Recent advances in Machine Learning for Physics Analysis at CMS

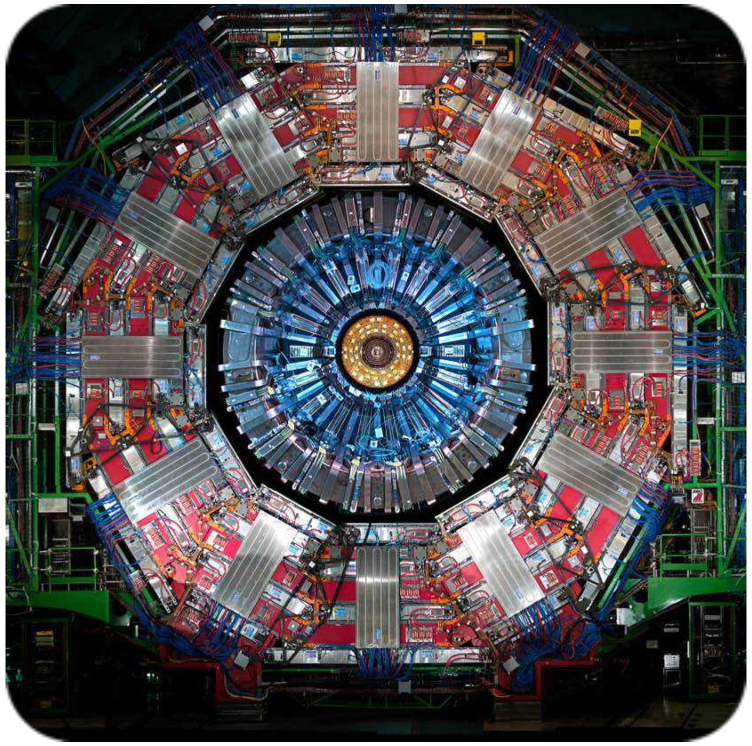
Raghav Kansal on behalf of the CMS Collaboration

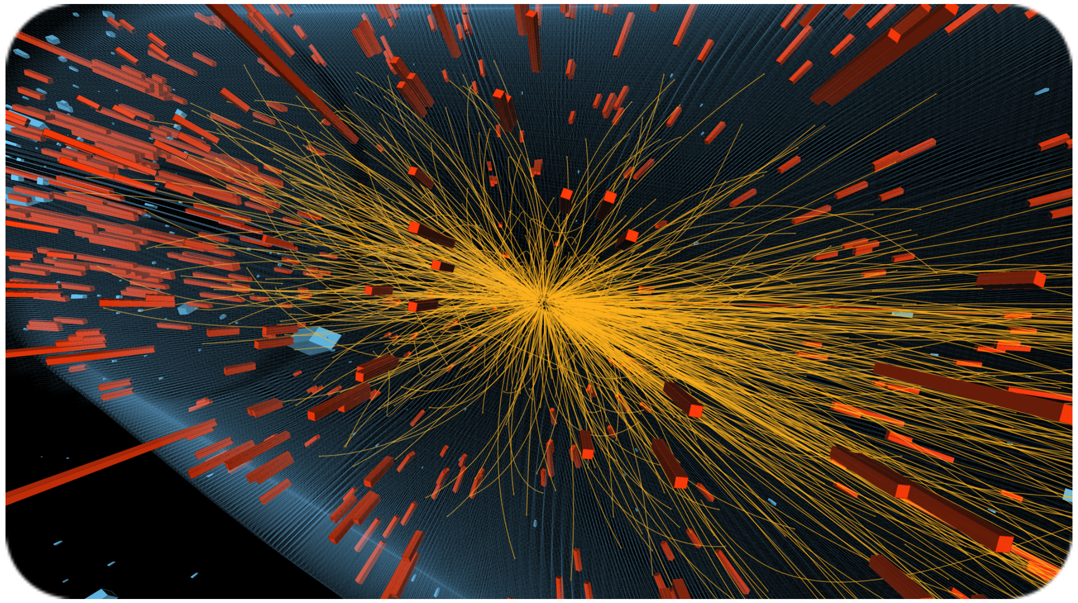
Caltech Fermilab

CIPANP 2025

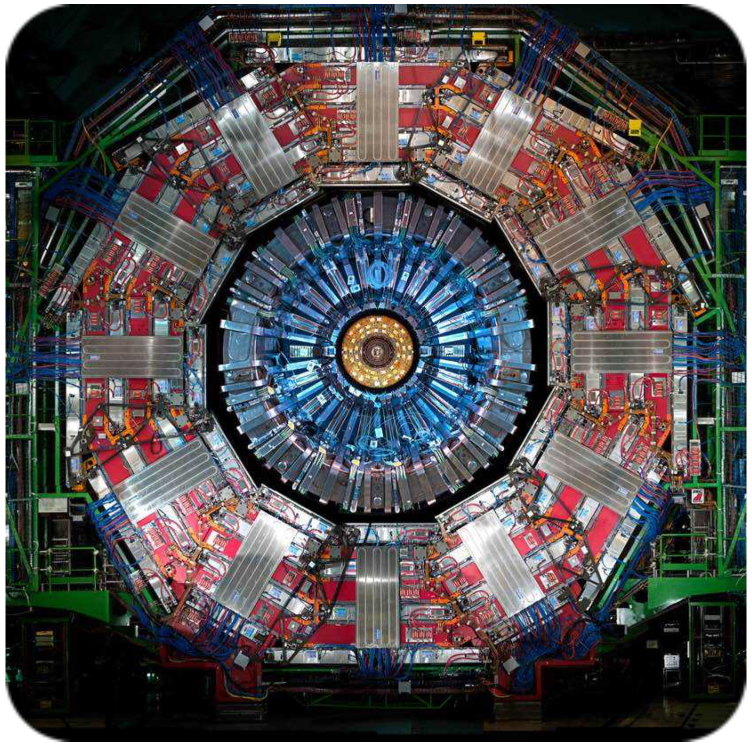
June 9th 2025

Data in the CMS experiment

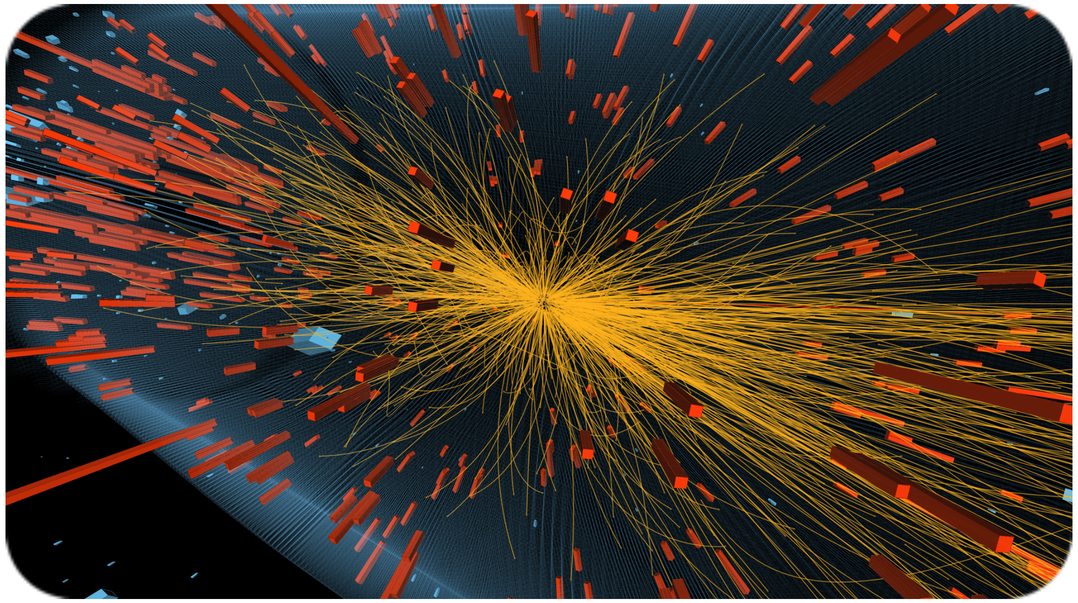




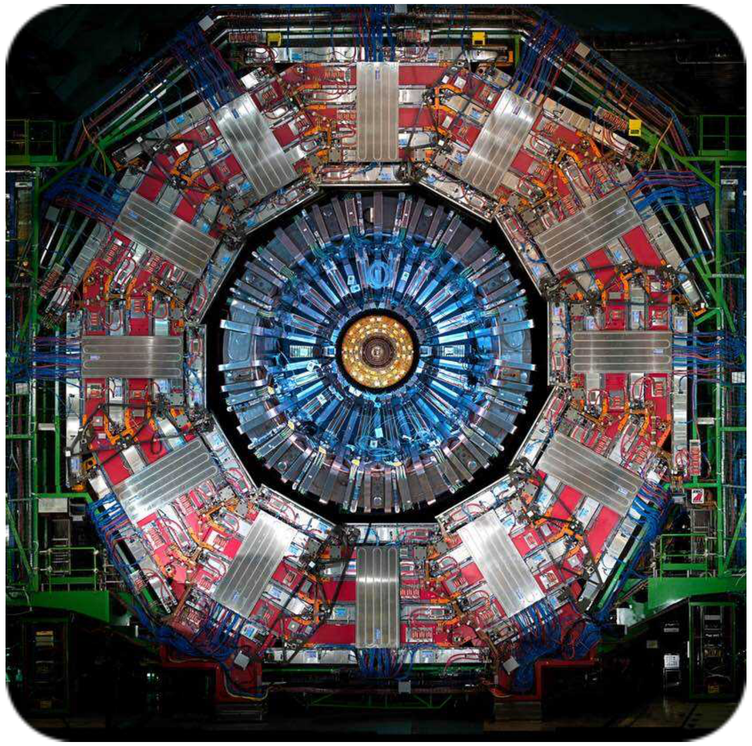
Data in the CMS experiment



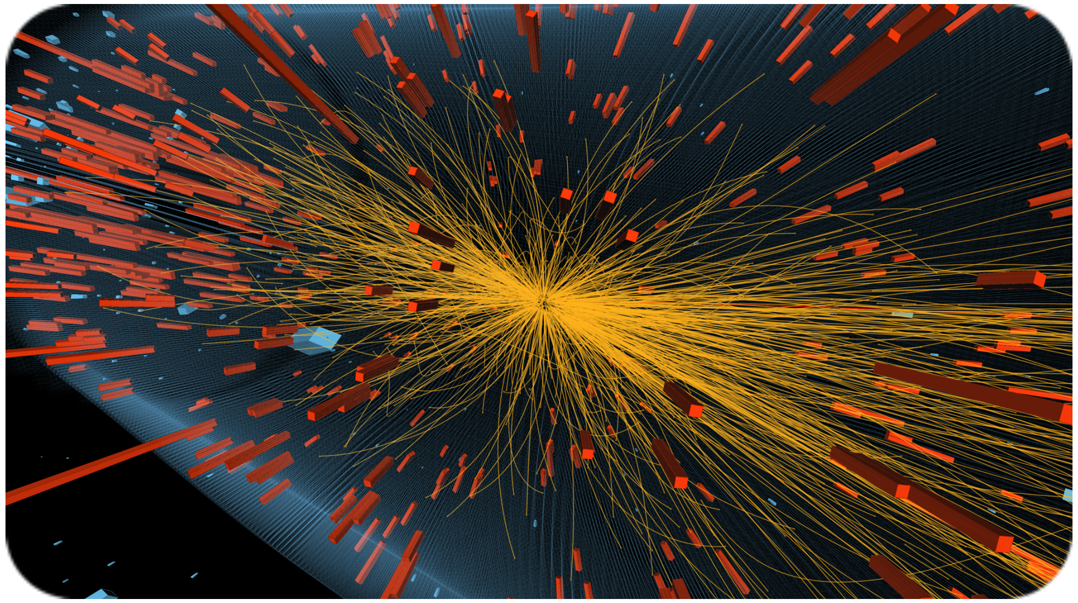
Collection
Triggers



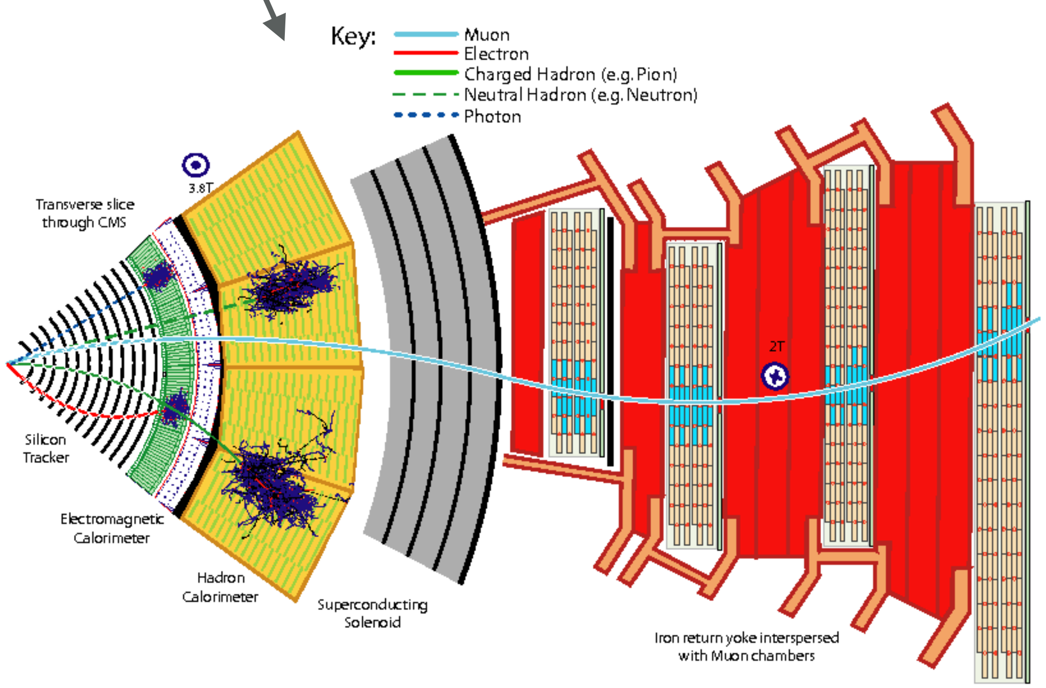
Data in the CMS experiment



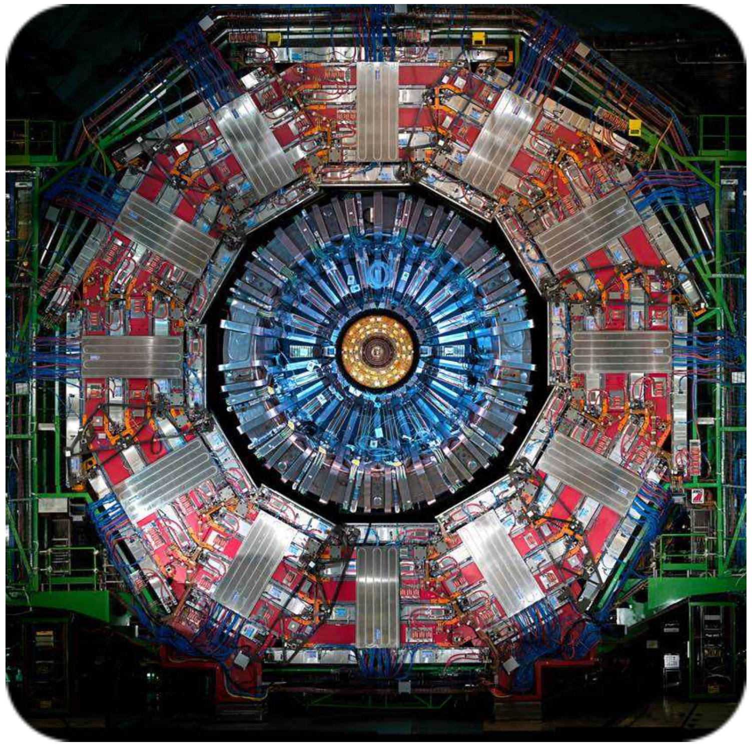
Collection
Triggers



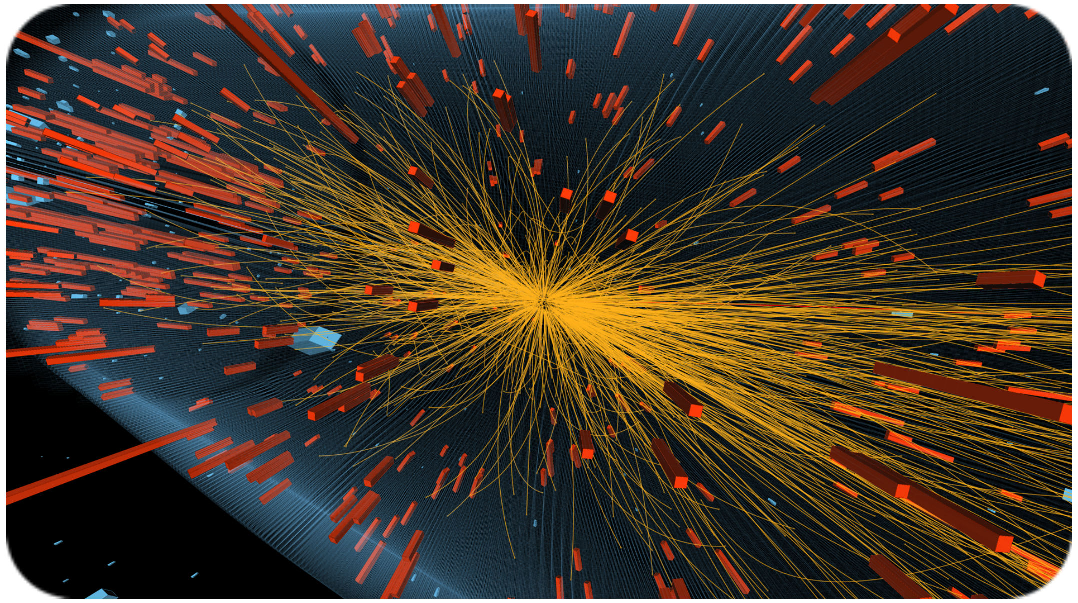
Reconstruction



Data in the CMS experiment

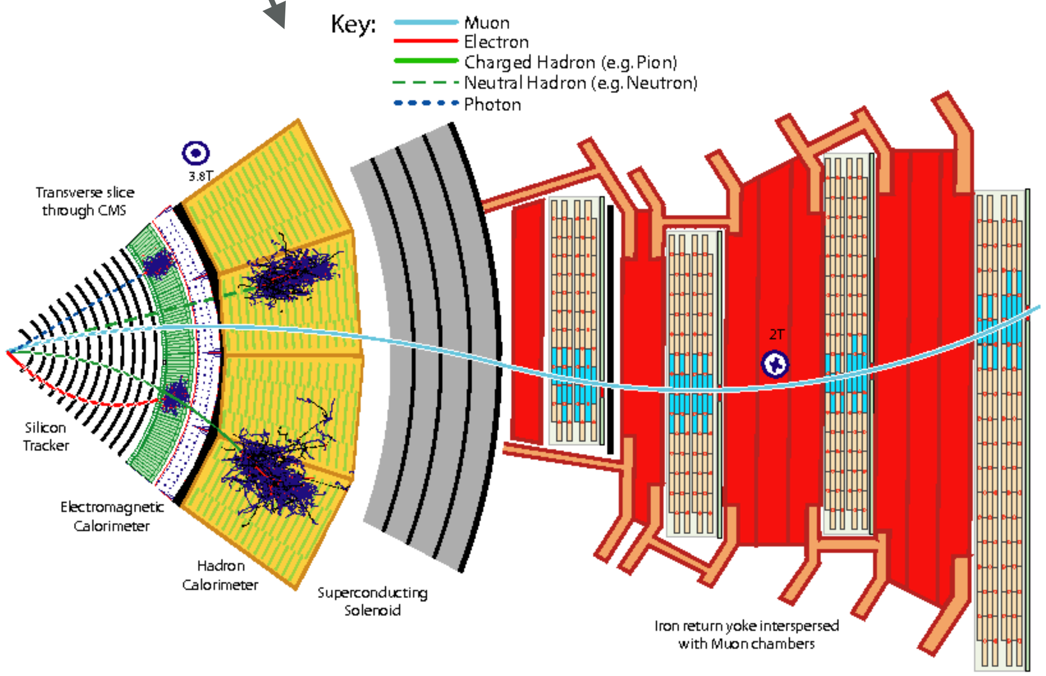


Collection
Triggers

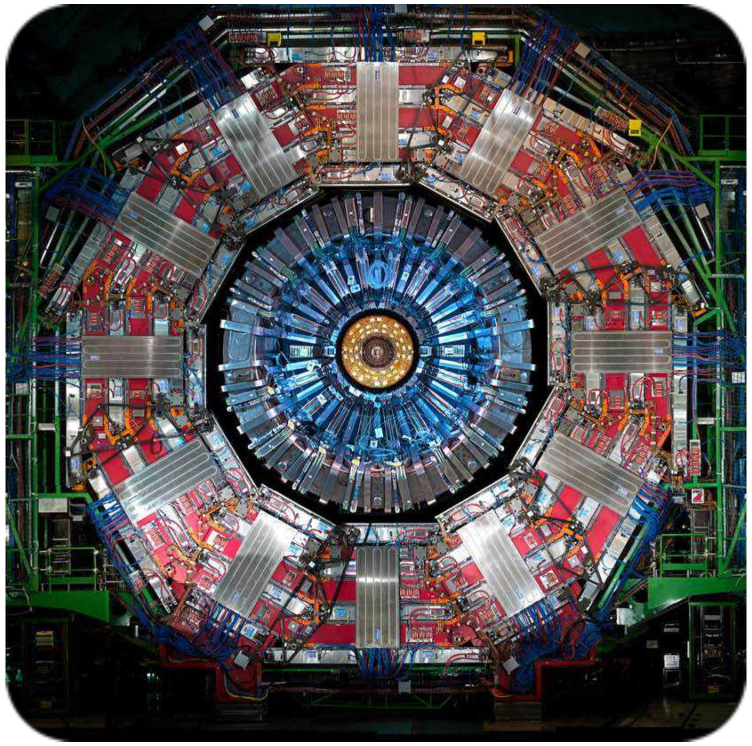


Simulation

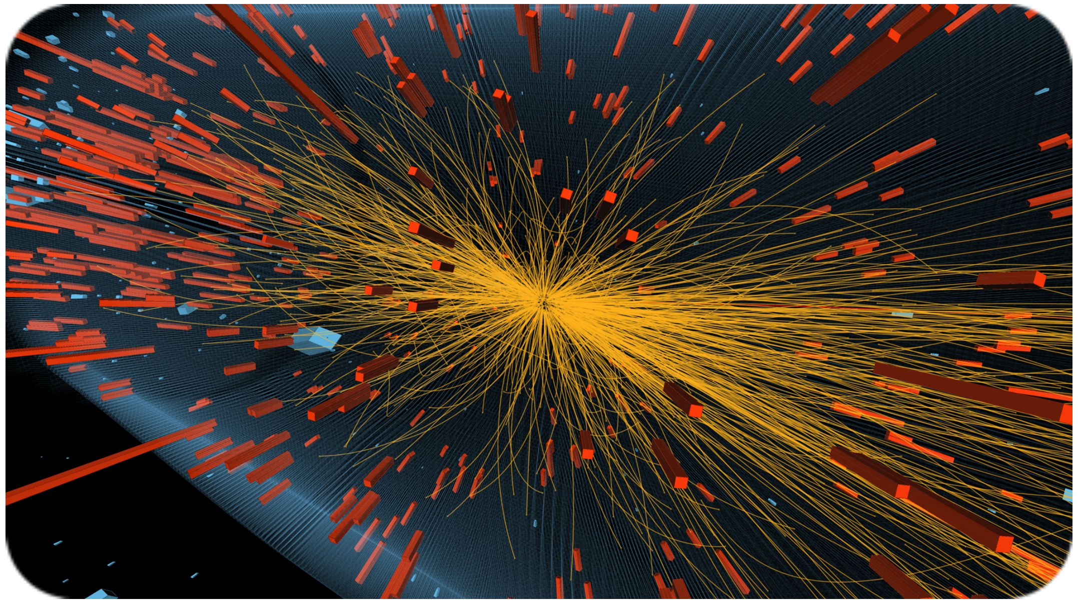
Reconstruction



Data in the CMS experiment

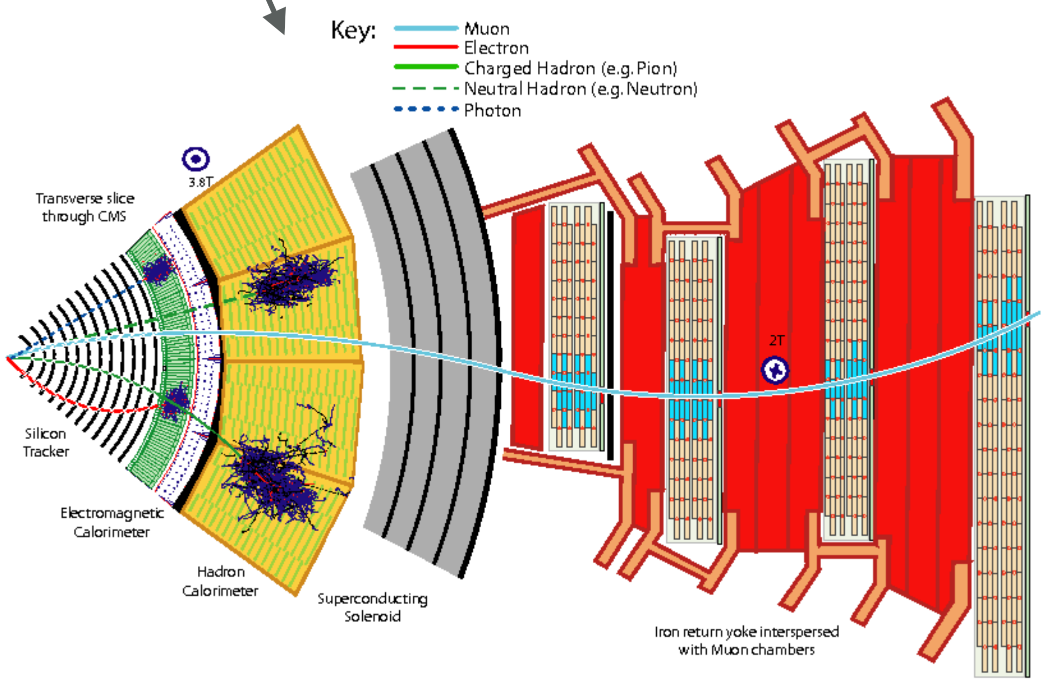


Collection
Triggers

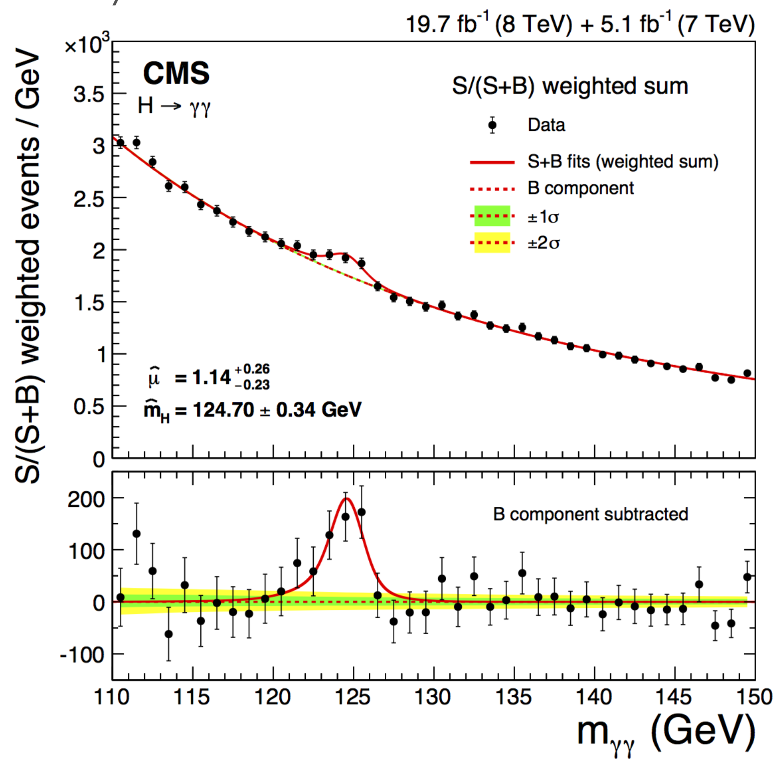


Simulation

Reconstruction

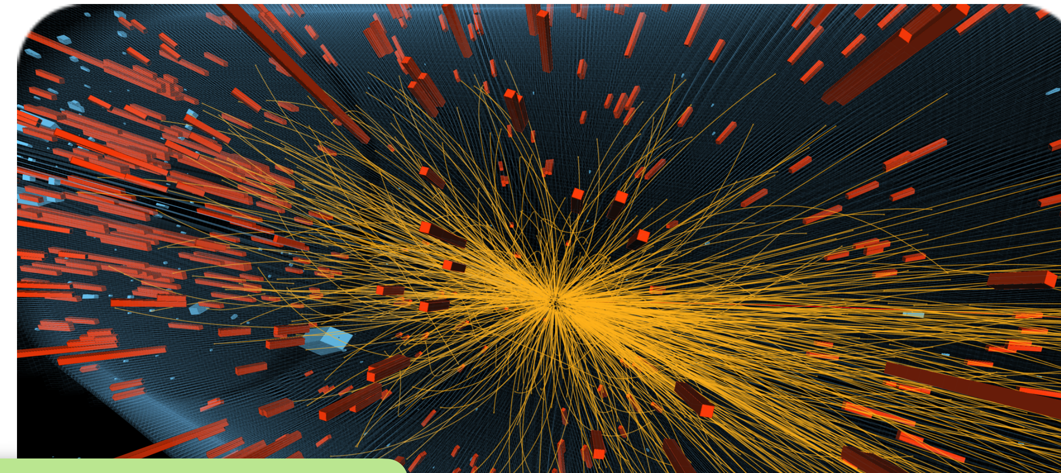


Analysis

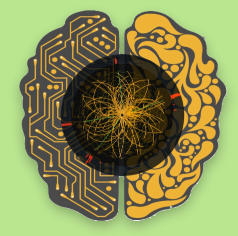


Data in the CMS experiment

Collection
Triggers

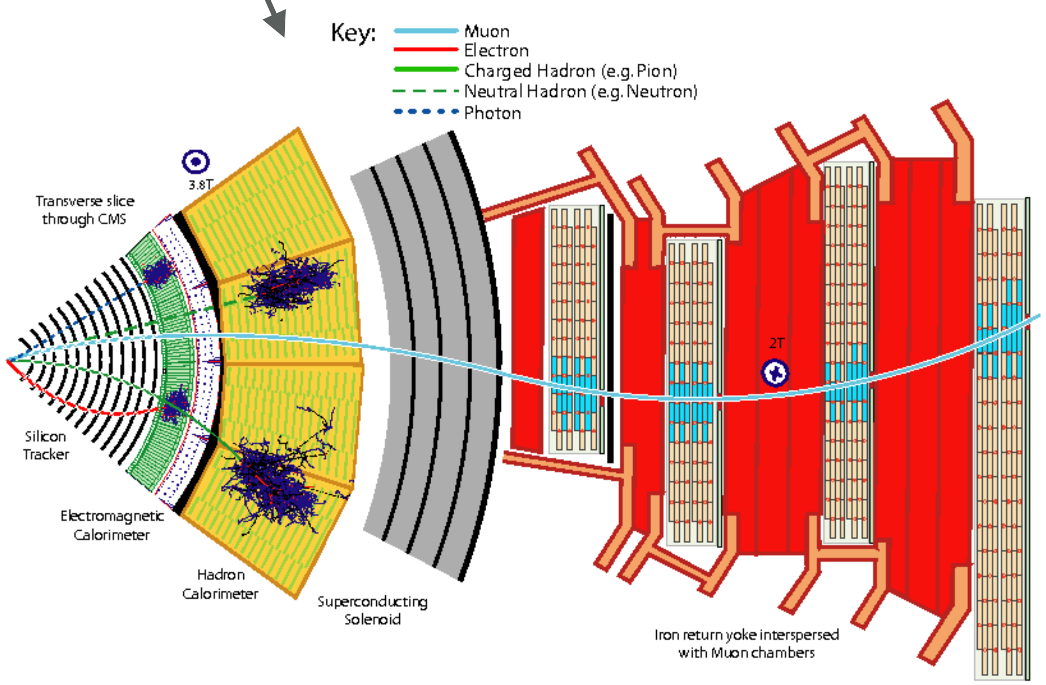


- 40M collisions/s, >1PB/s…
- Amazing opportunity to stress test the SM
- **AI necessary at all stages to maximize our physics reach**

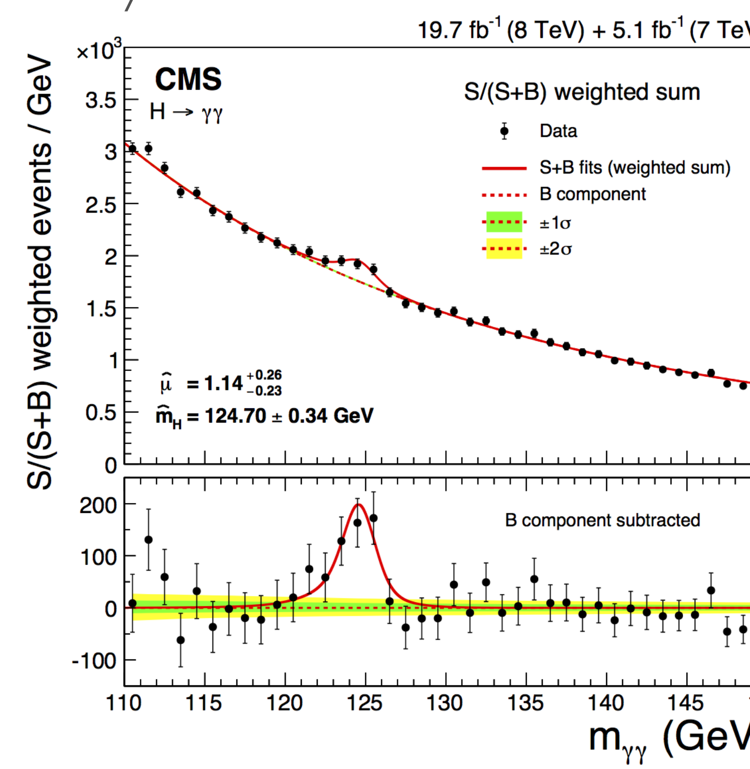


Simulation

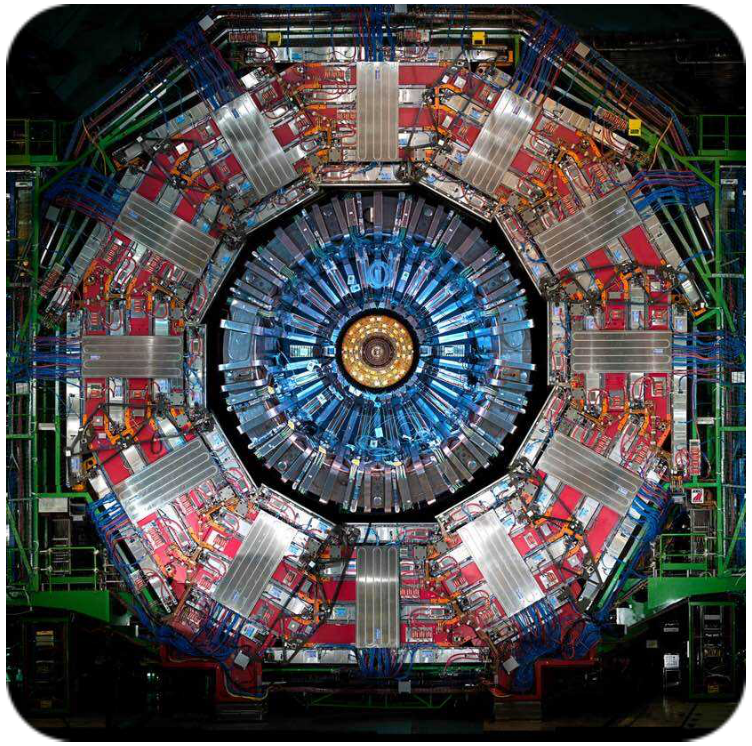
Reconstruction



Analysis



Data in the CMS experiment

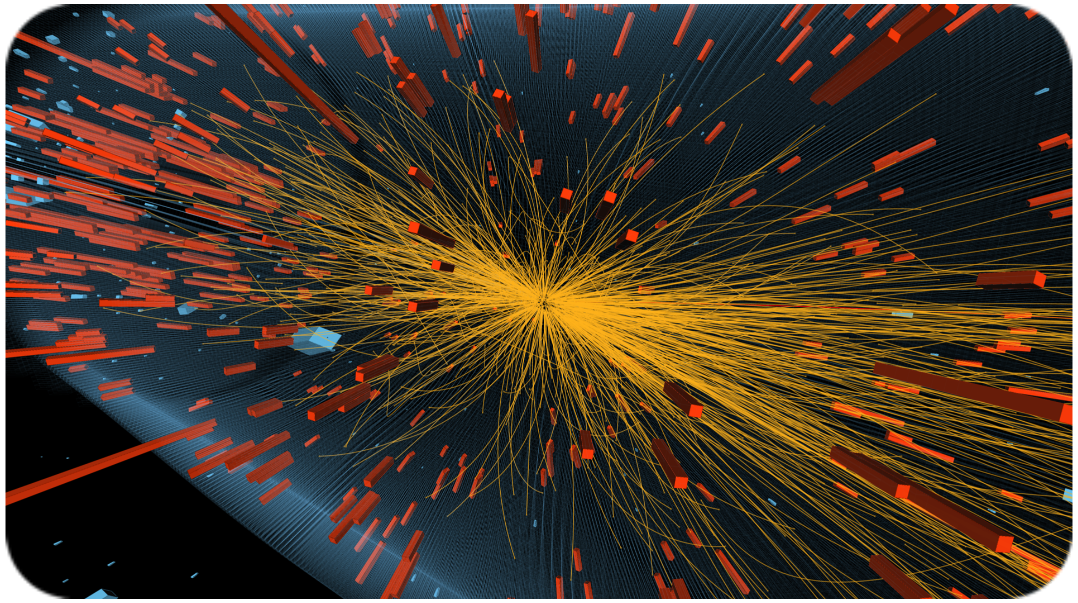


Collection

Triggers

hls4ml

Ultra-fast real time ML at the L1 trigger



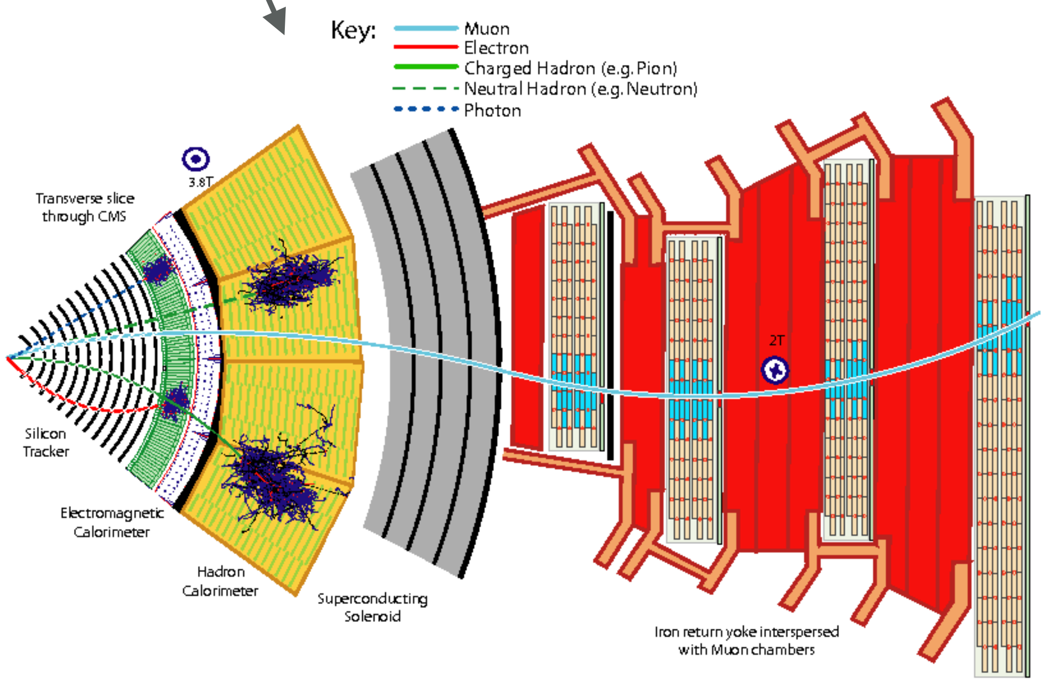
ML for simulations

Simulation

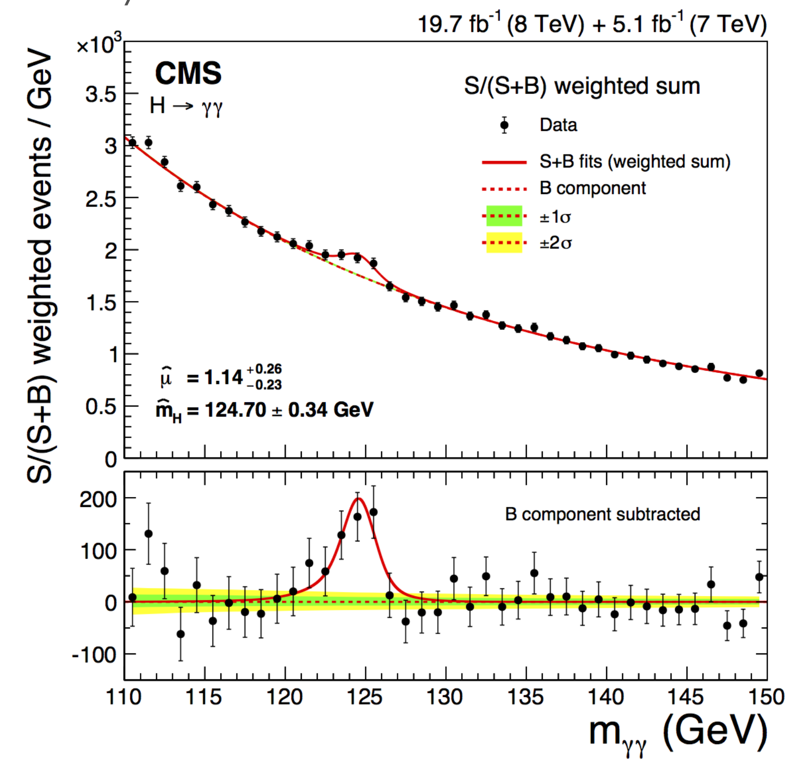
Reconstruction

ML for HGCAL reconstruction

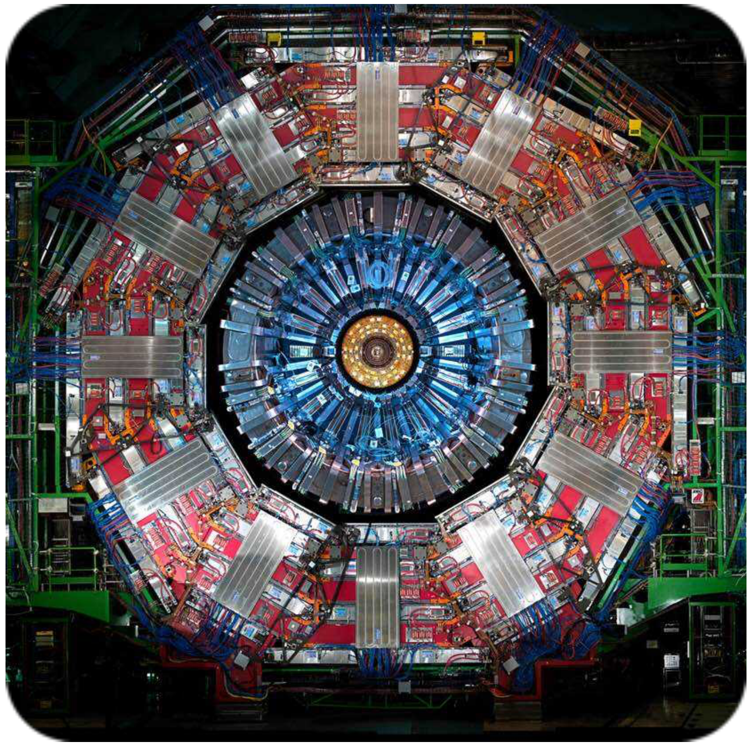
ML for particle flow



Analysis



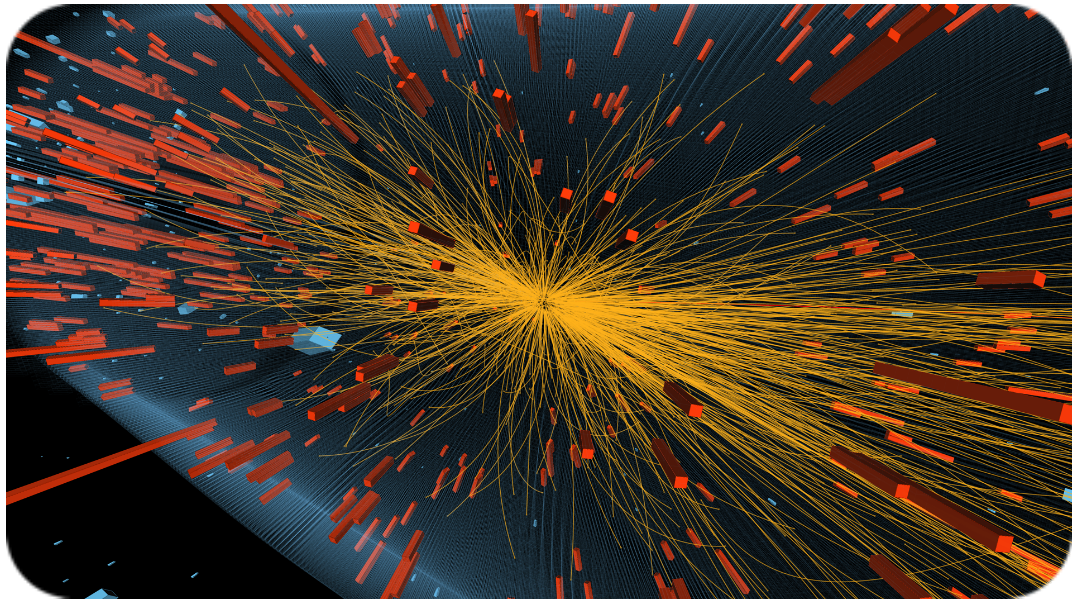
Data in the CMS experiment



Collection
Triggers

hls4ml

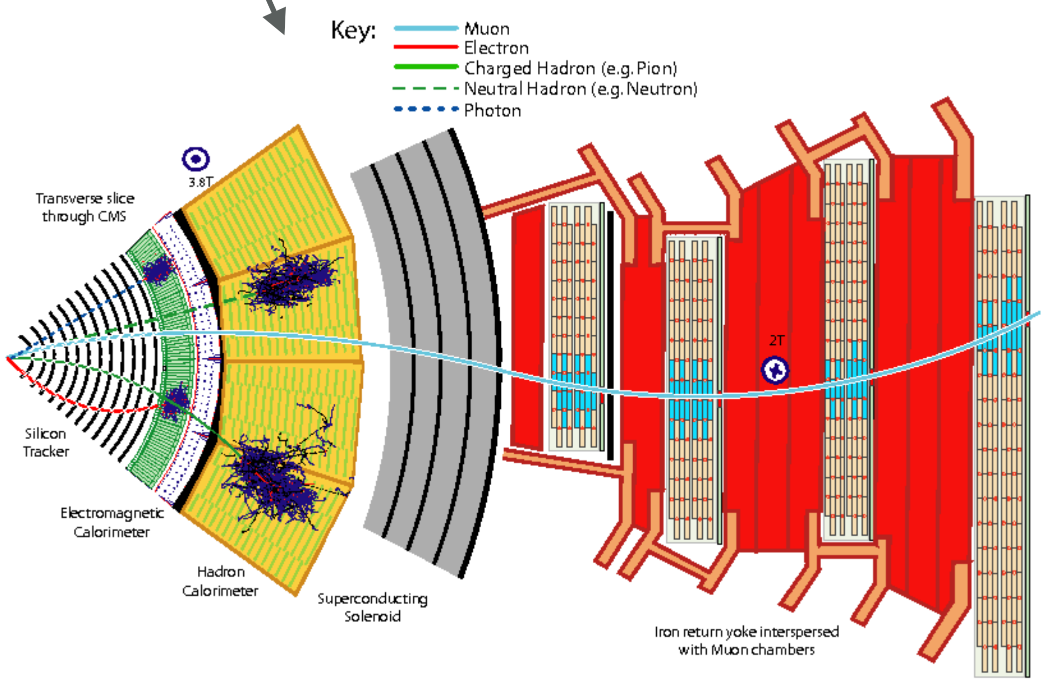
[Ultra-fast real time ML at the L1 trigger]



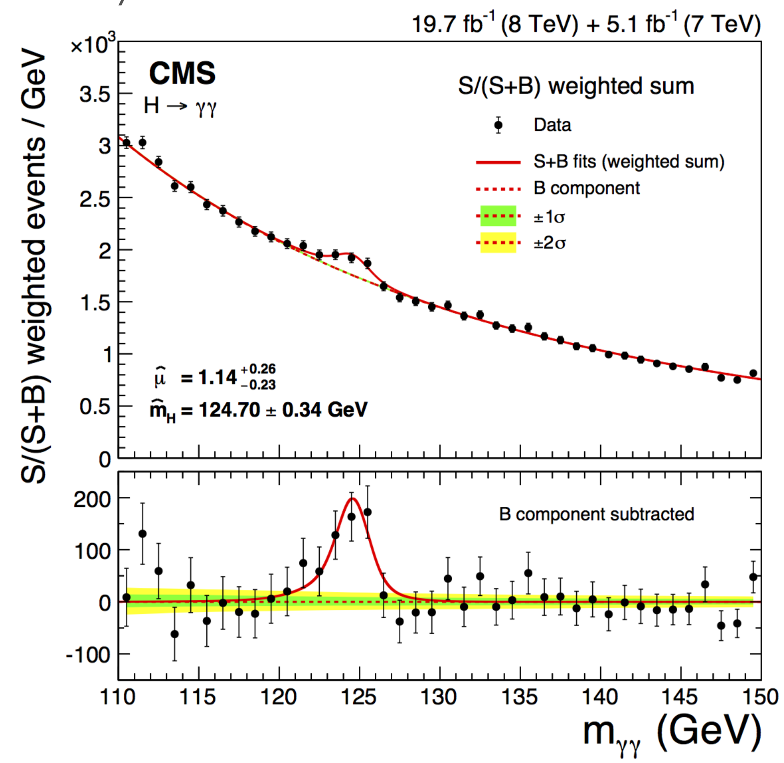
[ML for simulations] Simulation

Reconstruction

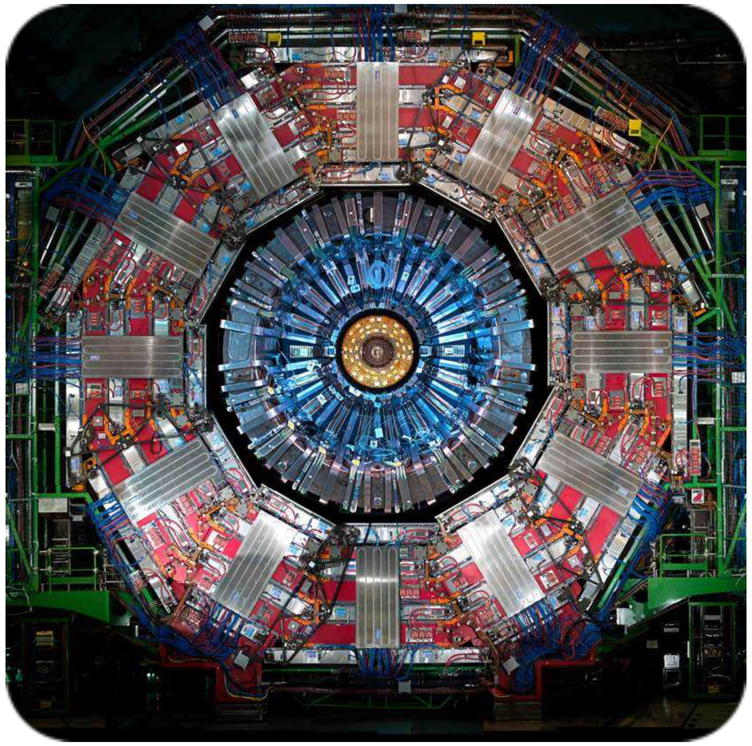
[ML for HGCAL reconstruction]
[ML for particle flow]



Analysis



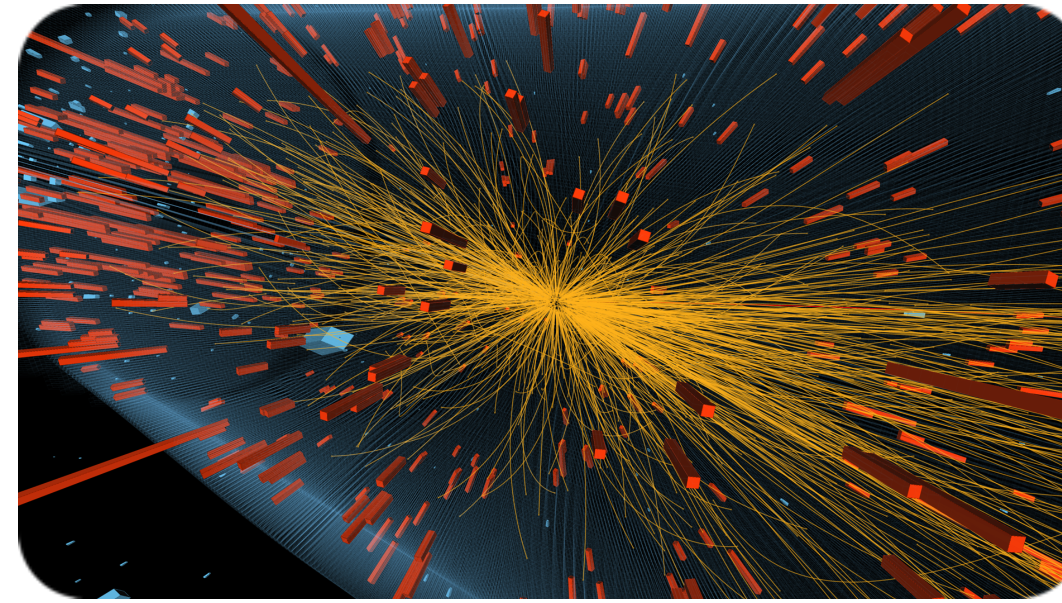
Data in the CMS experiment



Collection
Triggers

hls4ml

Ultra-fast real time ML at the L1 trigger

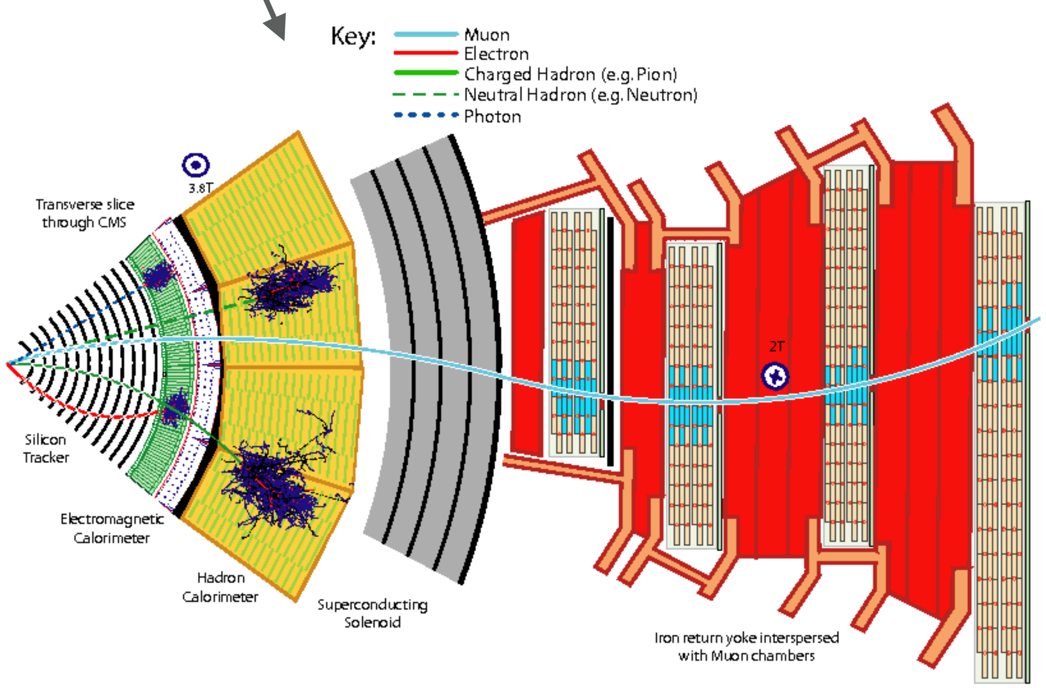


ML for simulations

Simulation

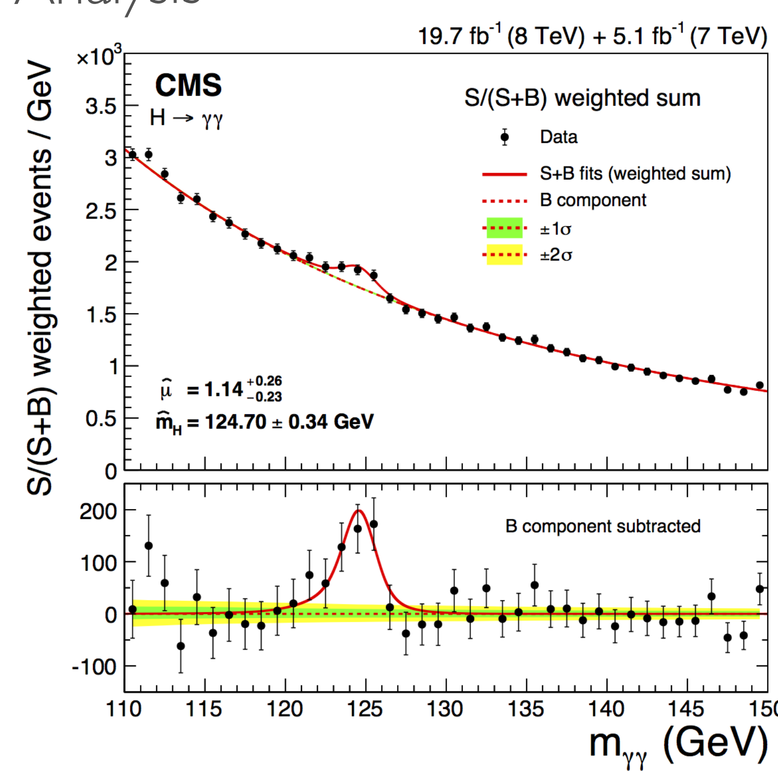
Reconstruction

ML for HGCAL reconstruction
ML for particle flow



This Talk

Analysis



Outline

Outline

- By no means a complete list - novel ML techniques are being developed every day for CMS analyses

Outline

- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups

Outline

- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups
- Beyond classic use case of improved S vs B discrimination…

Outline

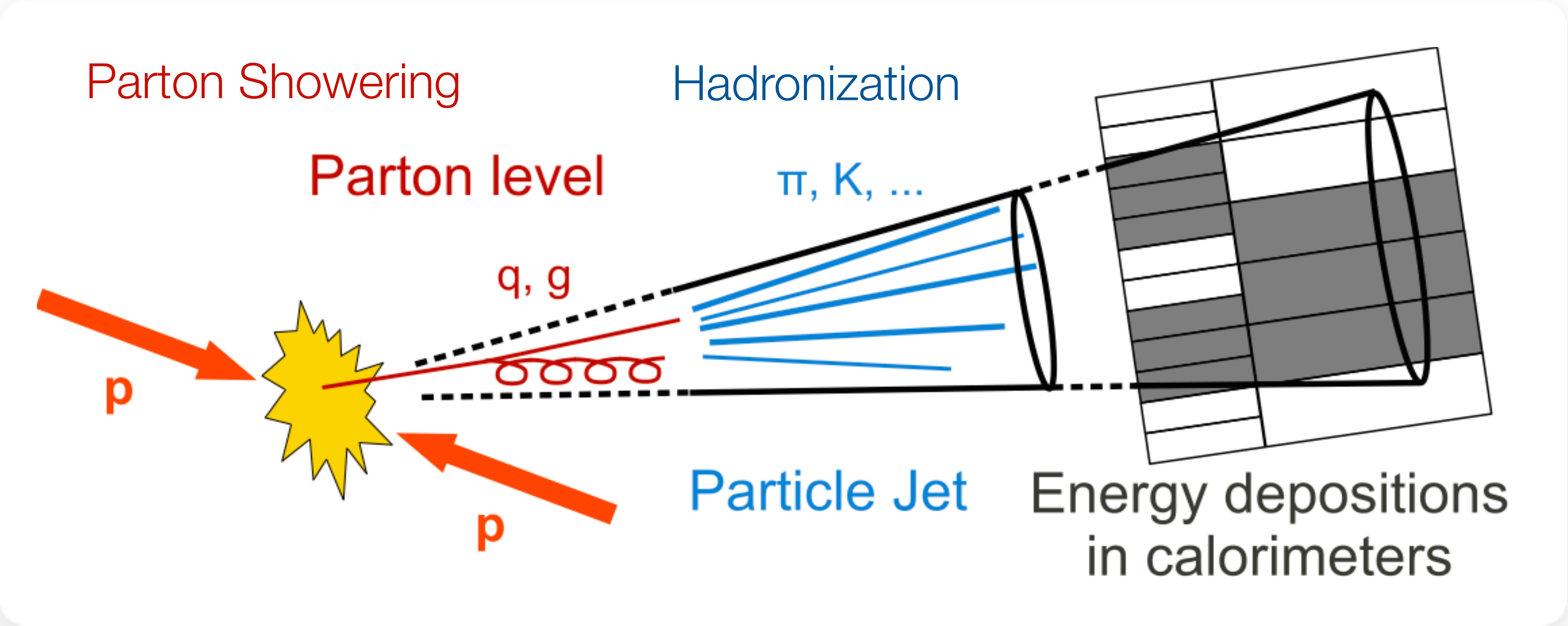
- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups
- Beyond classic use case of improved S vs B discrimination…
- …increased emphasis on challenging experimental aspects:
  - Calibrating deep taggers
  - Reweighting and systematics-aware training
  - Robust background estimation
  - GPU inference

Outline

- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups
- Beyond classic use case of improved S vs B discrimination…
- …increased emphasis on challenging experimental aspects:
  - Calibrating deep taggers
  - Reweighting and systematics-aware training
  - Robust background estimation
  - GPU inference

BTV-22-001

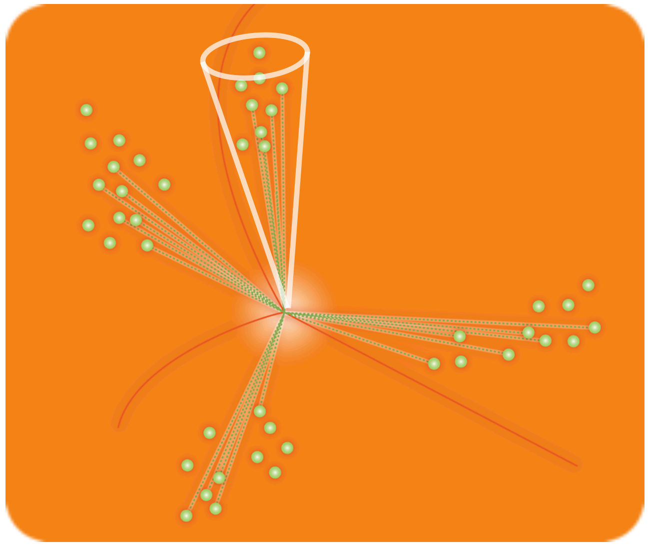
Quarks and gluons → *jets* in our detector



- Quarks and gluons "decay" too quickly to be observed directly
- We analyze the resulting "*jets*" to infer the originating parton
- Complex [O(100) particles] and noisy signal in identification and energy / mass resolution

BTV-22-001

Ripe playground for *deep* learning



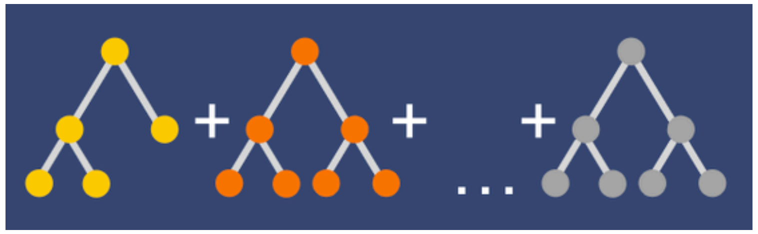
Ripe playground for *deep* learning

Time

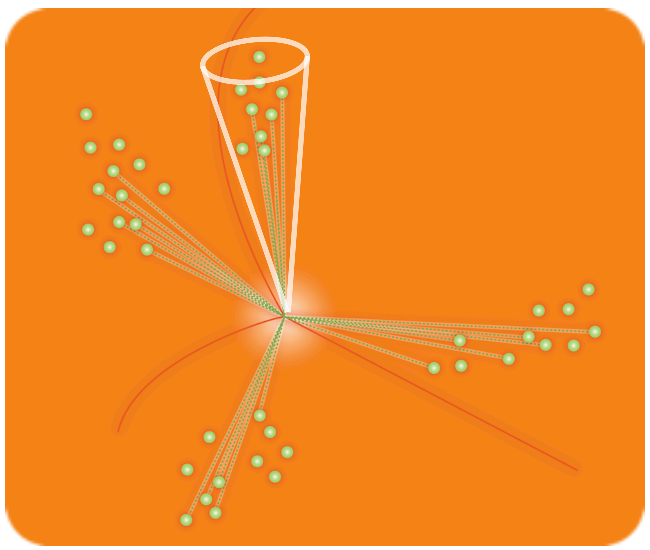
Start of LHC

High-level features
Jet mass, vertices etc.

BDT / Shallow NNs



e.g. double-B



Now

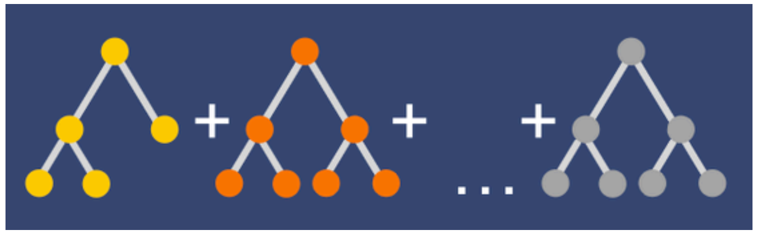
Ripe playground for *deep* learning

Time

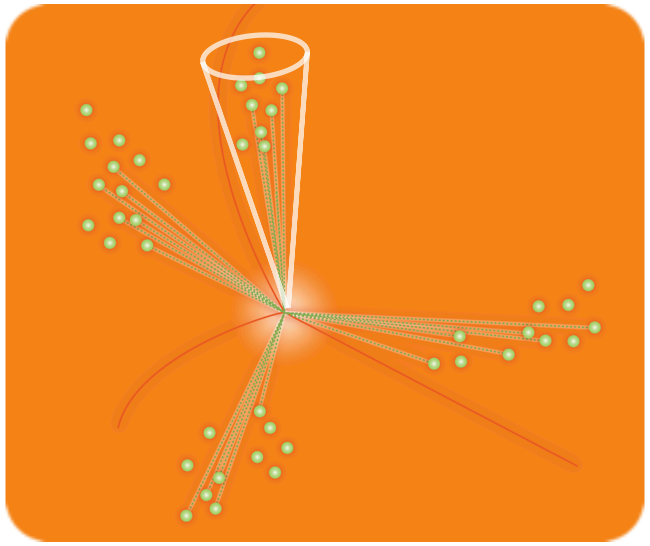
Start of LHC

High-level features
Jet mass, vertices etc.

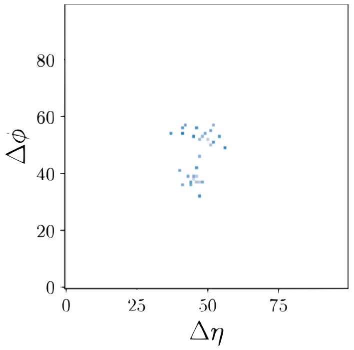
BDT / Shallow NNs



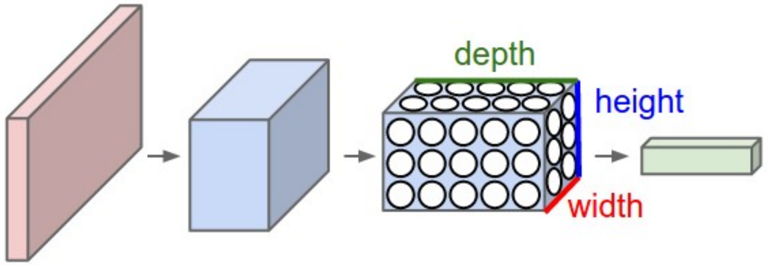
e.g. double-B



Jet images



Convolutional neural networks



e.g. DeepAK8

Now

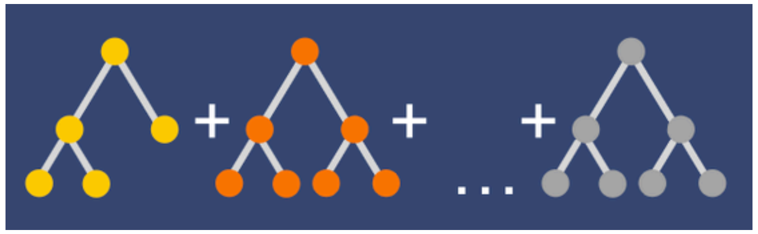
Ripe playground for *deep* learning

Time

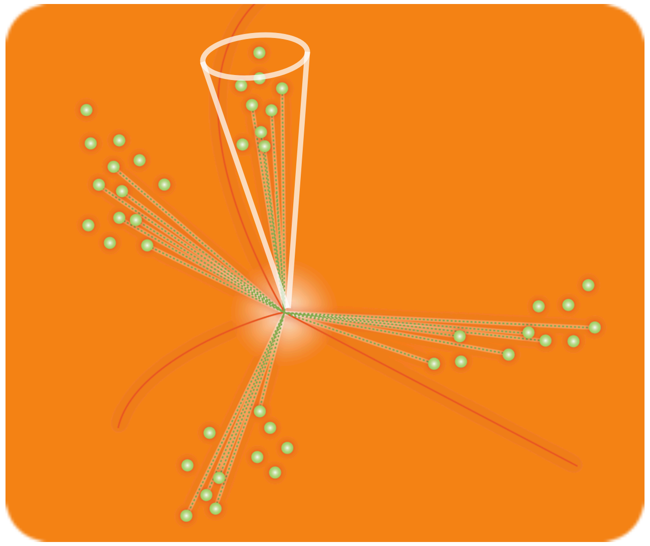
Start of LHC

High-level features
Jet mass, vertices etc.

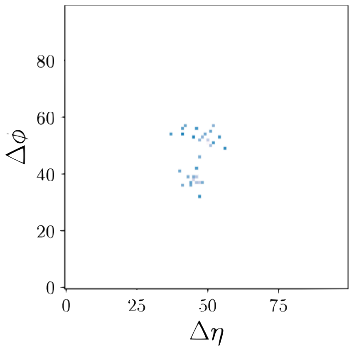
BDT / Shallow NNs



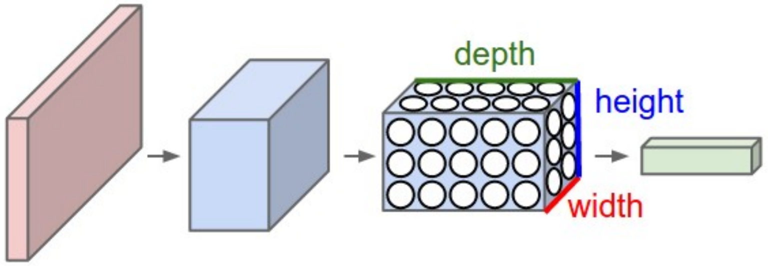
e.g. double-B



Jet images

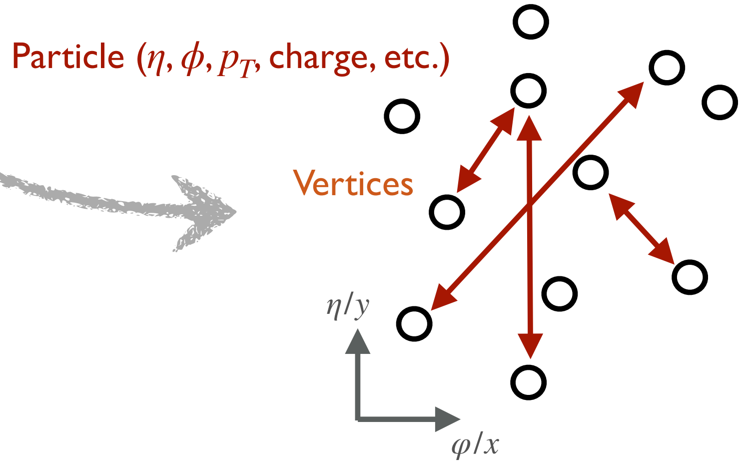


Convolutional neural networks



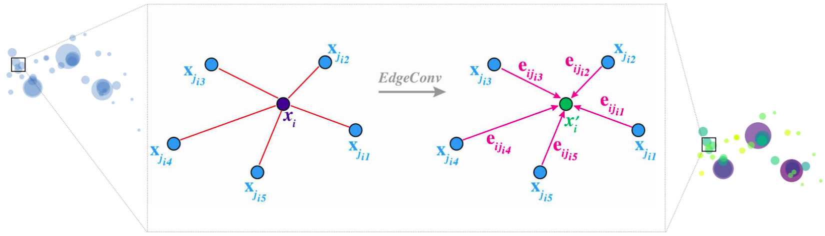
e.g. DeepAK8

Point / "particle" clouds



Now

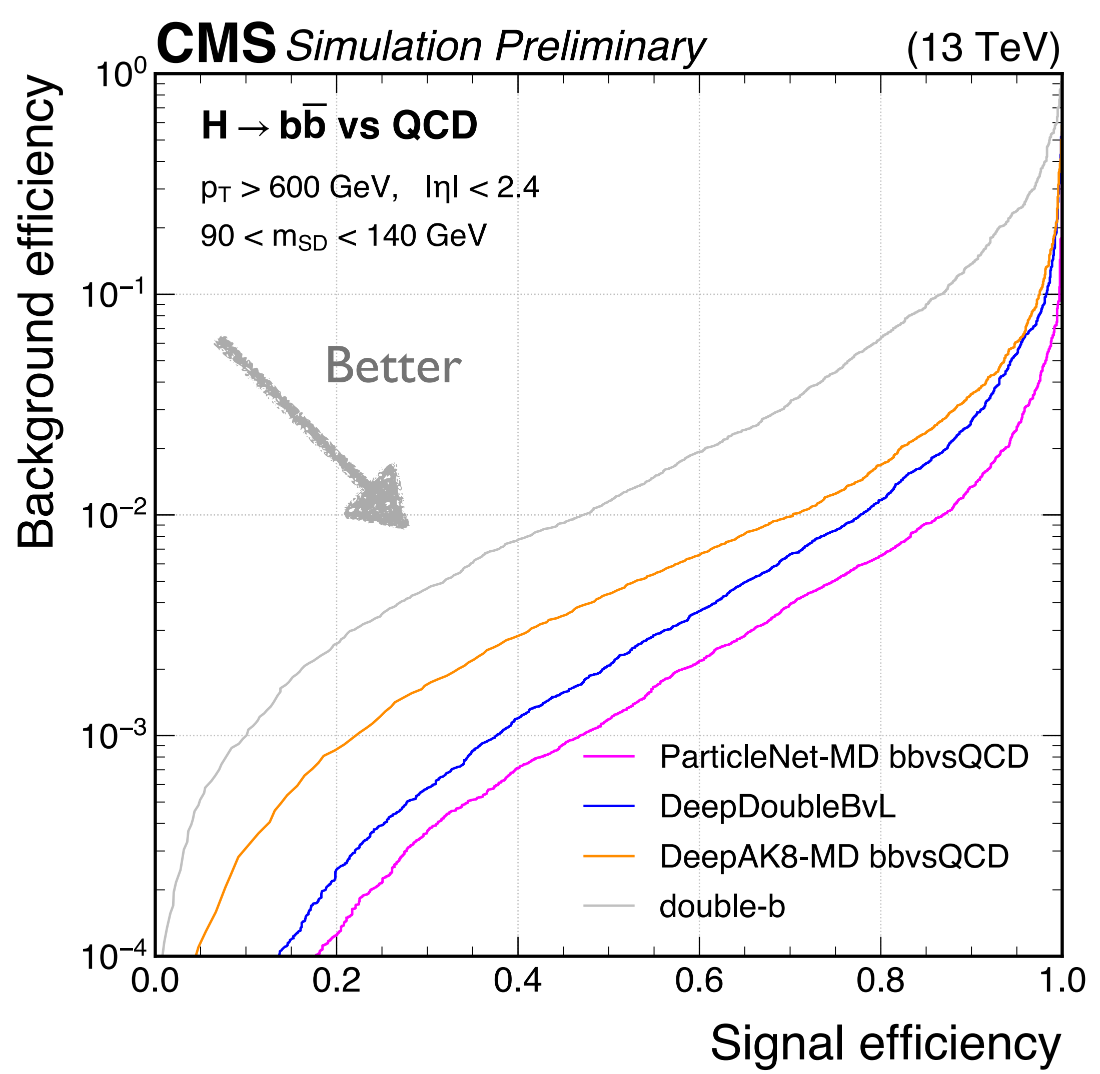
Graph neural networks / transformers



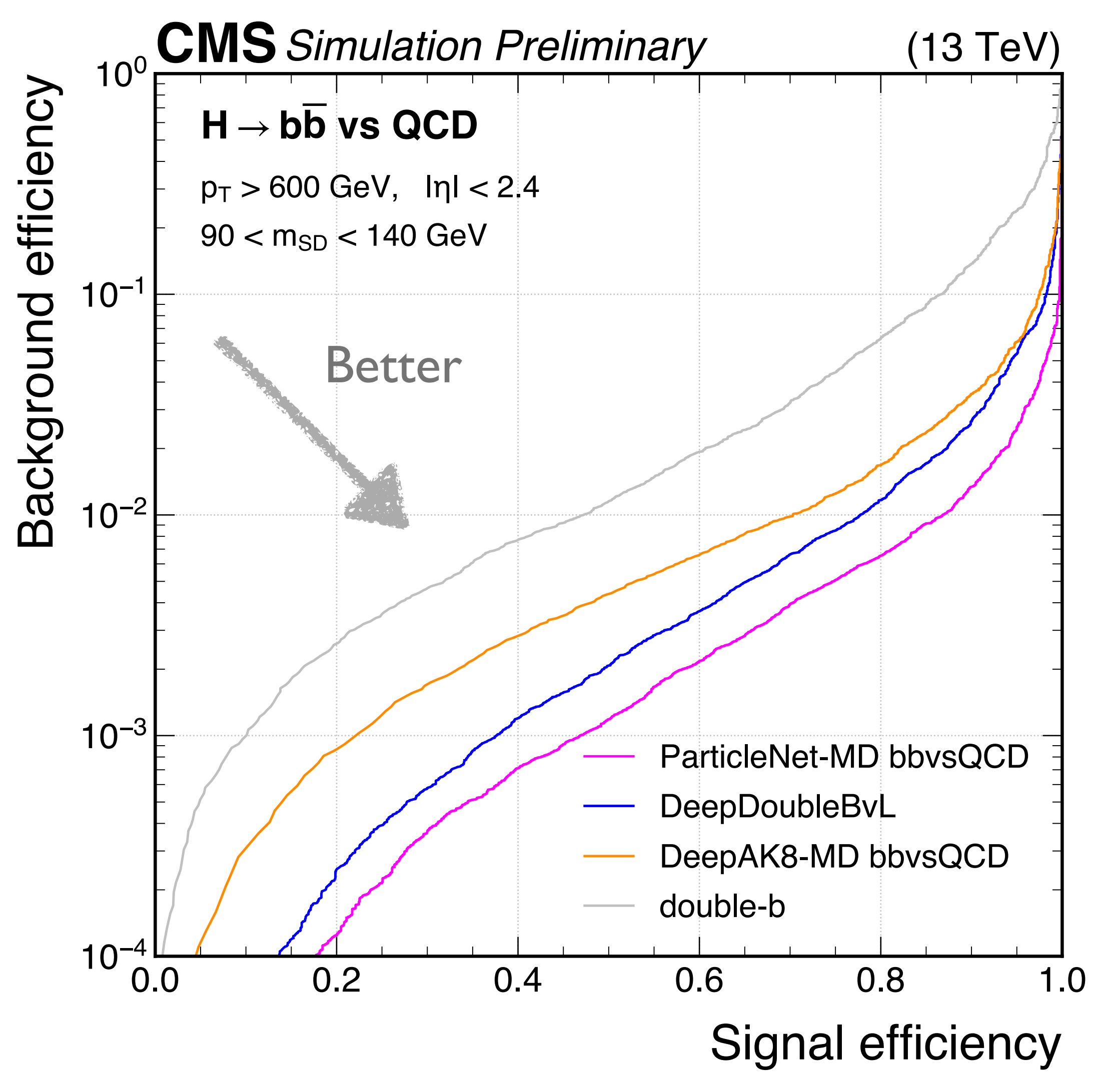
e.g. ParticleNet

BTV-22-001

Example: boosted Higgs→bb vs QCD background jets



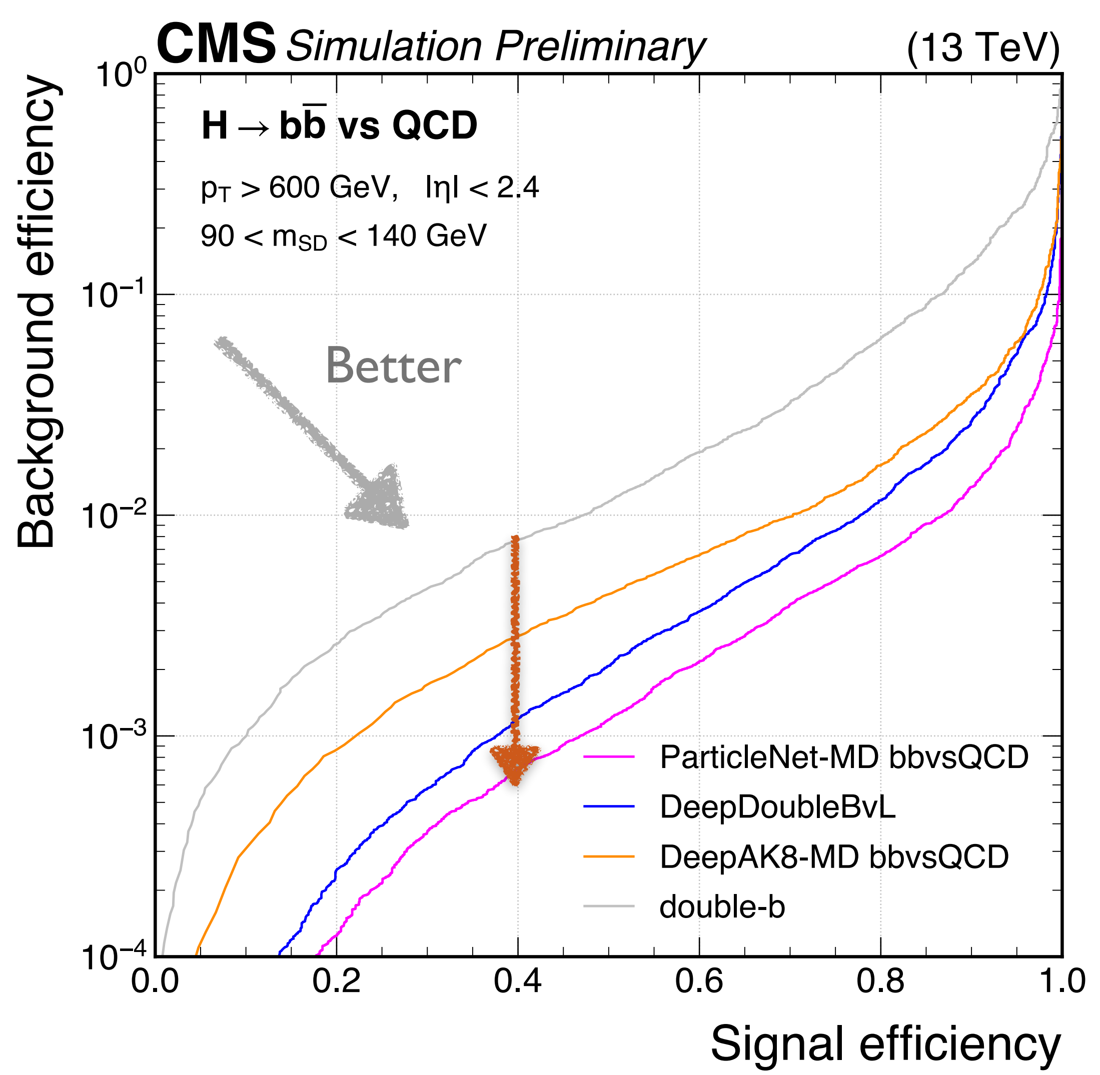
Example: boosted Higgs→bb vs QCD background jets



- Dominant Higgs decay mode, but assumed to drowned out by the LHC background

BTV-22-001

Example: boosted Higgs→bb vs QCD background jets

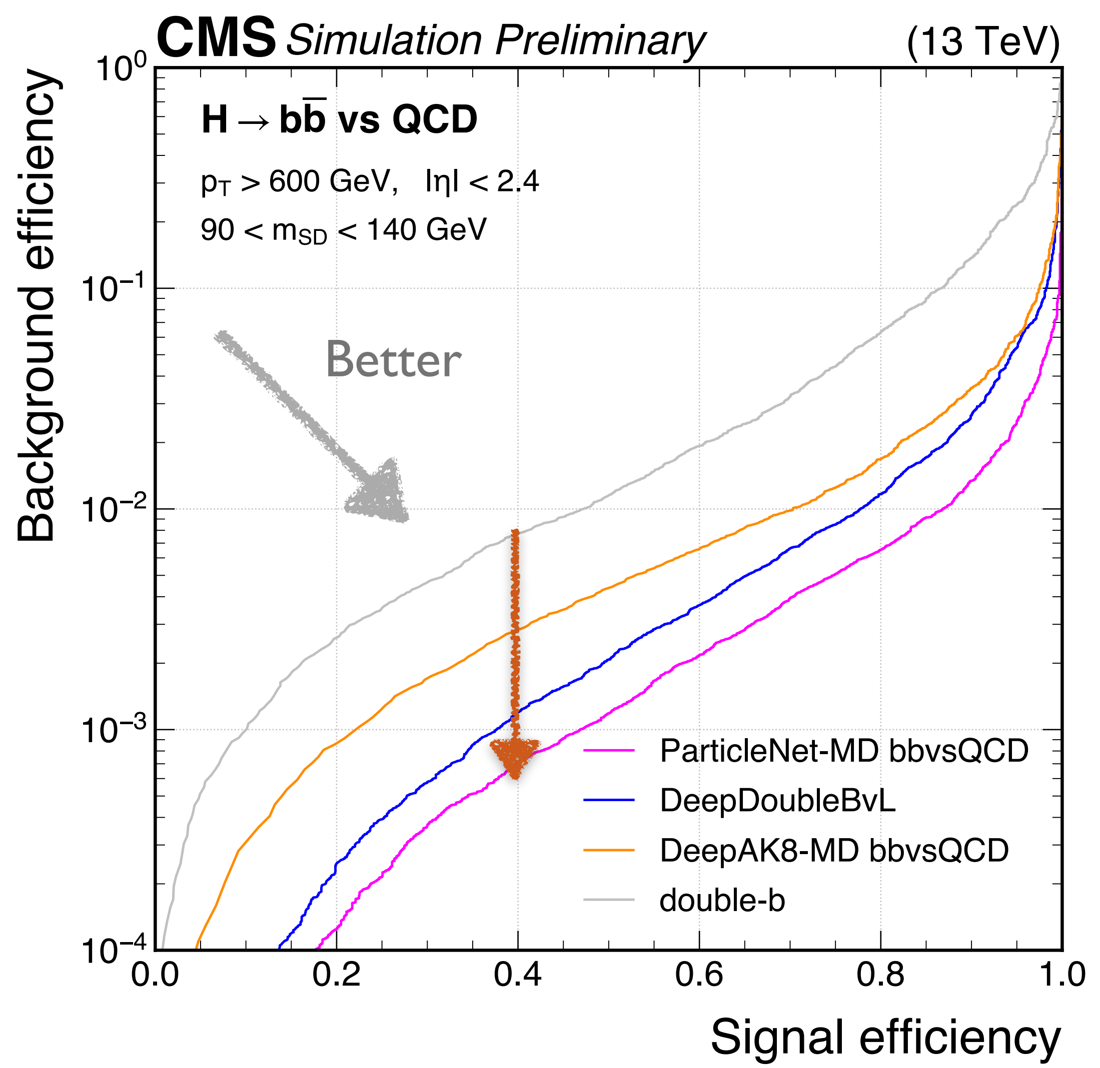


Through increasingly sophisticated, deeper algorithms, >20x better background rejection for the same signal efficiency

- Dominant Higgs decay mode, but assumed to drowned out by the LHC background

BTV-22-001

Example: boosted Higgs→bb vs QCD background jets



Through increasingly sophisticated, deeper algorithms, >20x better background rejection for the same signal efficiency

e.g. 99.9% BG rejection for 50% signal eff!

- Dominant Higgs decay mode, but assumed to drowned out by the LHC background
- DL made H→bb, HH→4b, Z'→bb, H→cc all possible in the last ~5 years!

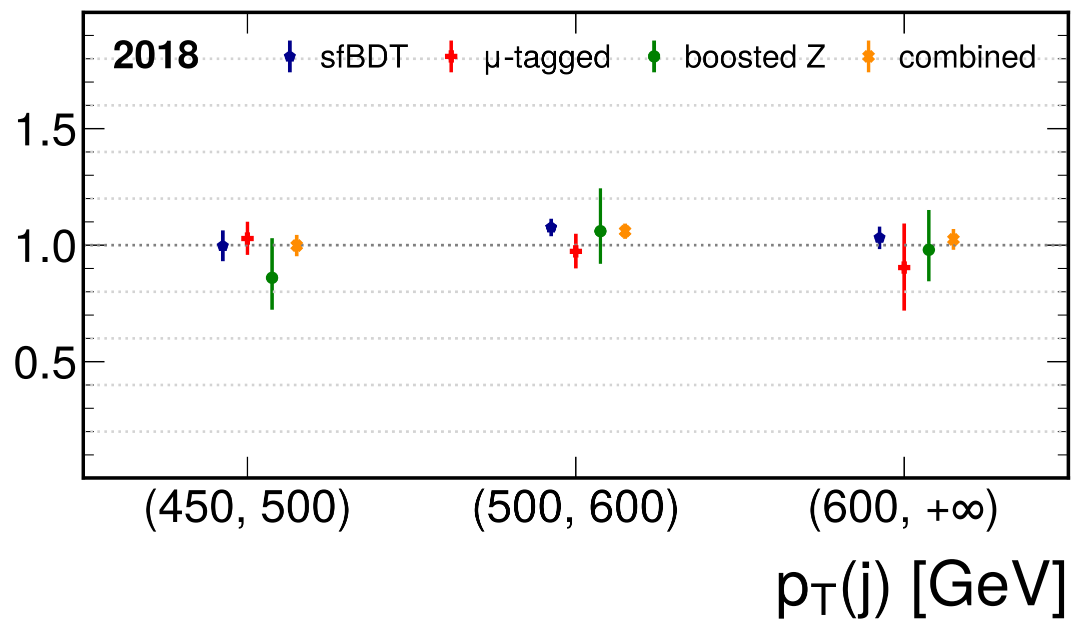
But tagging is the easy part! How to calibrate?

But tagging is the easy part! How to calibrate?

- Progress as well in innovative methods for measuring data / simulation agreement (often more important than the algorithm itself!)

But tagging is the easy part! How to calibrate?

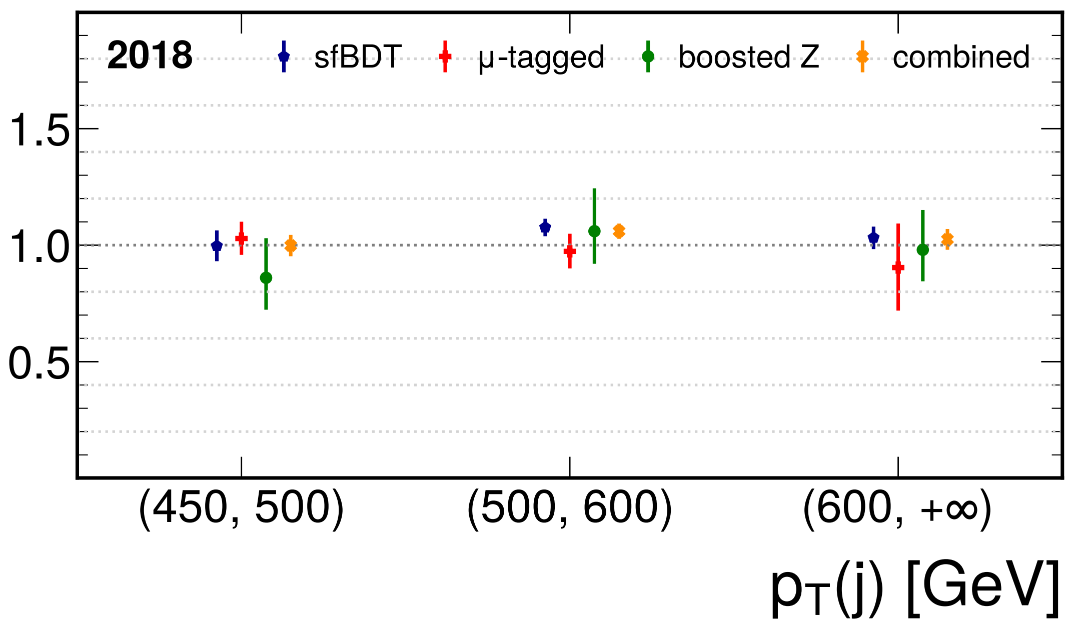
Using g→bb, Z→bb, μ-tagged b's as proxies



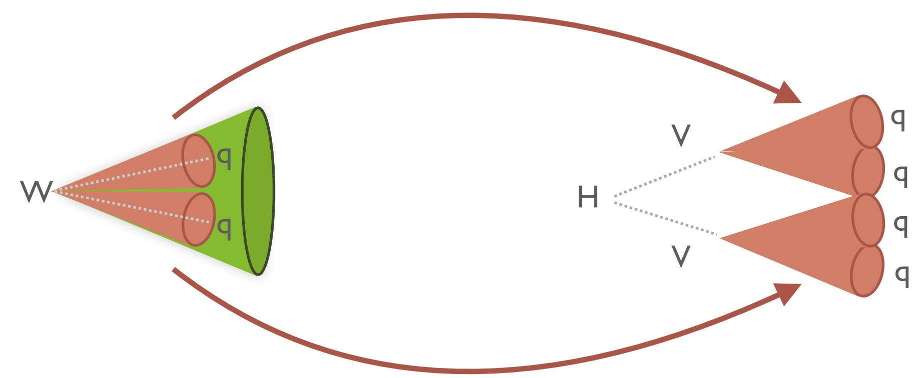
- Progress as well in innovative methods for measuring data / simulation agreement (often more important than the algorithm itself!)

But tagging is the easy part! How to calibrate?

Using g→bb, Z→bb, μ-tagged b's as proxies



Calibrating per subjet using the Lund jet plane for more complex topologies



- Progress as well in innovative methods for measuring data / simulation agreement (often more important than the algorithm itself!)

Outline

- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups
- Beyond classic use case of improved S vs B discrimination…
- …increased emphasis on challenging experimental aspects:
  - Calibrating deep taggers
  - Reweighting and systematics-aware training
  - Robust background estimation
  - GPU inference

ML reweighting for systematic variations

[MLG-24-001](#)
[DCTR method](#)

Can we learn variations instead of costly simulations?

Can we learn variations instead of costly simulations?

- Theory variation simulations are very computationally intensive
  Already limiting factor in some top measurements

Can we learn variations instead of costly simulations?

- Theory variation simulations are very computationally intensive
  Already limiting factor in some top measurements
- DCTR method: learn a NN-reweighting instead!

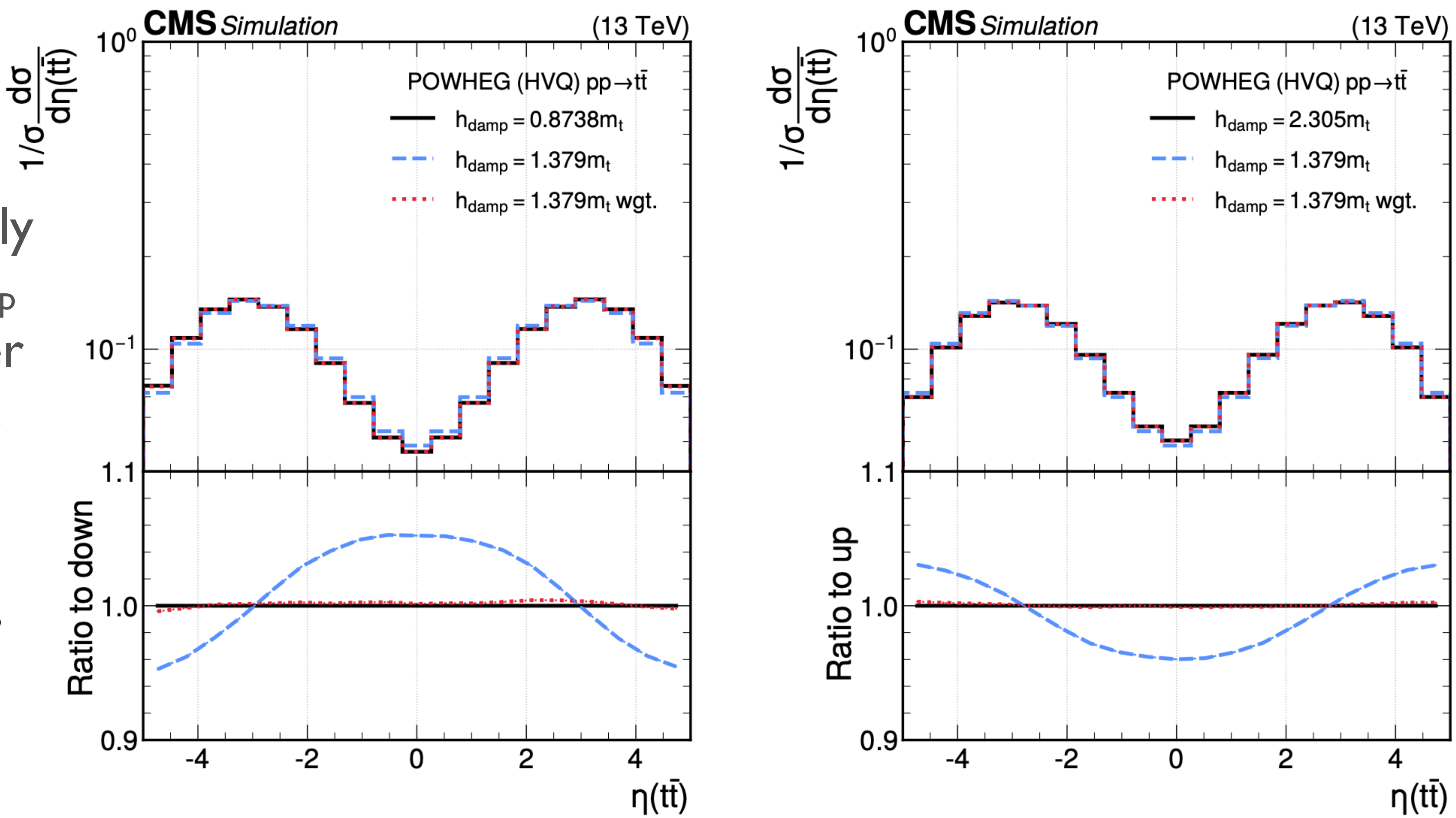
ML reweighting for systematic variations

MLG-24-001
DCTR method

Can we learn variations instead of costly simulations?

Varying empirically determined h_{damp} theory parameter in $t\bar{t}$ simulations

ML reweighting
h_{damp} = $1.38m_t$ to h_{damp} = $0.87m_t$



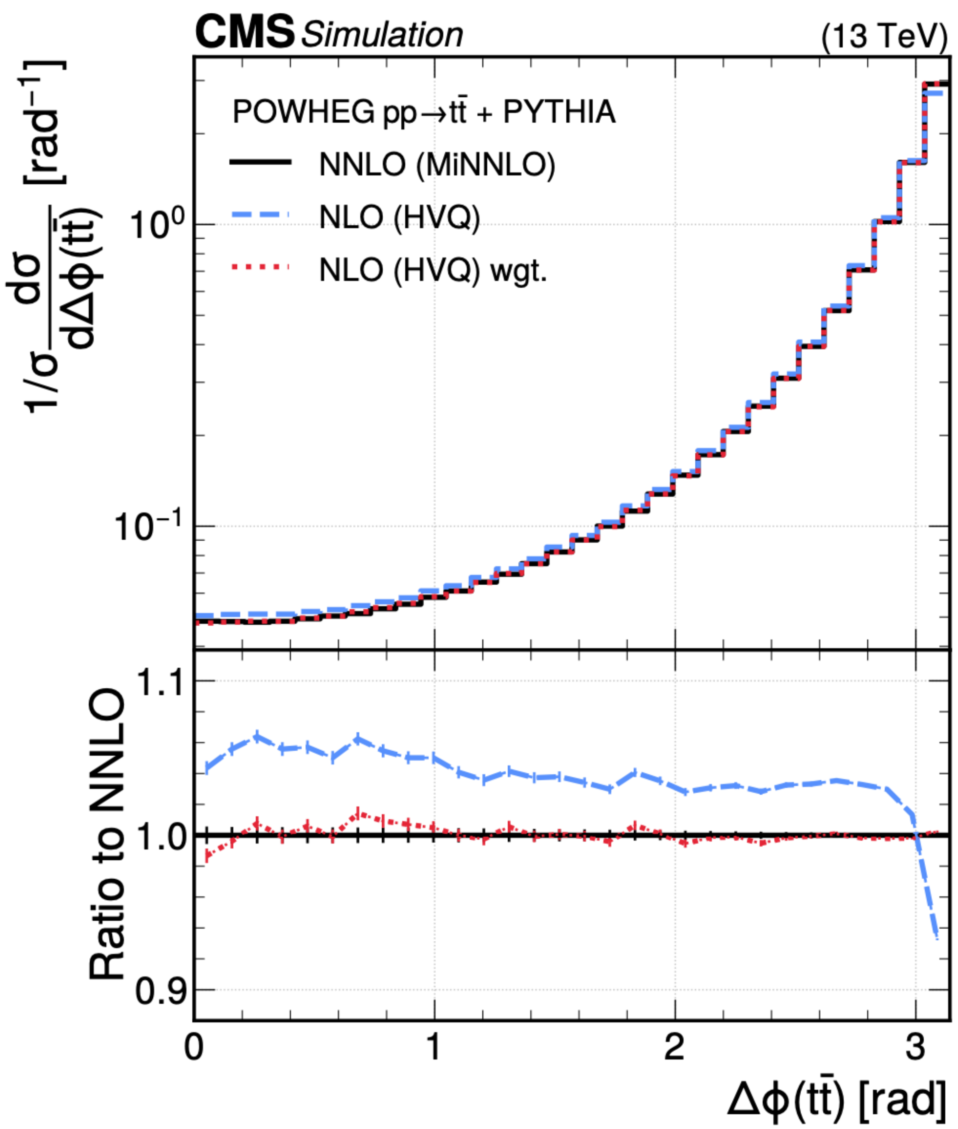
- Theory variation simulations are very computationally intensive
  Already limiting factor in some top measurements
- DCTR method: learn a NN-reweighting instead!

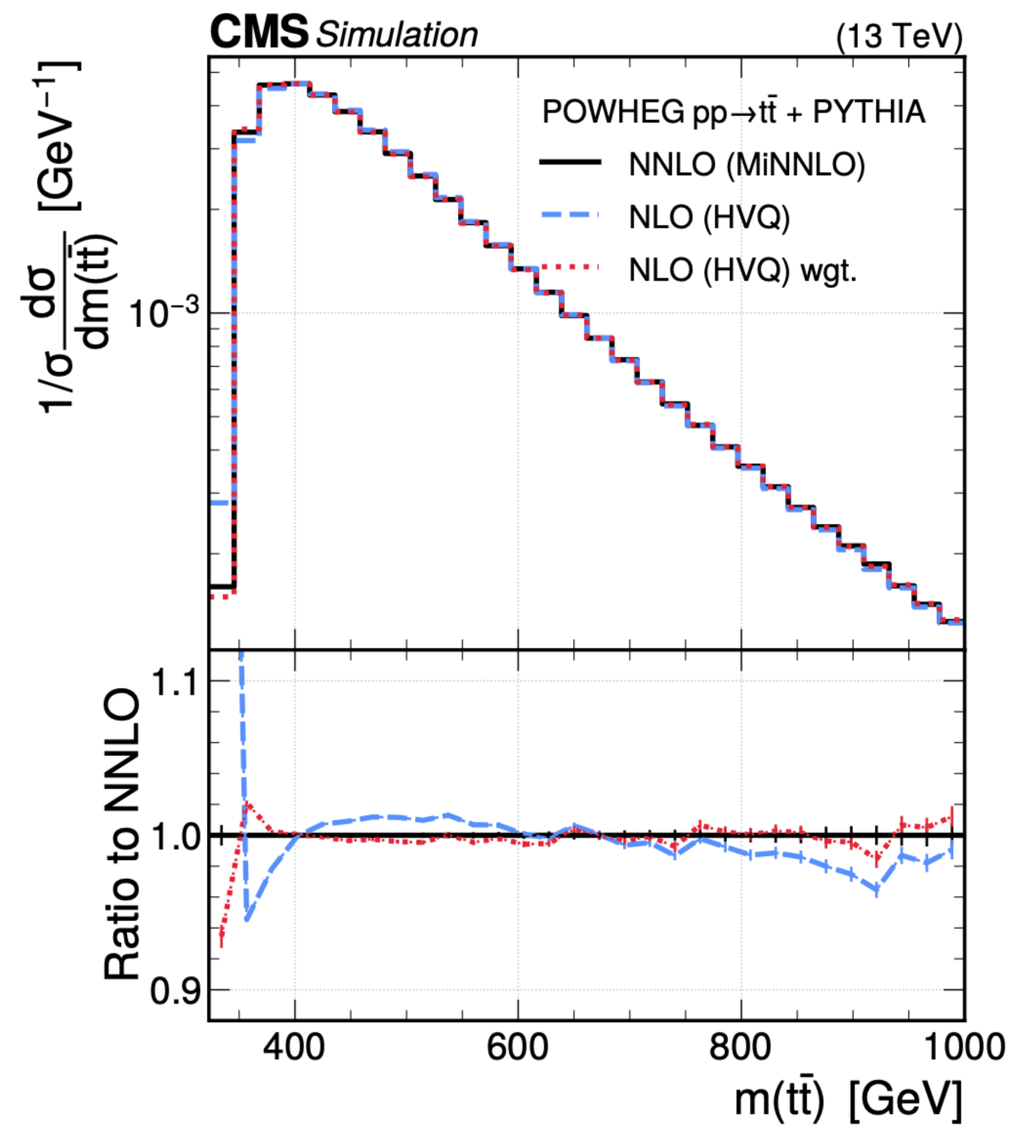
MLG-24-001
DCTR method

Can we learn NLO → NNLO corrections?

Learning reweighting in full, unbinned phase space

ML reweighting
NLO to NNLO



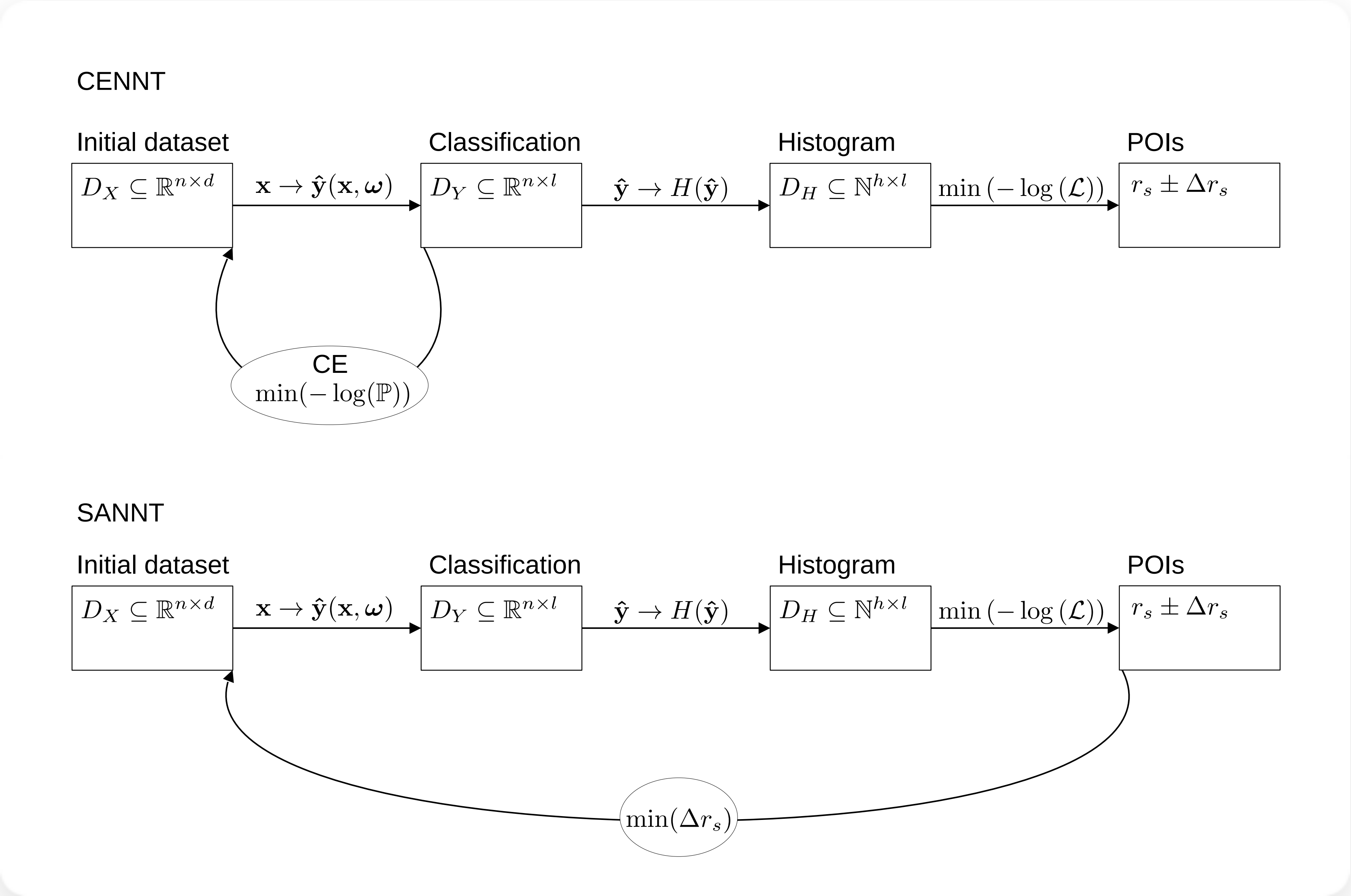


- Again, new simulations are very computationally intensive
- DCTR method: learn a NN-reweighting from existing simulations instead!

What if our tagger can take these systematics into consideration?

Systematics-aware neural network training

What if our tagger can take these systematics into consideration?

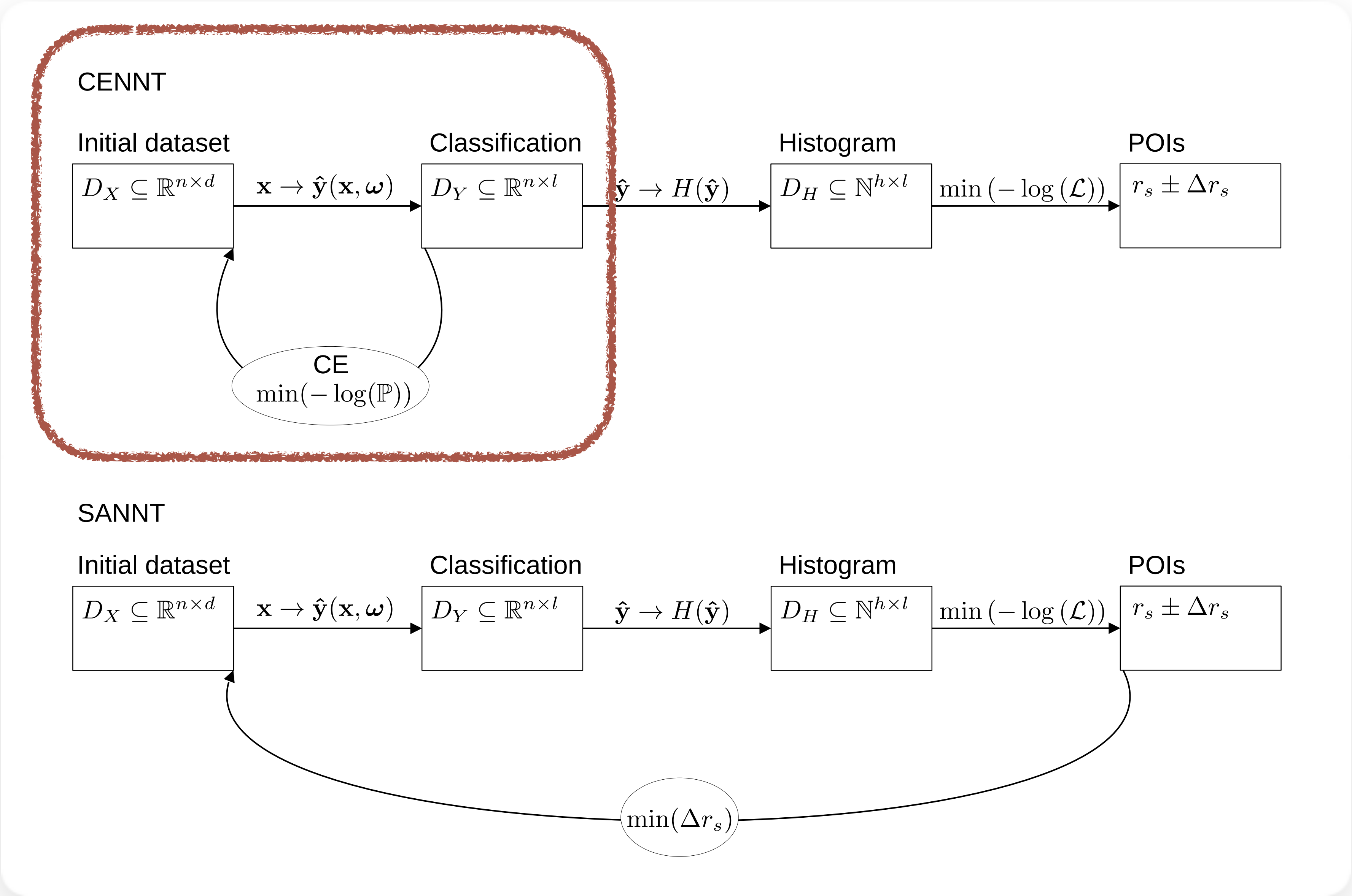


- SANNT: progress towards end-to-end, systematic-aware, analysis optimization

Systematics-aware neural network training

What if our tagger can take these systematics into consideration?

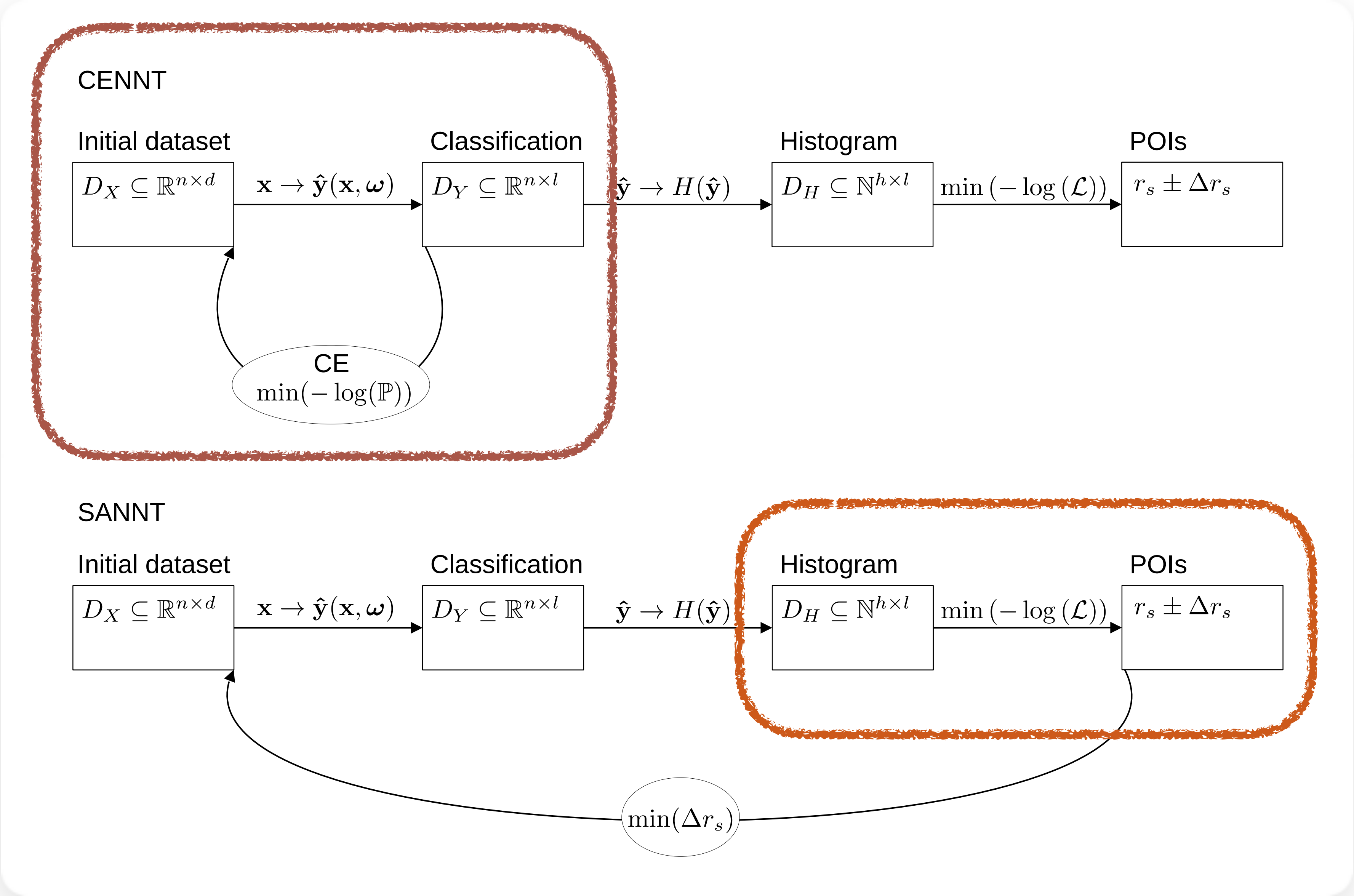
Instead of optimizing classification alone



- SANNT: progress towards end-to-end, systematic-aware, analysis optimization

What if our tagger can take these systematics into consideration?

Instead of optimizing classification alone

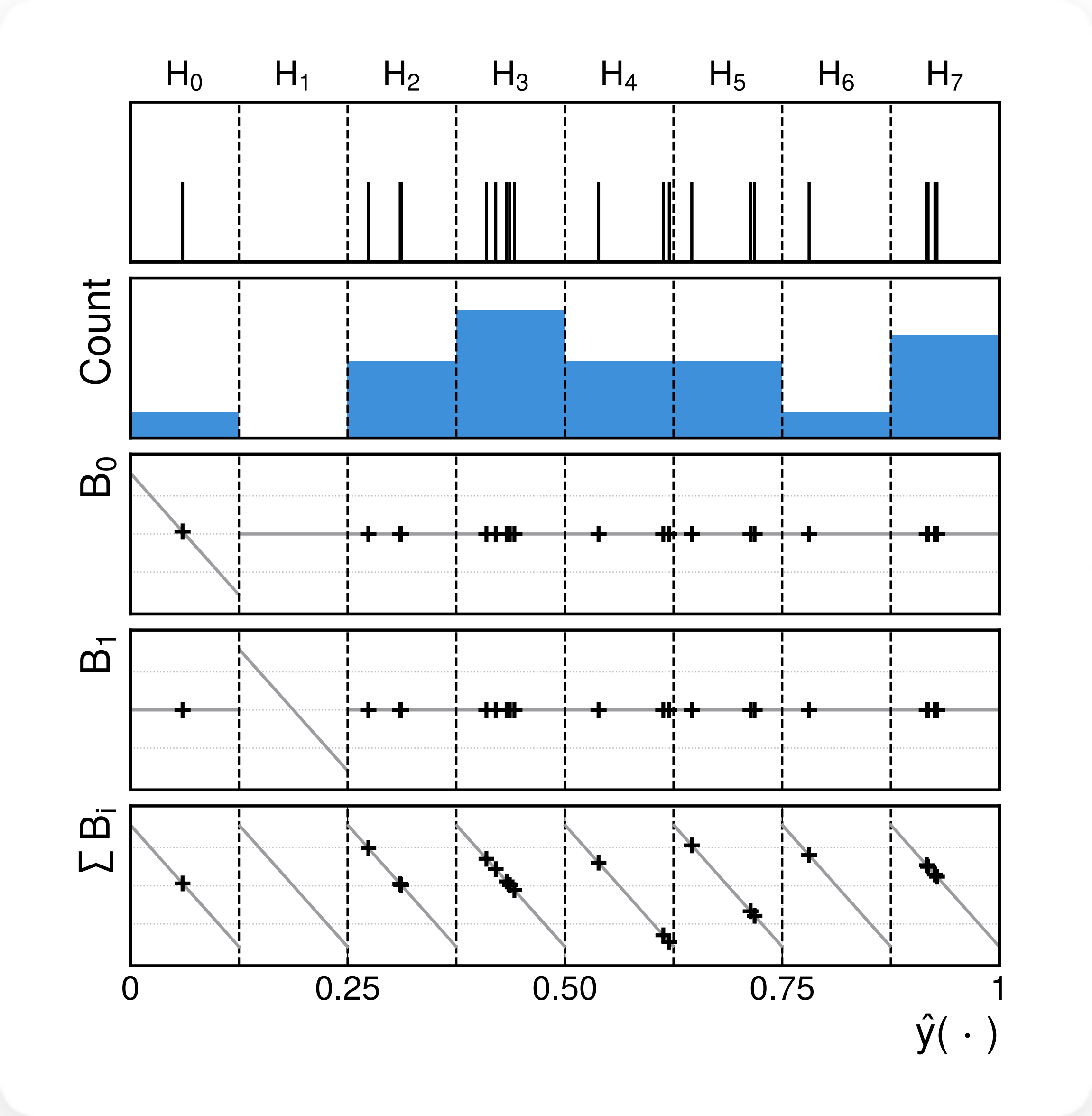


Backpropagate through the histogram and fit as well!

- SANNT: progress towards end-to-end, systematic-aware, analysis optimization

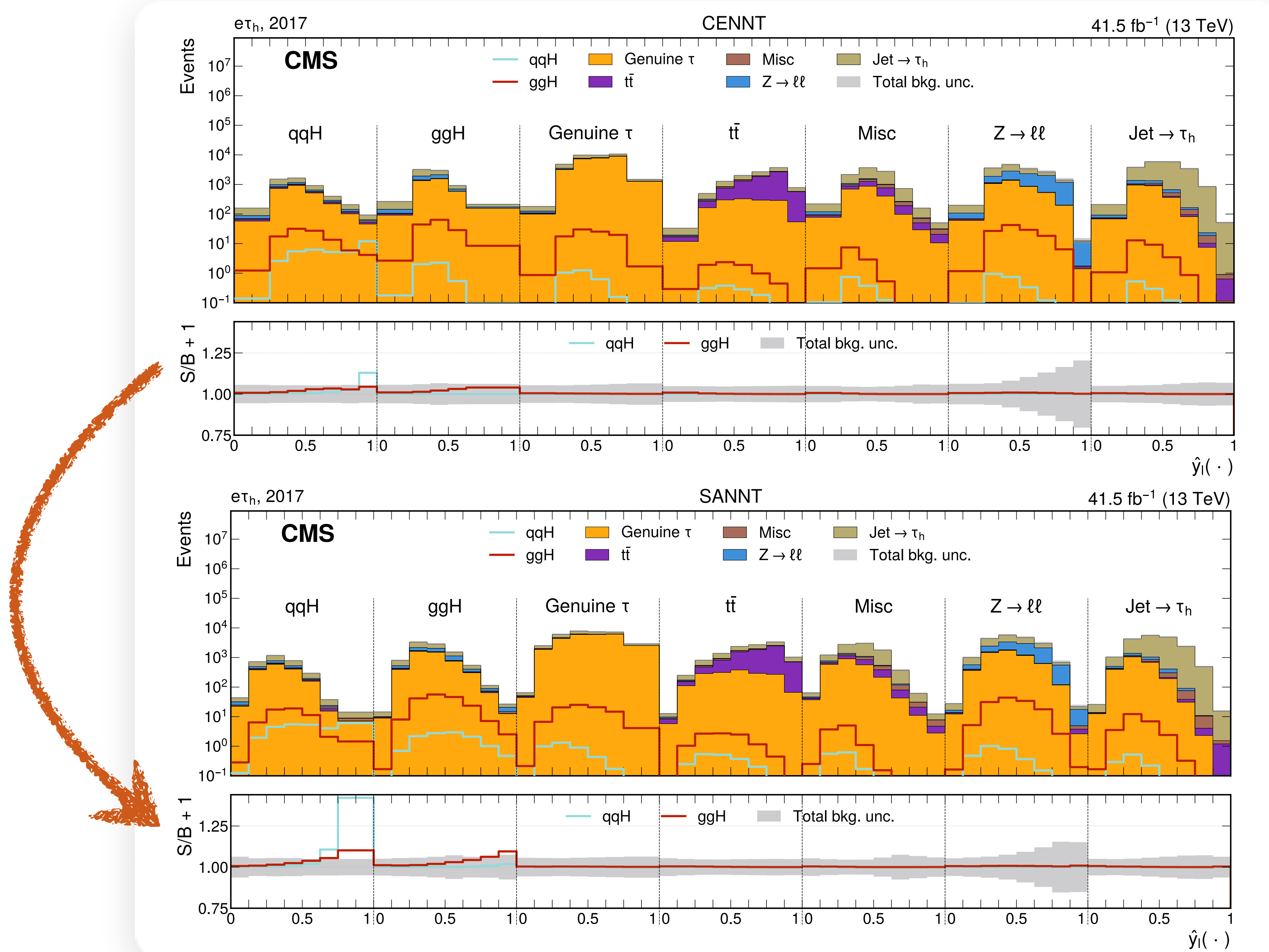
Systematics-aware neural network training

What if our tagger can take these systematics into consideration?



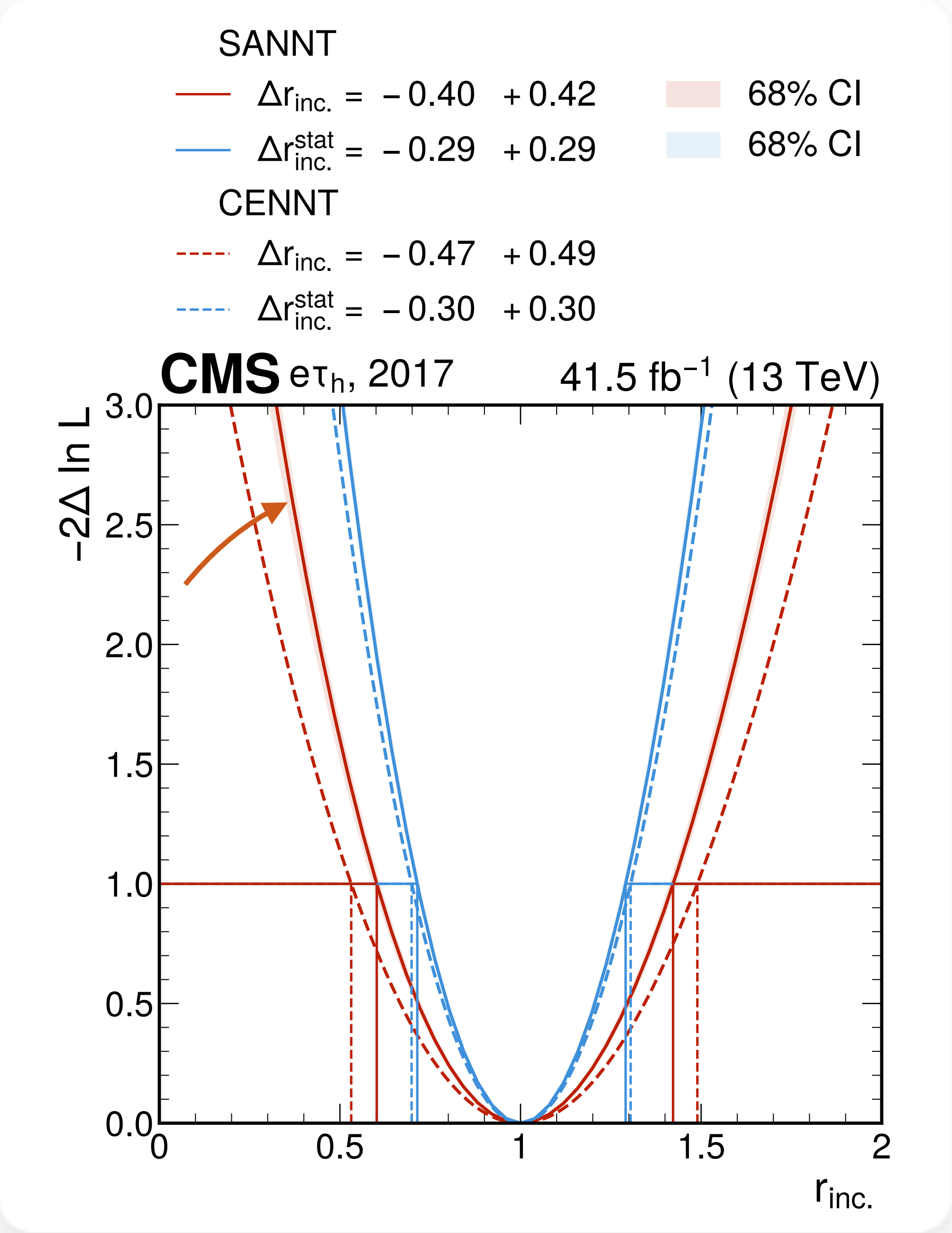
- SANNT: progress towards end-to-end, systematic-aware, analysis optimization
- Backpropagate through the histogram and fit

Significant improvement over traditional analysis



Applied to CMS H→ττ analysis

Significant improvement over traditional analysis



- Applied to CMS $H \rightarrow \tau\tau$ analysis
- 15% improvement in precision

Outline

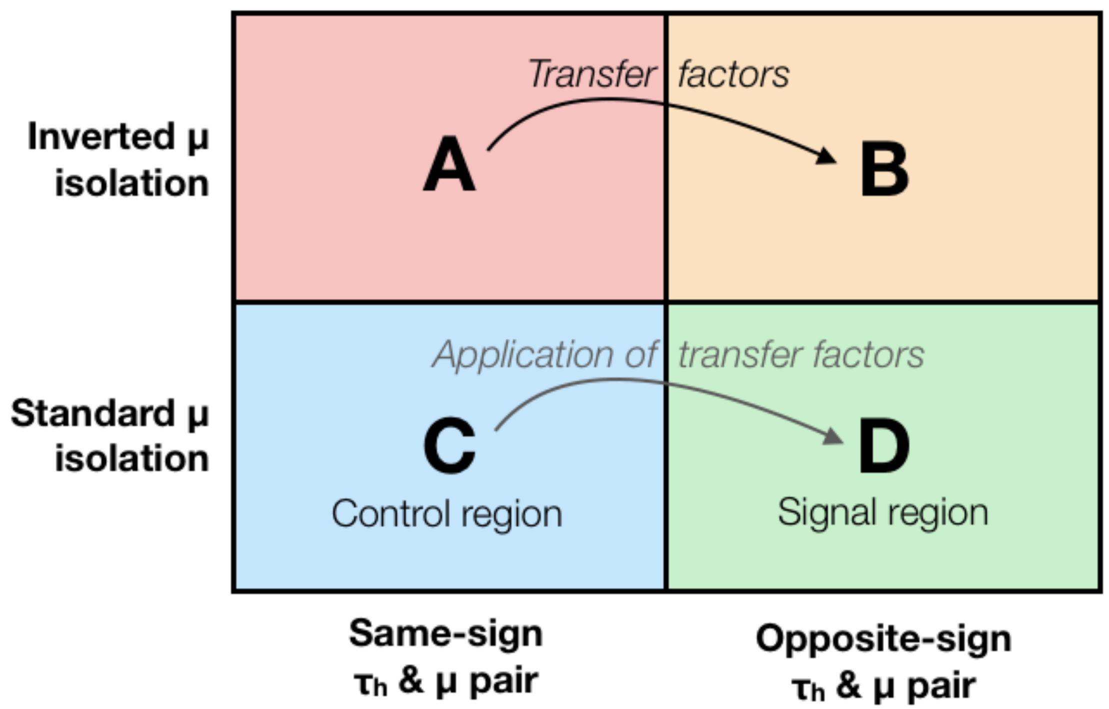
- By no means a complete list - novel ML techniques are being developed every day for CMS analyses
- Focus on CMS publications from the new CMS Machine Learning Group, as well as jet / b-tagging groups
- Beyond classic use case of improved S vs B discrimination…
- …increased emphasis on challenging experimental aspects:
  - Calibrating deep taggers
  - Reweighting and systematics-aware training
  - Robust background estimation
  - GPU inference

Automating the ABCD method with ML

*TEC: non-closure loss (backup)

Automating the ABCD method with ML

Two variables must be decorrelated!



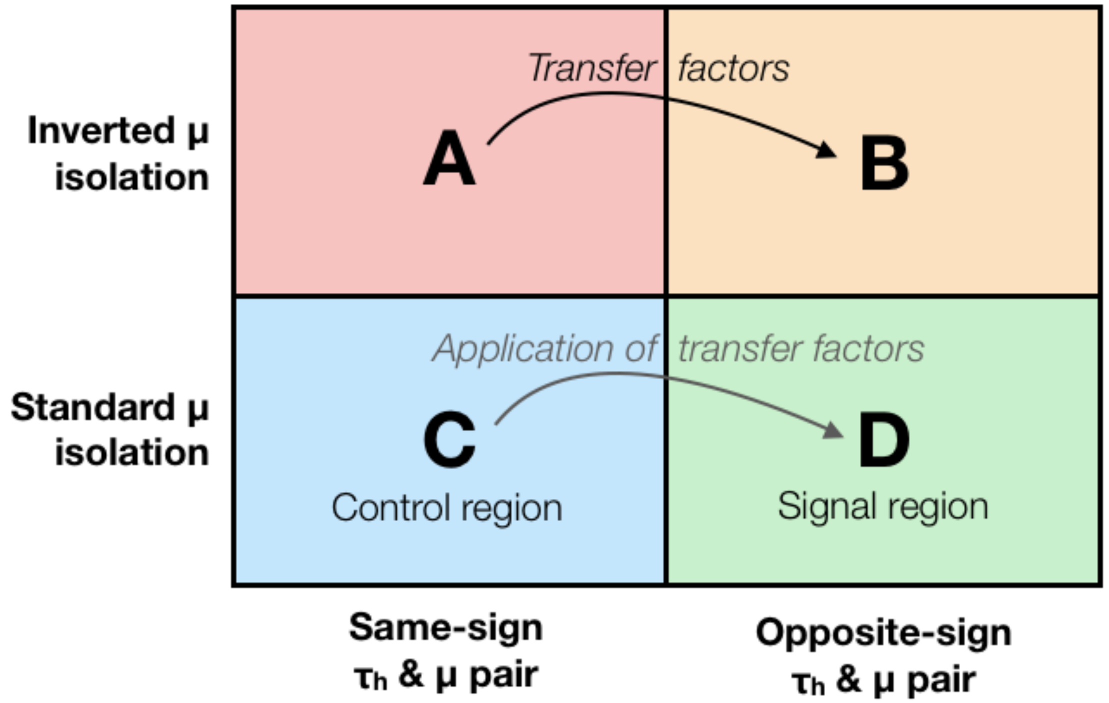
- ABCD: classic method for data-driven background estimation

*TEC: non-closure loss (backup)

MLG-23-003
ABCDisCo method

Automating the ABCD method with ML

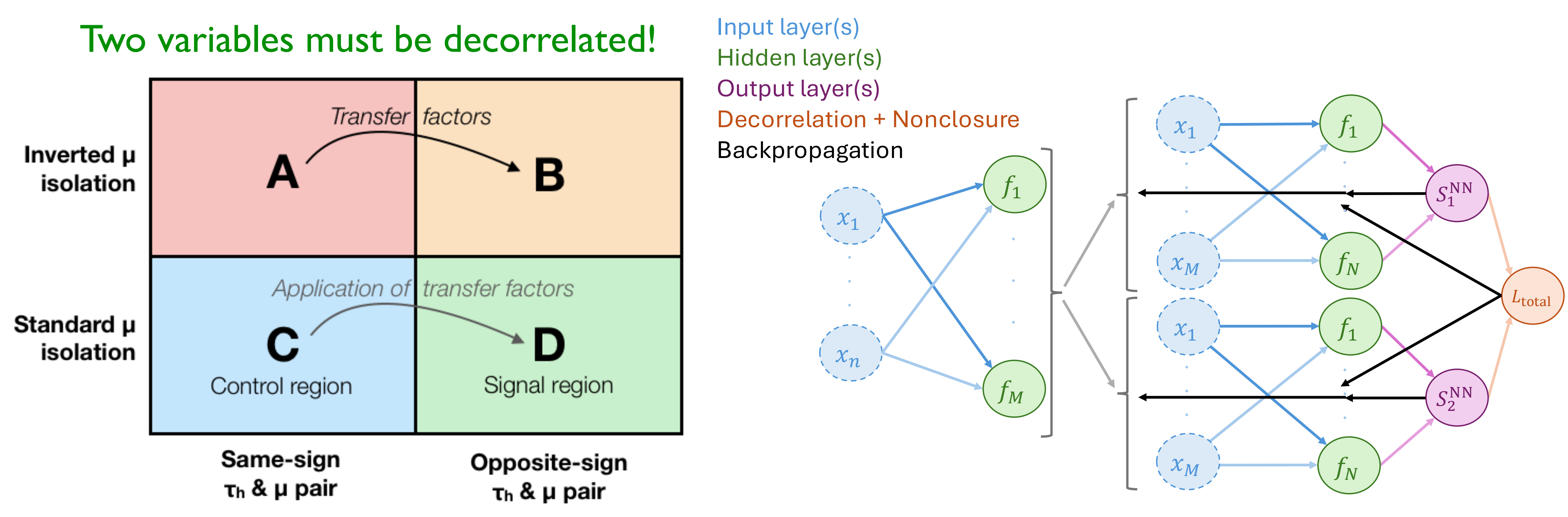
Two variables must be decorrelated!



- ABCD: classic method for data-driven background estimation
- But: needs two decorrelated variables - can be difficult with ML taggers

*TEC: non-closure loss (backup)

Automating the ABCD method with ML

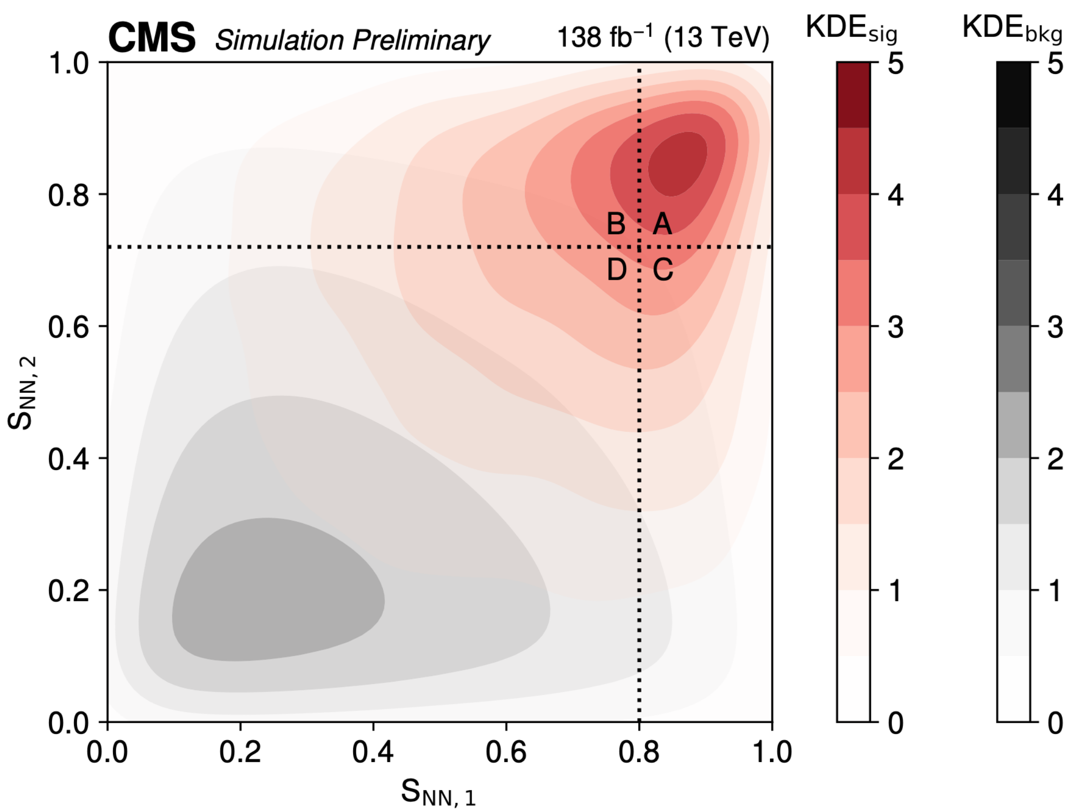


- ABCD: classic method for data-driven background estimation
- But: needs two decorrelated variables - can be difficult with ML taggers
- ABCDisCo: *learn* two decorrelated variables
  Enforced through distance correlation (DisCo) loss function

*TEC: non-closure loss (backup)

Application to SUSY search

Outputs are well decorrelated

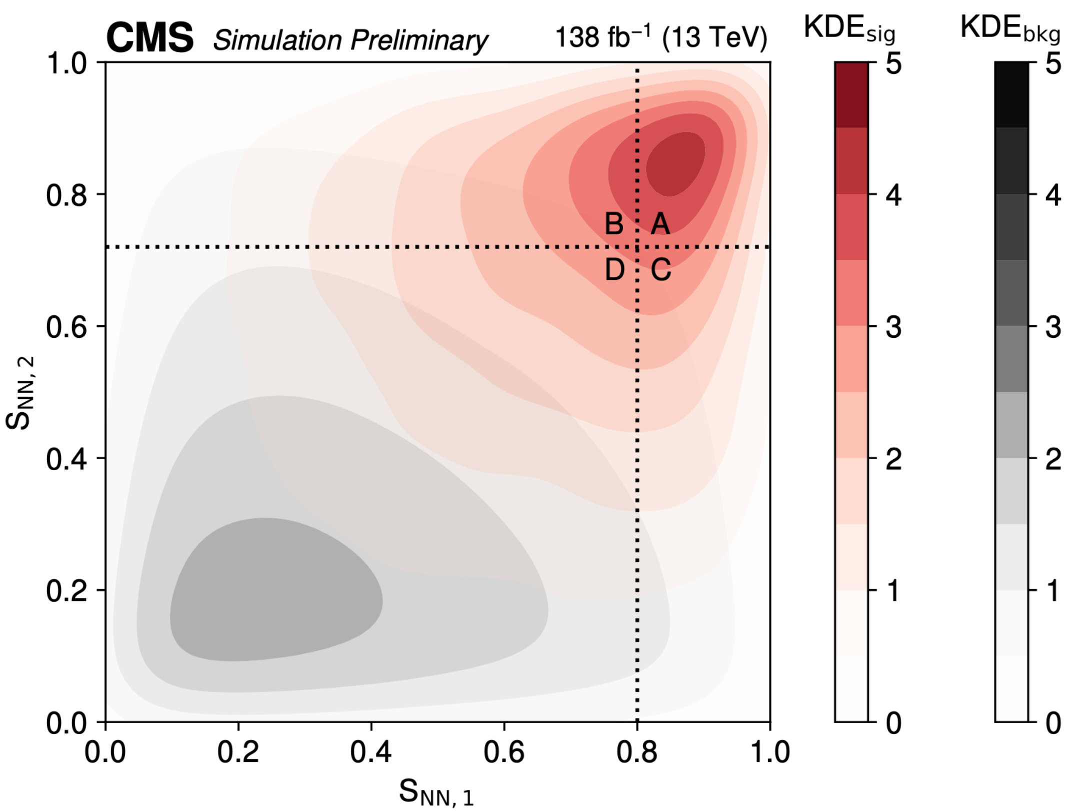


- ABCD: classic method for data-driven background estimation
- But: needs two decorrelated variables - can be difficult with ML taggers
- ABCDisCo: *learn* two decorrelated variables
  Enforced through distance correlation (DisCo) loss function

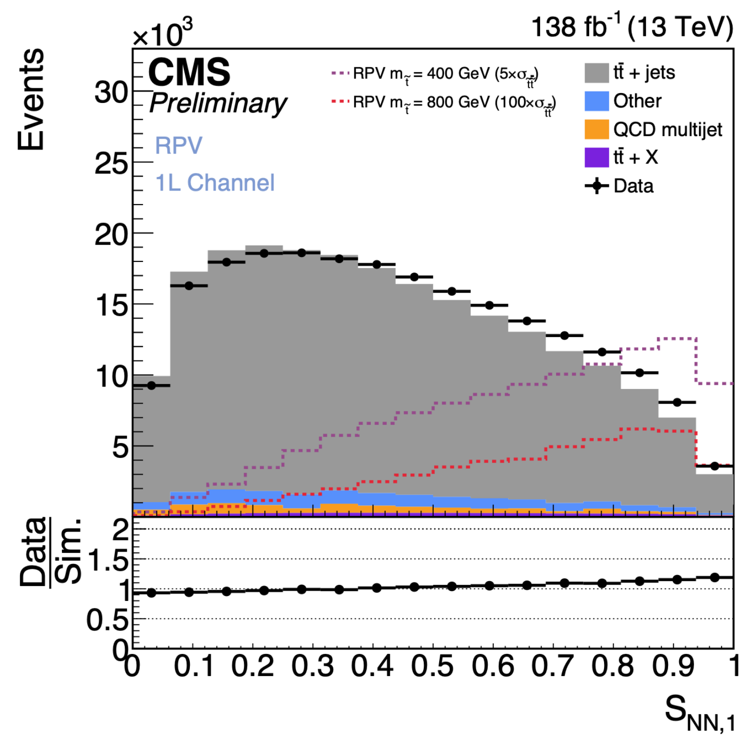
MLG-23-003
ABCDisCo method
SUS-23-001

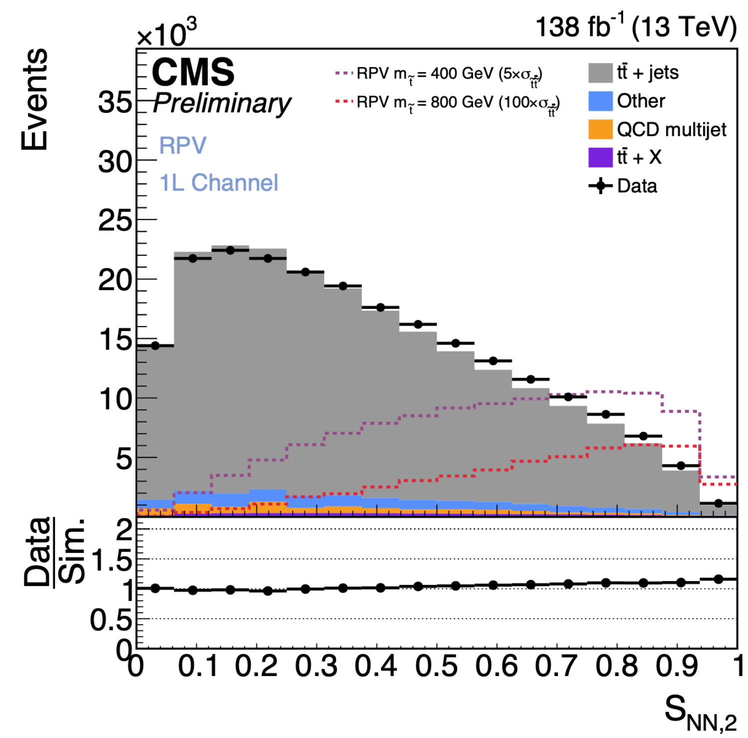
Application to SUSY search

Outputs are well decorrelated



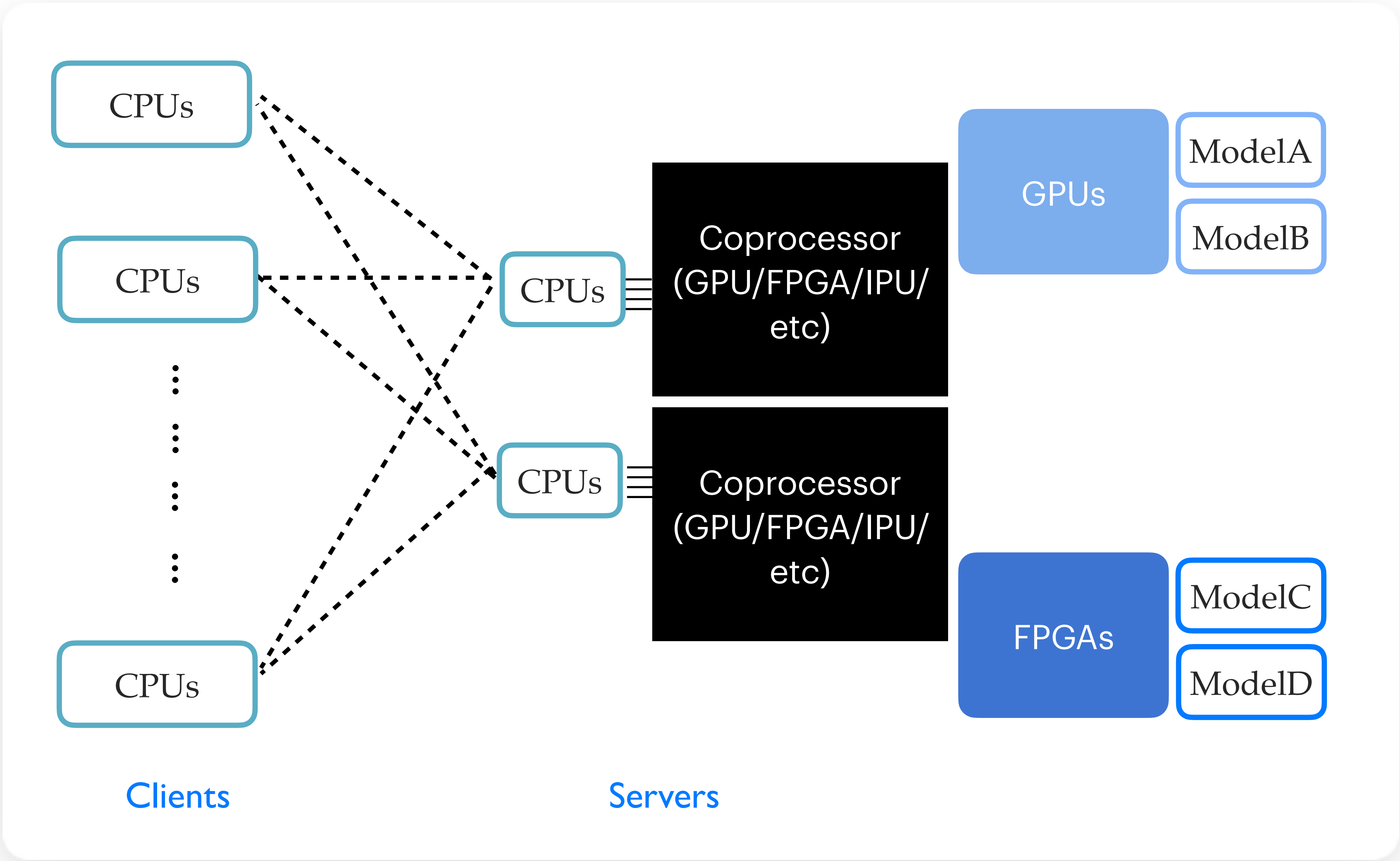
Strong S vs B discrimination as well





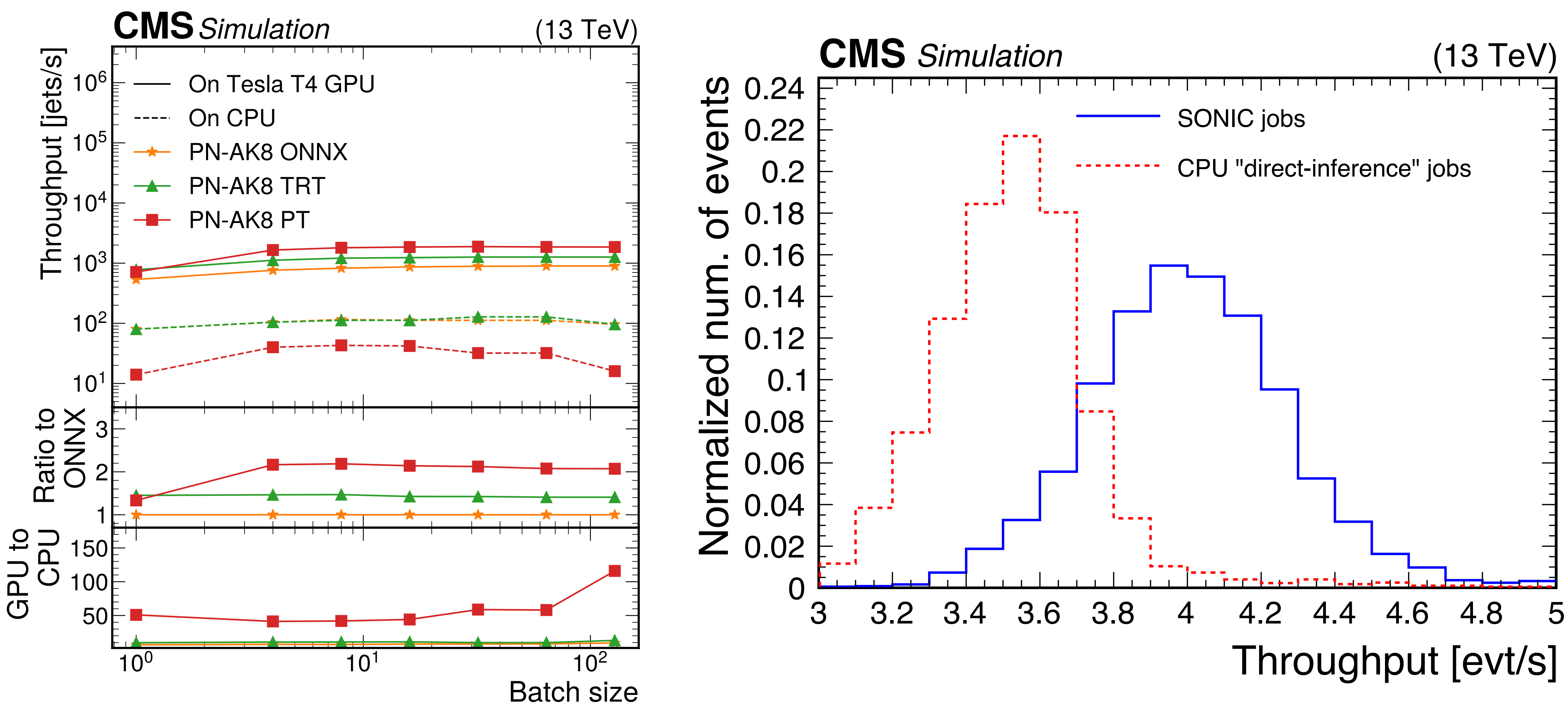
- ABCD: classic method for data-driven background estimation
- But: needs two decorrelated variables - can be difficult with ML taggers
- ABCDisCo: *learn* two decorrelated variables
  Enforced through distance correlation (DisCo) loss function

Framework for GPU inference within CMS SW

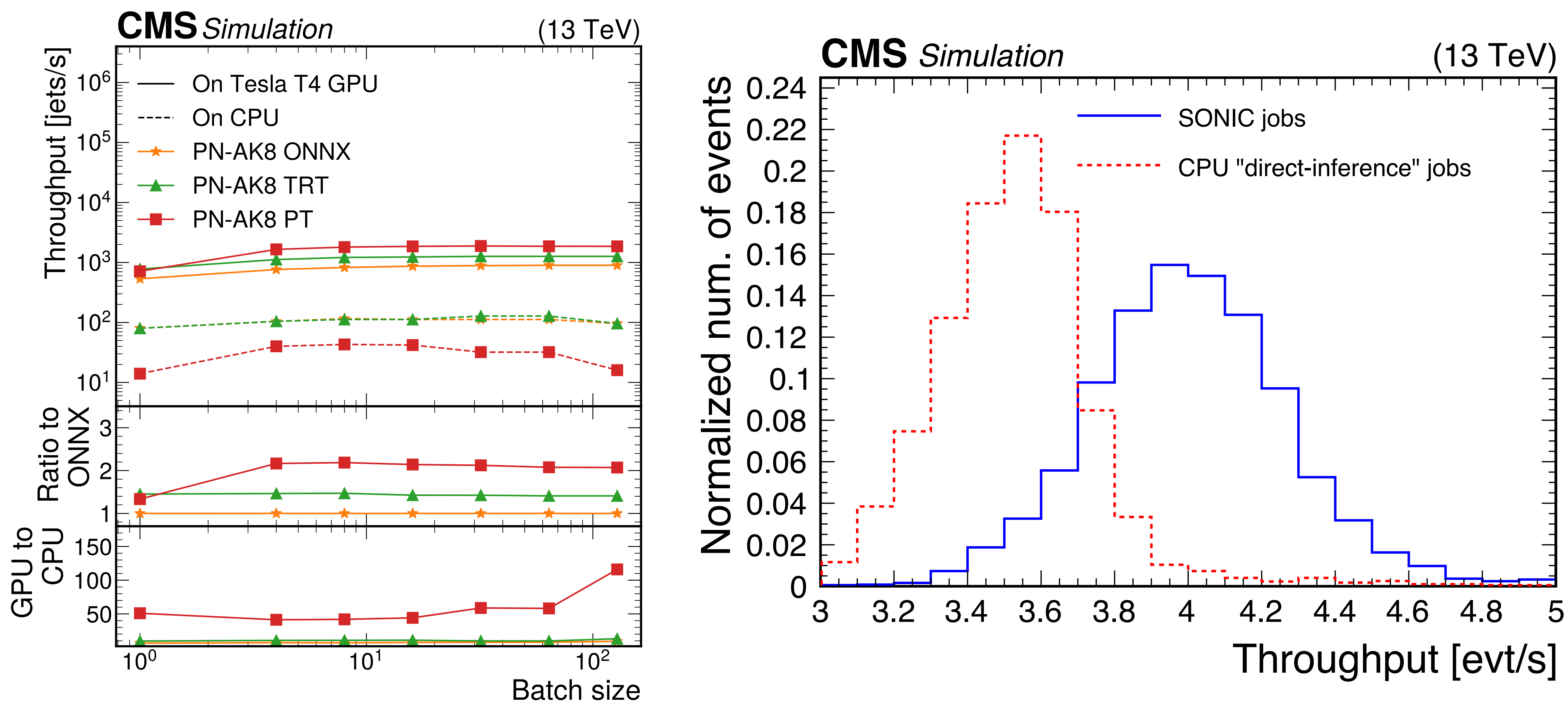


- Leveraging industry advancements in GPU processing power + increased ML in HEP
- Pathway to improving CMS computing efficiency beyond end of Moore's law

Framework for GPU inference within CMS SW



Framework for GPU inference within CMS SW



- Lower latency, higher throughput, than direct CPU inference for models like ParticleNet
- Will be vital to handle higher computing requirements in HL-LHC!

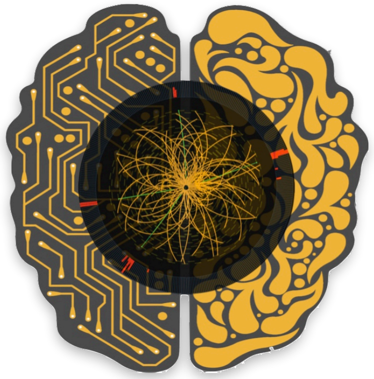
Summary

- Significant innovation in ML in CMS for analyses

- Powerful new algorithms for jet and event tagging

- Important work considering systematics, background estimation, and computing as well

- Stay tuned for more results this summer!



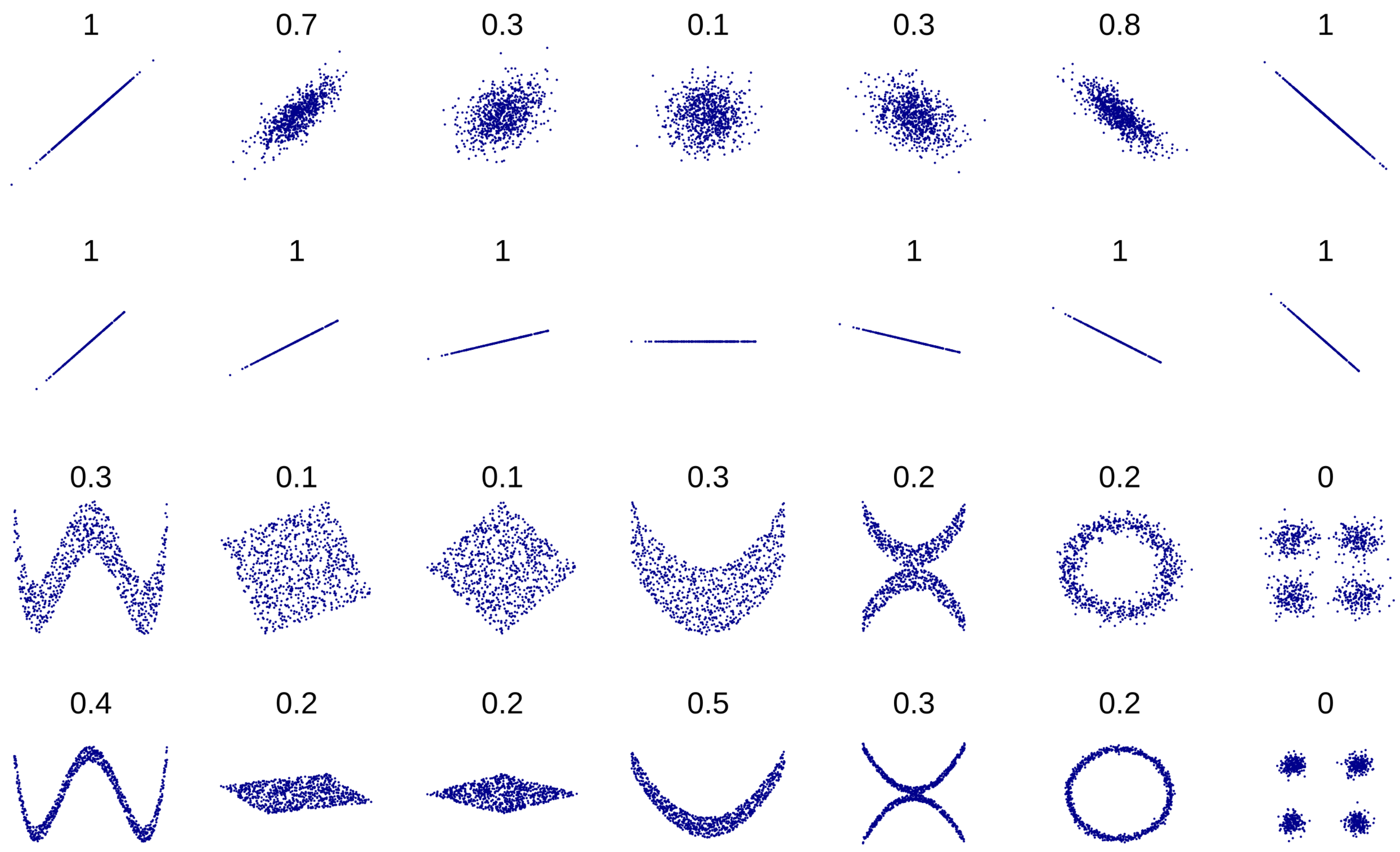
Backup

ABCDisCoTEC: Distance correlation loss

- DisCo loss for quantifying correlation

$$\mathcal{L}[f,g] = \mathcal{L}_{\text{classifier}}[f(X),y] + \mathcal{L}_{\text{classifier}}[g(X),y] + \lambda\,\mathrm{dCorr}^2_{y=0}[f(X),g(X)]$$

$$\mathrm{dCor}^2(X,Y) = \frac{\mathrm{dCov}^2(X,Y)}{\sqrt{\mathrm{dVar}^2(X)\,\mathrm{dVar}^2(Y)}}$$

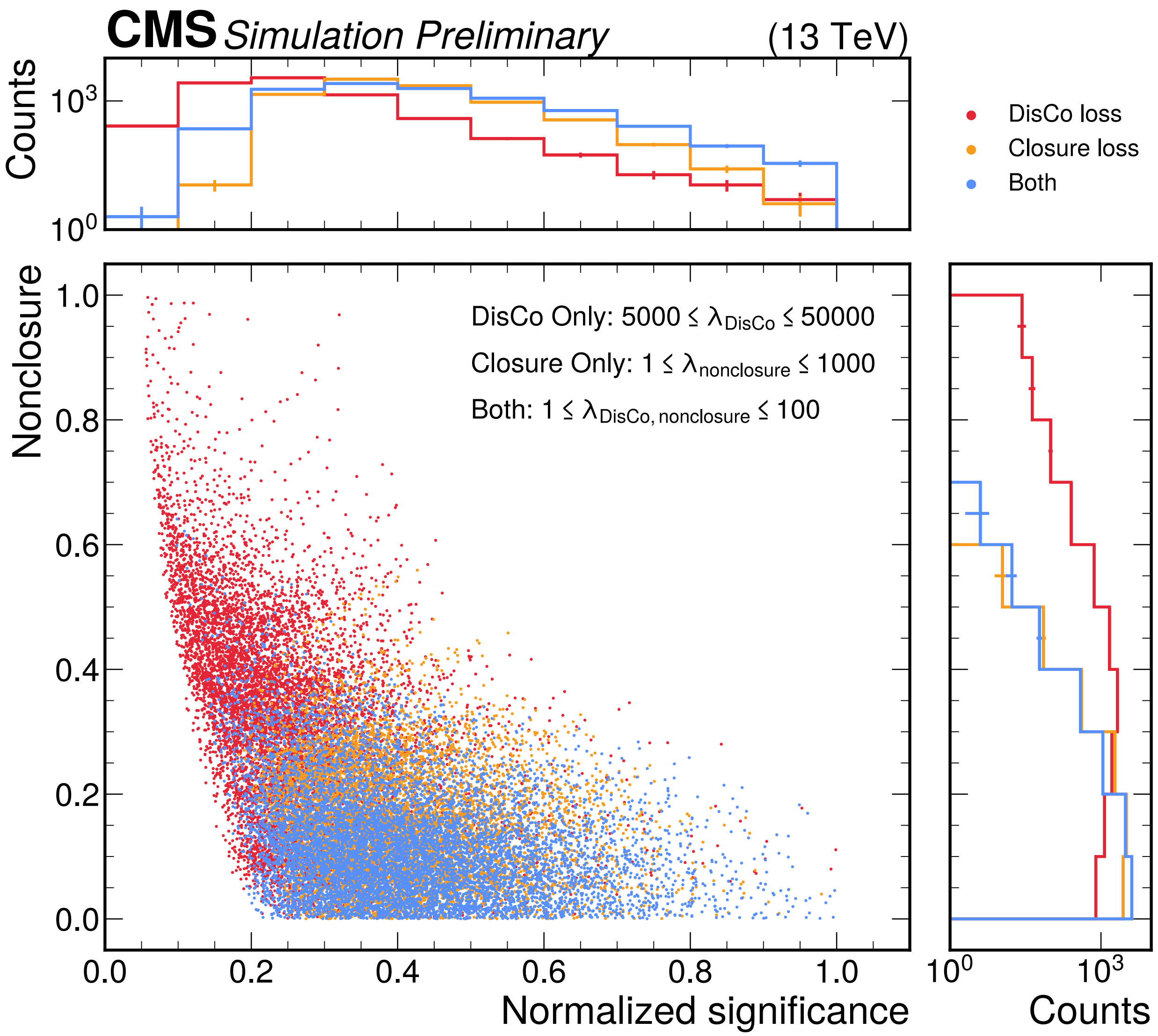


ABCDisCoTEC: non-closure loss

- ABCDisCo training enhanced with closure (ABCDisCoTEC)
- Improves non-closure of ABCDisCo method

$$L_{\mathrm{total}} = \lambda_{\mathrm{BCE}} L_{\mathrm{BCE}} + \lambda_{\mathrm{DisCo}} L_{\mathrm{DisCo}} + \lambda_{\mathrm{nonclosure}} L_{\mathrm{nonclosure}}$$

$$L_{\mathrm{nonclosure}} = \left(\frac{N_A N_D - N_B N_C}{N_A N_D + N_B N_C} \right)^2$$



ABCDisCoTEC: MDMM optimization

MLG-23-003
ABCDisCo method

- "Modified Differential Multiplier Method" allows better control with multiple training objectives