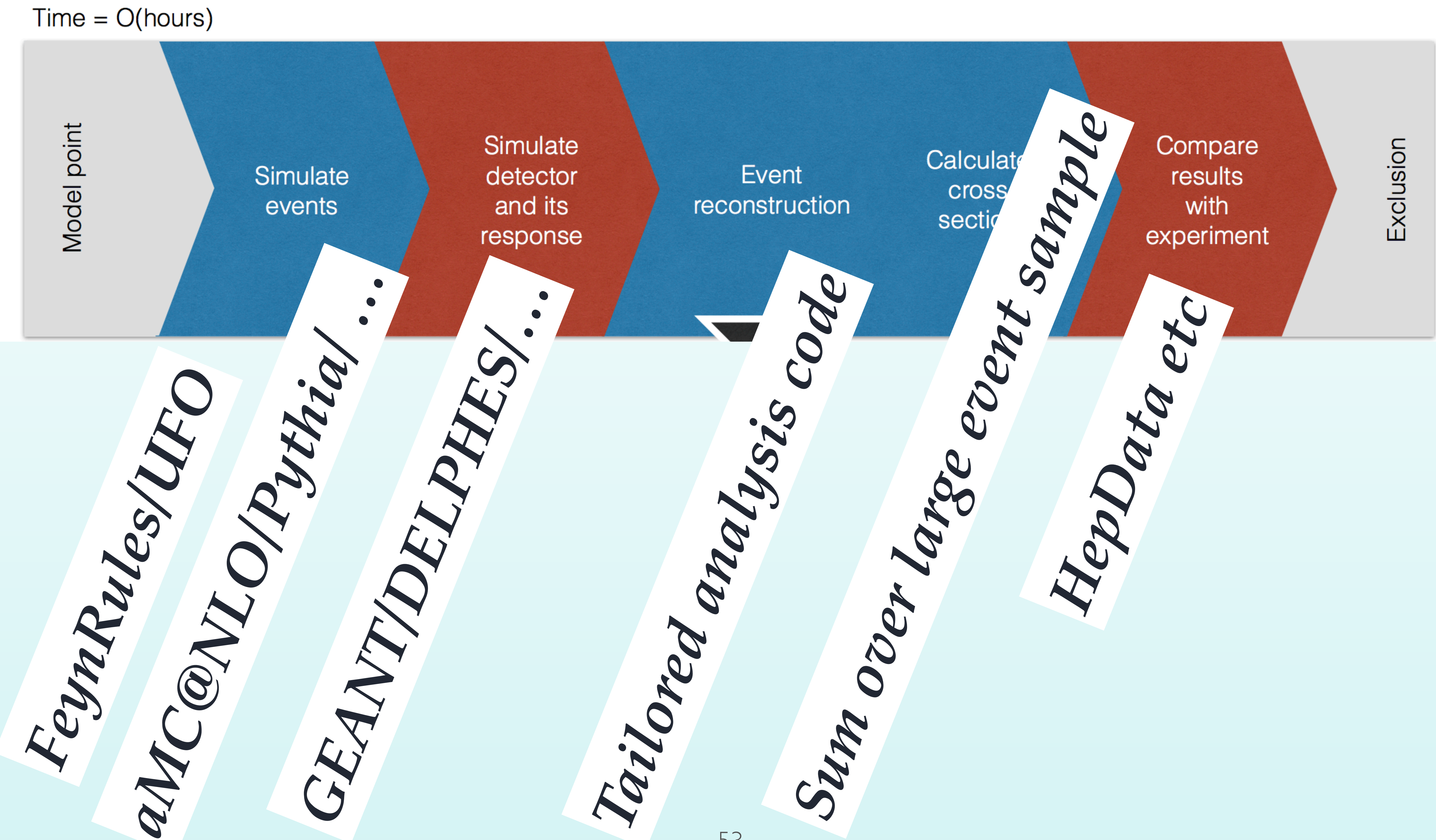
One problem is that **exploring the full parameter space** of the theory is in general very CPU time consuming

Time = O(hours)



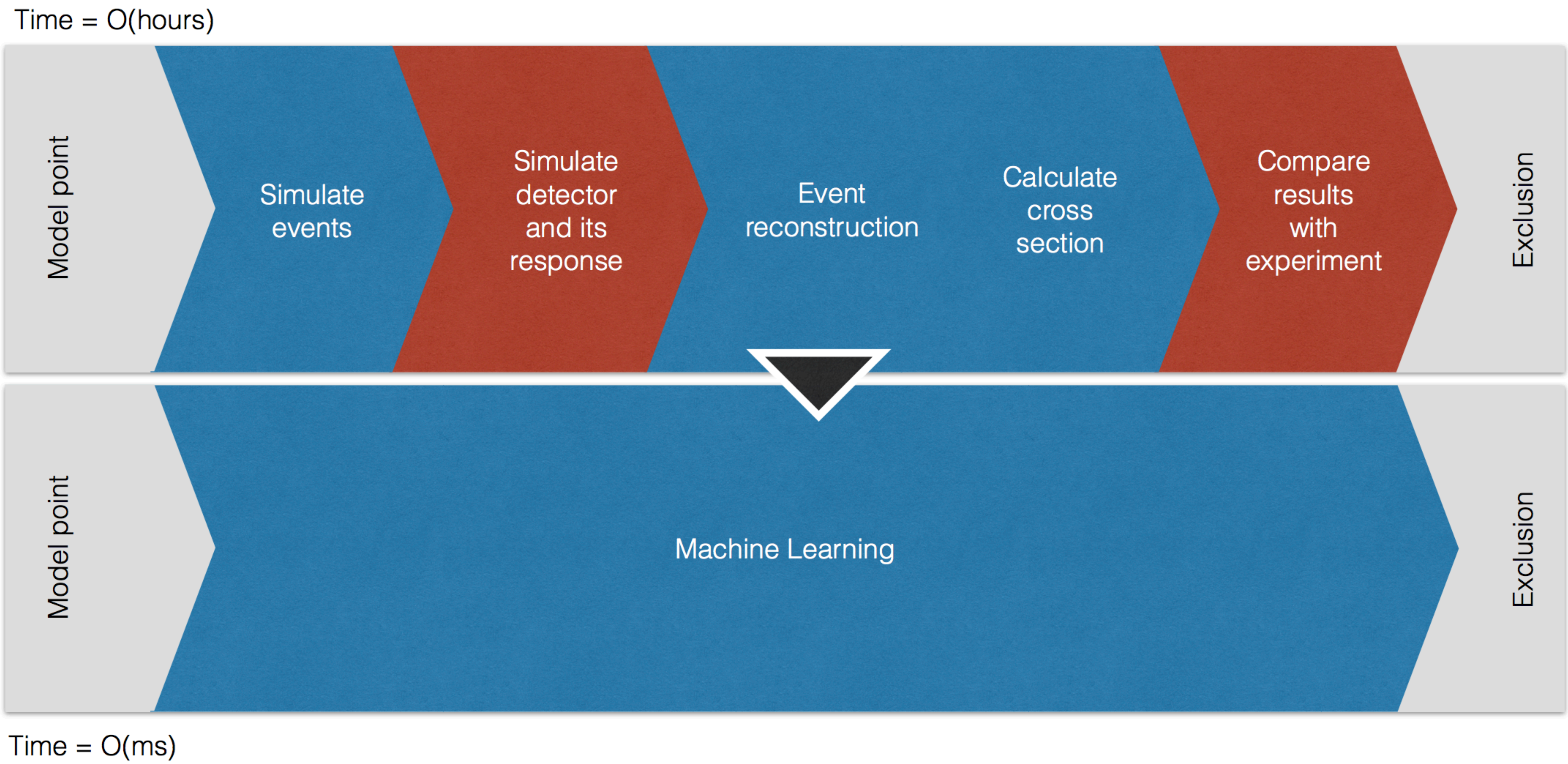
Harvesting the LHC data for BSM signals

One problem is that **exploring the full parameter space** of the theory is in general very CPU time consuming



Harvesting the LHC data for BSM signals

One problem is that **exploring the full parameter space** of the theory is in general very CPU time consuming

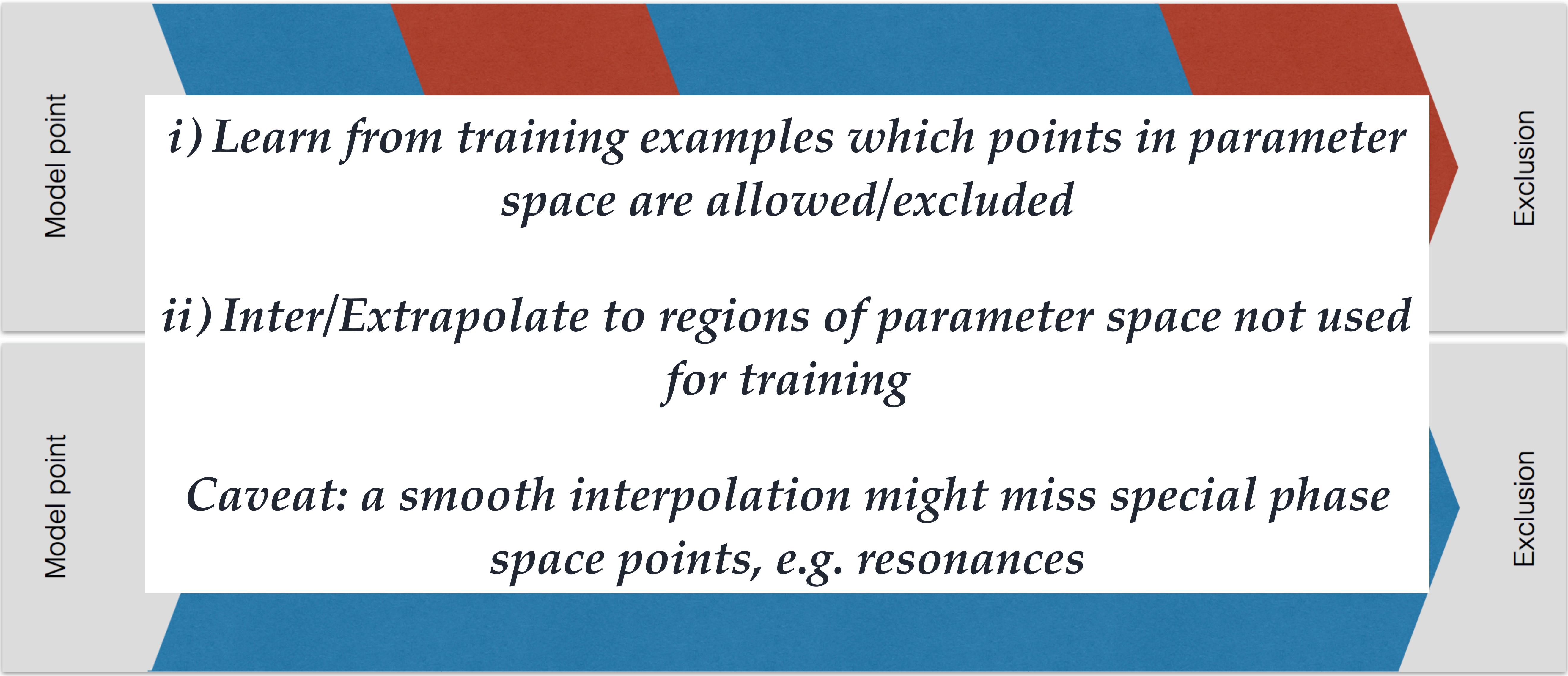


By using Machine Learning tools one can **speed-up the limit-setting procedure by orders of magnitude**, making possible an efficient exploration of the full parameter space of the theory

Harvesting the LHC data for BSM signals

One problem is that **exploring the full parameter space** of the theory is in general very CPU time consuming

Time = O(hours)



i) Learn from training examples which points in parameter space are allowed/excluded

ii) Inter/Extrapolate to regions of parameter space not used for training

Caveat: a smooth interpolation might miss special phase space points, e.g. resonances

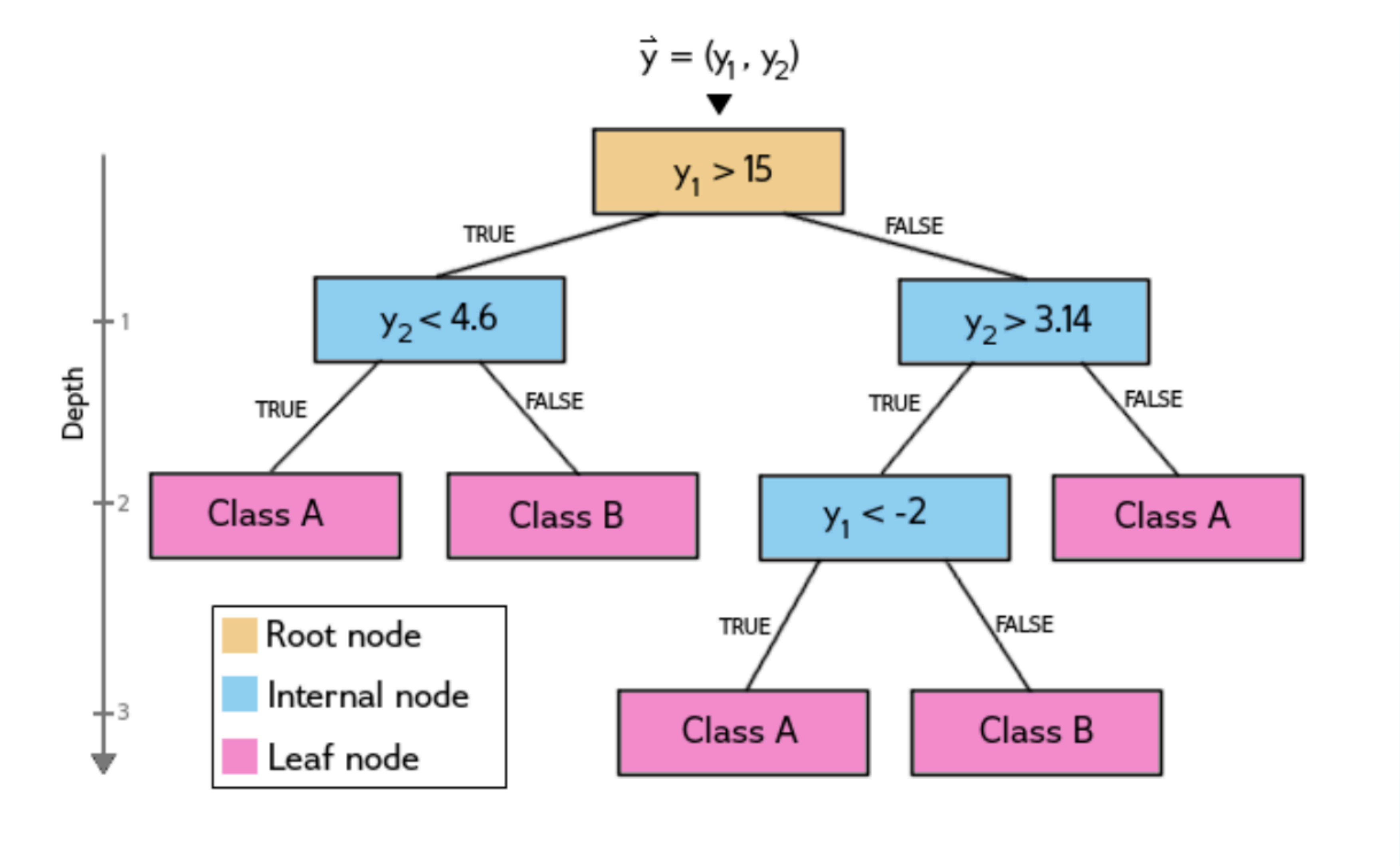
Time = O(ms)

By using Machine Learning tools one can **speed-up the limit-setting procedure by orders of magnitude**, making possible an efficient exploration of the full parameter space of the theory

Generalised BSM limits

- This is a classical example of a discrete Classification problem: a given point in bSM parameter space can be either **allowed** or **excluded**, with no options in between

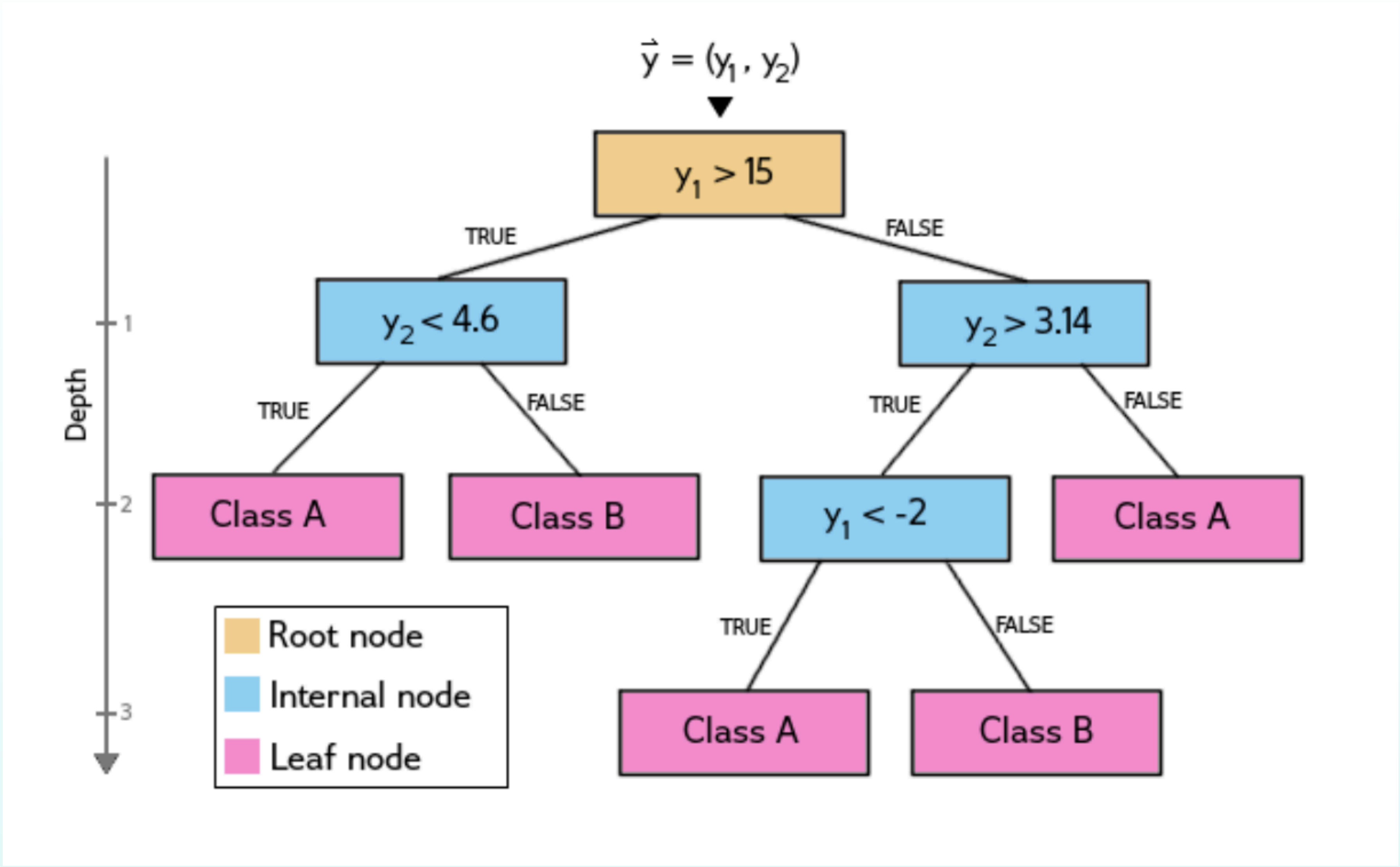
- **Decision Trees** classifiers, such as **Random Forest classifier**, exhibit good performance here



- Procedure starts by presenting parameter sets and class labels, to **learn patterns that the input data follow**.

- Same principle for all **classification algorithms**, specific implementation depending on the particular problem

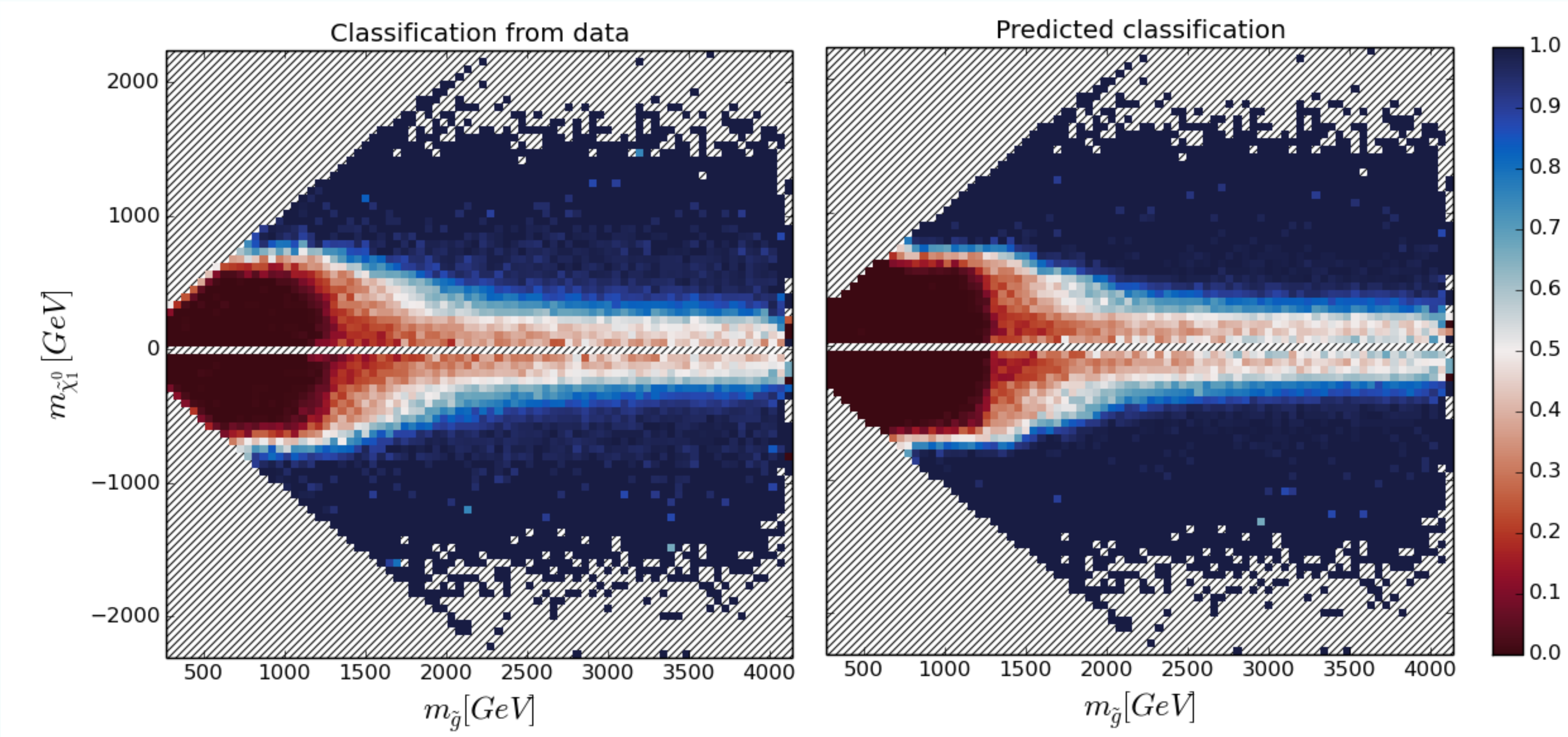
Generalised BSM limits



- A **Decision Tree** consists of multiple nodes, each node specifies a **test performed on the arriving attribute**
- The result of this test determines to which node the attribute set is sent next.
- Process is repeated until the **final leaf node is reached**, *i.e.* the node with no further nodes connected to it.
- At the final node a **class label is assigned to the set**, specifying its class according to the classifier.
- The tree works on the **entire parameter space**: every test performed interpreted as a cut in this space.
- The **parameter space is split into disjunct regions**, each having borders defined by the cuts in the root and internal nodes, and a classification defined by a leaf node.

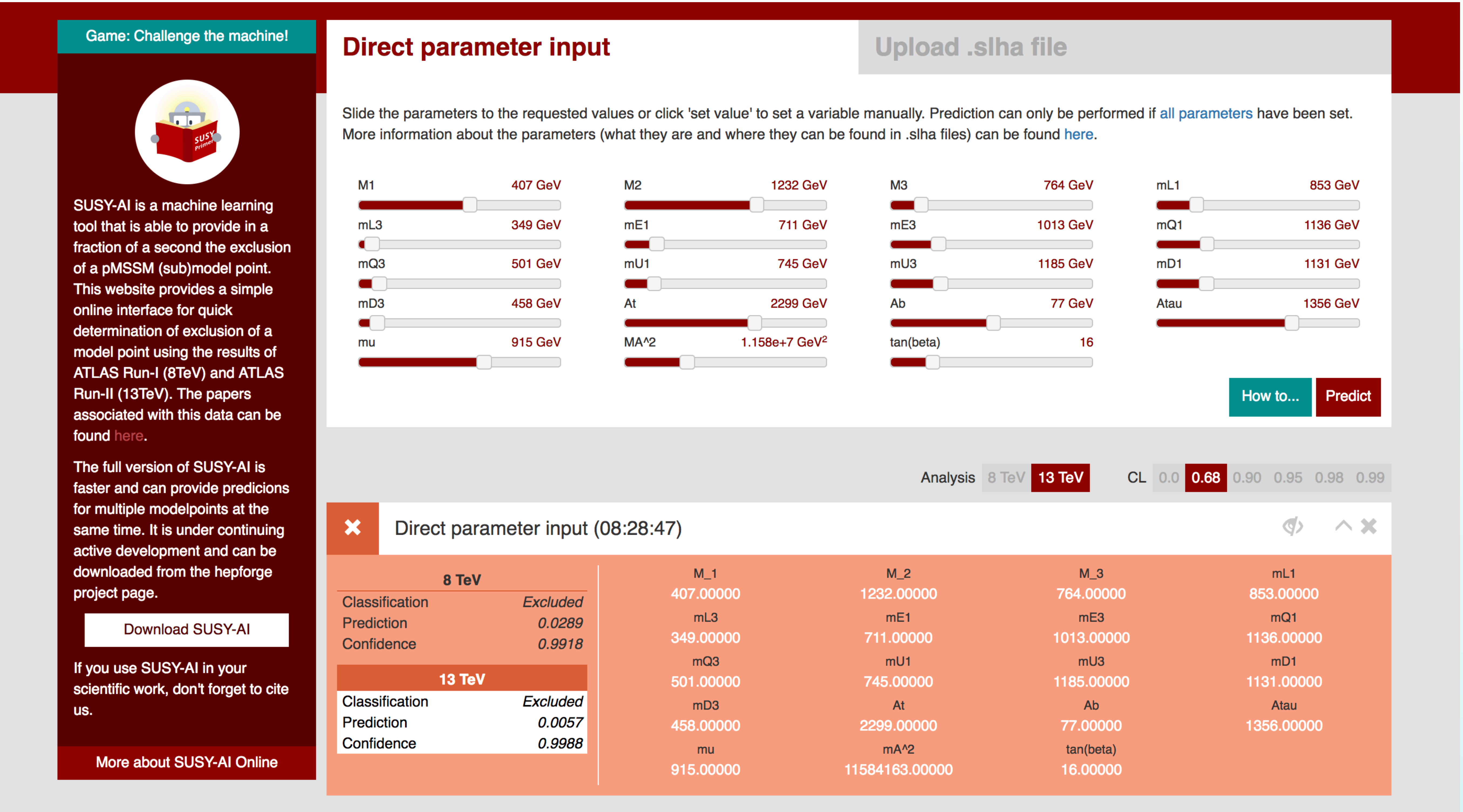
Harvesting the LHC data for BSM signals

Compare classification (allowed vs excluded) in **real data** vs the **ML-trained classifier**



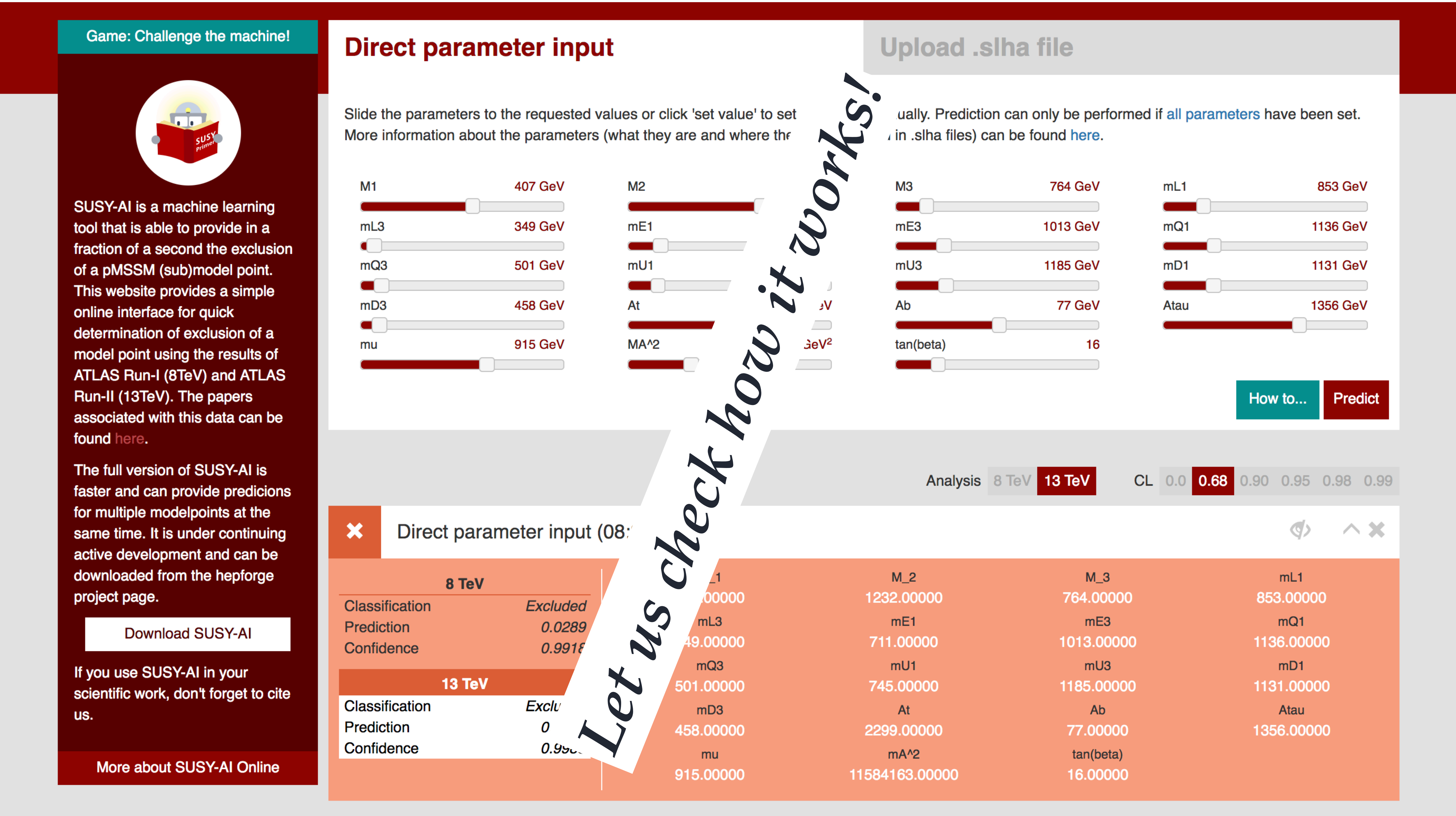
- Very efficient reproduction of the **full bSM parameter space**
- Can be **projected** in any of the dimensions of the 19-parameter space of the pMSSM
- Generalise to points of parameter space **not used for the classifier training** in O(ms) as opposed to O(h)

Harvesting the LHC data for BSM signals



www.susy-ai.com

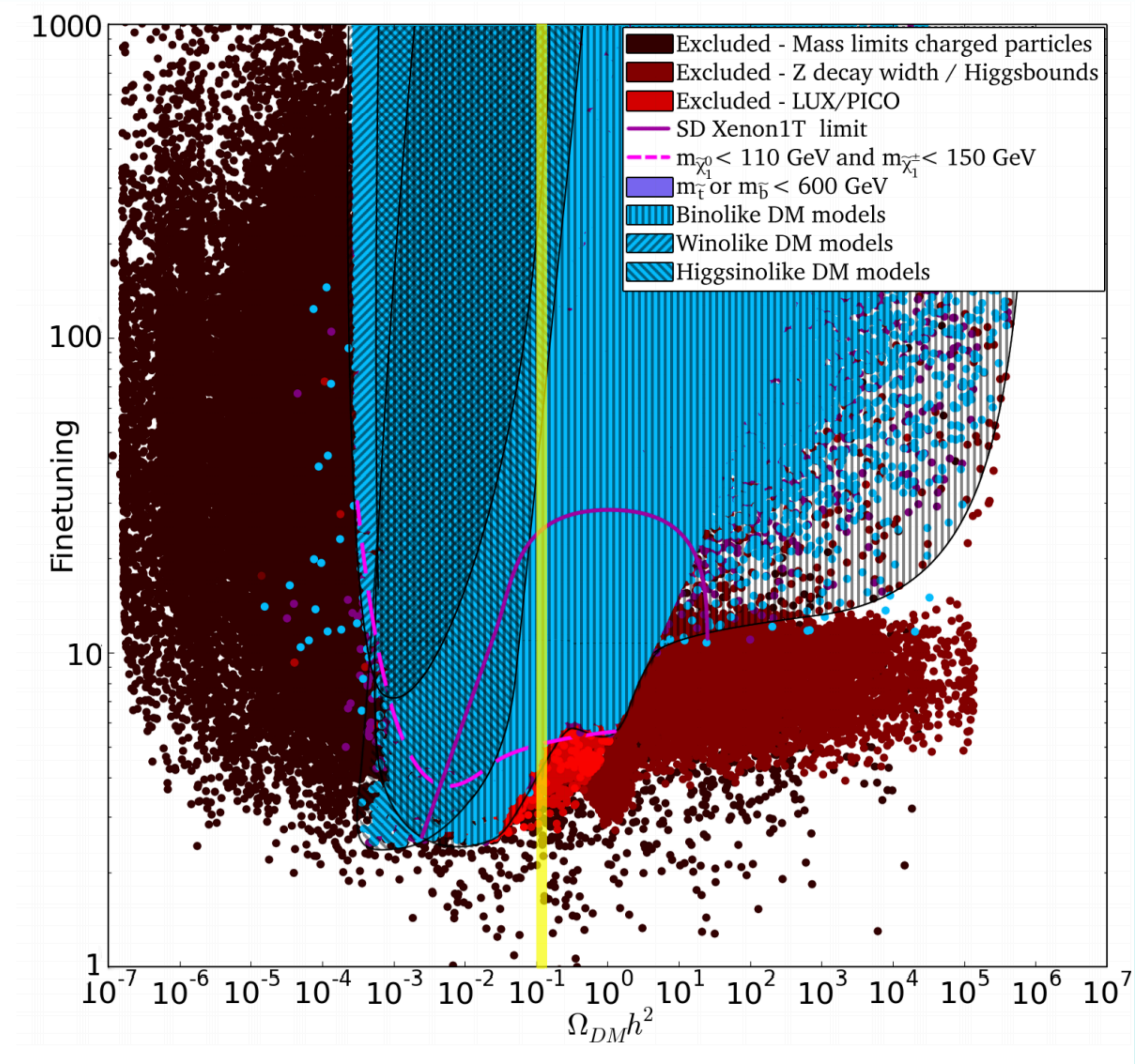
Harvesting the LHC data for BSM signals



Let us check how it works!

www.susy-ai.com

(Un)natural supersymmetry



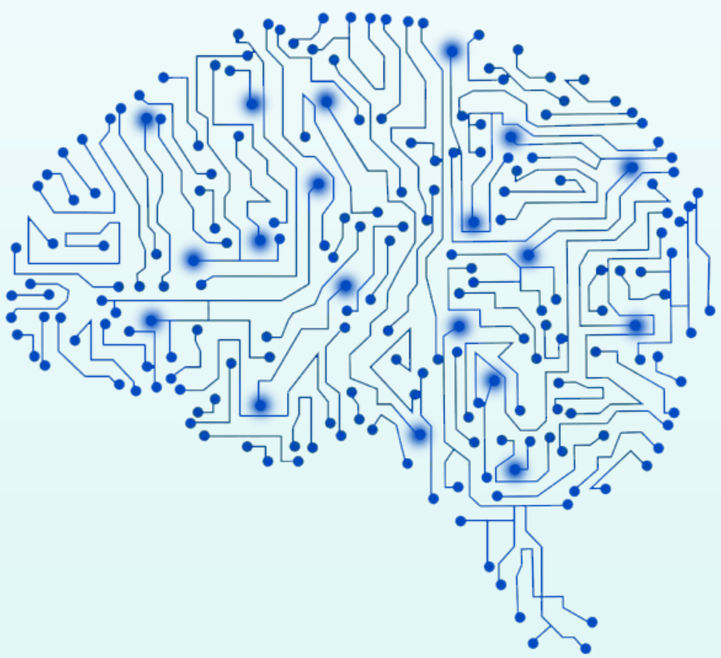
- ***Natural*** regions of the SUSY parameter space still allowed by **current constraints** …
- Being able to assess this requires a **full efficient exploration of the theory parameter space**

arXiv:1612.06333

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

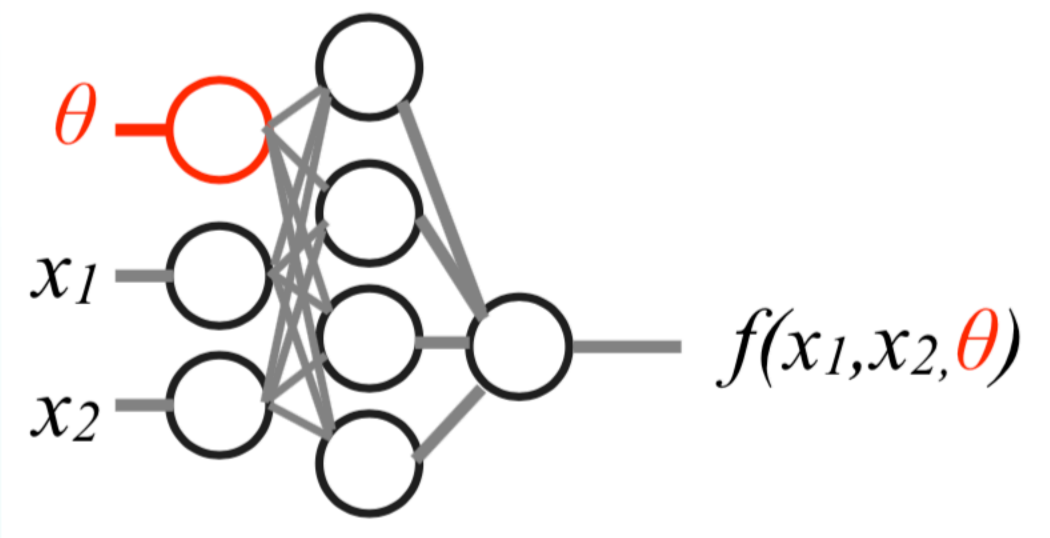


QCD-aware NNs For jet physics

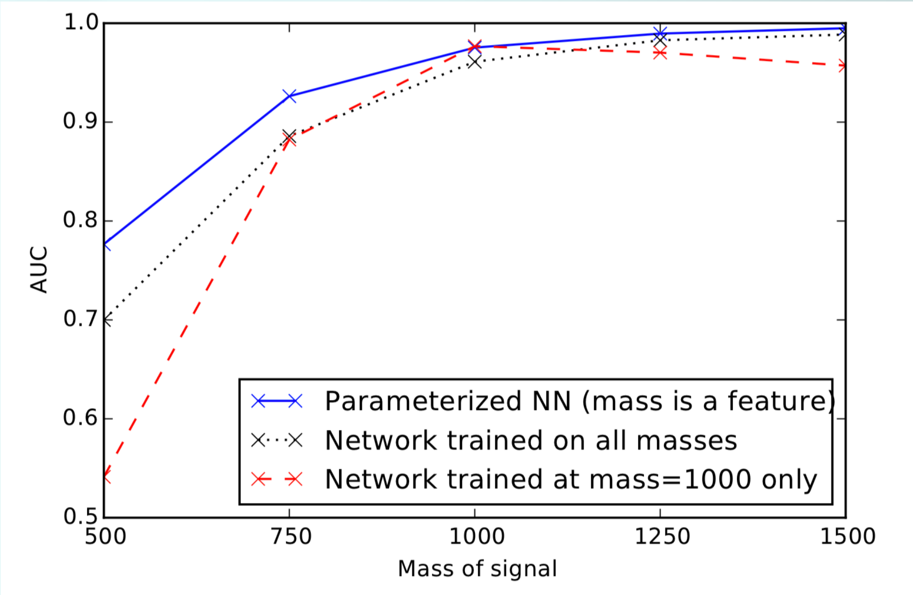
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation



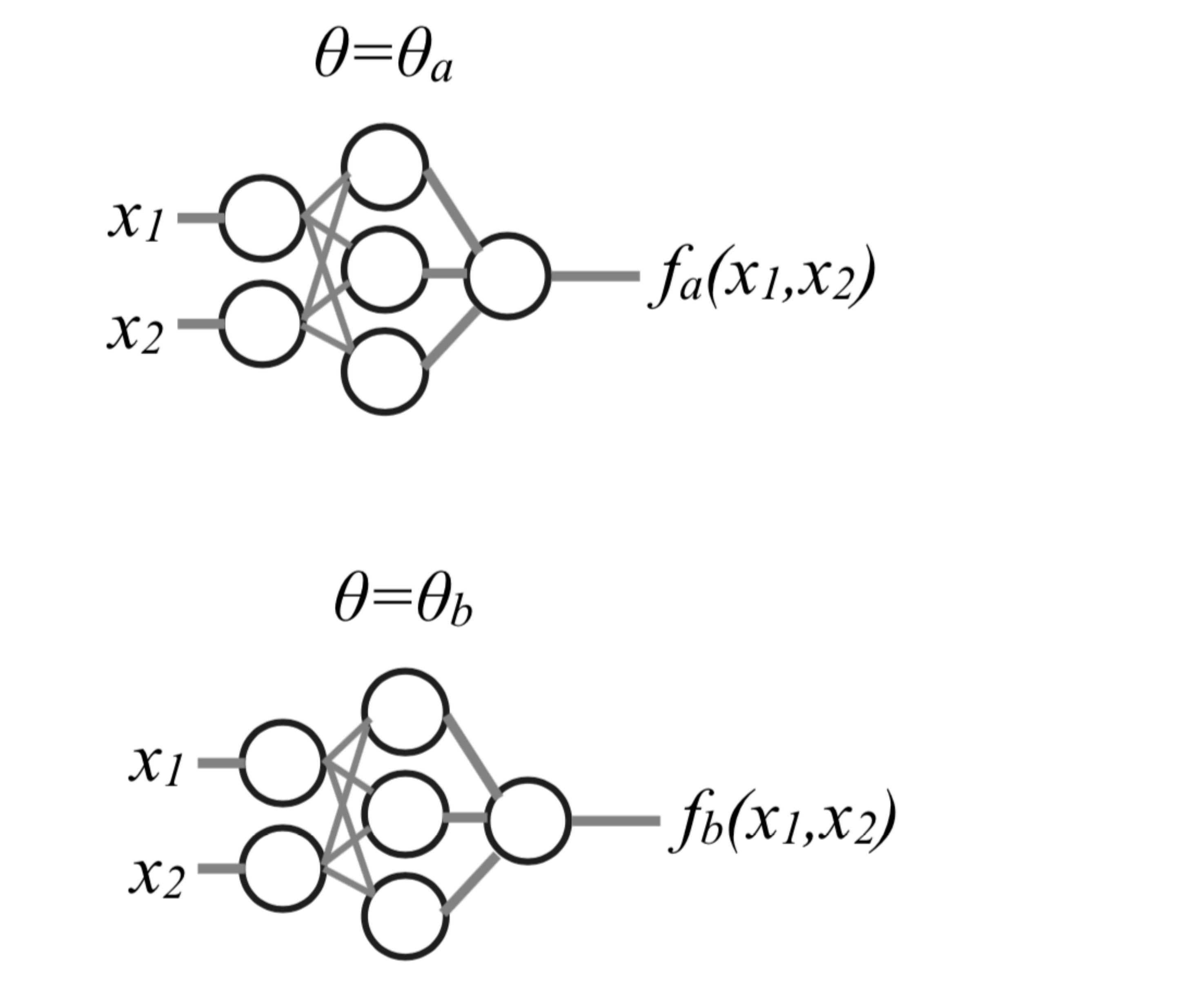
Parametrised Neural Networks for HEP



Pierre Baldo, Kyle Cranmer, Taylor Faucet,
Peter Sadowski, Daniel Whiteson

arXiv:1601.07913

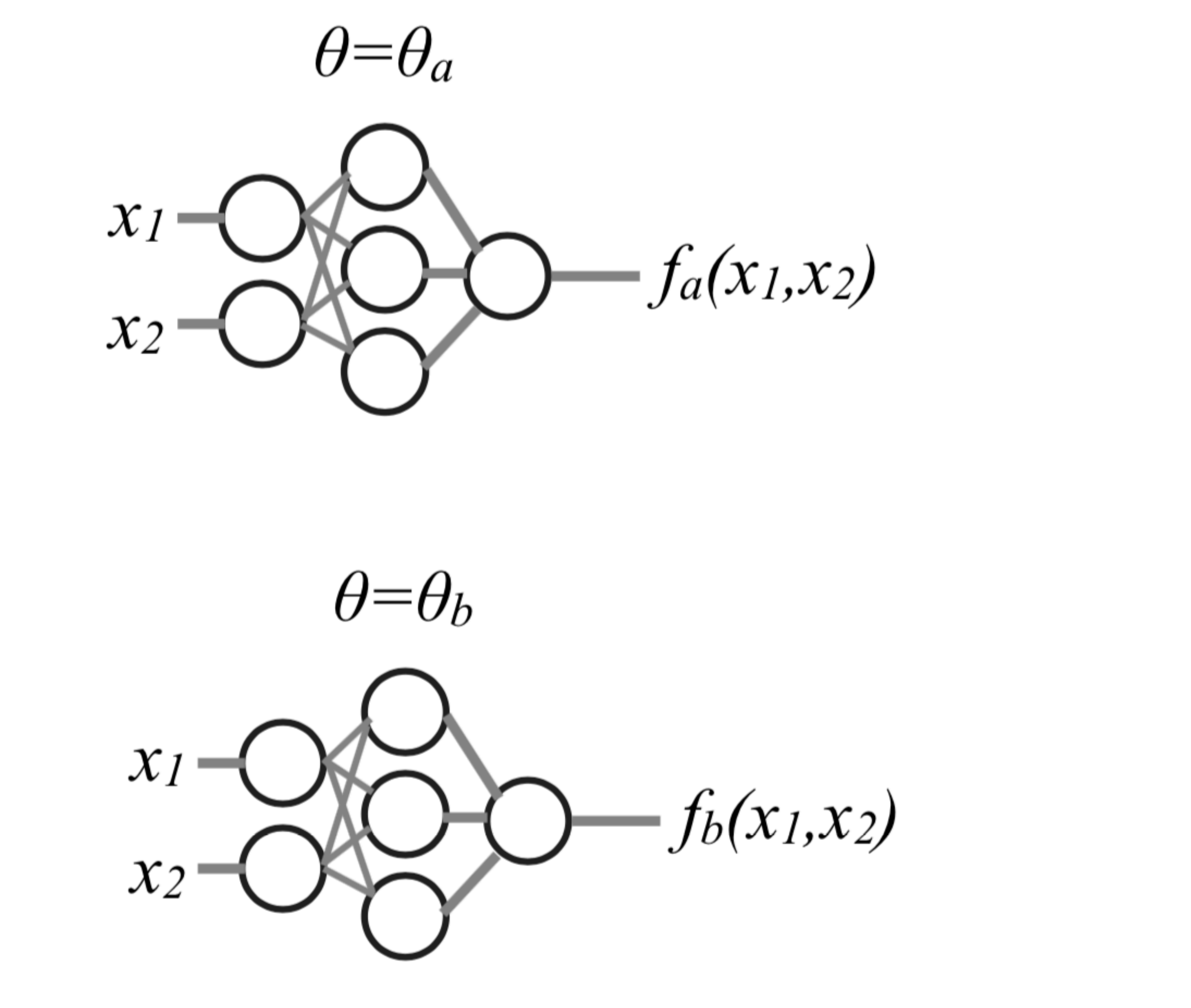
Parametrised ML for HEP



- As shown for the Higgs Pair Production case, **NNs are often used as classifiers**
- Input variables are **event kinematics**, ie, four-momenta or some other higher-level variables

- **Individual networks** trained with examples with a single value of some **parameter θ** (same the mass of some new BSM particle), x_i are the event kinematic variables
- The individual networks are purely functions of the input features.
- Problem: **performance for intermediate values of θ is not optimal** nor does it necessarily vary smoothly between the networks.

Parametrised ML for HEP



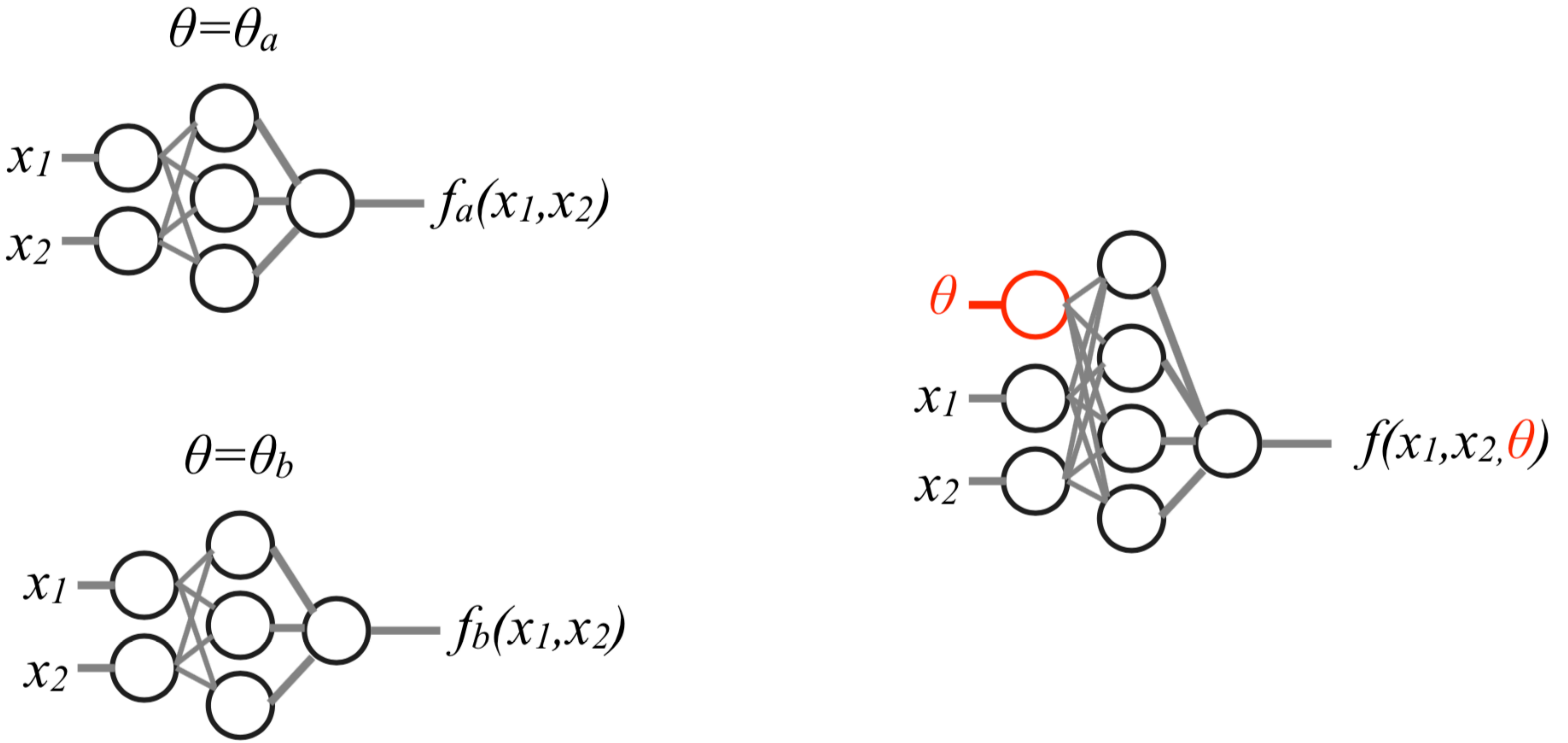
Signal: 5 TeV Z′
Background: QCD

Performance of classifiers will be very different in the two cases!

Signal: 50 GeV Z′ (dark photon)
Background: QCD

- **Individual networks** trained with examples with a single value of some **parameter θ** (same the mass of some new BSM particle), x_i are the event kinematic variables
- The individual networks are purely functions of the input features.
- Problem: **performance for intermediate values of θ is not optimal** nor does it necessarily vary smoothly between the networks.

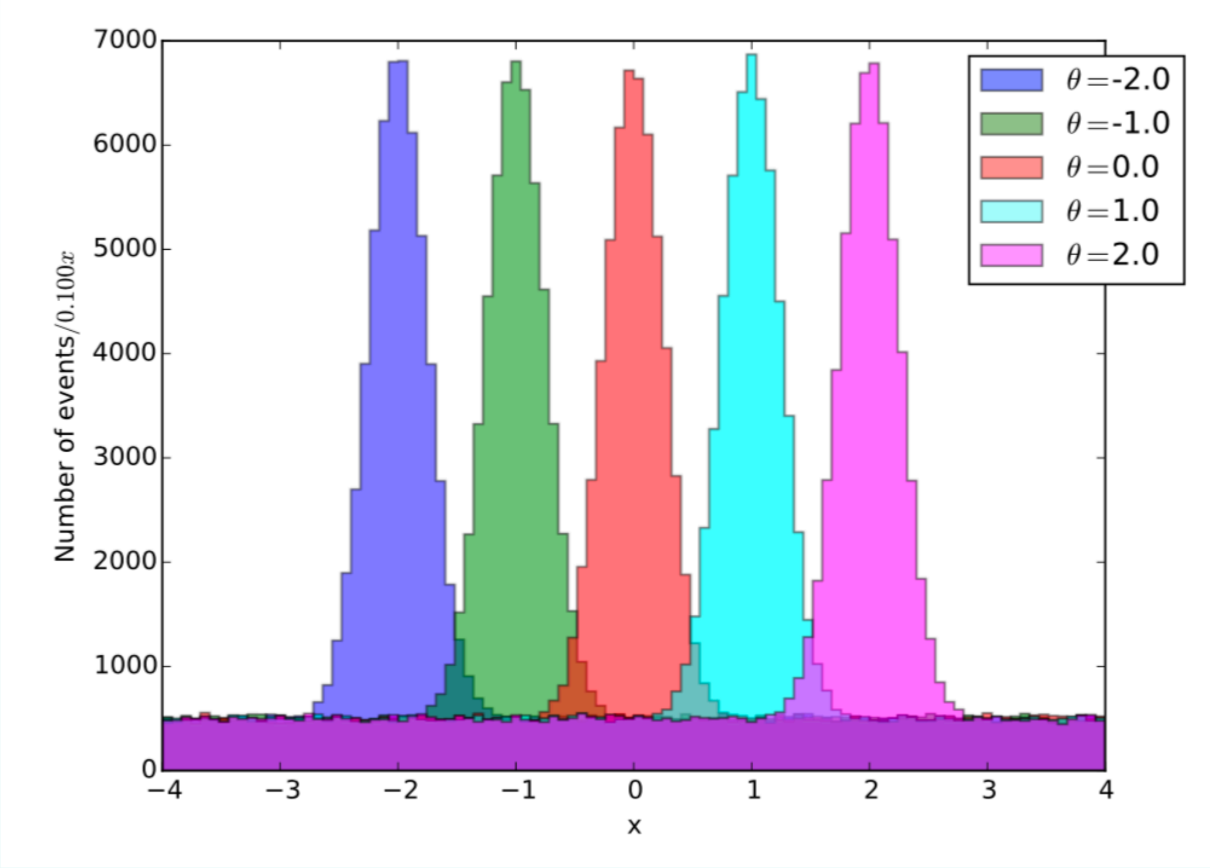
Parametrised ML for HEP

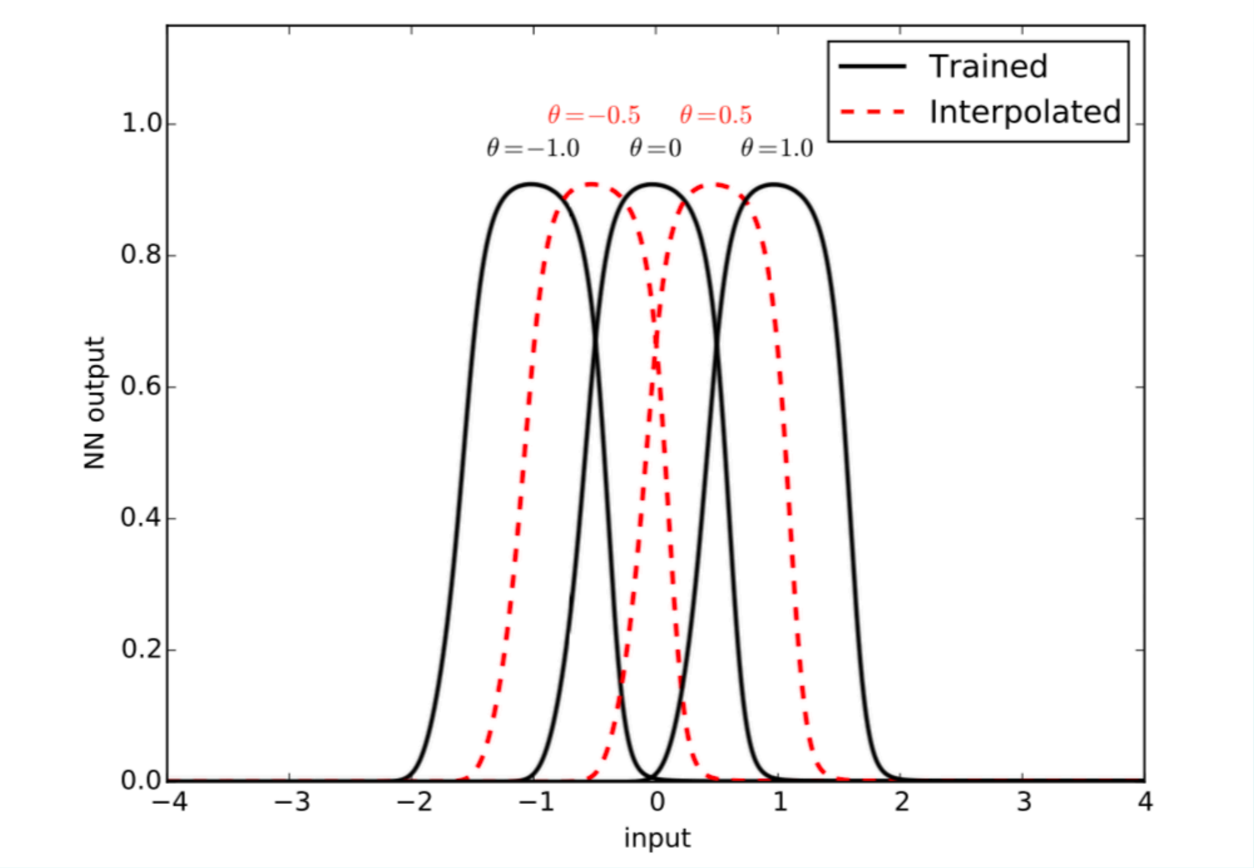


- A single network trained with input features **as well as an input parameter θ**
- Such a network is trained with examples at several values of the parameter θ
- Superior performance in **predicting/describing values of θ not used during the training**

also ensures smooth interpolation and allows one imposing physical constraints

Parametrised ML for HEP



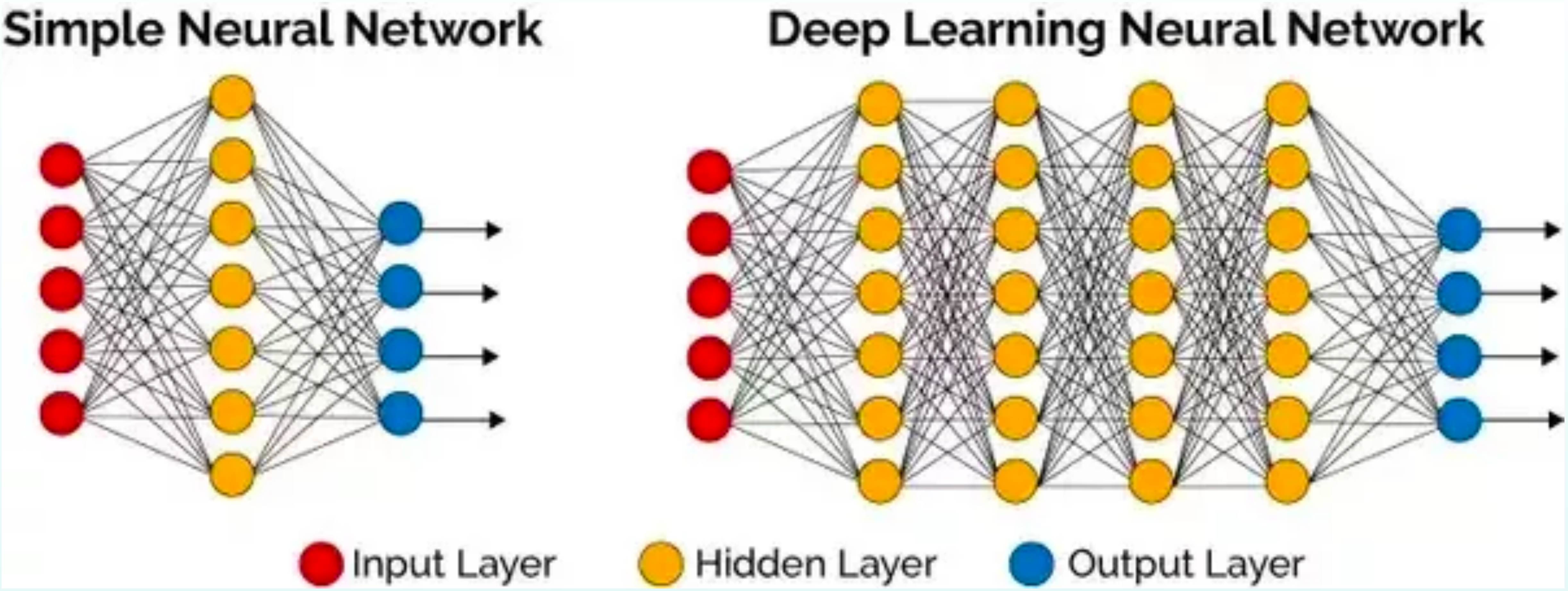


Toy example: the parametrised NN smoothly interpolates for values of the parameter not used in training

- The addition of the **input parameter θ** introduced extra considerations in the training
- The **distribution of θ** used for the training is only relevant in how it affects the quality of the resulting parameterized network: it does not imply that the **resulting inference is Bayesian**
- Also, for some components of the training sample **the values of θ might not be meaningful at all,** for example the specific value of a bSM particle mass does not affect the SM background samples

Use random values of the bSM parameters when training on the SM samples

A word on Deep Neural Networks



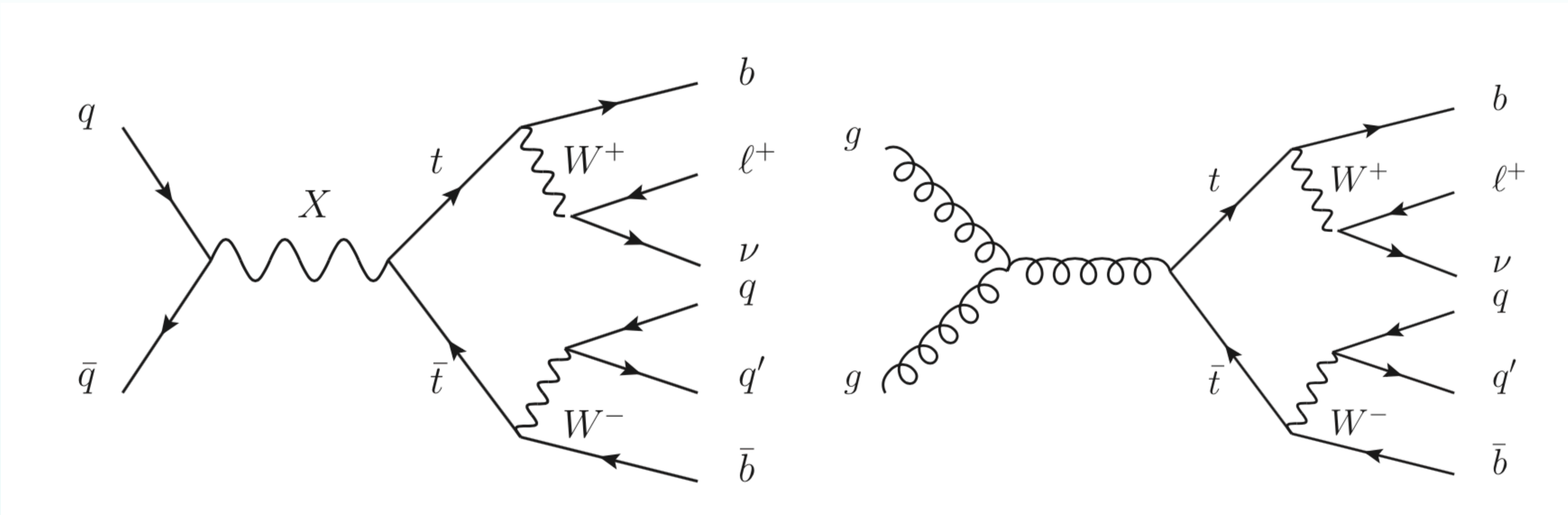
- A **Deep Neural Network (DNN)** is nothing but a standard multi-layer feed-forward perceptron with a large number of internal layers

- All types of neural nets eg **Recursive, Convolutional, Parametrised** etc can be made "deep" by adding more hidden layers

- For several applications, the **increased complexity** achieved this way leads to a significant improvement in performance

Application to bSM searches

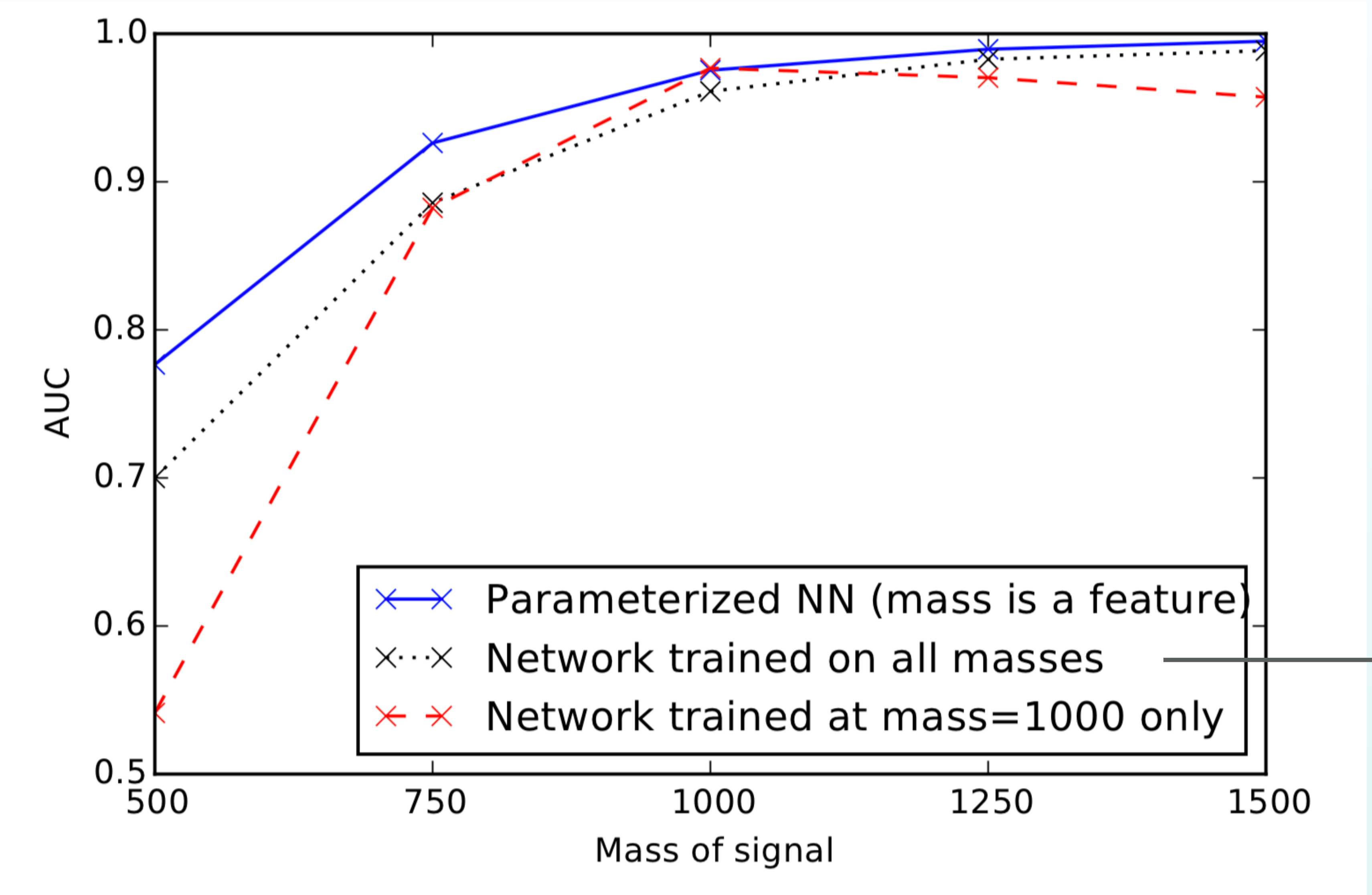
*Consider a **new heavy bSM particle** that decays into top quarks pairs*



- The goal here is to produce an optimised classifier that allows discriminating between signal and background events, **without any prior assumption on the value of m_X**
- **Parameterized deep neural network models** were trained on GPUs using the **Blocks framework**
- The architectures contain five hidden layers of **500 hidden rectified linear units** with a logistic output unit, with **stochastic gradient descent** used for the NN training

Application to bSM searches

Compare performance of parametrised NNs with traditional NNs



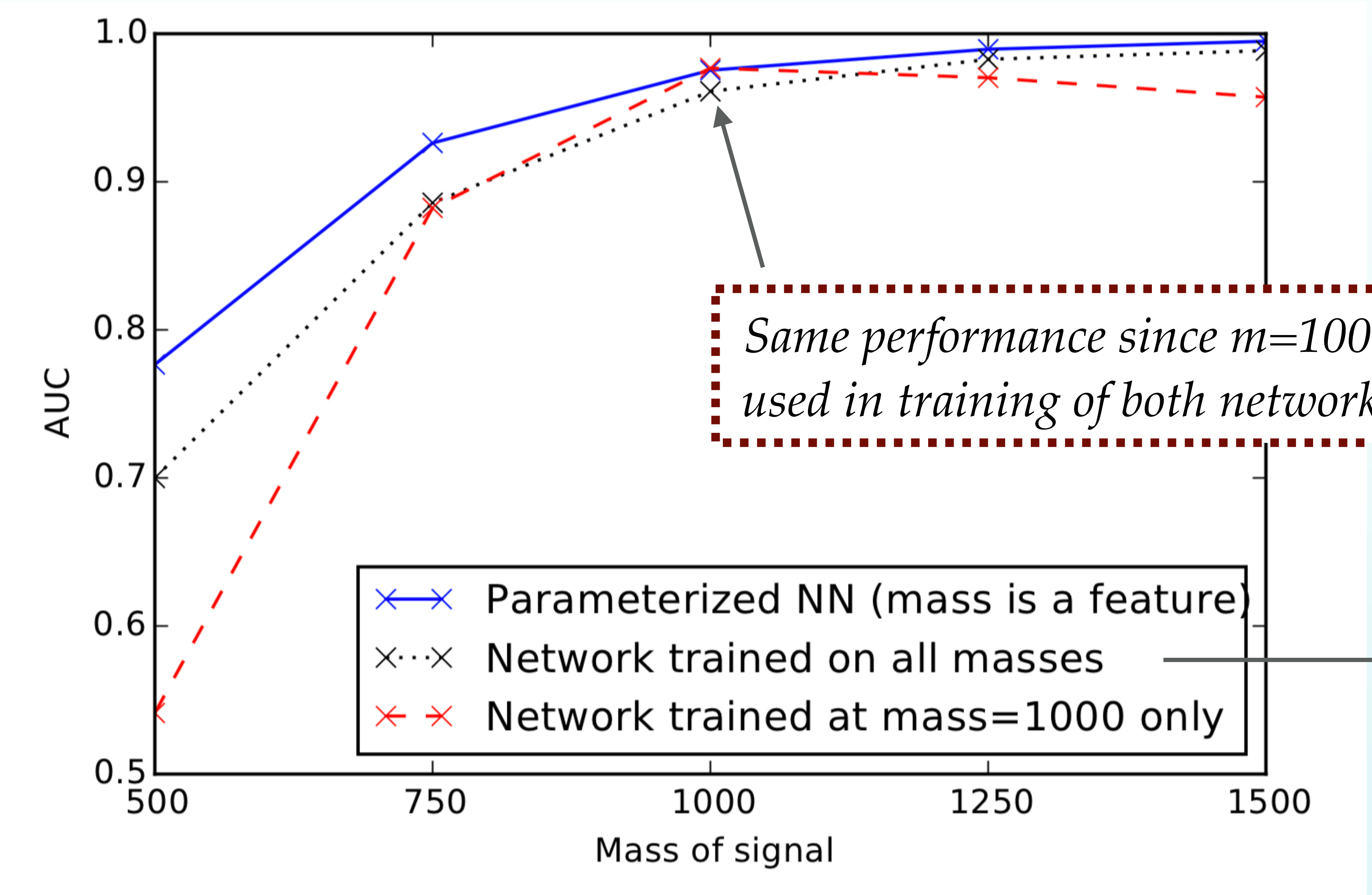
AUC: area under ROC

Unlabelled mass in training samples

- Compare **discrimination** (AUC) for parameterised network and single network trained at 1000 GeV
- The AUC score **decreases for single network as the mass deviates from the trained value**, parameterised network improved performance;
- Performance a single network trained with an unlabelled mixture of signal samples at all masses is inferior to that of the parametrised network

Application to bSM searches

Compare performance of parametrised NNs with traditional NNs



Same performance since m=1000 used in training of both networks

AUC: area under ROC

Unlabelled mass in training samples

- Compare **discrimination** (AUC) for parameterised network and single network trained at 1000 GeV
- The AUC score **decreases for single network as the mass deviates from the trained value**, parameterised network improved performance;
- Performance a single network trained with an unlabelled mixture of signal samples at all masses is inferior to that of the parametrised network

More on DNNs for BSM searches

Searching for Exotic Particles in High-Energy Physics with Deep Learning

P. Baldi,[1] P. Sadowski,[1] and D. Whiteson[2]

[1] *Dept. of Computer Science, UC Irvine, Irvine, CA 92617* *

[2] *Dept. of Physics and Astronomy, UC Irvine, Irvine, CA 92617* †

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine learning approaches are often used. Standard approaches have relied on 'shallow' machine learning models that have a limited capacity to learn complex non-linear functions of the inputs, and rely on a pain-staking search through manually constructed non-linear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Using benchmark datasets, we show that deep learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep learning approaches can improve the power of collider searches for exotic particles.

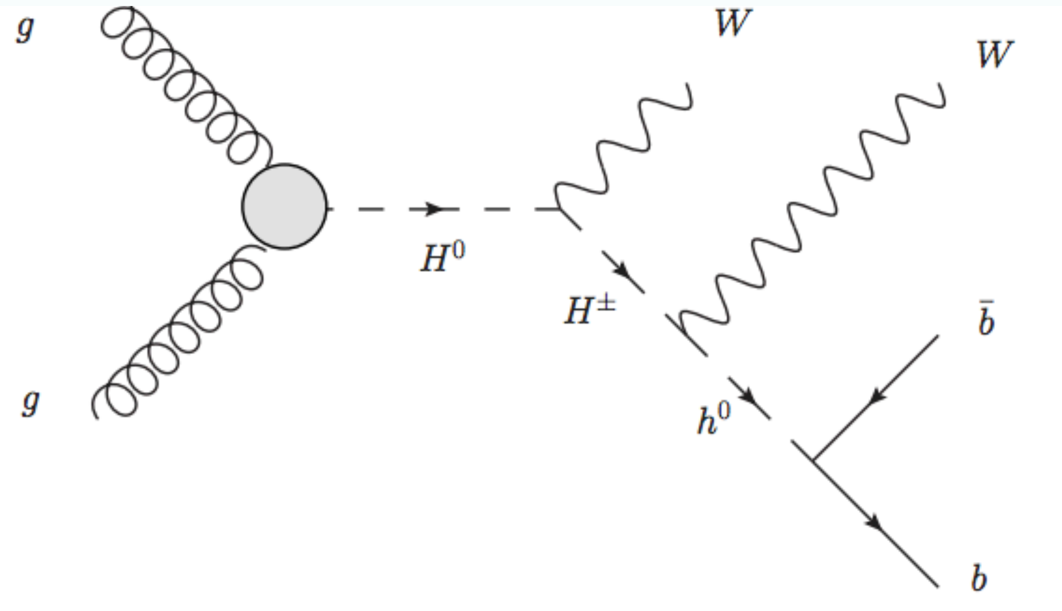
Compare the performance of the discrimination between **shallow and deep neural networks**

Here use a **five-layer NNs** with **300 hidden units** in each layer

Compare also with the performance of the Boosted Decision Tree

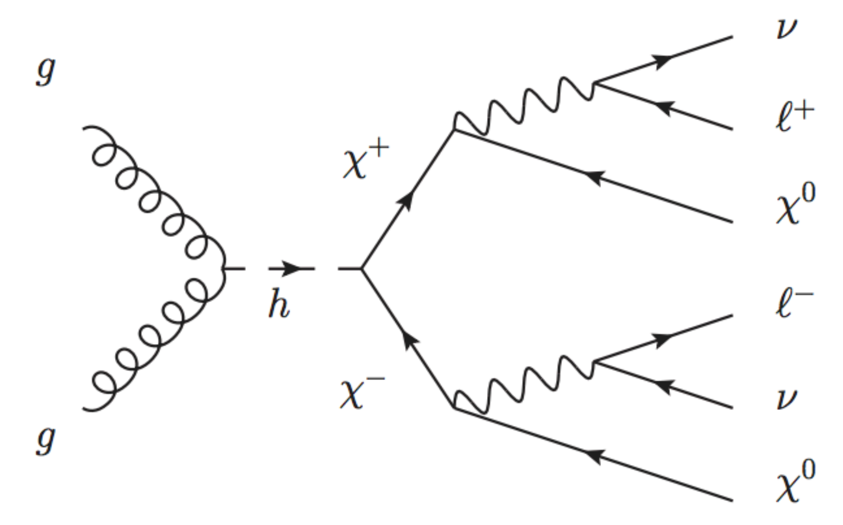
More on DNNs for BSM searches

Signal: heavy Higgs production (eg 2HDM)



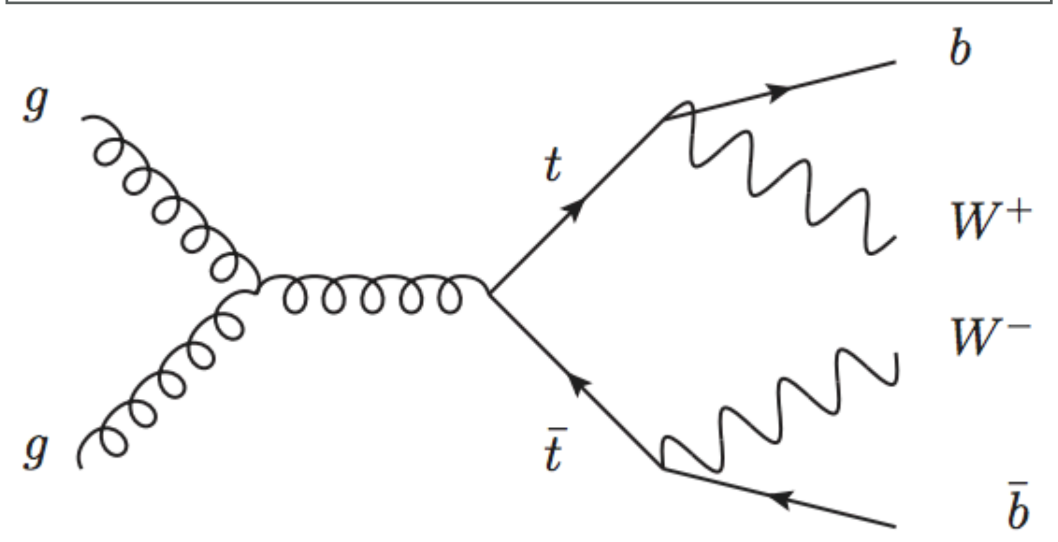
(a)

Signal: Chargino pair production in SUSY



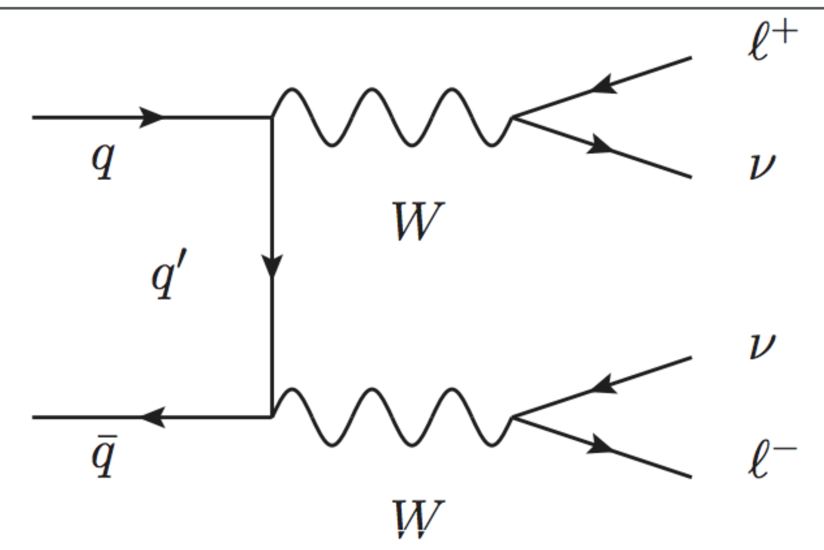
(a)

Background: QCD top-quark pair



(b)

Background: W pair production



(b)

More on DNNs for BSM searches

Higgs benchmark scenario

	AUC		
Technique	Low-level	High-level	Complete
BDT	0.73 (0.01)	0.78 (0.01)	0.81 (0.01)
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
DN	0.880 (0.001)	0.800 (< 0.001)	0.885 (0.002)
	Discovery significance		
Technique	Low-level	High-level	Complete
NN	2.5σ	3.1σ	3.7σ
DN	4.9σ	3.6σ	5.0σ

Using the right classifier can make a difference between ``evidence" and ``discovery"

SUSY benchmark scenario

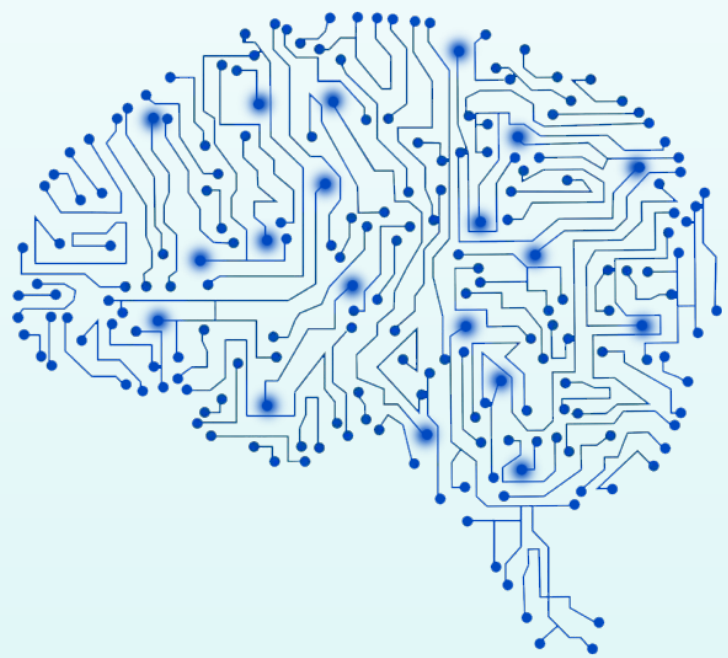
	AUC		
Technique	Low-level	High-level	Complete
BDT	0.850 (0.003)	0.835 (0.003)	0.863 (0.003)
NN	0.867 (0.002)	0.863 (0.001)	0.875 (< 0.001)
$NN_{dropout}$	0.856 (< 0.001)	0.859 (< 0.001)	0.873 (< 0.001)
DN	0.872 (0.001)	0.865 (0.001)	0.876 (< 0.001)
$DN_{dropout}$	0.876 (< 0.001)	0.869 (< 0.001)	0.879 (< 0.001)
	Discovery significance		
Technique	Low-level	High-level	Complete
NN	6.5σ	6.2σ	6.9σ
DN	7.5σ	7.3σ	7.6σ

The improvement by using deep networks is moderate here

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

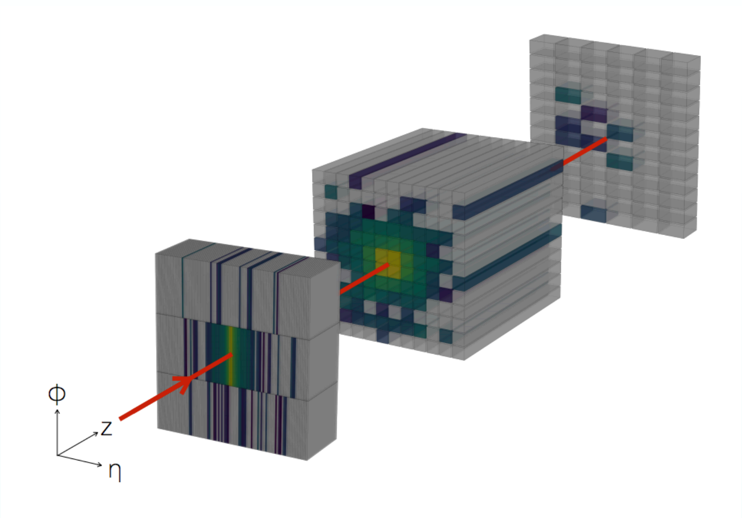


QCD-aware NNs For jet physics

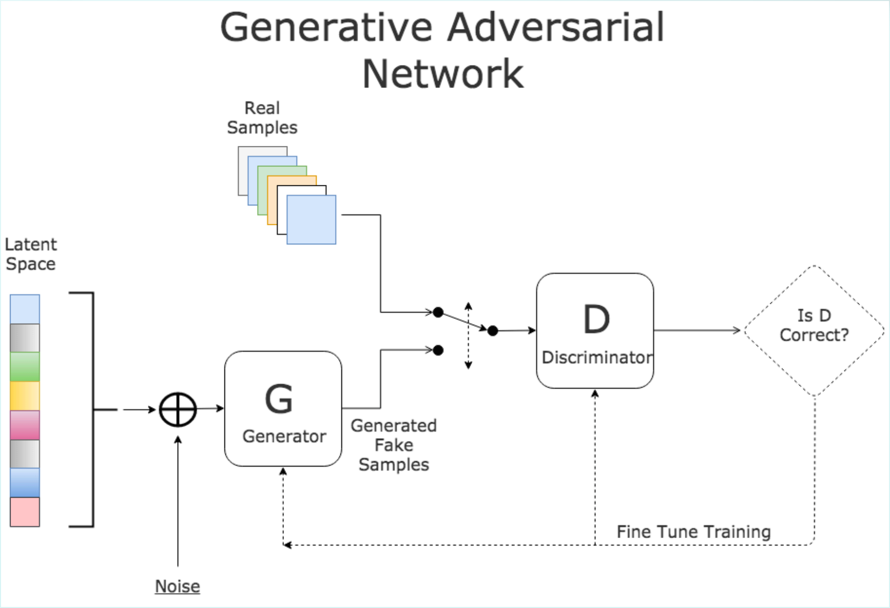
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation



Simulating HEP detectors with Generative Adversarial Networks

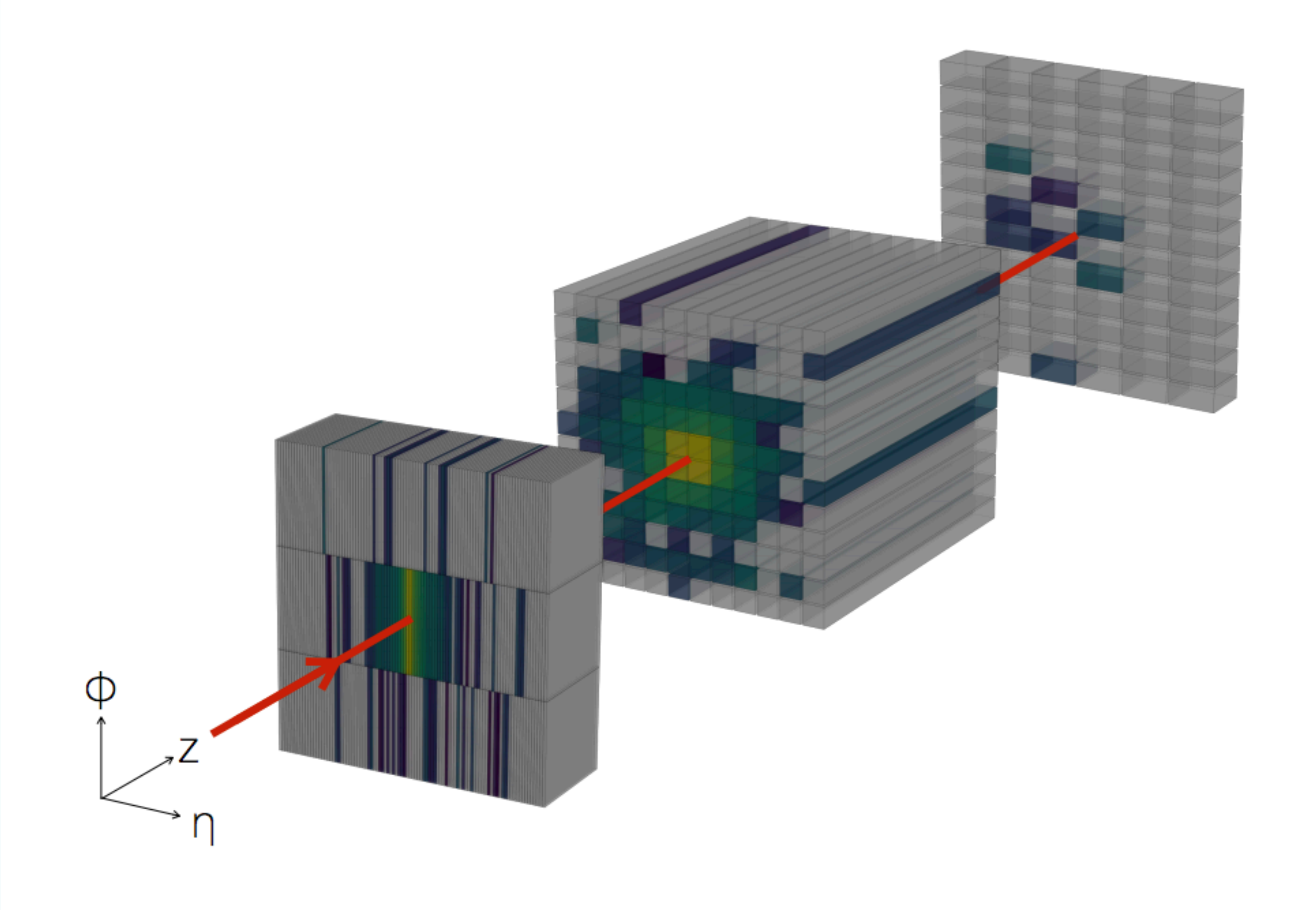


Paganini, de Oliveira, Nachman

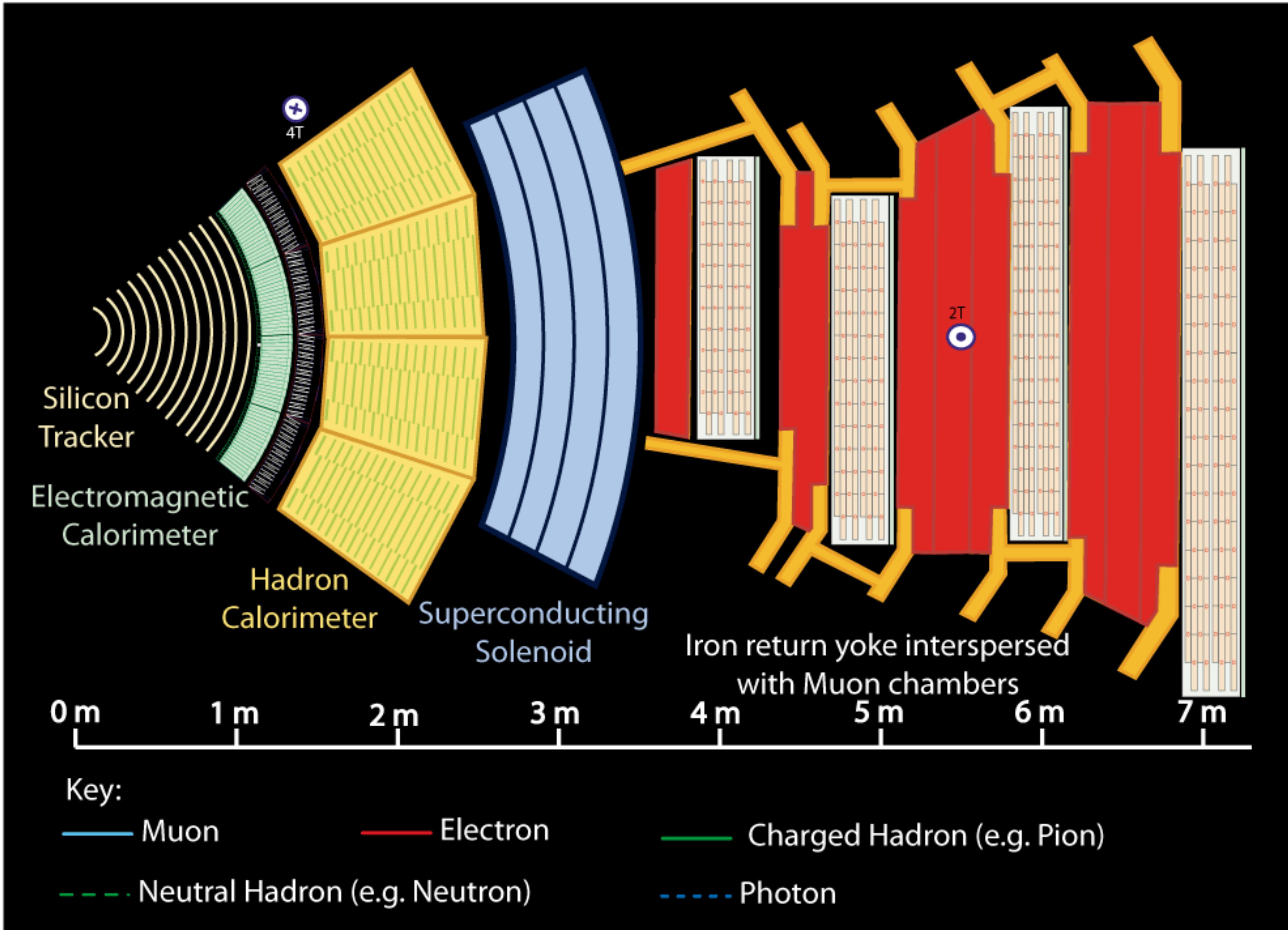
arxiv:1712.10321

The name of the game

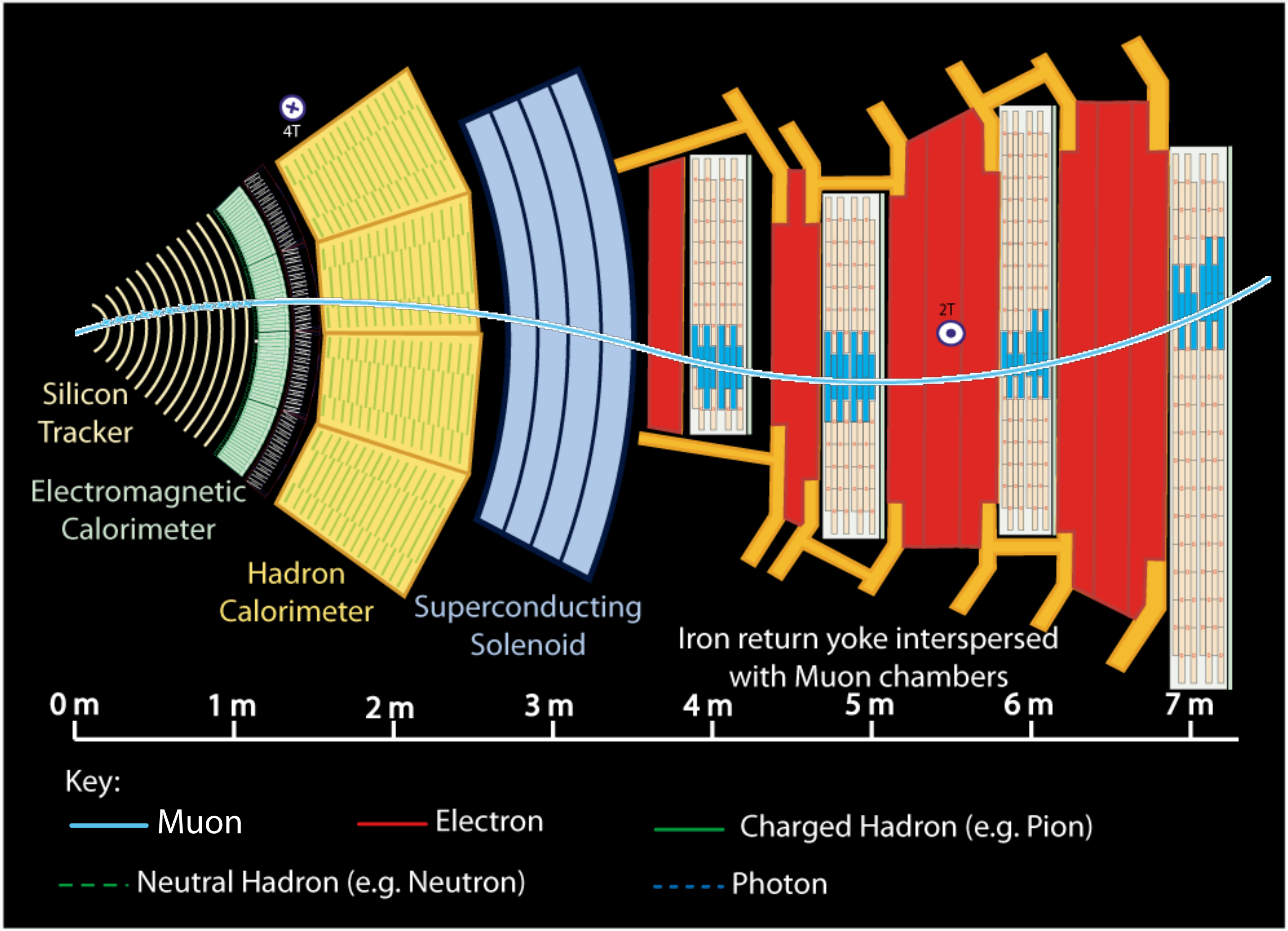
- Modelling accurately the response of detectors with the **propagation of high energy particles** is an essential task for present and future HEP experiment
- **Detector simulation at the LHC** is a very CPU-intensive task, dominated by modelling of particle showers inside calorimeters
- **Generative Adversarial Networks** can speed up detector simulation by orders of magnitude



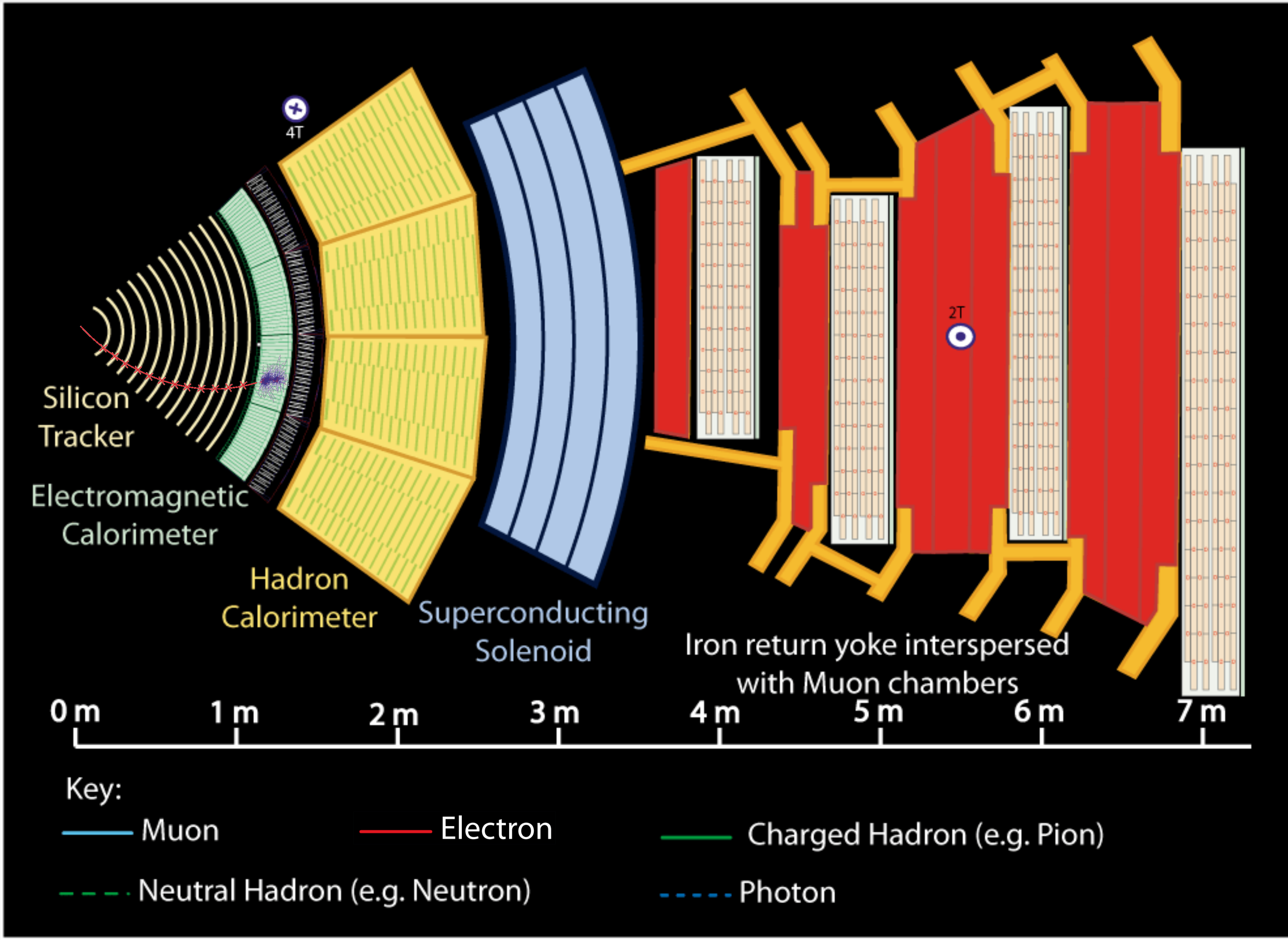
*Task: to efficiently model the **propagation of high energy particles** (and their interaction) within the layers of electromagnetic and hadronic calorimeters*

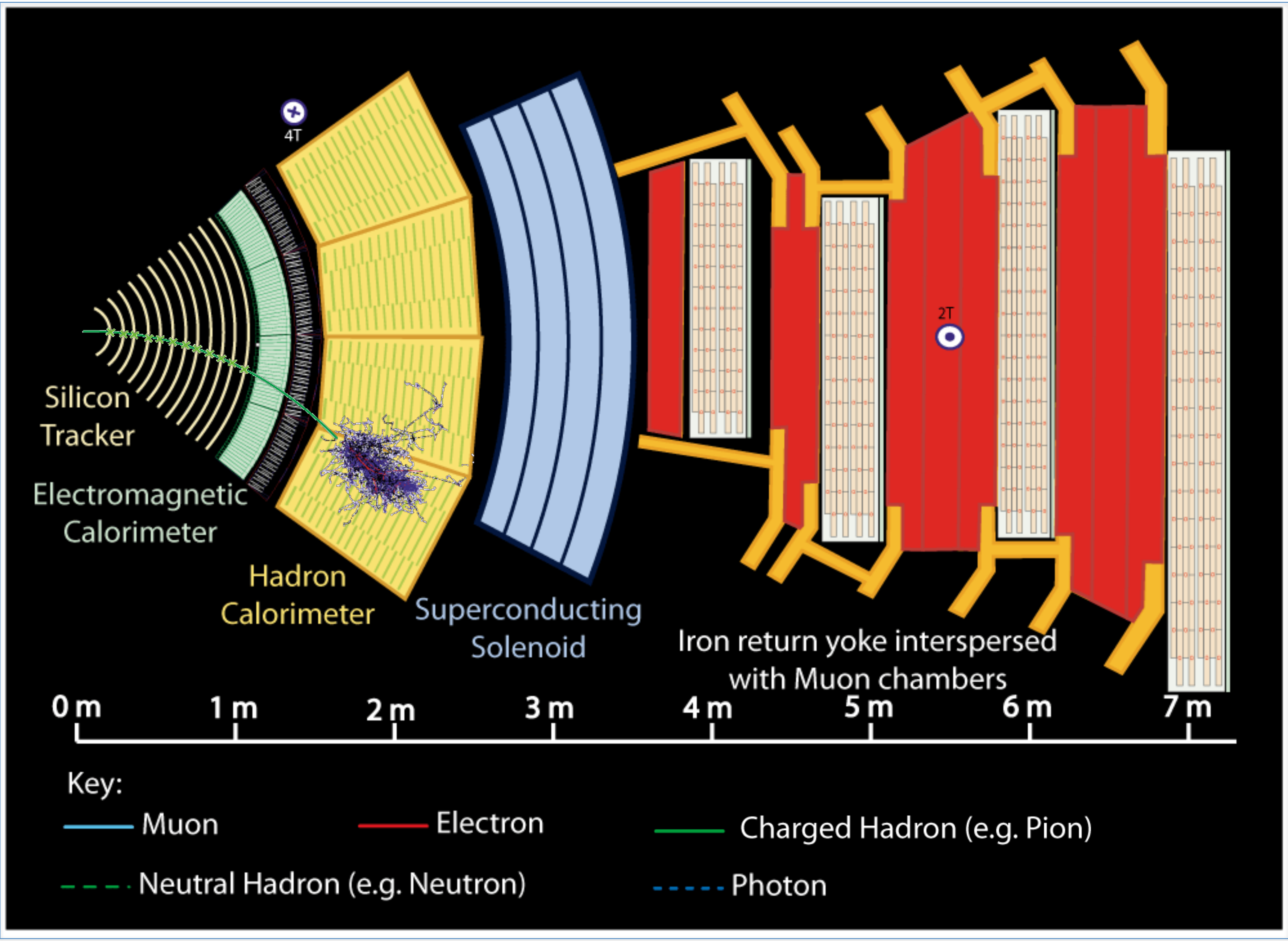


4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

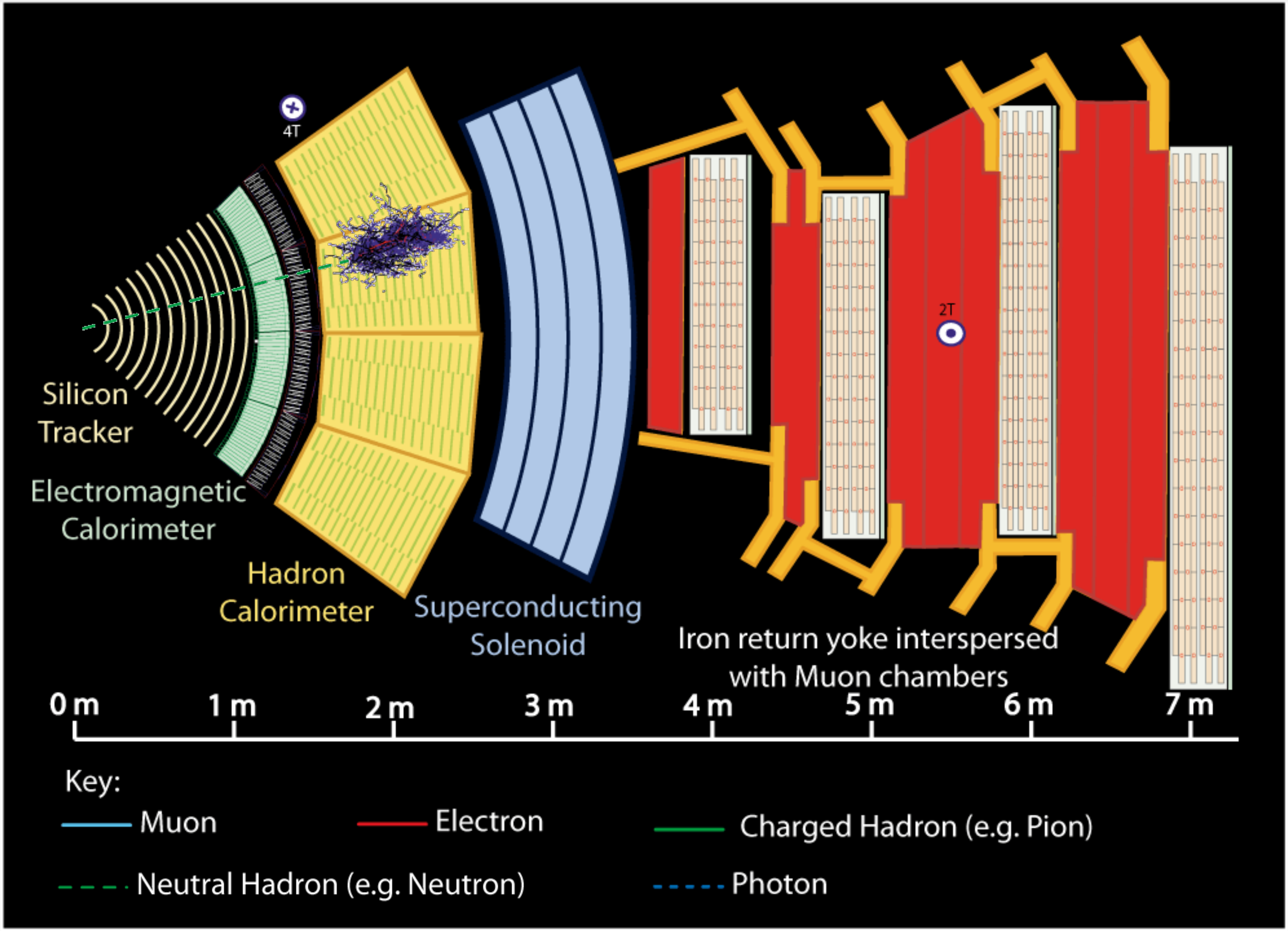


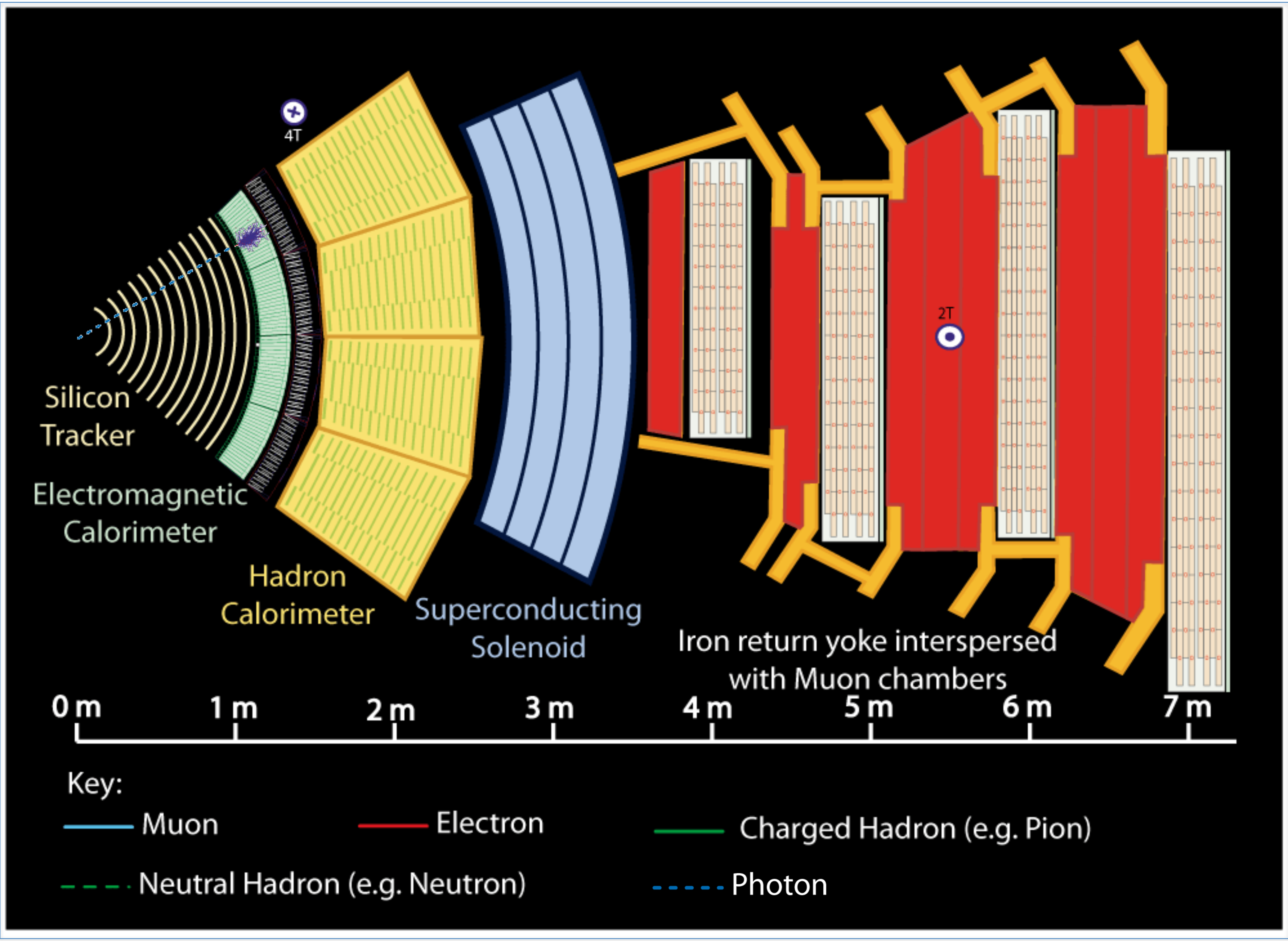
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon


4T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
2T
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon



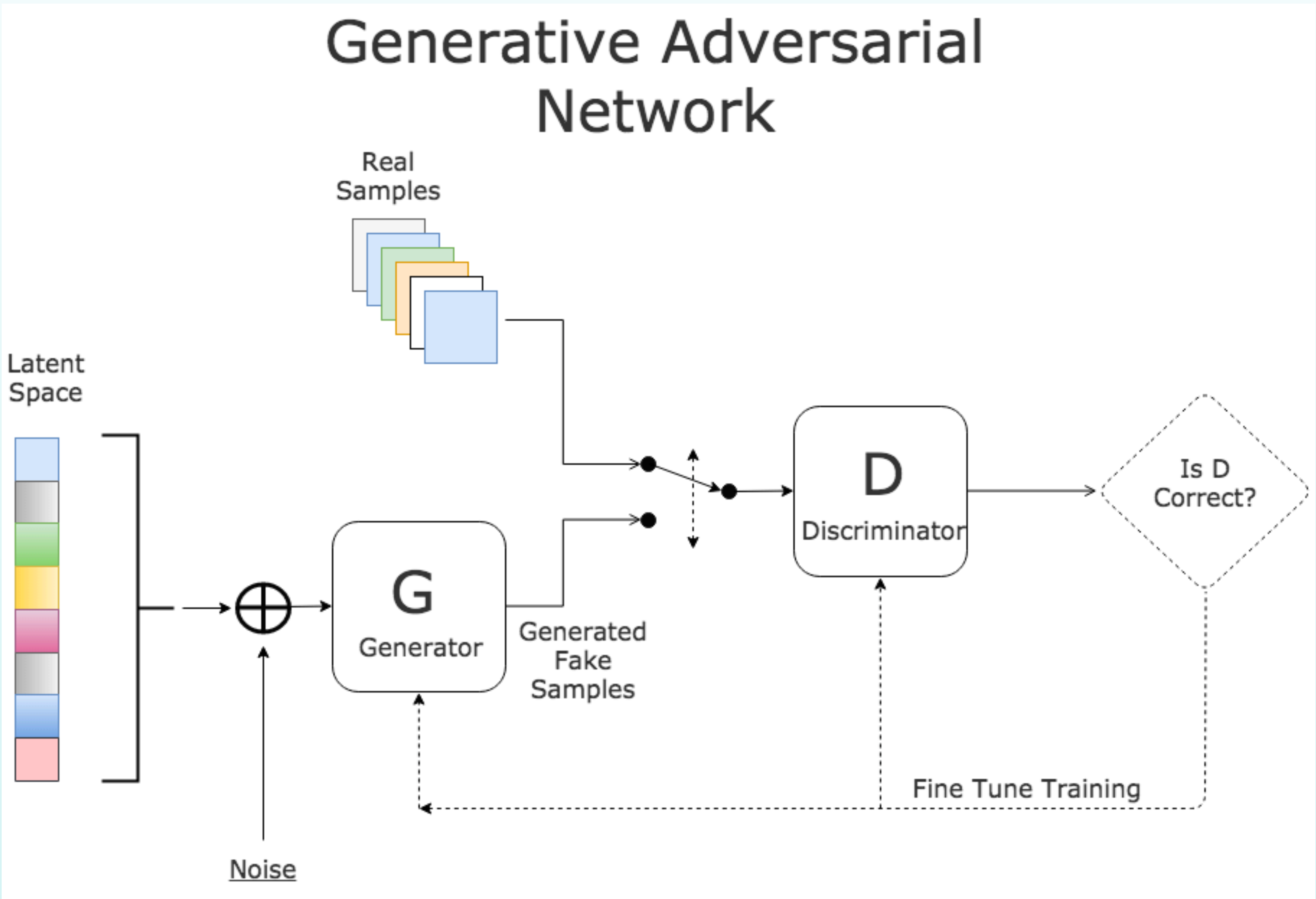
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon


4T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
2T
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon


4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

A word on GANs

- New architecture for an **unsupervised neural network training** (unlabelled samples)
- Based on two **independent nets** that work separately and act as adversaries:
  - the **Discriminator (D)** undergoes training and plays the role of classifier, and
  - the **Generator (G)** and is tasked to generate random samples that **resemble real samples** with a twist rendering them as fake samples.

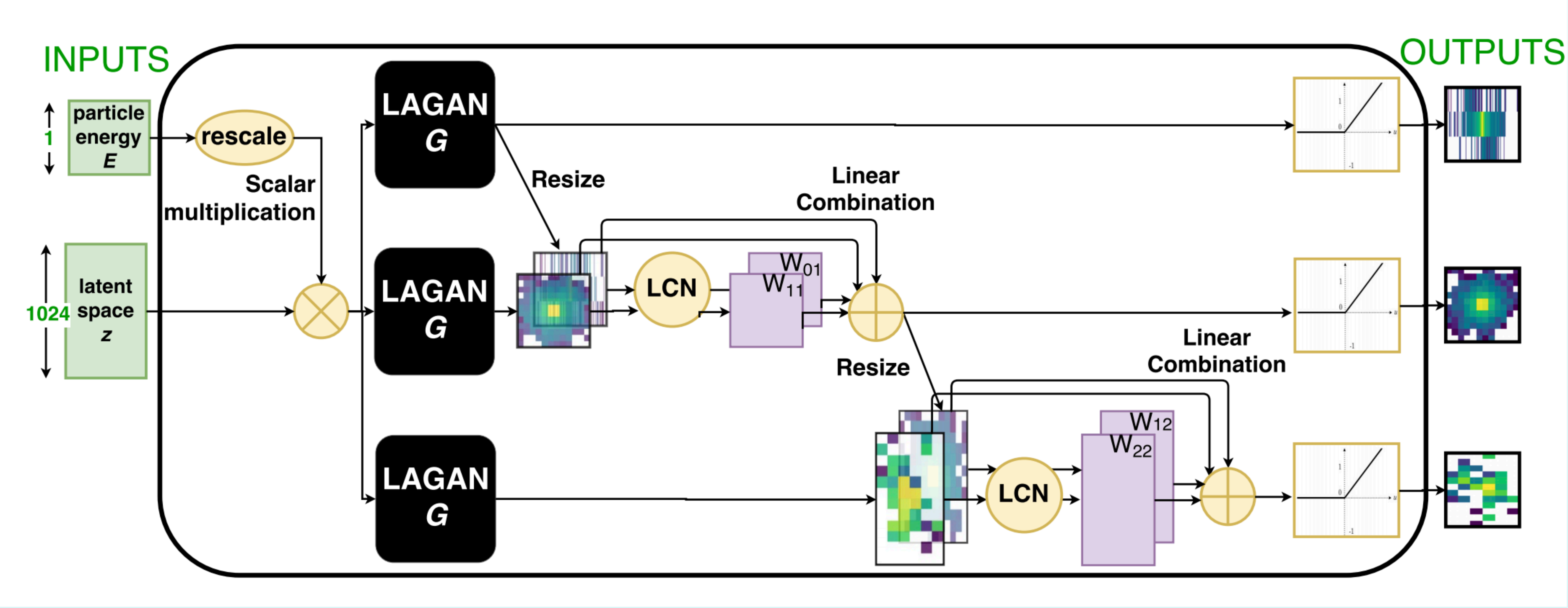


CaloGANs

- Use GANs as a tool to **speed up full simulation of particle showers** in a HEP calorimeter
- The generator G learns a map from **a latent space** to the space of **generated samples** for training
- Carefully understanding the **underlying physics of particle propagation** in a detector is crucial to set up and optimise the training strategy, *e.g.* relationships between neighbouring detector layers

Generator Network

*Output: **generated samples** to be used for training D*



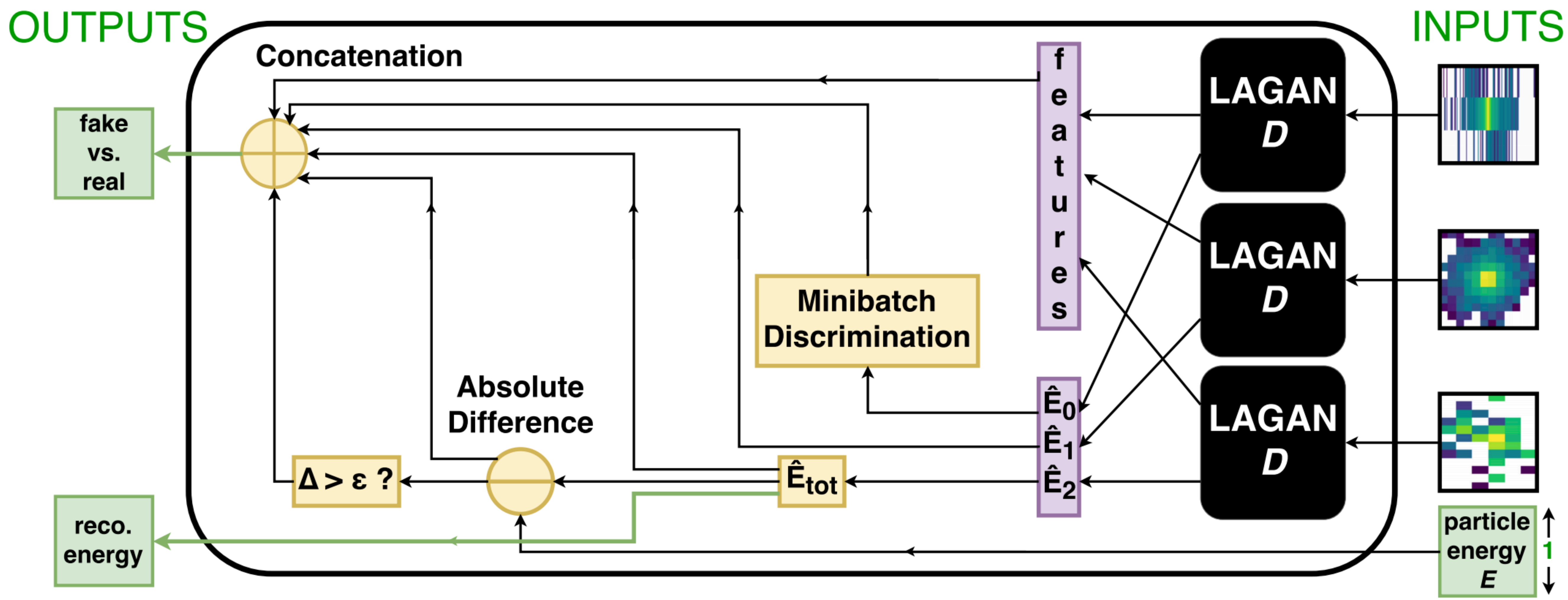
CaloGANs

- Use GANs as a tool to speed up full simulation of particle showers in a HEP calorimeter
- The generator G learns a map from **a latent space** to the space of **generated samples** for training
- Carefully understanding the underlying physics of particle propagation in a detector is crucial to set up and optimise the training strategy, *e.g.* relationships between neighbouring detector layers

Output:
i) Is sample real or fake?
ii) What is the reconstructed particle energy as function of true energy?

Discriminator Network

Input: generated samples



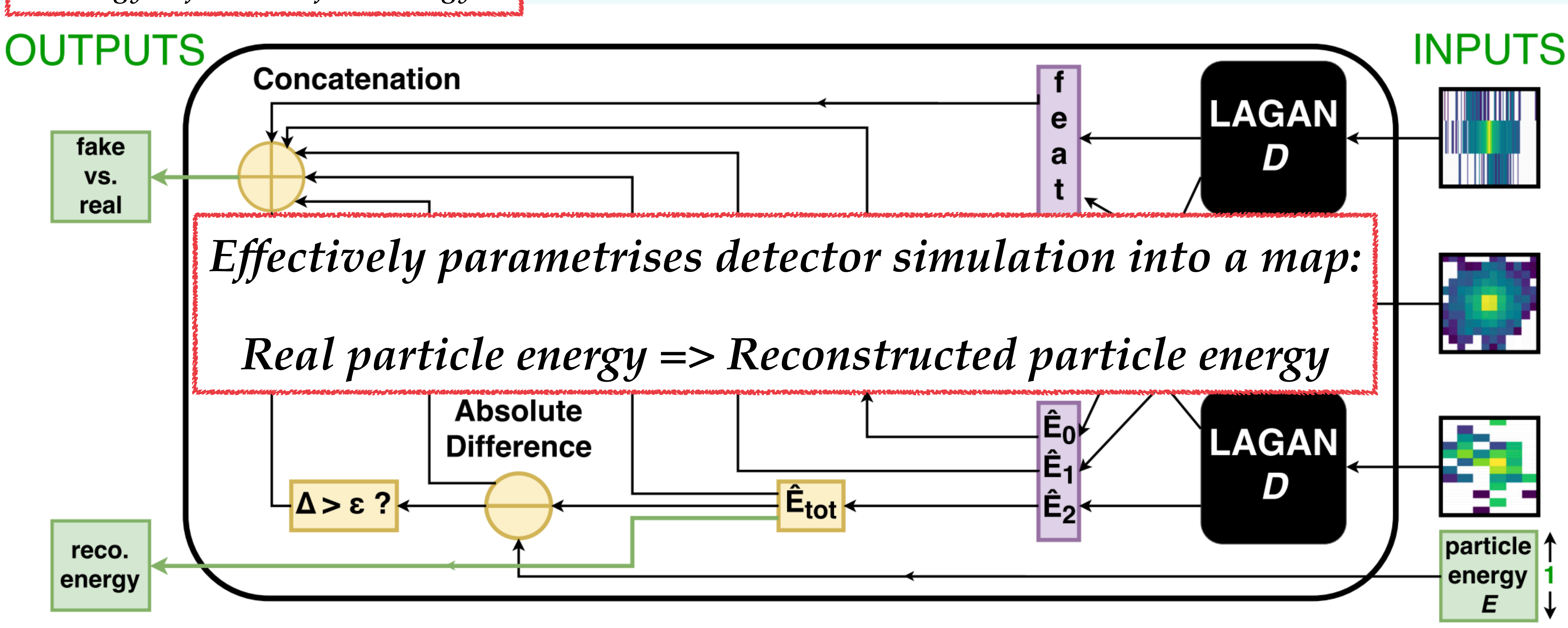
CaloGANs

- Use GANs as a tool to speed up full simulation of particle showers in a HEP calorimeter
- The generator G learns a map from **a latent space** to the space of **generated samples** for training
- Carefully understanding the underlying physics of particle propagation in a detector is crucial to set up and optimise the training strategy, *e.g.* relationships between neighbouring detector layers

Output:
i) Is sample real or fake?
ii) What is the reconstructed particle energy as function of true energy?

Discriminator Network

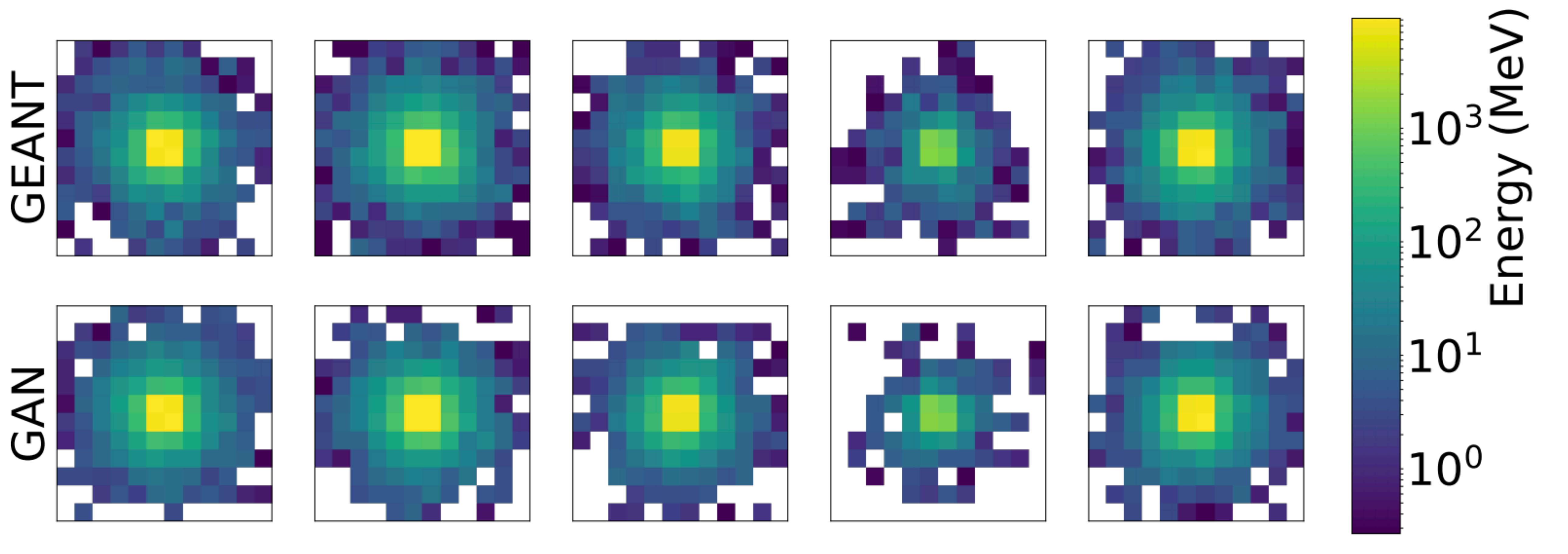
Input: generated samples



Effectively parametrises detector simulation into a map:

Real particle energy => Reconstructed particle energy

CaloGANs

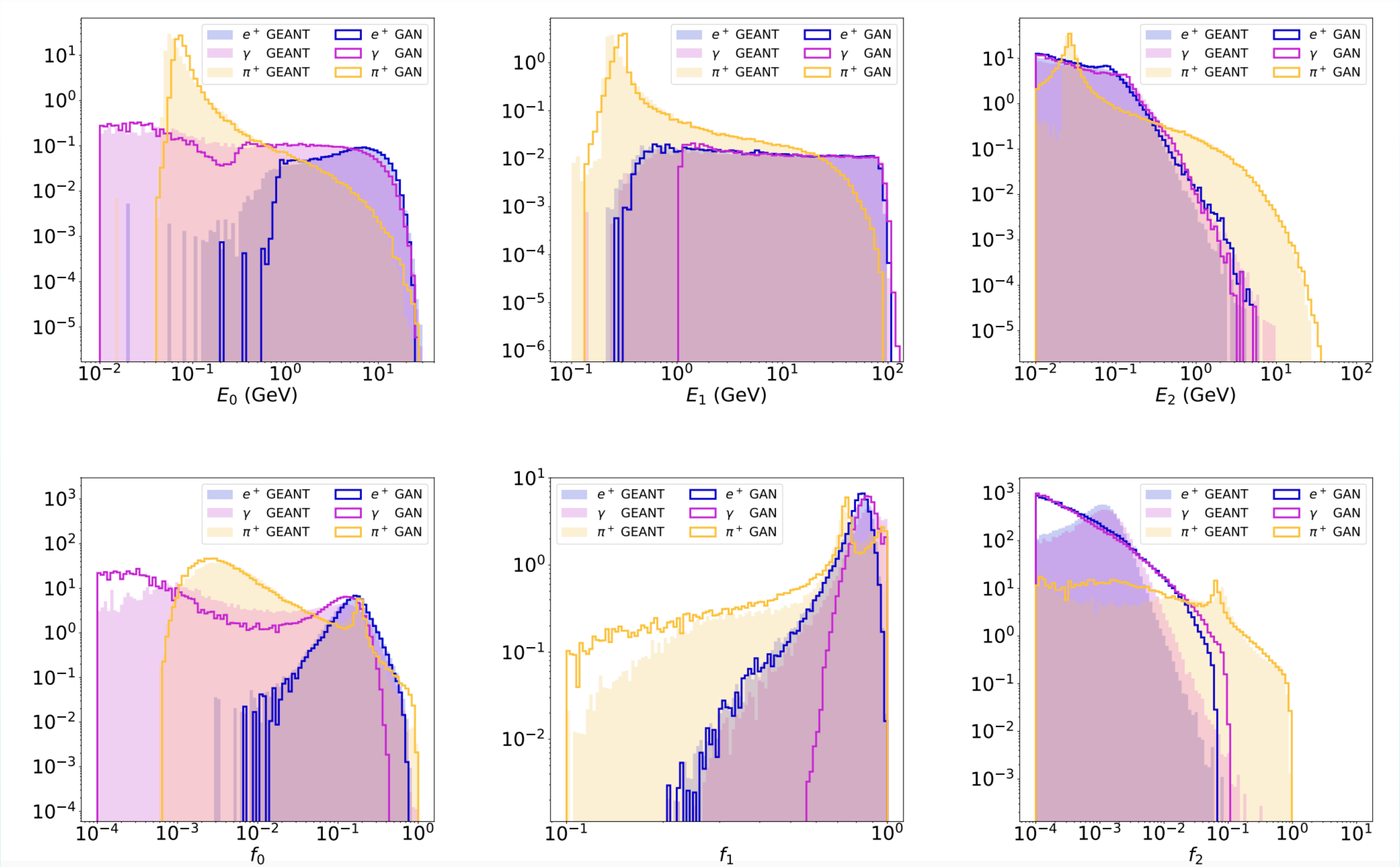


- Compare **real training samples** with the corresponding ``fake'' (generated) samples from the GAN, identified with some nearest neighbours criterion

- To optimisation of the GAN aims to **generate fake samples indistinguishable close to the real ones**

- The training of the GANs ends when the **Classifier** is not able any more from tell apart the real samples from the fake ones that the **Generator** is producing

CaloGANs vs full detector simulation



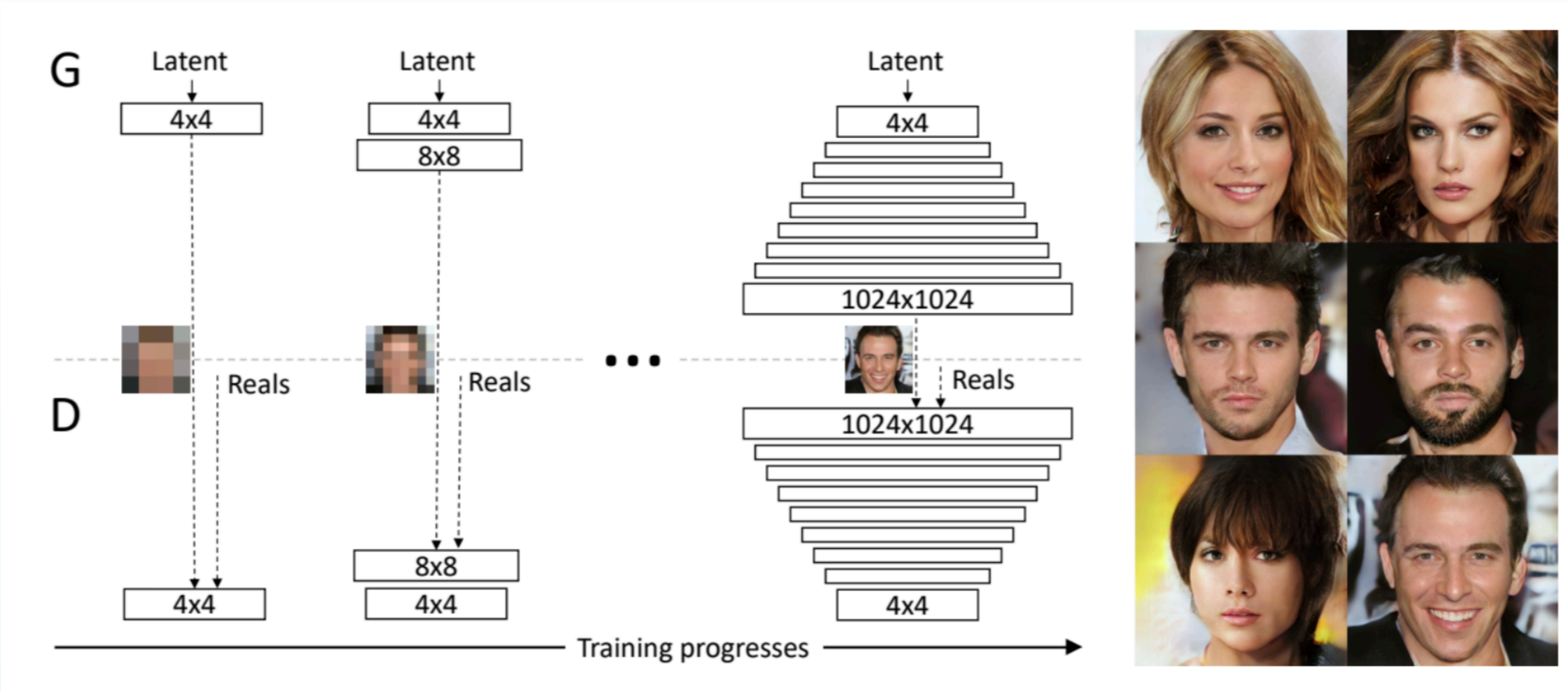
Still room for improvement but very promising results!

CaloGANs vs full detector simulation

Simulator	Hardware	Batch Size	ms/shower
GEANT4	CPU	N/A	1772
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012

Speed-up by several orders of magnitude, specially when running in GPUs

The many uses of GANs



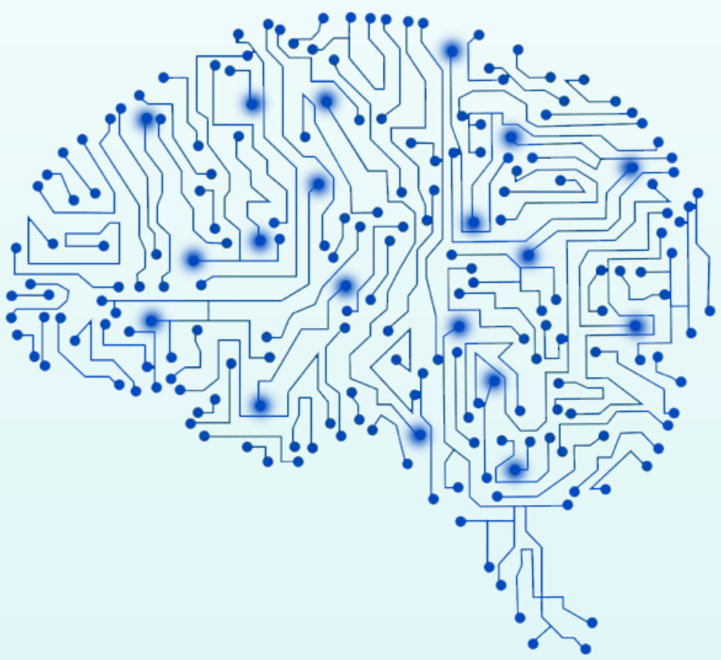
arXiv:1710.10196

Which one of these images are real and which ones are fake (generated by the GANs)?

Machine Learning for HEP

The structure of the proton at the LHC

Higgs self-interactions

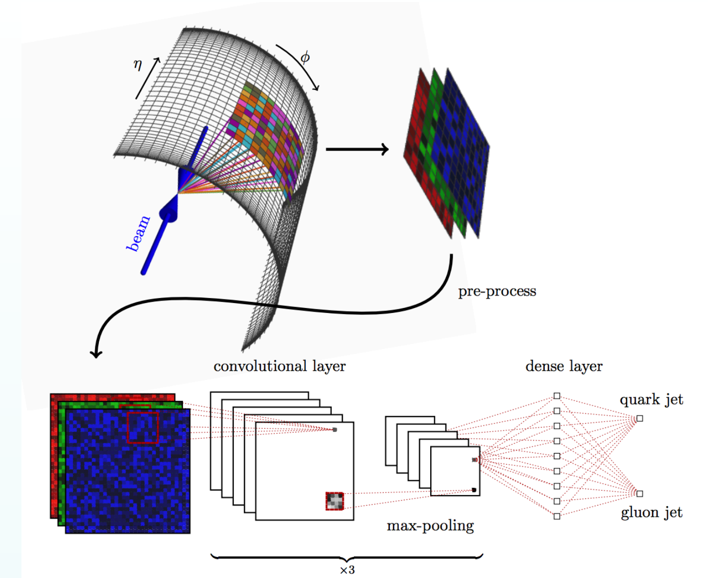


QCD-aware NNs for jet physics

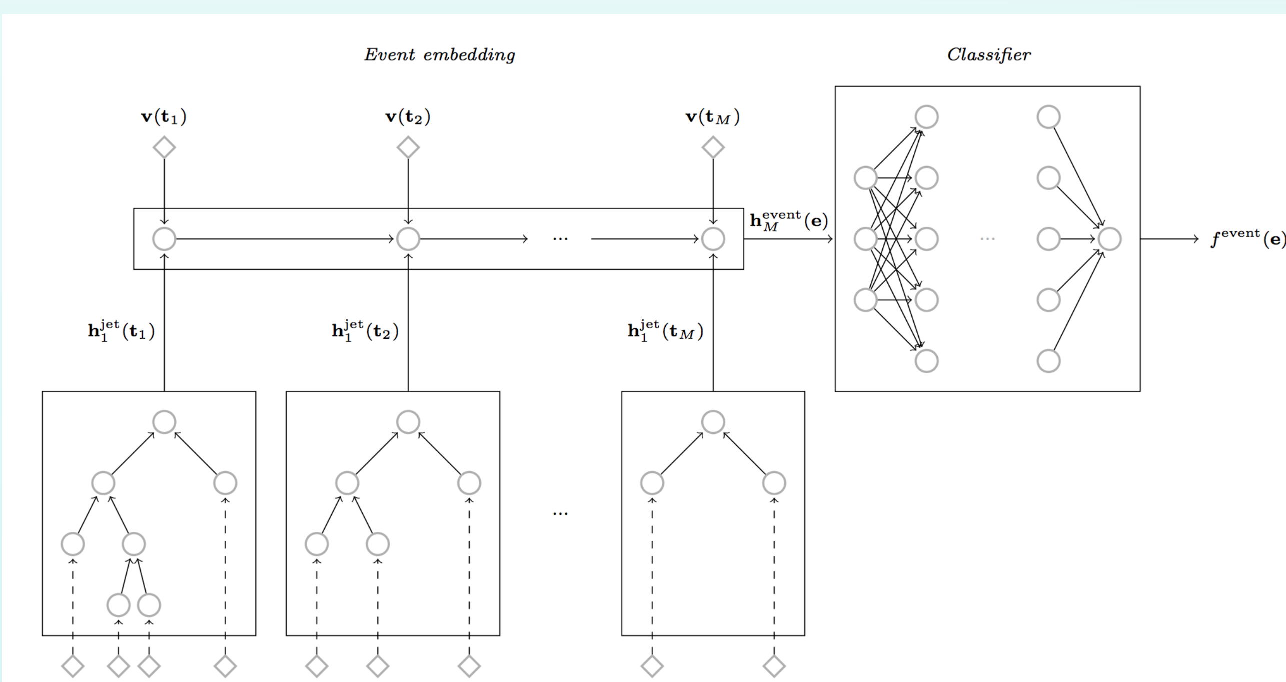
Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation



QCD-aware recursive neural networks

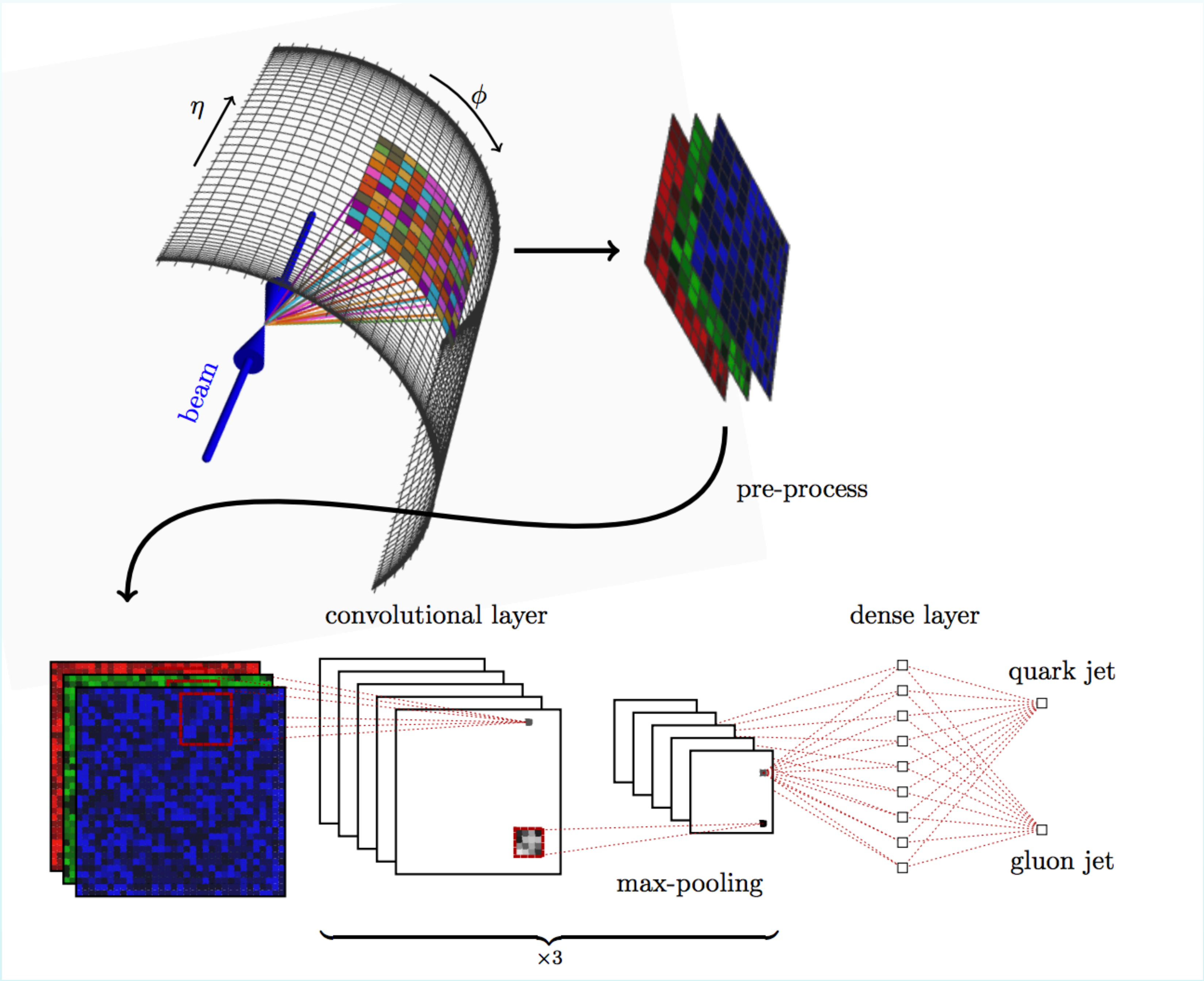


Louppe, Cho, Becot, Cranmer

1702.00748

From image recognition to jetography

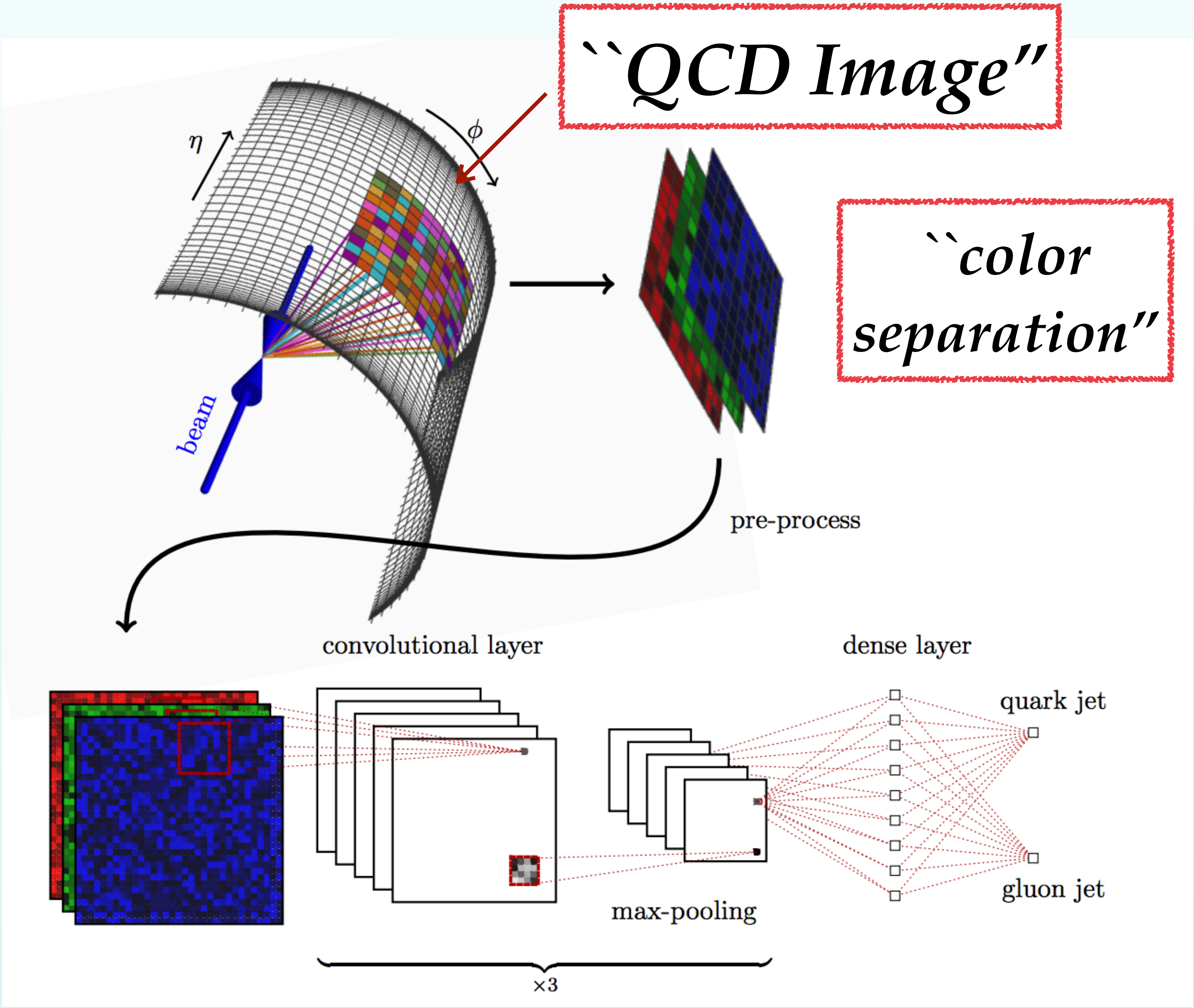
- In the context of HEP applications of Machine Learning, **jets from hadron collisions** have been extensively studied
- Topics include **quark/gluon discrimination, jet substructure, jet charge**, and other jet properties
- Progress in these ML applications has been driven by analogy between **images and hadron calorimeters**



Eg arxiv:1612.01551
Deep convolutional NNs for quark/gluon jet discrimination

From image recognition to jetography

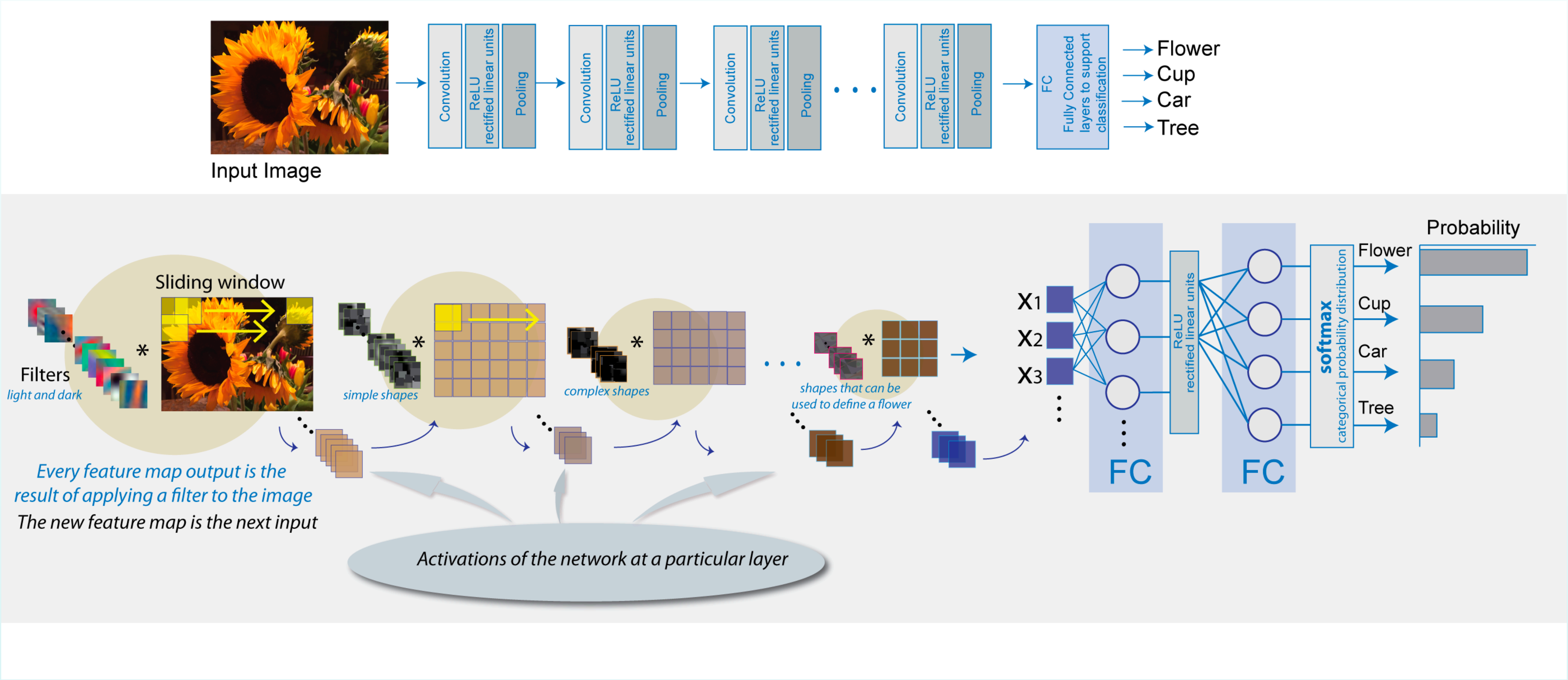
- In the context of HEP applications of Machine Learning, **jets from hadron collisions** have been extensively studied
- Topics include **quark/gluon discrimination, jet substructure, jet charge**, and other jet properties
- Progress in these ML applications has been driven by analogy between **images and hadron calorimeters**



Eg arxiv:1612.01551
Deep convolutional NNs for quark/gluon jet discrimination

Convolutional Neural Networks

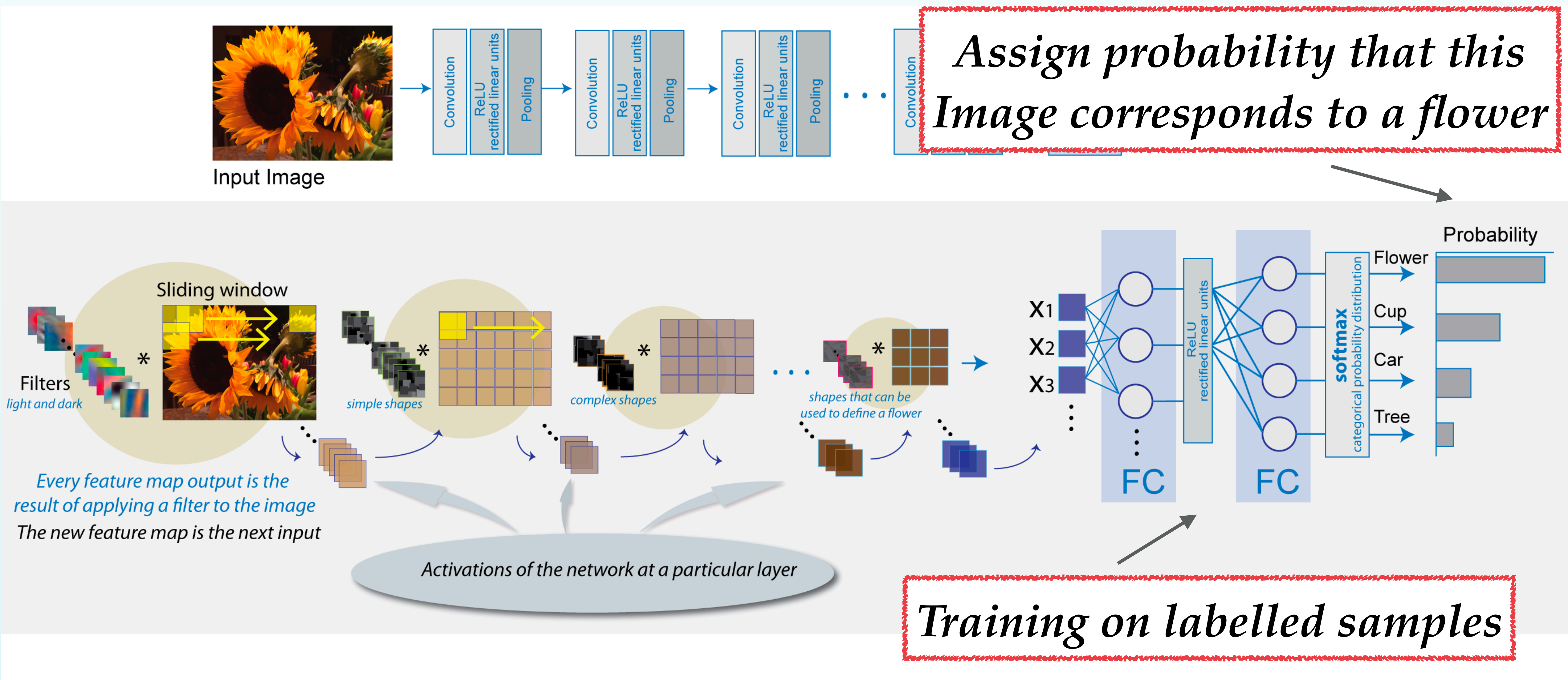
- Convolutional Neural Networks (CNNs) have convolutional layers based on **filters**
- Each **filter** maps a group of numbers into a number, reducing the dimensionality of the data
- Specially useful for **pattern recognition** (eg for self-driving vehicles)



mathworks.com

Convolutional Neural Networks

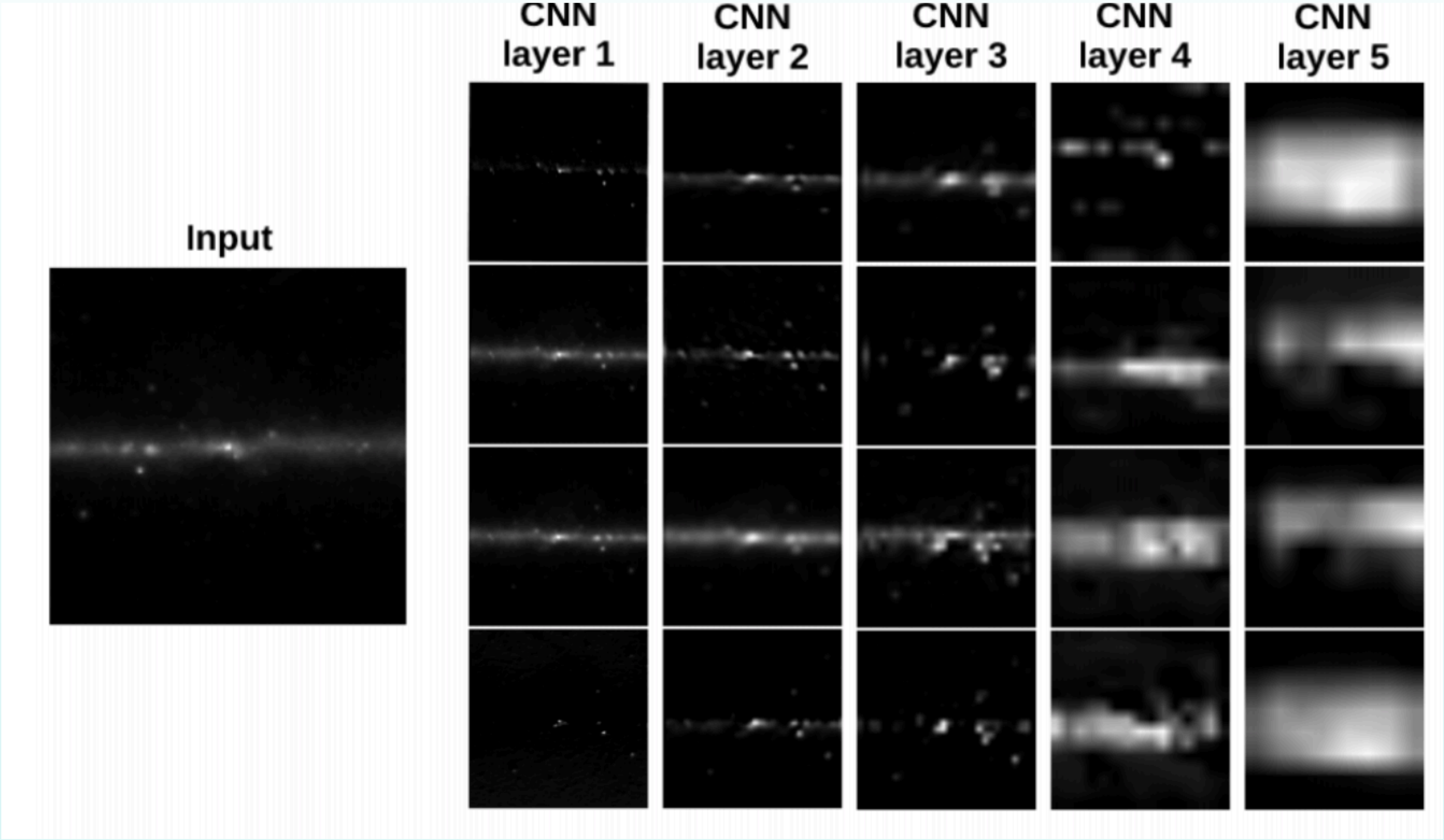
- **Convolutional Neural Networks** (CNNs) have convolutional layers based on **filters**
- Each **filter** maps a group of numbers into a number, reducing the dimensionality of the data
- Specially useful for **pattern recognition** (eg for self-driving vehicles)



mathworks.com

CNNs for Dark Matter Searches

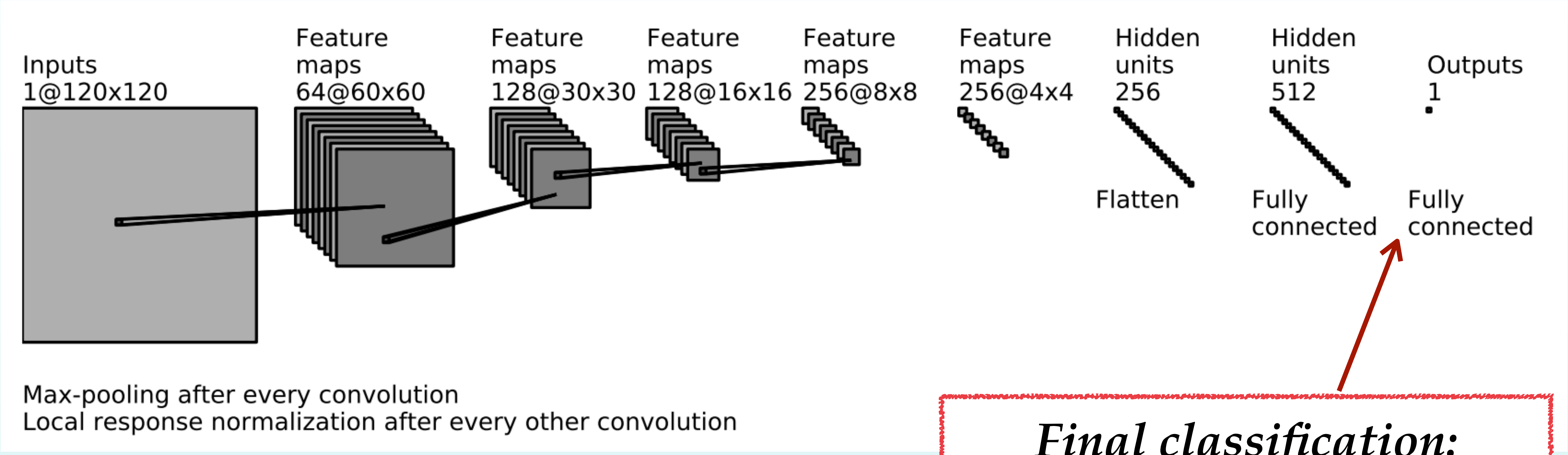
Use CNNs to discriminate **point sources** (astrophysical origin) versus **diffuse flux** (dark matter) in galactic centre images



Caron, Hendriks et al arXiv:1708.06706

CNNs for Dark Matter Searches

Use CNNs to discriminate **point sources** (astrophysical origin) versus **diffuse flux** (dark matter) in galactic centre images



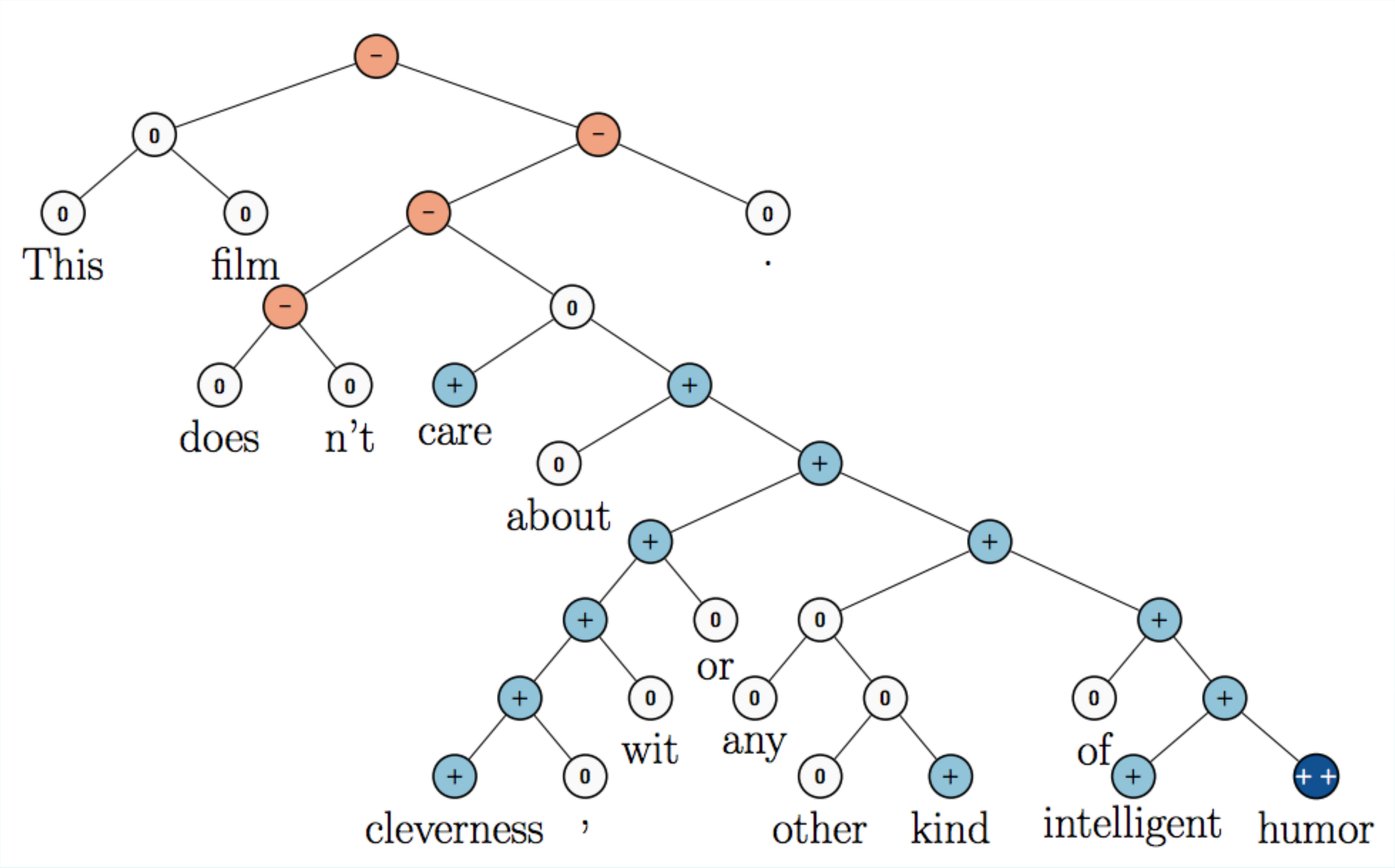
Final classification: point sources vs diffuse flux

Caron, Hendriks et al arXiv:1708.06706

Recursive Neural Networks

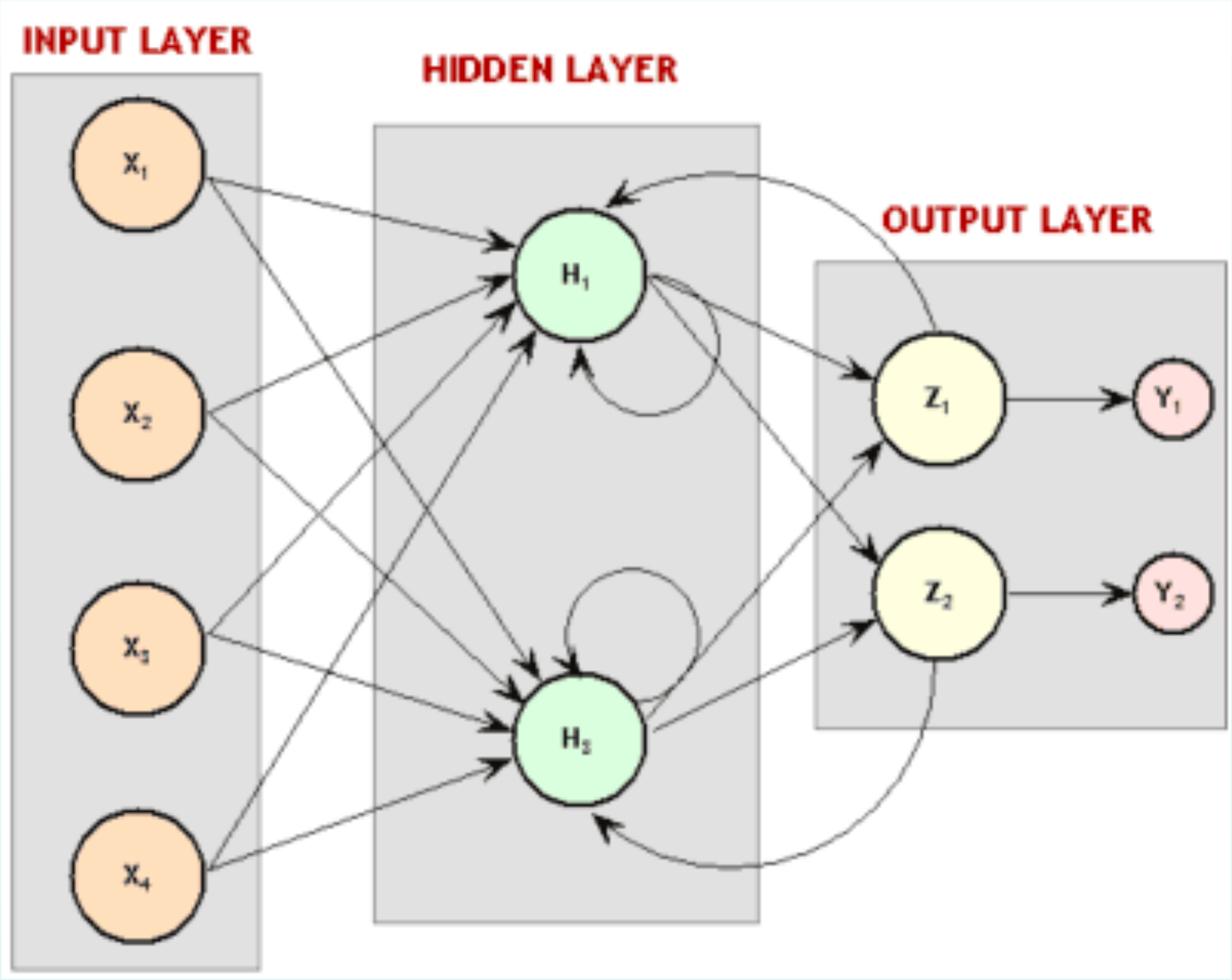
- One can build recursive NNs that are ``aware'' of the fact that QCD is the correct theory of the strong force in Nature

- **Recursive Neural Networks** (RNNs) are **deep neural networks** where the same set of weights are applied **recursively following a structured input**



Recurrent Neural Networks

Not to mix with **Recurrent Neural Networks,** which use the output of the current node as the input to the next node



Recurrent Neural Networks

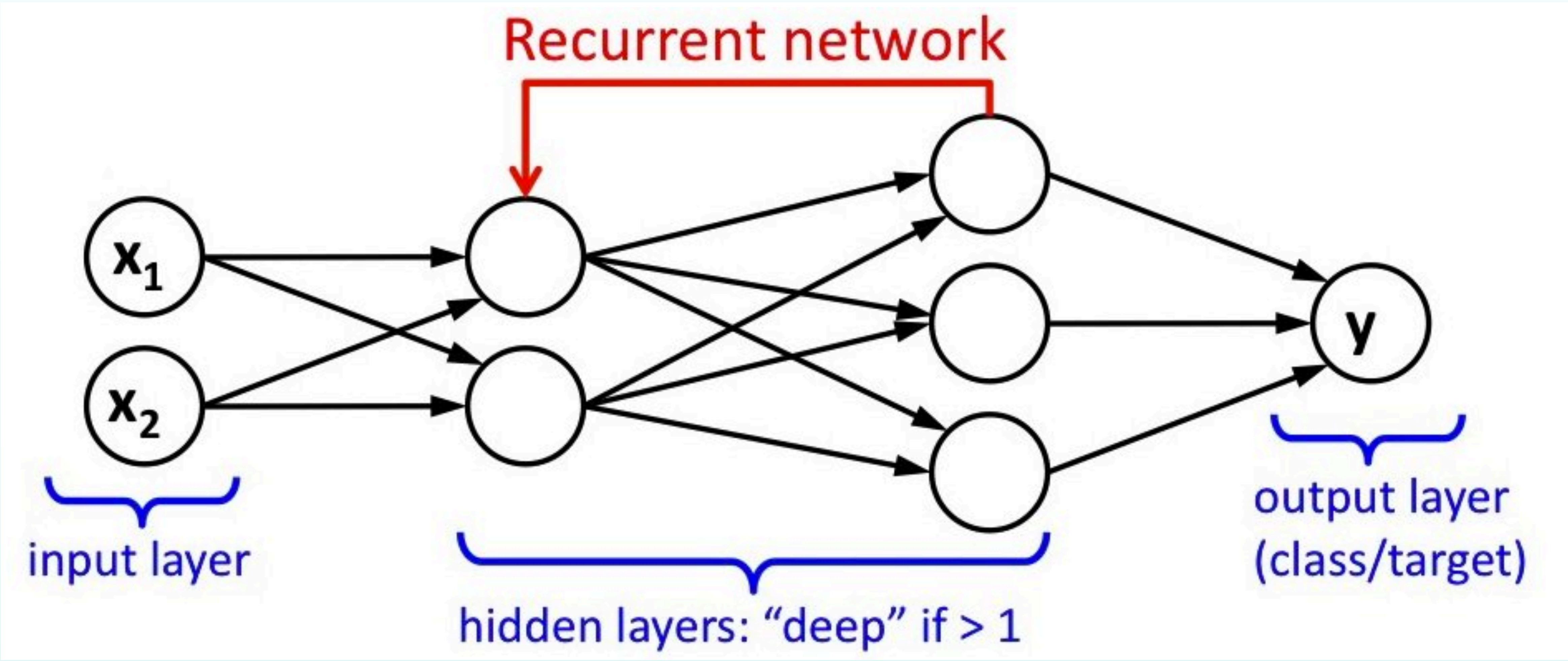
Lead to truly game-changer applications, such as **random generation of country song lyrics**

```
Tied right now
I got life now he never thought I got by the all
Going up like a house four boy
Nothing his thing out of hands
No one with the danger in the world
I love my black fire as I know
But the short knees just around me
Fun the heart couldnes fall to back
I see a rest of my wild missing far
When I was missing to wait
And if I think
It's a real tame
I say I belong is every long night
Maybe lovin' you
```

http://www.mattmoocar.me/blog/RNNCountryLyrics/

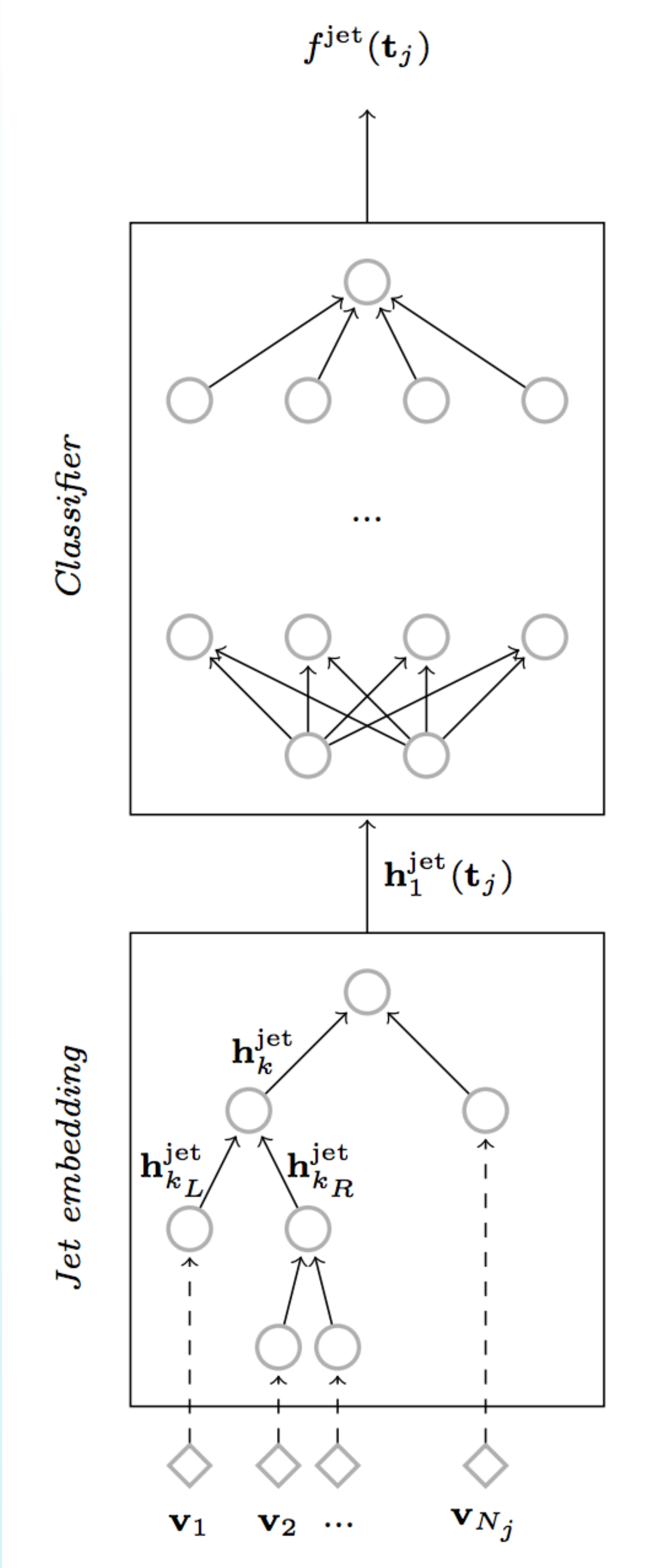
Recurrent Neural Networks

RNNs use as inputs not just the current "training examples" but also **what they have perceived previously**: they have a **built-in notion of time ordering** useful for time-dependent functions



The output of a RNN at time time, ***y(t)***, depends both on the current input example ***x(t)*** as well as of its previous output ***y(t-1)*** (or activation states of hidden neutrons at ***t-1***)

QCD-aware NNs for jet physics



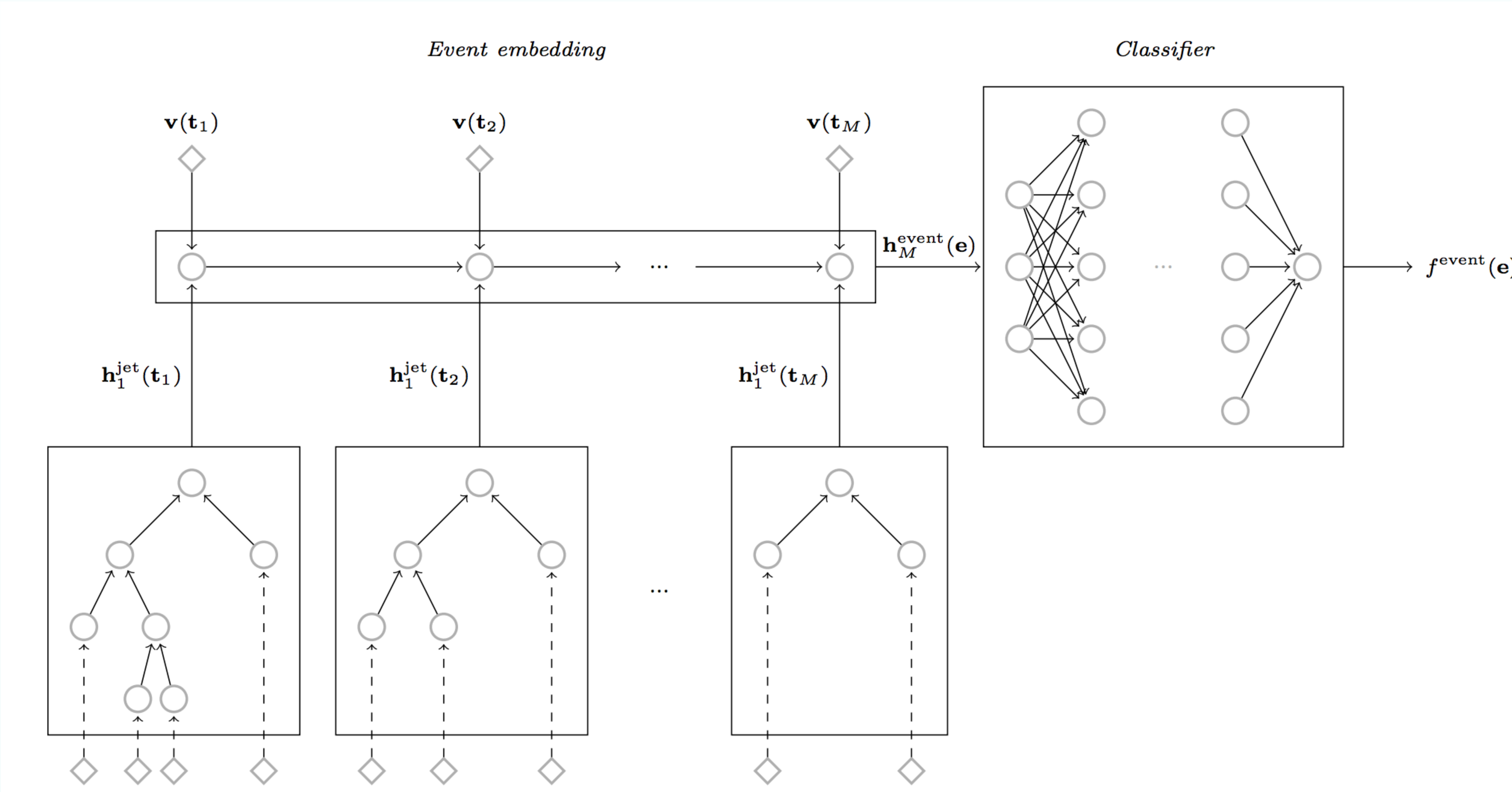
QCD-motivated recursive jet embedding for classification

- Each particle is represented by **four-momentum $\mathbf{v}_i$**
- For each individual jet, the **embedding $\mathbf{h}_1^{\text{jet}}(\mathbf{t}_j)$** is computed recursively from the root node down to the outer nodes of the **binary tree $\mathbf{t}_j$**
- The resulting embedding is chained to a **subsequent classifier**
- The **topology of the network is distinct for each jet** and is determined by a **sequential recombination jet algorithm**

e.g. the anti-kT jet clustering algorithm leads to a different NN topology than the Cambridge/Aachen one

QCD-aware NNs for jet physics

The same strategy can be applied to the **full event composed by many jets**



Machine learning classification based on recursive neural networks can implement physical features such as that the ***reconstructed jets will be infrared and collinear safe***

QCD-aware NNs for jet physics

v_i: particle four-momentum
$h_1^{jet}(t_j)$: jet embedding, including "QCD" clustering information

Performance of classification of **QCD vs non-QCD jets** with different settings

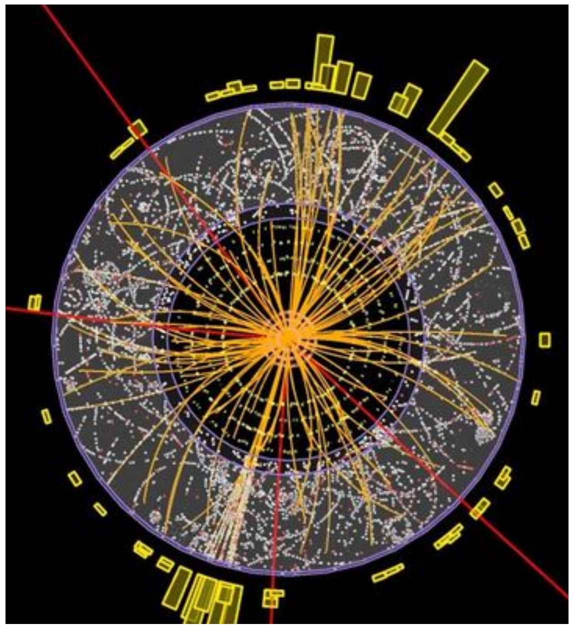
Best results are achieved through **nested recurrence** over the jets and over its constituents, as motivated by QCD

Jet clustering here can be understood as a **preprocessing of the input-level data**

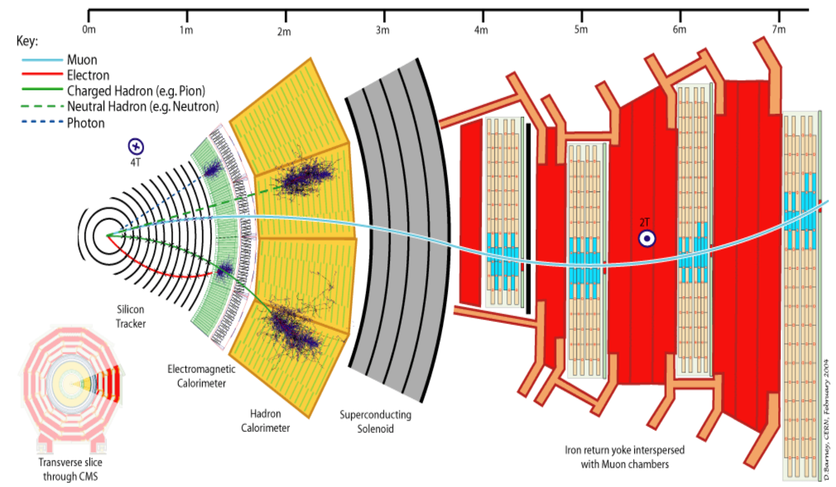
Jet clustering also essential to isolate soft and semi-hard QCD physics that complicate the classification

Input	ROC AUC	$R_{\epsilon=80\%}$
Hardest jet		
$\mathbf{v}(\mathbf{t}_j)$	0.8909 ± 0.0007	5.6 ± 0.0
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(k_t)}$	$\mathbf{0.9602 \pm 0.0004}$	$\mathbf{26.7 \pm 0.7}$
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(\text{desc}-p_T)}$	0.9594 ± 0.0010	25.6 ± 1.4
2 hardest jets		
$\mathbf{v}(\mathbf{t}_j)$	0.9606 ± 0.0011	21.1 ± 1.1
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(k_t)}$	0.9866 ± 0.0007	156.9 ± 14.8
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(\text{desc}-p_T)}$	$\mathbf{0.9875 \pm 0.0006}$	$\mathbf{174.5 \pm 14.0}$
5 hardest jets		
$\mathbf{v}(\mathbf{t}_j)$	0.9576 ± 0.0019	20.3 ± 0.9
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(k_t)}$	0.9867 ± 0.0004	152.8 ± 10.4
$\mathbf{v}(\mathbf{t}_j)$, $\mathbf{h}_j^{\text{jet}(\text{desc}-p_T)}$	$\mathbf{0.9872 \pm 0.0003}$	$\mathbf{167.8 \pm 9.5}$
No jet clustering, desc-p_T on $\mathbf{v}_i$		
$i = 1$	0.6501 ± 0.0023	1.7 ± 0.0
$i = 1, \ldots, 50$	$\mathbf{0.8925 \pm 0.0079}$	$\mathbf{5.6 \pm 0.5}$
$i = 1, \ldots, 100$	0.8781 ± 0.0180	4.9 ± 0.6
$i = 1, \ldots, 200$	0.8846 ± 0.0091	5.2 ± 0.5
$i = 1, \ldots, 400$	0.8780 ± 0.0132	4.9 ± 0.5

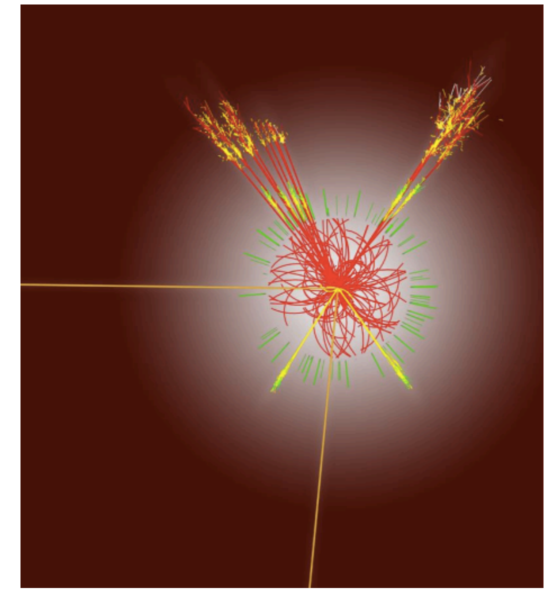
Machine Learning tools are everywhere!



Deep Kalman RNNs

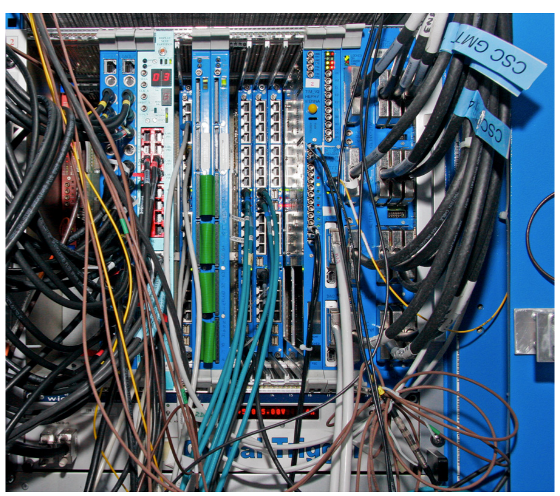


Generative Models, Adversarial Networks

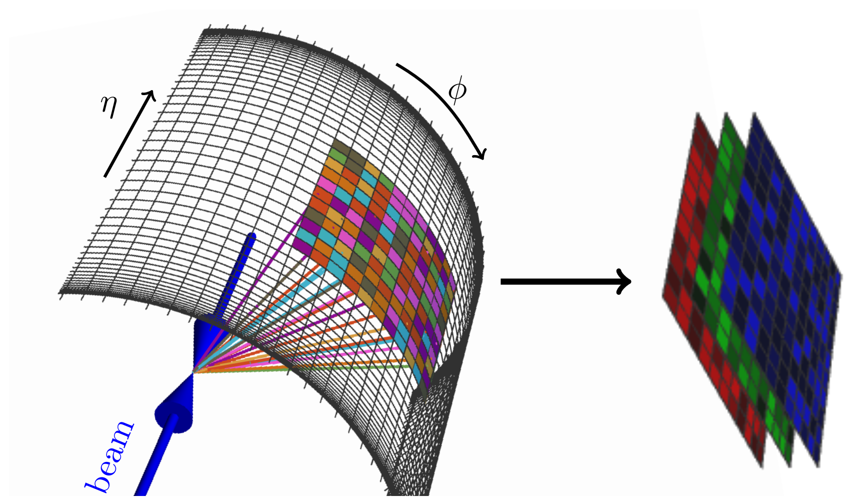


FCN, Recurrent, LSTM NN

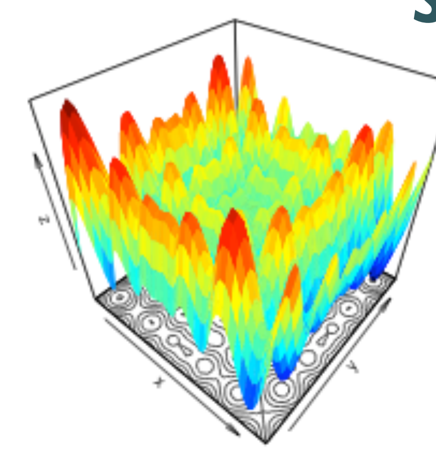
S. Glayzer



Deep ML +FPGA



Convolutional DNN

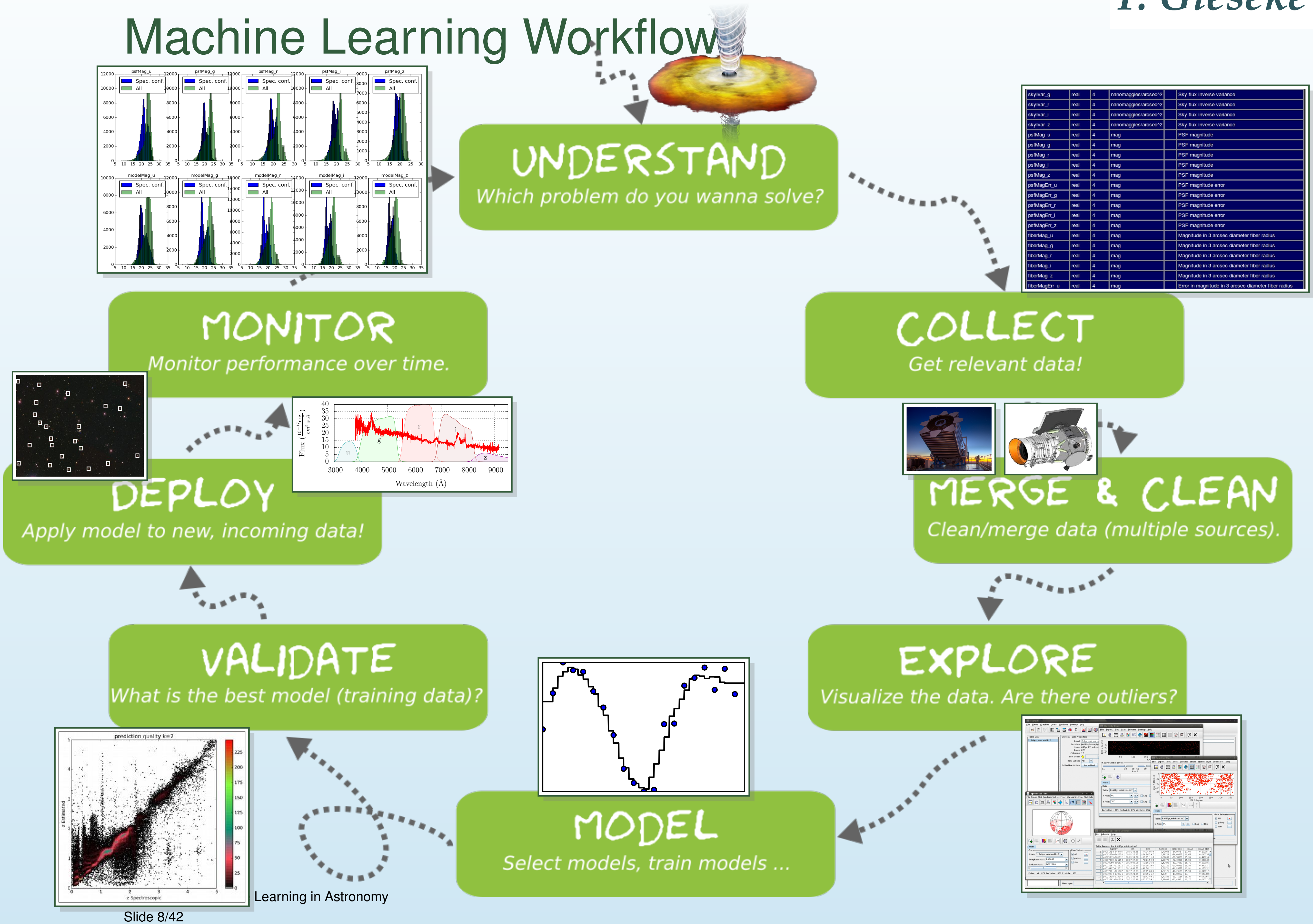


Multiobjective Regression

For many crucial applications, ML tools not just one option, but **the only option**

ML cheat sheet

F. Gieseke



Endless possibilities - but also many non-trivial hurdles to overcome

Take-away message

Andy Buckley
@agbuckley

Following

I'm all for technical sophistication, but it's depressing how many young scientists we're training in little more than how to press the Go button on TMVA and TensorFlow black boxes

3:11 PM - 4 Apr 2018 from Glasgow, Scotland

Andy Buckley @agbuckley · 17h
Too much is glorified data entry and algorithm-babysitting. I'm reminded of Yuval Noah Harari in Sapiens, on how -- in contrast to our conventional telling -- wheat domesticated *us* smh.com.au/opinion/slaves… Who's the boss, the ML or the scientists?

Proficiency in ML applications requires a deep understanding of both the physical problem being addressed as well as of the inner workings of the specific algorithms used!

ANNs and LHC phenomenology

- **Machine Learning algorithms** are already **transforming our world**, from the way we move, shop and heal ourselves, to our understanding of what makes us unique as humans

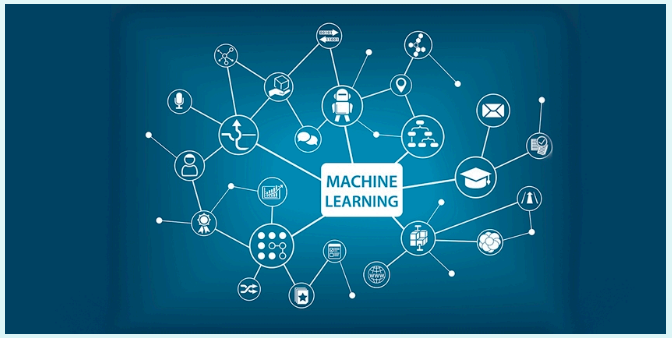
- In the context of **LHC data analysis and interpretation, ML tools are ubiquitous**, from event selection deep in the detector chain (triggering) to bottom-quark tagging and automated BSM models classification (and exclusion)

- Avoid using ML tools as black boxes: a detailed understanding of both the **physical and the algorithmic aspects of the problem is essential**

Higgs self-interactions

QCD-aware NNs For jet physics

The structure of the proton at the LHC

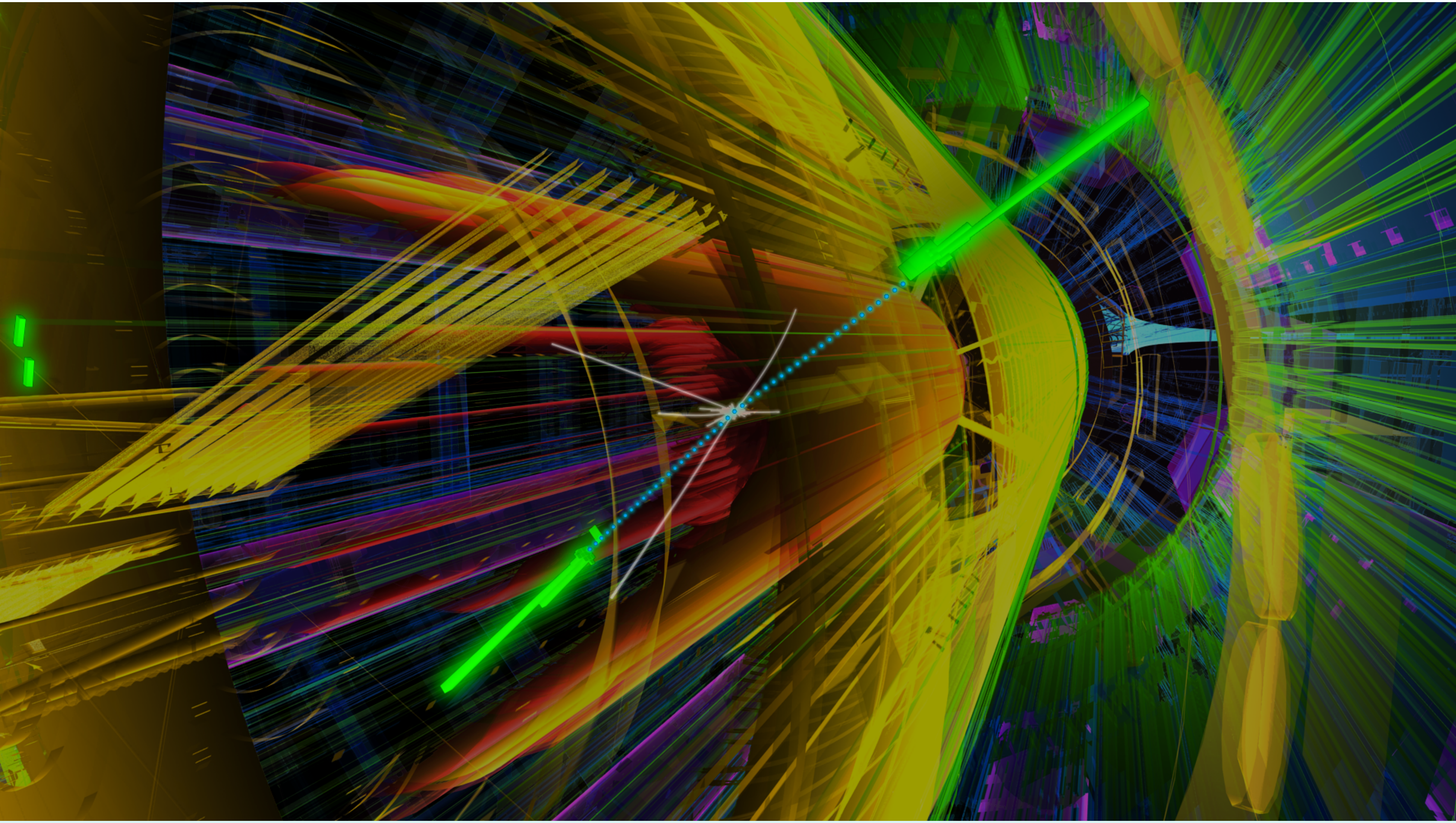


Automated bSM exclusion limits

Boosting bSM searches

HEP detector simulation

Fascinating times ahead at the high-energy frontier!



Ready to be exploited with our Machine Learning toolbox!

Fascinating times ahead at the high-energy frontier!

Ready to be exploited with our Machine Learning toolbox!