Machine learning for LLP searches at the LHC and beyond: Overview and future

Gregor Kasieczka
gregor.kasieczka@uni-hamburg.de
Twitter: @GregorKasieczka

LLPX - November 11th, 2021

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

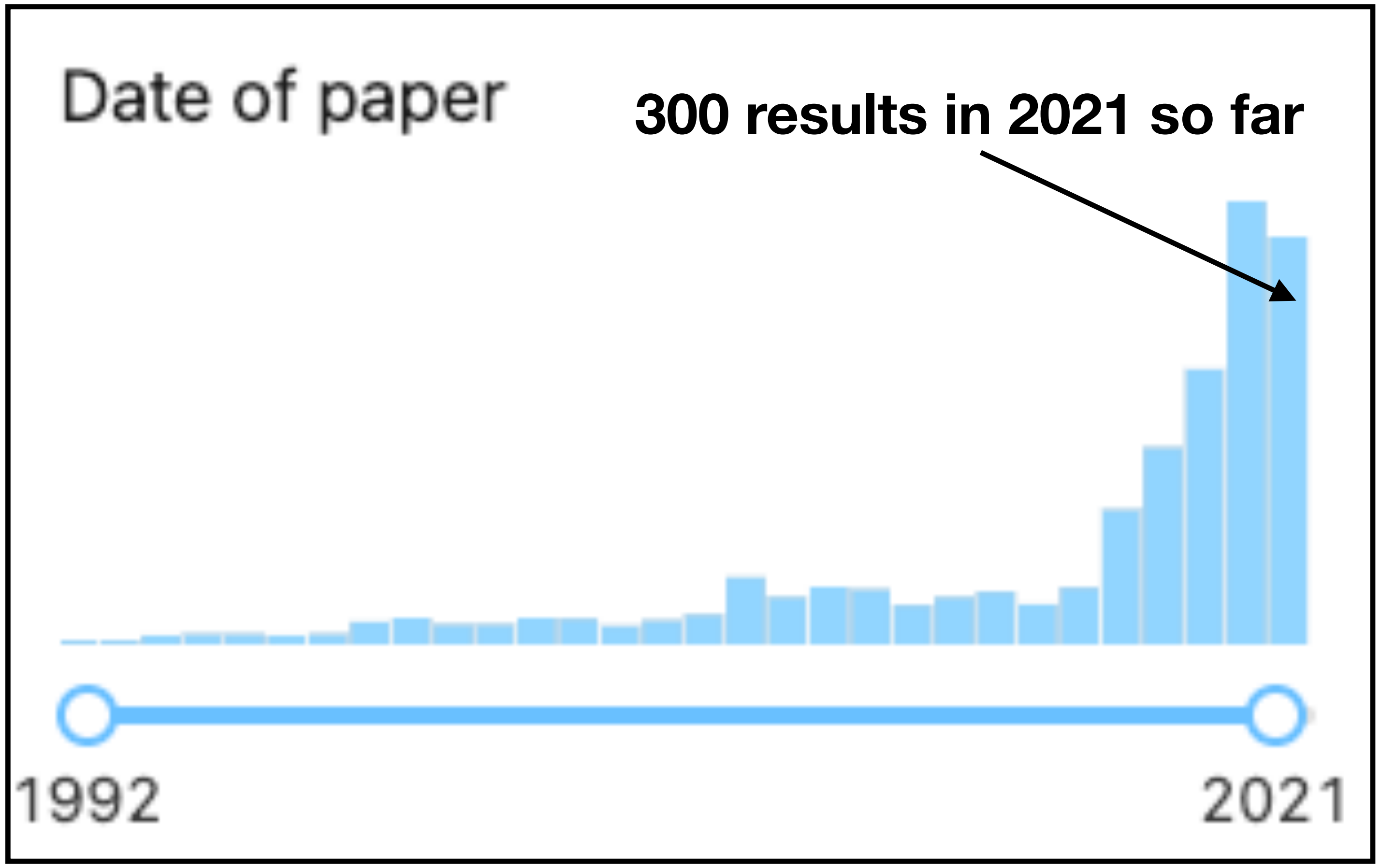
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

PIER
Partnership of Universität Hamburg and DESY

Bundesministerium für Bildung und Forschung

Overview: In general..



Inspire Search:

("machine learning" or "deep learning" or neural) and (hep-ex or hep-ph or hep-th)

Extremely active adaptation of machine learning to particle physics

..and long-lived

ATLAS & CMS Run 2 Publications

- O(40) on long-lived/disappearing/emerging jets
- Using machine learning: 8
    - Standard b-tagging: 3 (2107.06092, 2104.13474, 1909.03460)
    - Boosted Decision Trees (BDT) for signal identification
    - New architecture development: DNN + Decorr (1912.12238)

Pheno

- O(10) publications

Much slower adaptation to long-lived searches

***Goal of this talk**: Highlight status, discuss reasons, see possible ways forward*

Start with supervised learning

Supervised Learning:
Attempt to infer some target *(truth label)*: **classification**, regression

Use training data with known labels
(often from **Monte Carlo simulation**)

$\mathbf{x}$

y

Learn to predict:
(Neural network f with parameters θ)

$$\hat{y} = f_\theta(\mathbf{x})$$

$\hat{y}$

observable features such as kinematics, vertices,

truth label (e.g. LLP or Not)

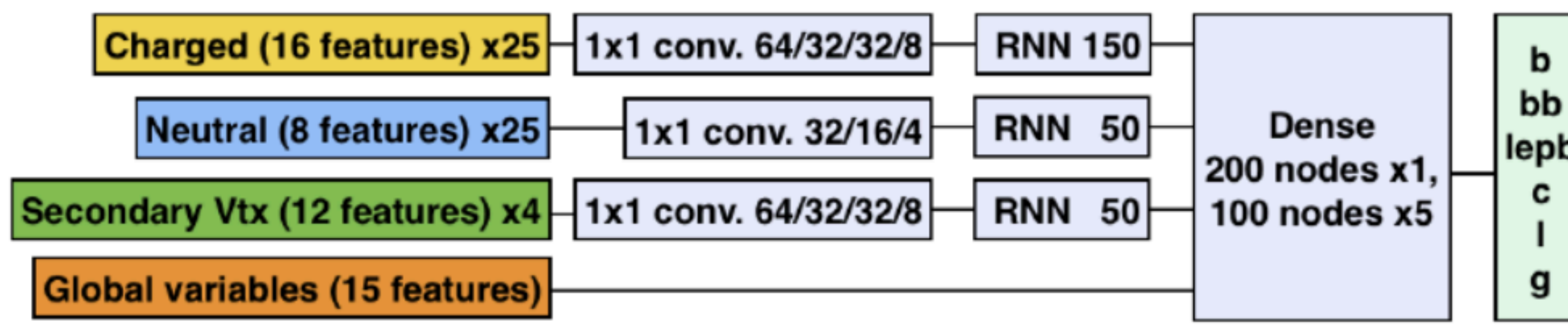
e.g. predicted LLP score

Target: For classification, find θ values that minimise cross-entropy:

$$\mathcal{L} = -y \log\left(\hat{y}\right) - (1-y) \log\left(1-\hat{y}\right)$$

How we can use it?

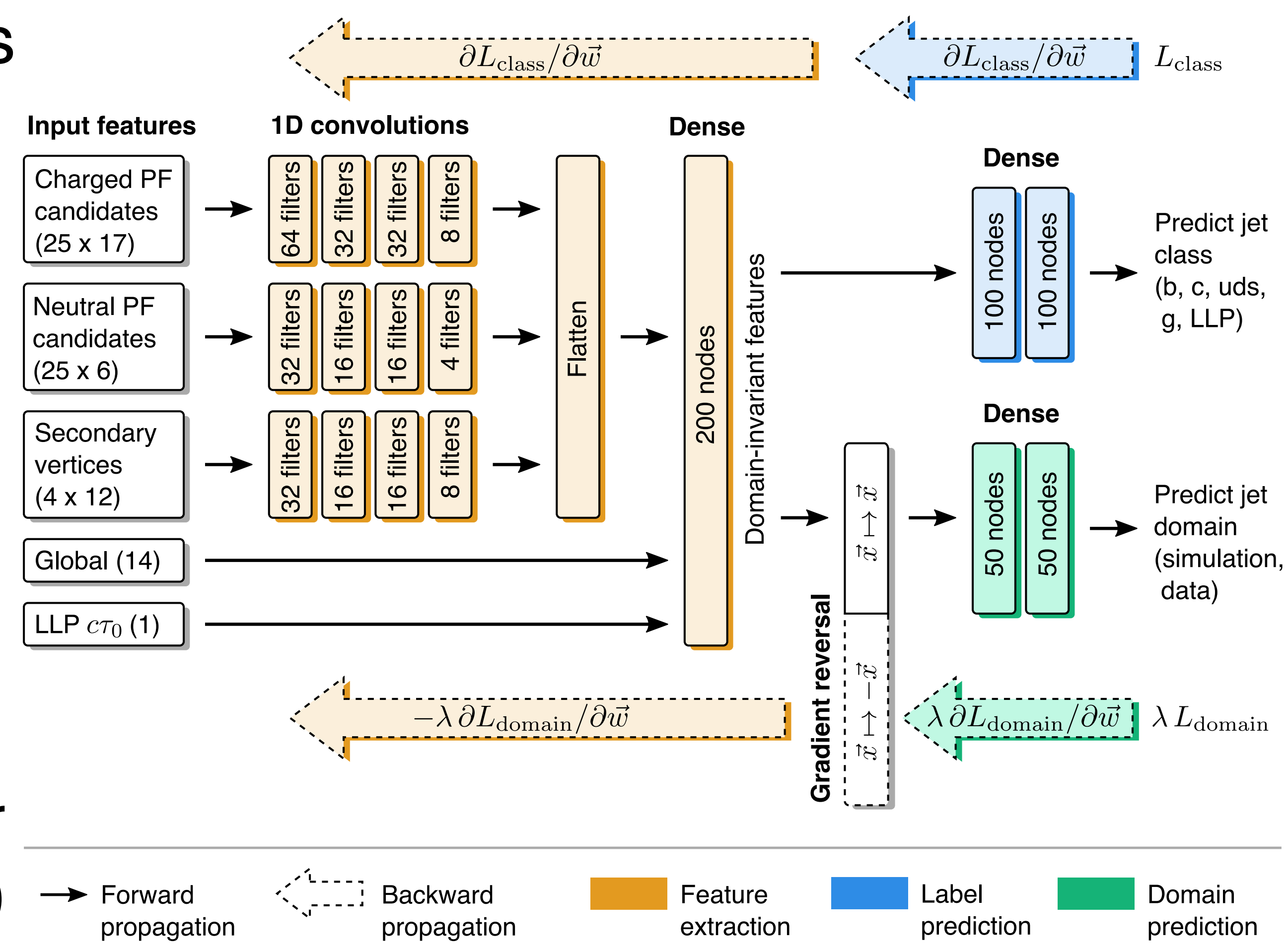
- **Tagging of known SM particles**
  - Use case: Assume associated production LLP+X or use to define control regions
  - Rely on default flavour/resonance/.. taggers



e.g. CMS DeepJet architecture (CMS-DP-2018-058, 2008.10519.)

How we can use it?

- **Tagging of known SM particles**
- **Reconstruction and tagging of unknown LLP particles**
  - Use case: Produce/identify LLP candidates to define signal regions on object-level
  - Offers several interesting challenges (discussed later)

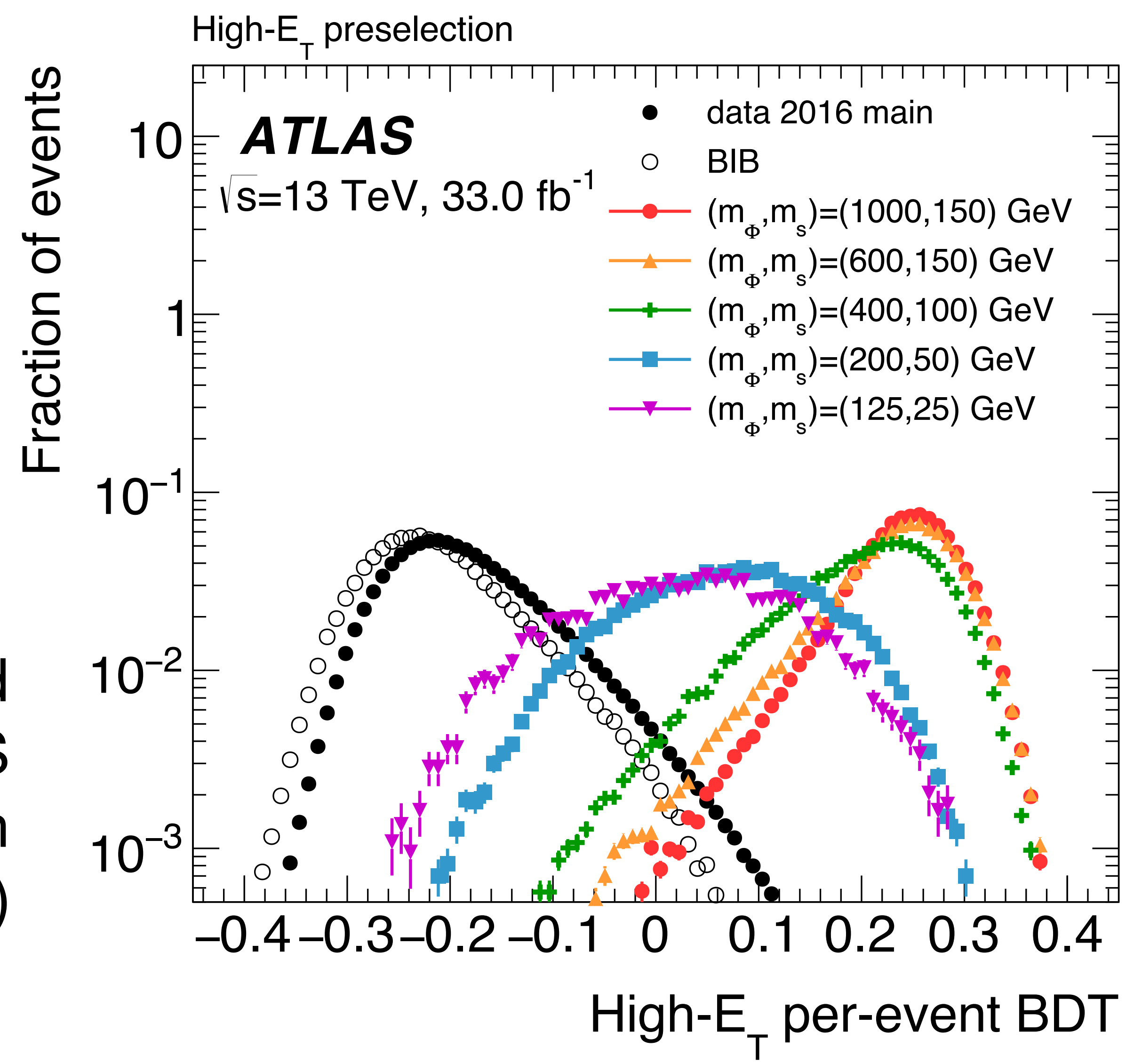


e.g. Displaced vertex tagger by CMS (1912.12238)

How we can use it?

- **Tagging of known SM particles**
- **Reconstruction and tagging of unknown LLP particles**
- **Tagging of complete signal topologies**
  - Obtain global signal score
  - Similar issues at per-particle taggers

e.g. signal-identification Boosted Decision Tree (BDT) from ATLAS displaced hadronic jet search (1902.03094)



Status

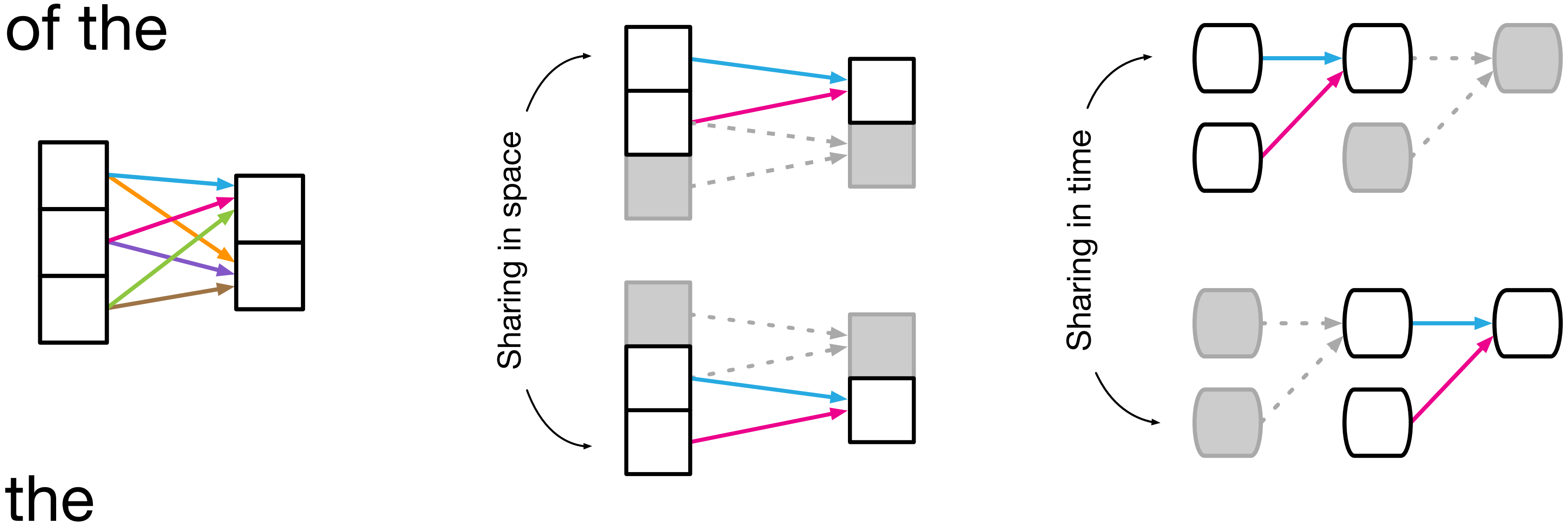
- Which ML techniques are used in LLP searches (for signal identification)?
- Experimental results dominated by BDT-based tagging
  - e.g. 2012.01581,1909.01246, 1902.03094, 1806.07355
- This means decision functions using a relatively small number (~10) of high-level features (e.g.) as input
- Not bad per-se
  (but I will still argue why architecture matters)

Inputs to hadronic-BDT from ATLAS search 1909.01246

- jet width, defined as the p_T-weighted sum of the ΔR between each energy cluster and the jet axis;
- jet vertex tagger (JVT) output [91];
- $E_{\mathrm{ECAL}}/E_{\mathrm{total}}$;
- jet mass, as defined by the jet clustering algorithm [92];
- jet charge, defined as the momentum-weighted charge sum constructed from tracks associated with the jet; tracks are associated with jets using ghost association [93];
- jet timing, defined as the energy-weighted average of the timing for each cell in the jet.

Question of Architecture

- Integration of symmetries of the data (e.g. permutation invariance) can increase performance of machine learning models
- See 1806.01261 for an excellent discussion form the side of computer science



(a) Fully connected (b) Convolutional (c) Recurrent

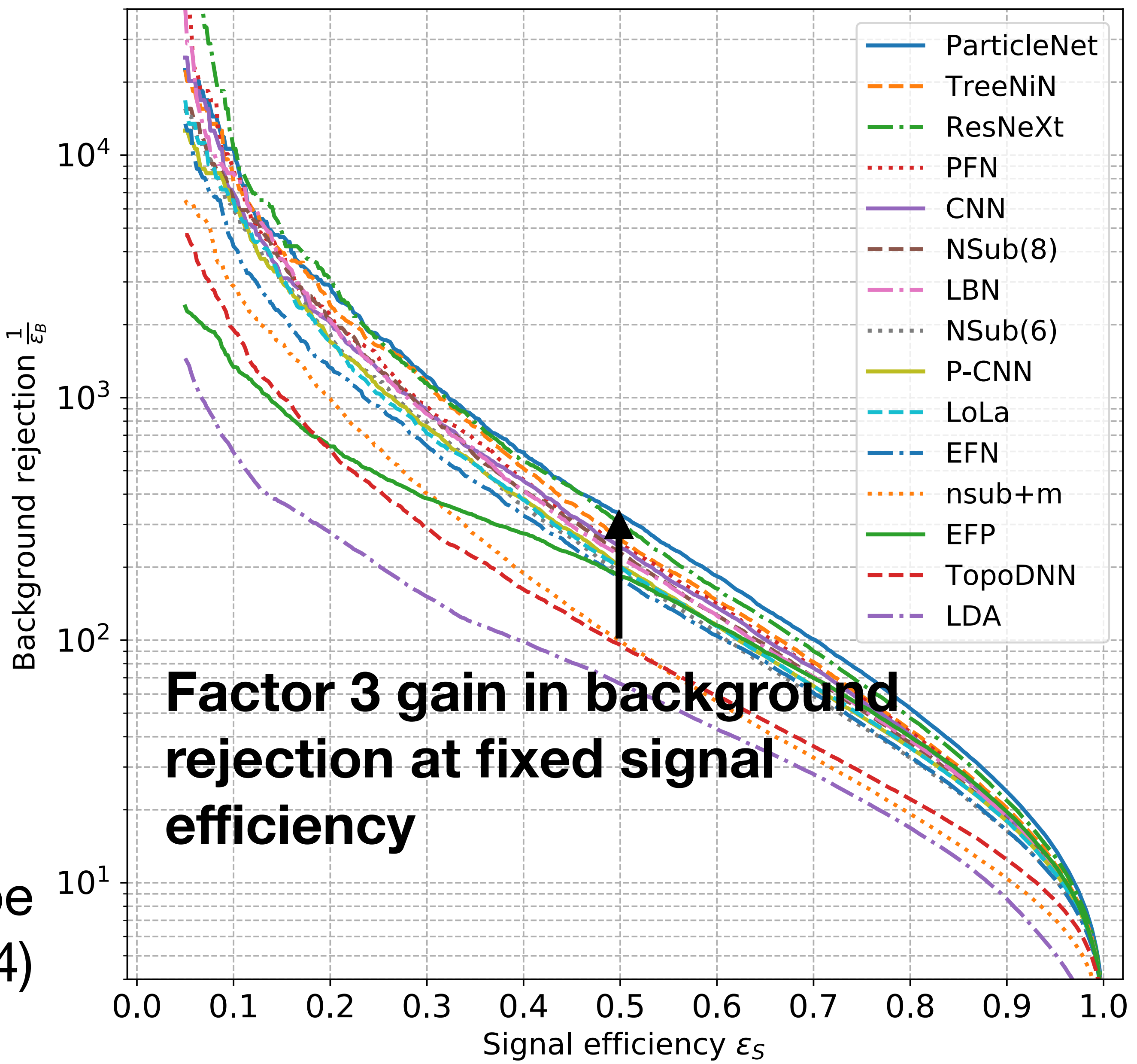
Component	Entities	Relations	Rel. inductive bias	Invariance
Fully connected	Units	All-to-all	Weak	-
Convolutional	Grid elements	Local	Locality	Spatial translation
Recurrent	Timesteps	Sequential	Sequentiality	Time translation
Graph network	Nodes	Edges	Arbitrary	Node, edge permutations

Inductive biases of standard ML architectures (1806.01261)

Consider BDT equivalent to fully-connected NN in this regard

Question of Architecture

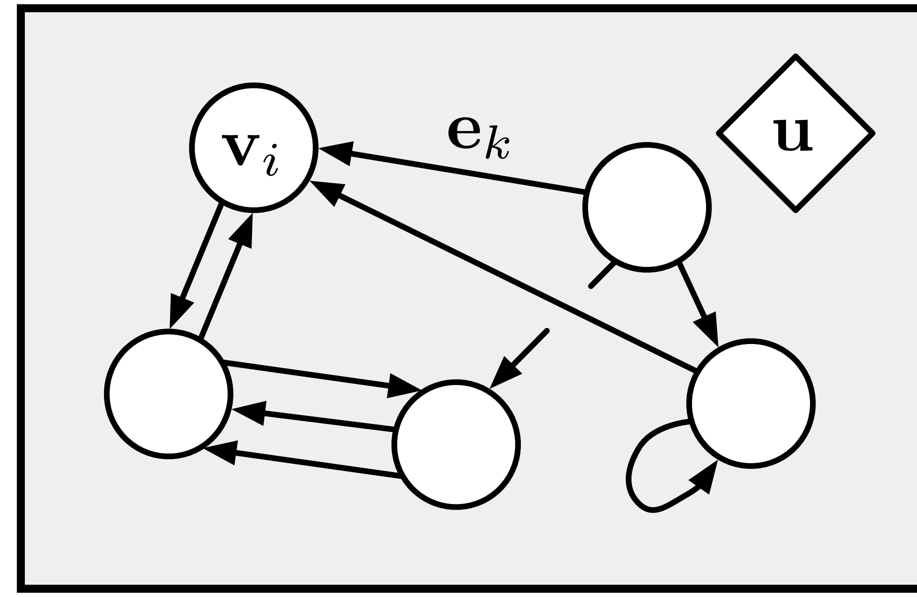
- Integration of symmetries of the data (e.g. permutation invariance) can increase performance of machine learning models
- See 1806.01261 for an excellent discussion form the side of computer science
- Also observed for HEP application (e.g. top tagging benchmark)



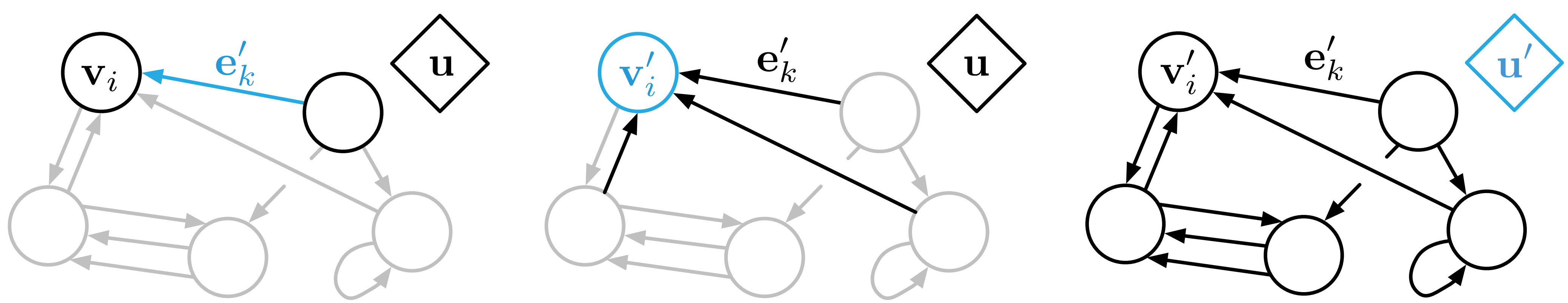
Top Tagging Landscape paper (1902.09914)

Popular Choice: Graphs

- Consists of
  Vertex: particle (e.g., four-vector)
  Edge: distance (for example geometric)
- Works with:
  - Data that naturally comes as a graph (e.g. a decay sequence)
  - Data embedded in some geometric space (point cloud)
- Active development of graphs on CS side, increasing number of HEP applications: 1902.08570, 1902.07987, 1908.05318, 2008.03601, 2103.16701, 2101.08578, …
  See **2007.13681** for a review



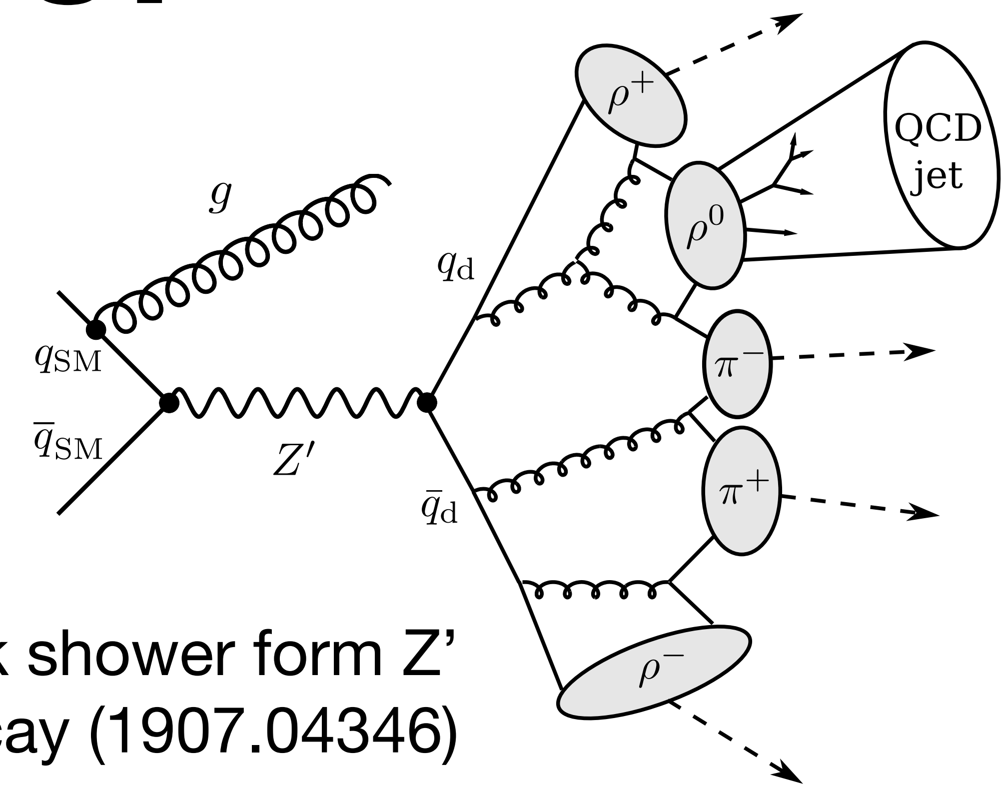
Graph and update rules from 1806.01261



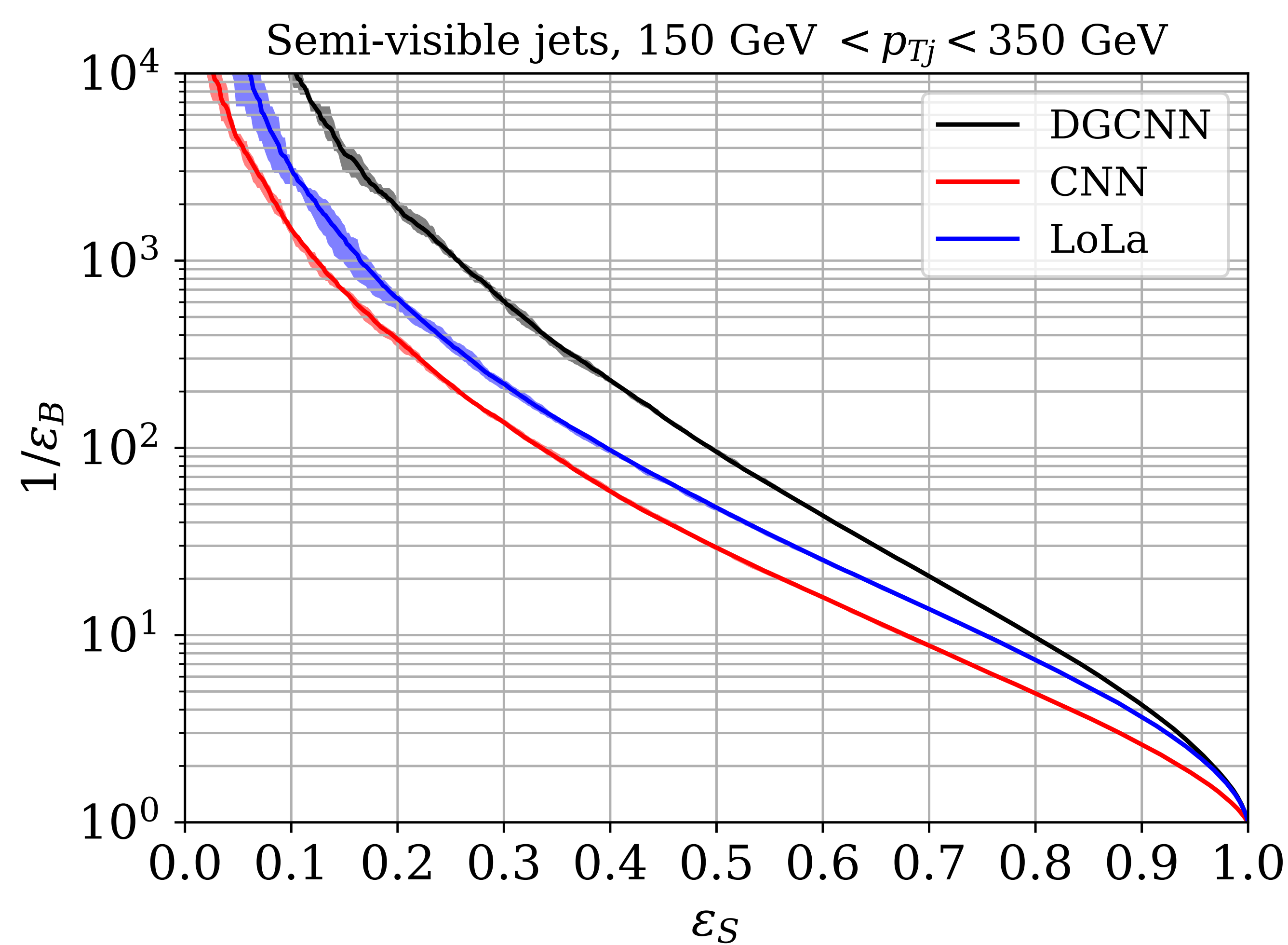
(a) Edge update (b) Node update (c) Global update

And for LLPs?

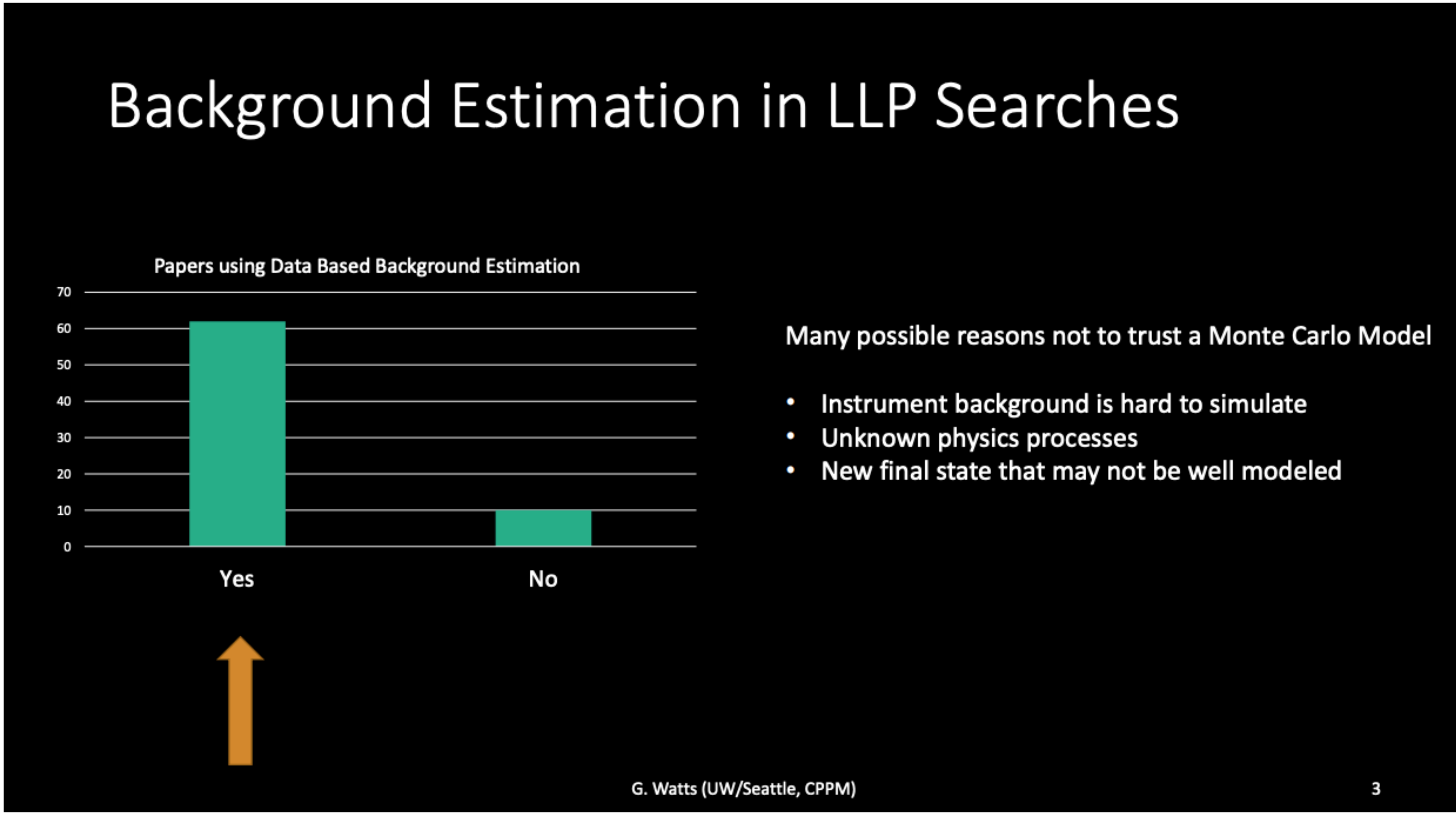
- Application to semi-visible jets by Bernreuther, Finke, Kahlhoefer, Kraemer, Alexander Mueck (2006.08639)
- Succesfully trained graph-based network to distinguish semi-visible jets from QCD jets
- Use jet-constituents kinematics as inputs



Dark shower form Z' decay (1907.04346)



Challenge: Background Estimation



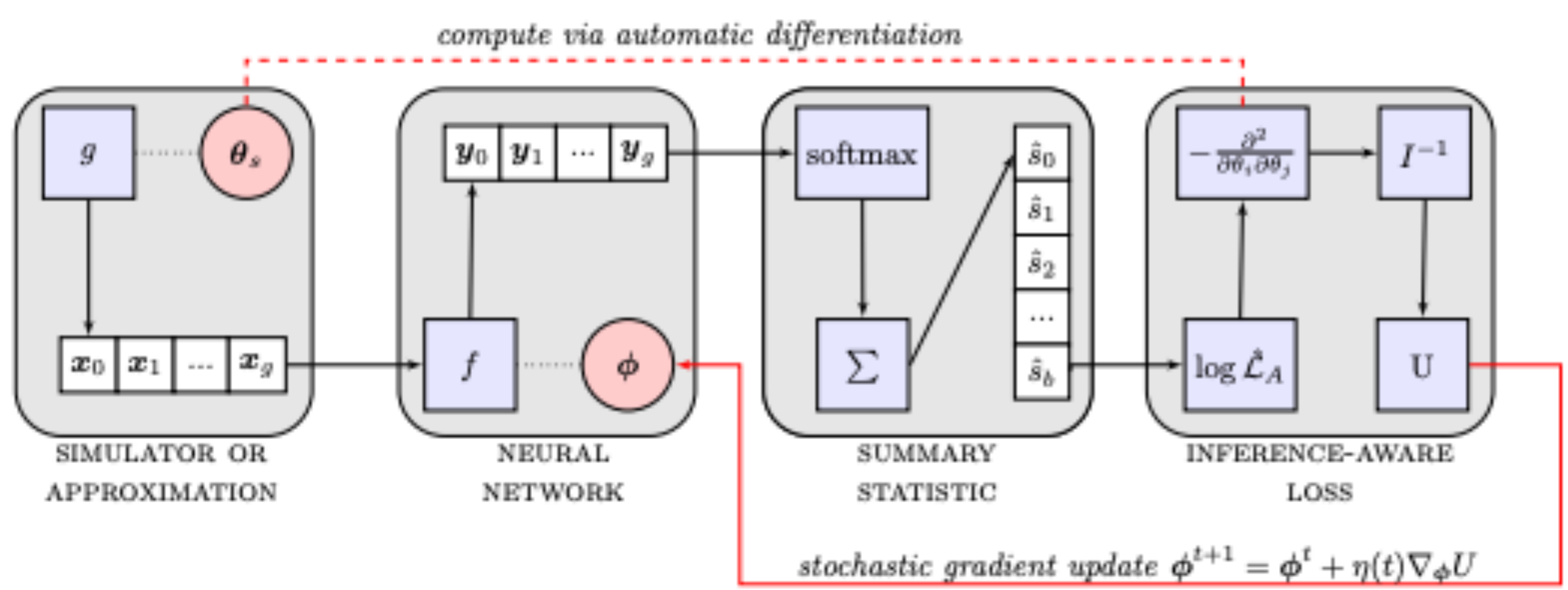
-> See talk by Gordon right afterwards

Challenge: Uncertainties

- Need to assign uncertainty to classifier outputs / simulation data differences

- Once trained, a ML model is a deterministic function of its inputs
  - **Classical techniques of uncertainty quantification still work!**
    - Propagation of input uncertainties or measurement in data

- Additional ML aspects
  - **Include** uncertainties in optimisation to maximise sensitivity (1806.00322, 1806.04743, 2110.00810)



Challenge: Uncertainties

- Need to assign uncertainty to classifier outputs / simulation data differences

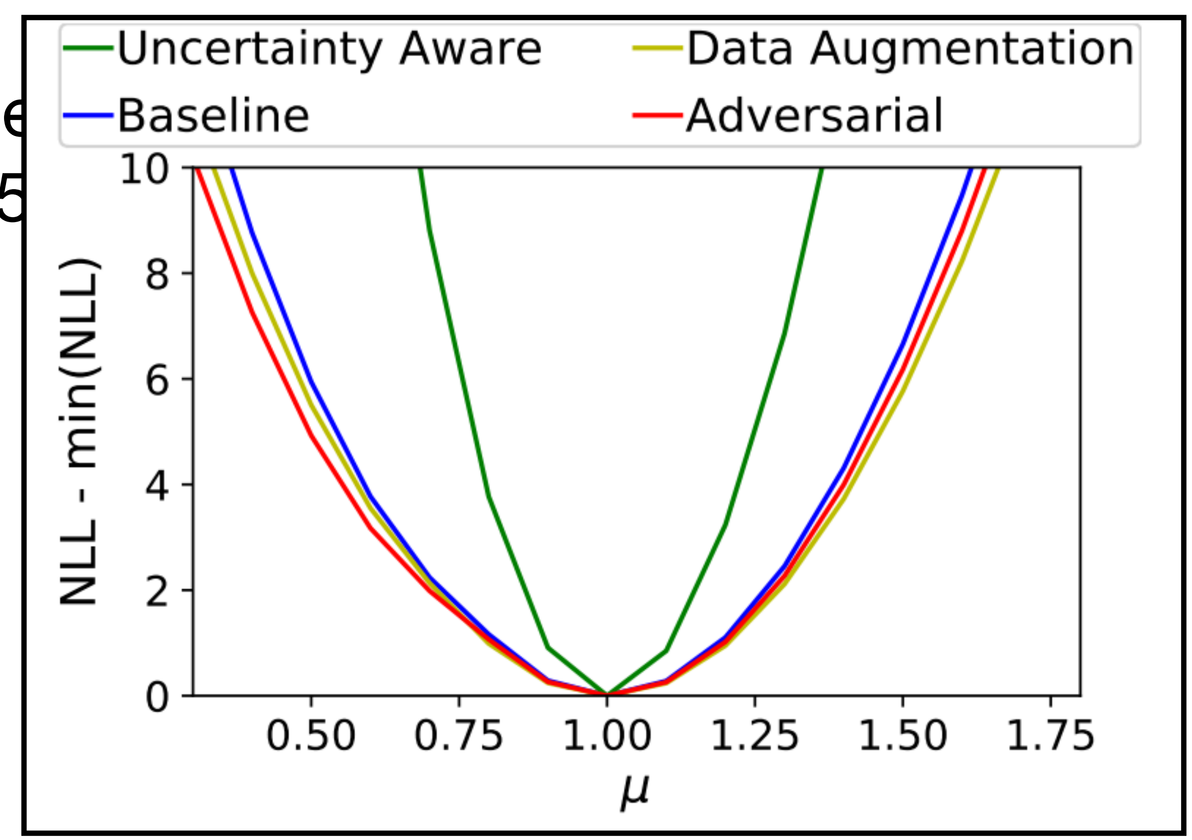
- Once trained, a ML model is a deterministic function of its inputs
  - **Classical techniques of uncertainty quantification still work!**
    - Propagation of input uncertainties or measurement in data

- Additional ML aspects
  - Include uncertainties in optimisation to maximise sensitivity (1806.00322, 1806.04743, 2110.00810)
  - Apply **decorrelation** to reduce effect of uncertainties / simulation difference (1611.01046,1703.03507,2001.05310,…)

$$L = L_{classifier}(\vec{y}, \vec{y}_{true}) + \lambda \, \mathrm{dCorr}^2_{y_{true}=0}(\vec{m}, \vec{y})$$

Challenge: Uncertainties

- Need to assign uncertainty to classifier outputs / simulation data differences
- Once trained, a ML model is a deterministic function of its inputs
  - **Classical techniques of uncertainty quantification still work!**
    - Propagation of input uncertainties or measurement in data
- Additional ML aspects
  - Include uncertainties in optimisation to maximise sensitivity (1806.00322, 1806.04743, 2110.00810)
  - Apply decorrelation to reduce effect of unce
    simulation difference (1611.01046,1703.035
  - Use parametrised networks to improve treatment of nuisance parameters (2105.08742,2109.08159)



Challenge: Uncertainties

- Need to assign uncertainty to classifier outputs / simulation data differences

- Once trained, a ML model is a deterministic function of its inputs
  - **Classical techniques of uncertainty quantification still work!**
    - Propagation of input uncertainties or measurement in data

- Additional ML aspects
  - Include uncertainties in optimisation to maximise sensitivity (1806.00322, 1806.04743, 2110.00810)
  - Apply decorrelation to reduce effect of uncertainties / simulation difference (1611.01046,1703.03507,2001.05310,…)
  - Use parametrised networks to improve treatment of nuisance parameters (2105.08742,2109.08159)
  - Build ML models that provide uncertainties along with predictions (e.g. Bayesian architectures, 1904.10004 2003.11099)

Challenge: Diversity of Models

- Common issue of ML for searches:
  - Different parameters of new physics model yield different signal properties *(e.g. kinematics as function of resonance mass; dark shower properties as function of coupling strength)*
  - Affects performance of ML-based selection *(Also true for cut-based approaches but higher sensitivity of ML-taggers will mean this effect is larger)*

- **Idea I - Pragmatic:**
  - Do “nothing”, accept different sensitivity to different signal models
  - Not *wrong*, but in general not optimal either
  - Time-efficient

Challenge: Diversity of Models

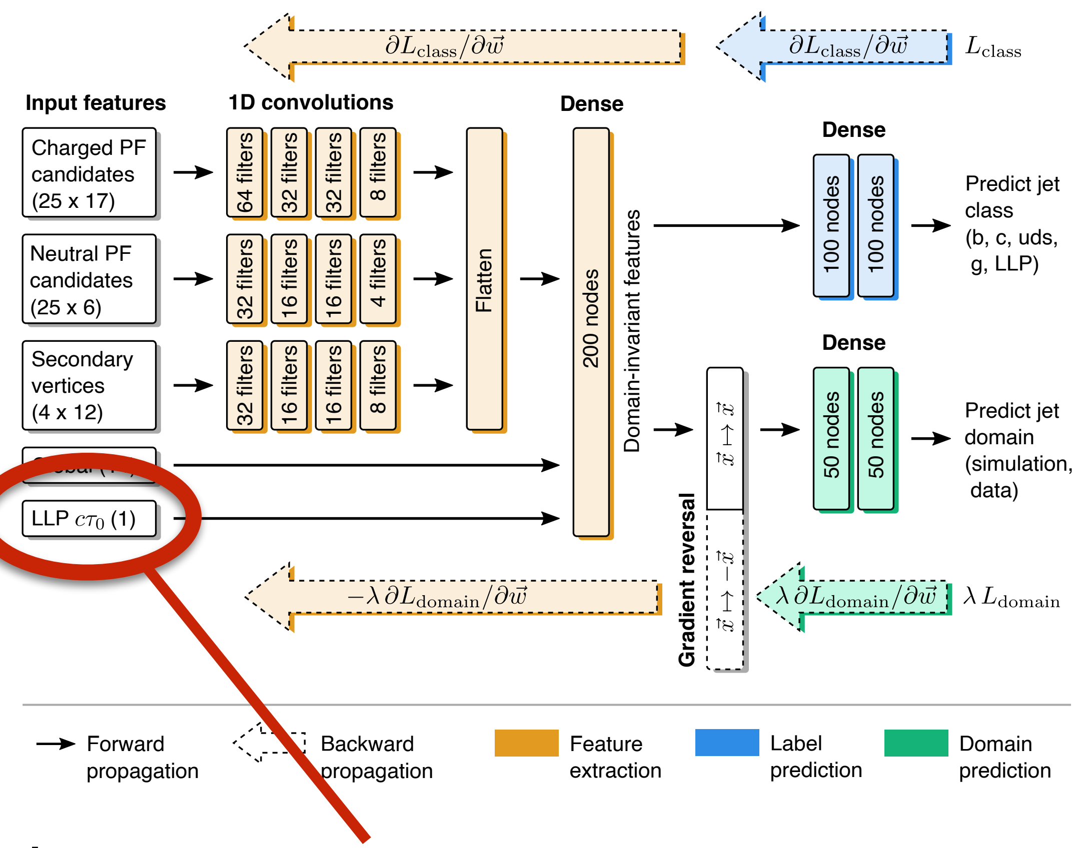
- Common issue of ML for searches:
  - Different parameters of new physics model yield different signal properties *(e.g. kinematics as function of resonance mass; dark shower properties as function of coupling strength)*
  - Affects performance of ML-based selection *(Also true for cut-based approaches but higher sensitivity of ML-taggers will mean this effect is larger)*

- **Idea I - Pragmatic**
- **Idea II - Extensive**
  - Train separate ML classifier for each signal-parameter value
  - Better classification performance expected
  - Overhead of book-keeping and validation

Challenge: Diversity of Models

- Common issue of ML for searches:
  - Different parameters of new physics model yield different signal properties *(e.g. kinematics as function of resonance mass; dark shower properties as function of coupling strength)*
  - Affects performance of ML-based selection *(Also true for cut-based approaches but higher sensitivity of ML-taggers will mean this effect is larger)*

- **Idea I - Pragmatic**
- **Idea II - Extensive**
- **Idea III - Parametrisation (see 1601.07913)**
  - Use signal-parameter as additional input in training; sample randomly for backgrounds;
  - Expect similar performance as II, but with one network
  - Less bookkeeping, validation for different signal parameter values still needed



Lifetime conditioning in 1912.12238

Challenge: Diversity of Models

- Common issue of ML for searches:
  - Different parameters of new physics model yield different signal properties *(e.g. kinematics as function of resonance mass; dark shower properties as function of coupling strength)*
  - Affects performance of ML-based selection *(Also true for cut-based approaches but higher sensitivity of ML-taggers will mean this effect is larger)*

- **Idea I - Pragmatic**
- **Idea II - Extensive**
- **Idea III - Parametrisation**
- **Idea IV - Anomaly searches**
  Discuss next

Anomaly Searches

- **Motivation**: Develop search-strategies that are less dependent on specific model assumptions
- Data analysis based on distributions over measured events
- Single outliers are statistically irrelevant, look for systematic over-densities

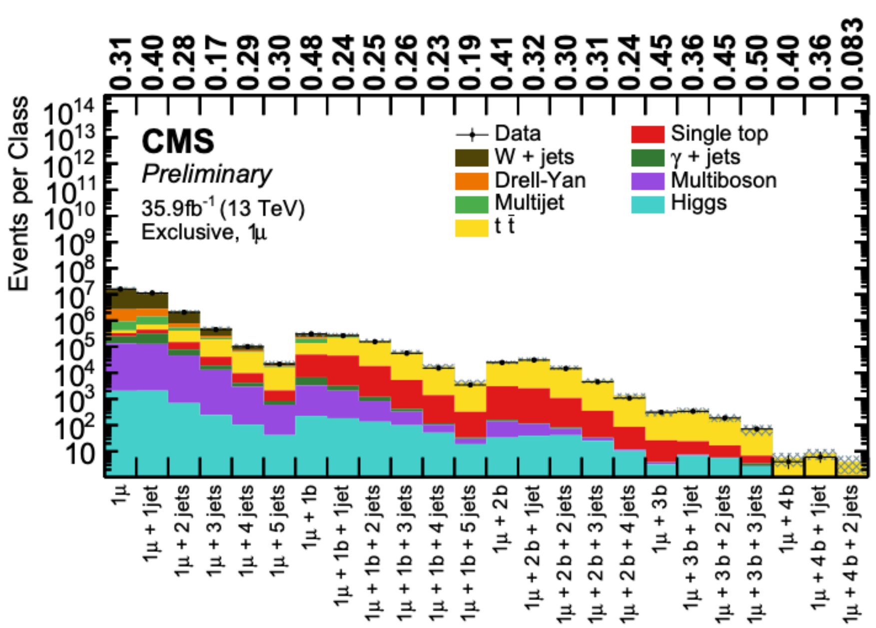
Can we use simulation to estimate backgrounds?

Yes

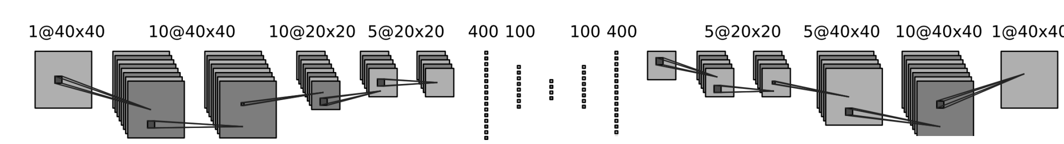
- Systematically compare simulation and recorded data, look for differences
- Con: Relies on imperfect simulation
- Pro: Sensitive to all types of anomalies

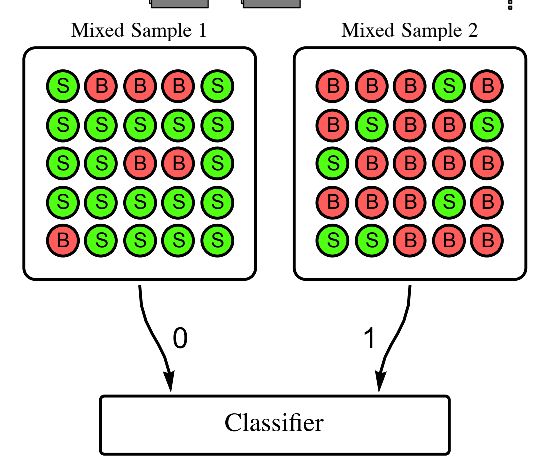
No

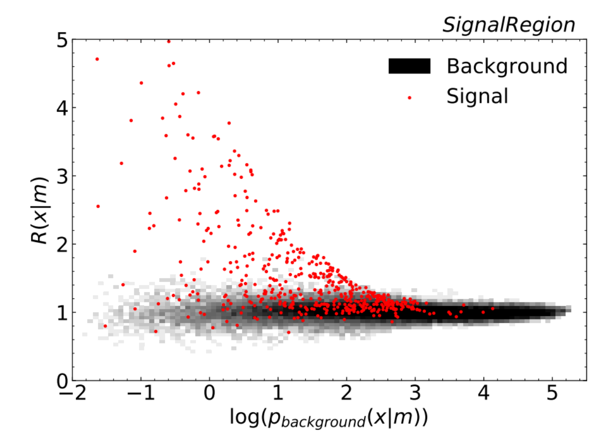
- Estimate background from data
- Con: Need to make assumptions about signal model
- Pro: No reliance on simulation



MuSic search (2010.02984)



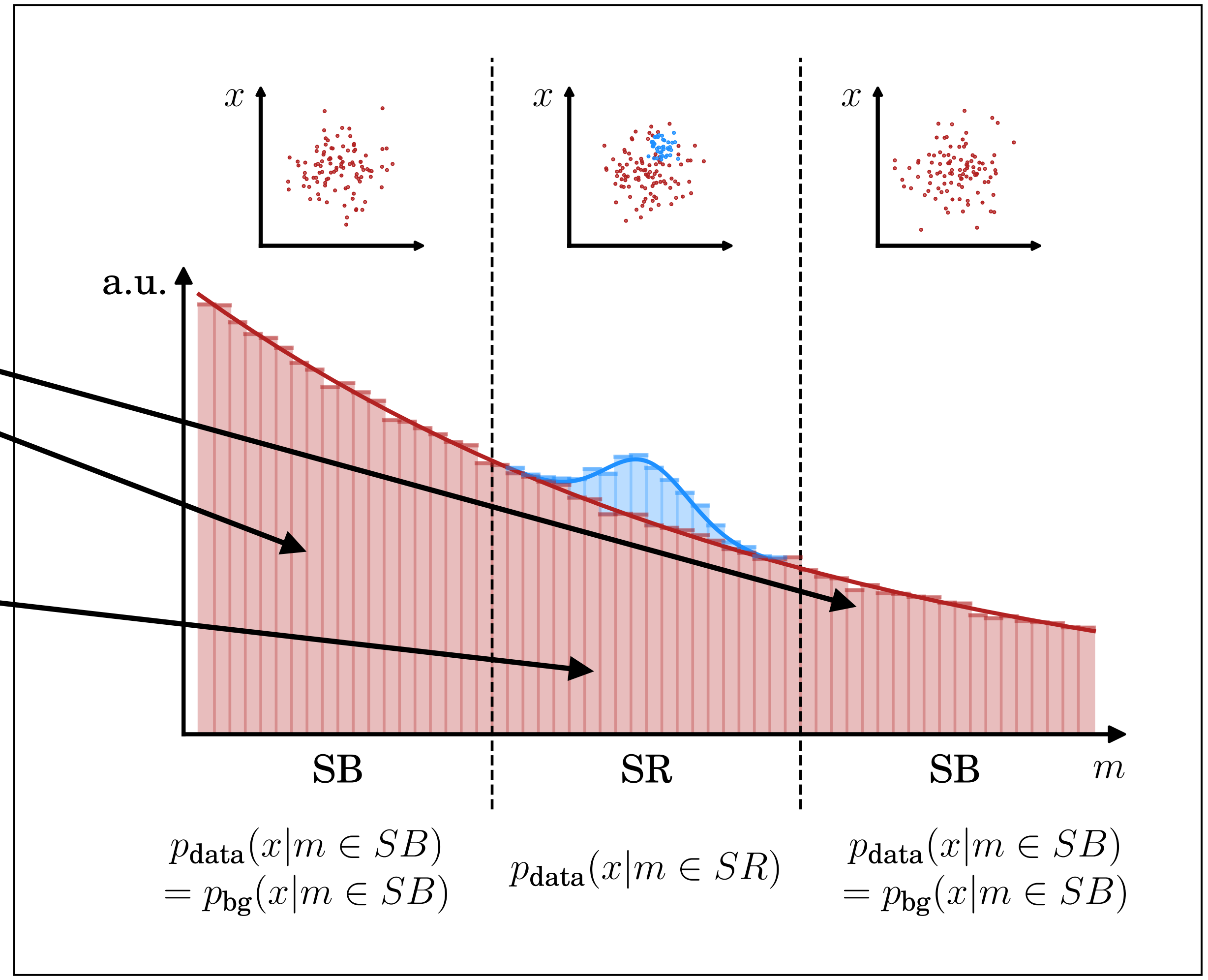




Example: Anomaly-enhanced bump hunt (CATHODE)

Train density estimator here

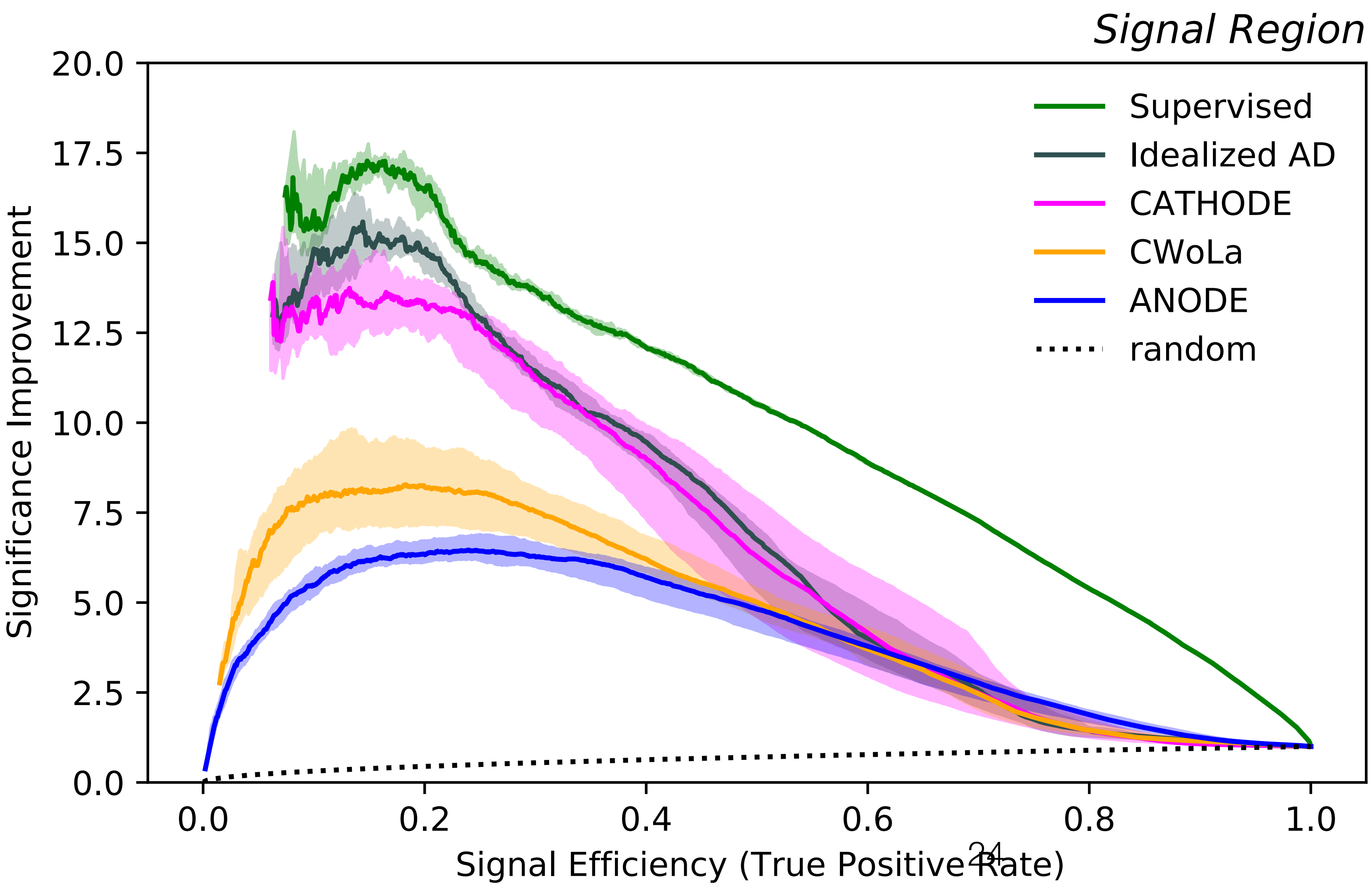
and sample here



- Train density estimator (a class of powerful and flexible generative model) in sideband
- Interpolate to signal region
- Sample data there
- This produces *'extrapolated-background'*

Example: Anomaly-enhanced bump hunt (CATHODE)

- Train classifier to distinguish *data* from *extrapolated background in signal region*
- If these can be distinguished: potential signal present
- Excellent performance and stability compared to other methods, close to supervised classifier



Performance of Classifying Anomalies THrough Outer Density Estimation (CATHODE) algorithm (2109.00546)

Anomaly Searches for LLP

- In general: Trade-off between coverage and sensitivity
- For overviews see LHC Olympics (2101.08320) and DarkMachines (2105.14027) community papers
- Open challenges:
  - Difficult if signal is not a bump and backgrounds are hard to estimate
  - Generalisation to higher number of observables and systematic understanding of sensitivity

Application of unsupervised **anomaly detection to LLP** search (2107.12379)
See Aris' talk in this session

Prepared for submission to JHEP

Unsupervised Hadronic SUEP at the LHC

Jared Barron,[a] David Curtin,[a] Gregor Kasieczka,[b] Tilman Plehn,[c] and Aris Spourdalakis[a]

[a] *Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7*
[b] *Institut für Experimentalphysik, Universität Hamburg, Germany*
[c] *Institut für Theoretische Physik, Universität Heidelberg, Germany*

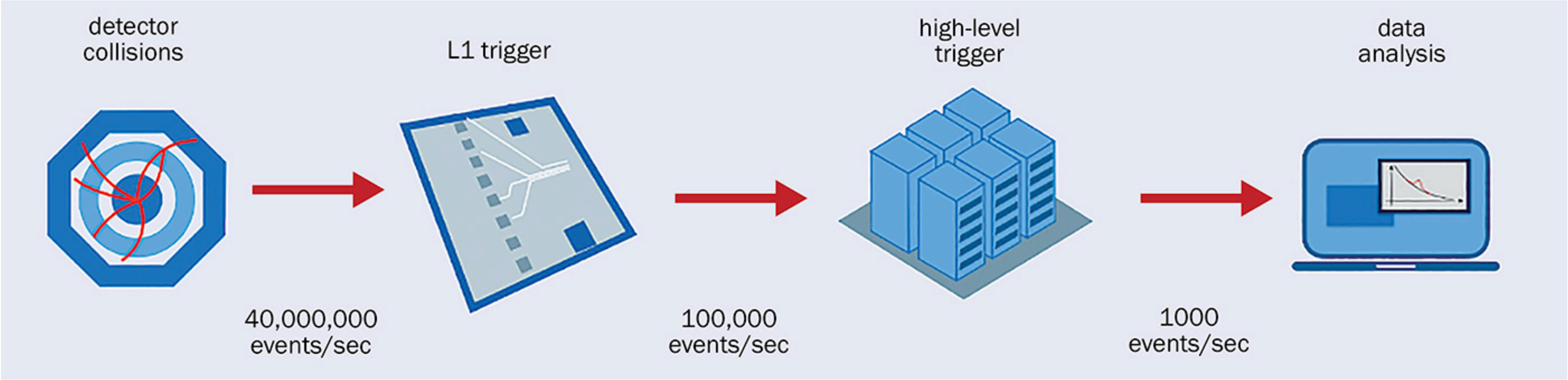
E-mail: jared.barron@mail.utoronto.ca, aspourda@physics.utoronto.ca, gregor.kasieczka@cern.ch , plehn@uni-heidelberg.de, aspourda@physics.utoronto.ca

Abstract: Confining dark sectors with pseudo-conformal dynamics produce SUEPs, or Soft Unclustered Energy Patterns, at colliders: isotropic dark hadrons with soft and democratic energies. We target the experimental nightmare scenario, SUEPs in exotic Higgs decays, where all dark hadrons decay promptly to SM hadrons. First, we identify three promising observables: the charged particle multiplicity, the event ring isotropy, and the matrix of geometric distances between charged tracks. Their patterns can be exploited through a cut-and-count search, supervised machine learning, or an unsupervised autoencoder. We find that the HL-LHC will probe exotic Higgs branching ratios at the per-cent level, even without a detailed knowledge of the signal features. Our techniques can be applied to other SUEP searches, especially the unsupervised strategy, which is independent of overly specific model assumptions and the corresponding precision simulations.

arXiv:2107.12379v2 [hep-ph] 4 Nov 2021

Final Aside: Trigger!

- Focused on strategies for offline analysis
- Additional challenge of recording potential LLP signal events
- Both model-specific LLP triggers (2004.10744, 2103.08620) as well as anomaly based strategies (1811.10276,2005.01598) considered
- Crucial Run 3 / HL-LHC development!

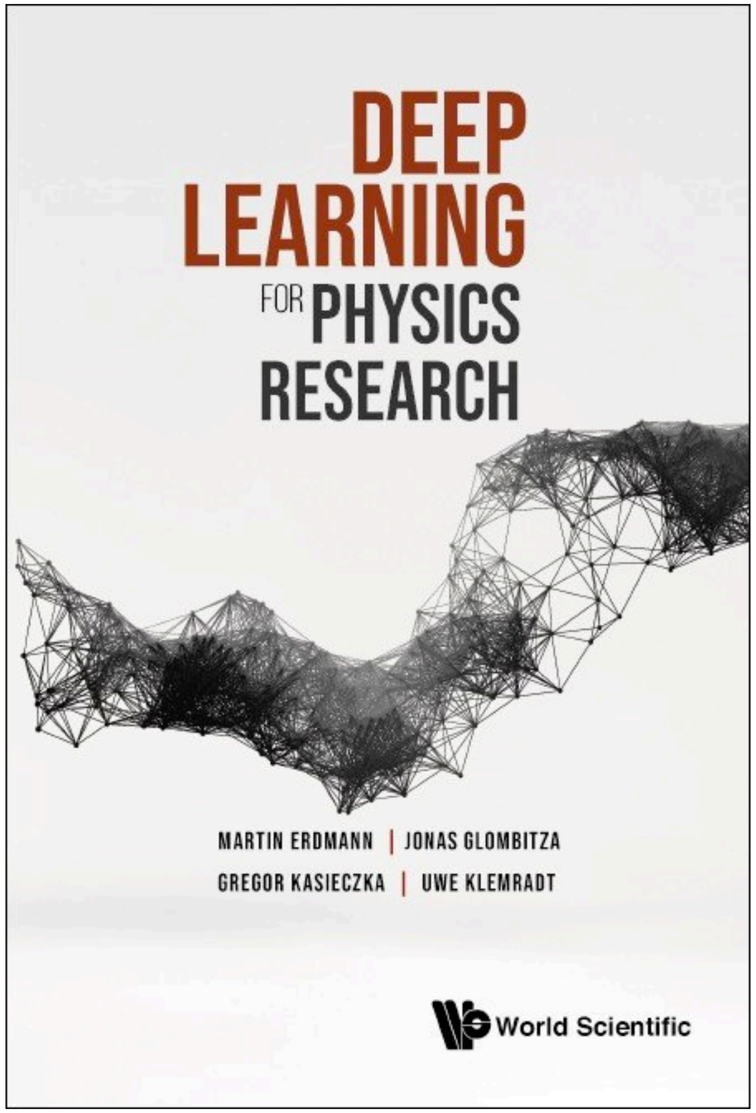


CERN Courier / Pierini (https://cerncourier.com/a/hunting-anomalies-with-an-ai-trigger/)

Conclusions

- Deep Learning for particle physics is rapidly developing solutions to a wide range of problems
- Long-lived analyses amplify existing challenges:
  - Inexact simulation / background estimation
  - Deluge of signal models
  - Recording data
- The tools and ideas are there, but tailoring them to long-lived analyses will take work
  - Trade-off coverage and sensitivity
- Overview of ML in HEP papers: https://iml-wg.github.io/HEPML-LivingReview/

Our intro book:



https://worldscientific.com/worldscibooks/10.1142/12294

Thank you!