Finetuning Foundation Models for Analysis Optimization

arXiv:2401.13536

Lukas Heinrich, Nicole Hartman, Matthias Vigl

6th IML Workshop, CERN 29.01.2024 – 02.02.2024

Technische Universität München

ORIGINS Excellence Cluster

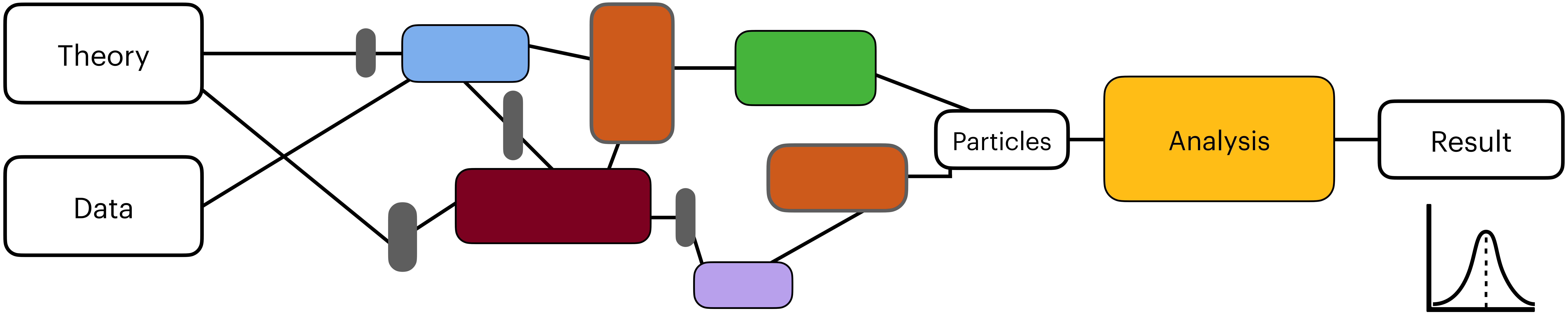
ATLAS EXPERIMENT

MAX-PLANCK-INSTITUT FÜR PHYSIK

Analysis pipeline at the LHC

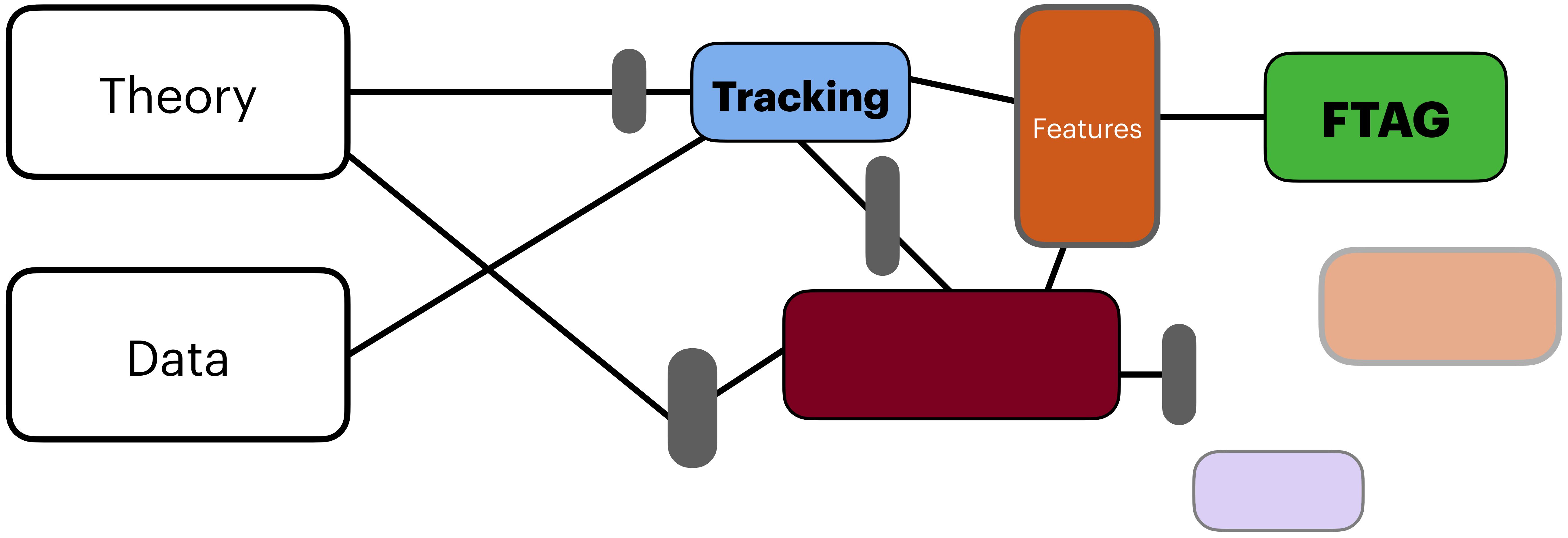
Lots of (also ML) components in our analysis pipeline

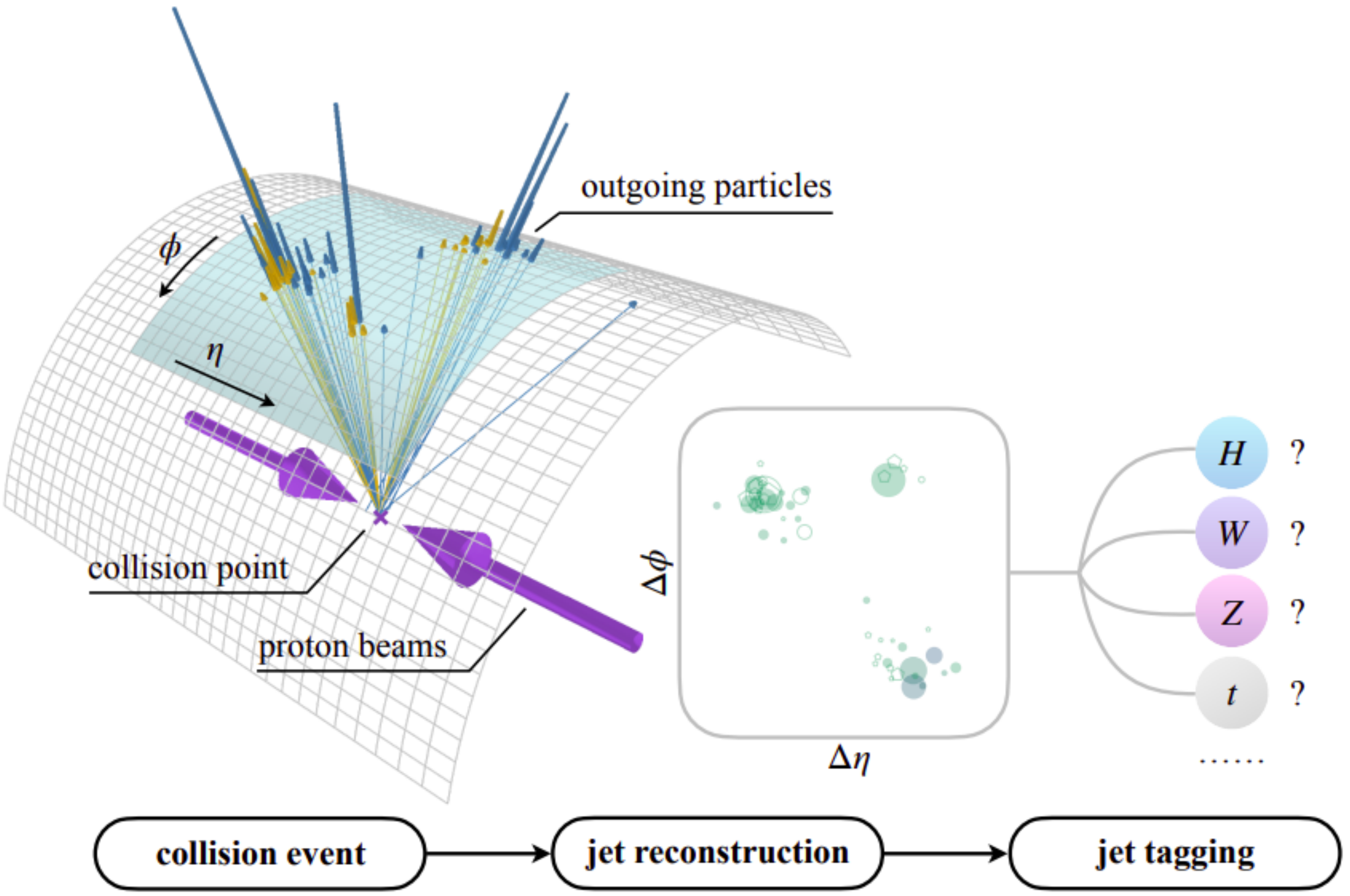
But each optimized separately and downstream components are optimized based on the steps prior to it



Analysis pipeline at the LHC

e.g. jet flavour-tagging can only be optimized after tracking, but we rarely re-optimize tracking for flavor tagging or jet classification



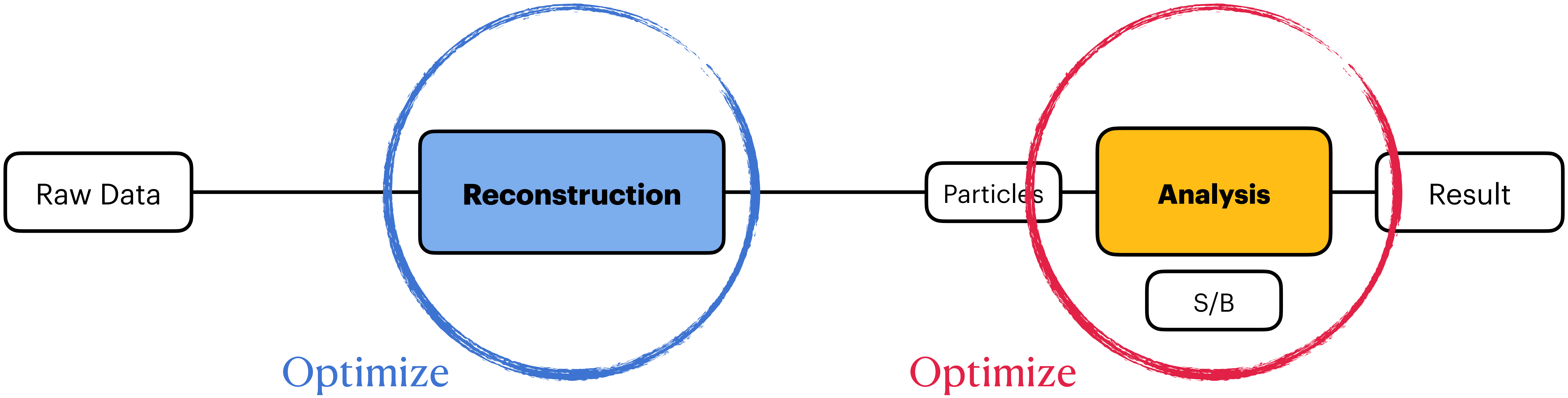


[arXiv:2202.03772]

Lukas Heinrich, Nicole Hartman, Matthias Vigl

Analysis pipeline at the LHC

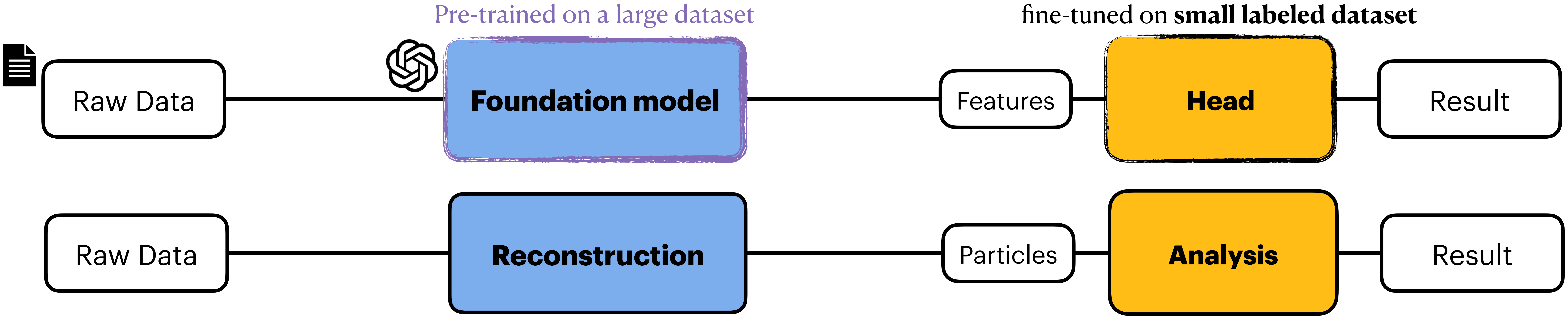
The optimization of the sensitivity is primarily the job of the **analysis**, given a fixed **reconstruction** - mostly common for all analysis



Modern ML with Foundation Models

ML and HEP setups are fortunately very aligned

Also often split in two parts, but key difference is that backbone can be fine-tuned w/ gradient descent



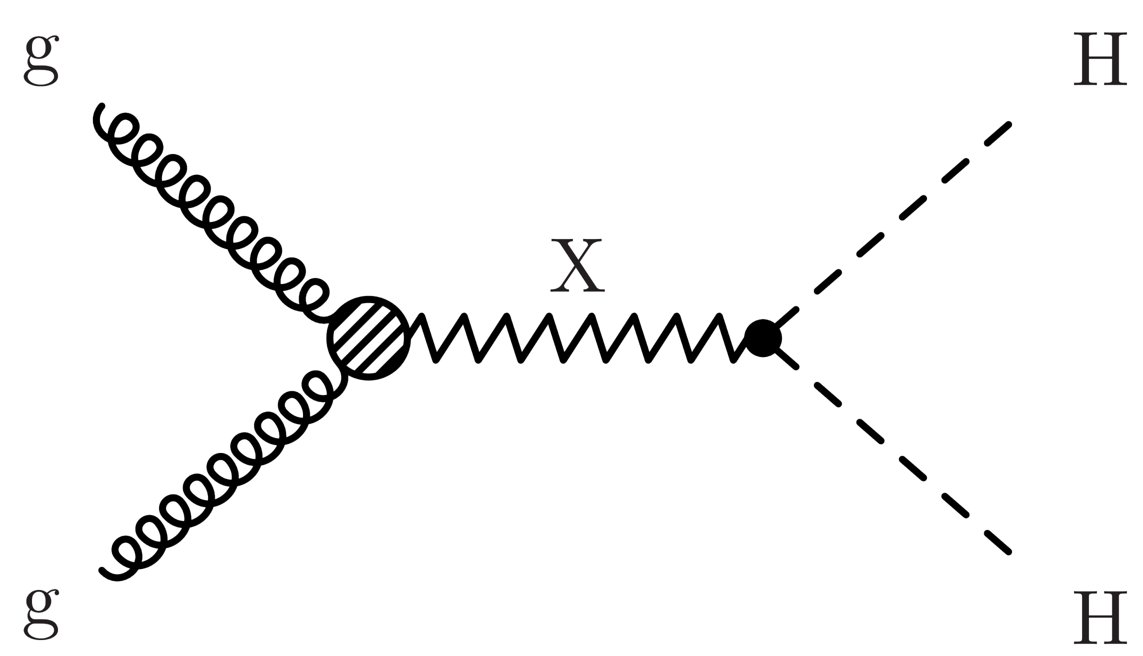
Q: Could this workflow also work in HEP?

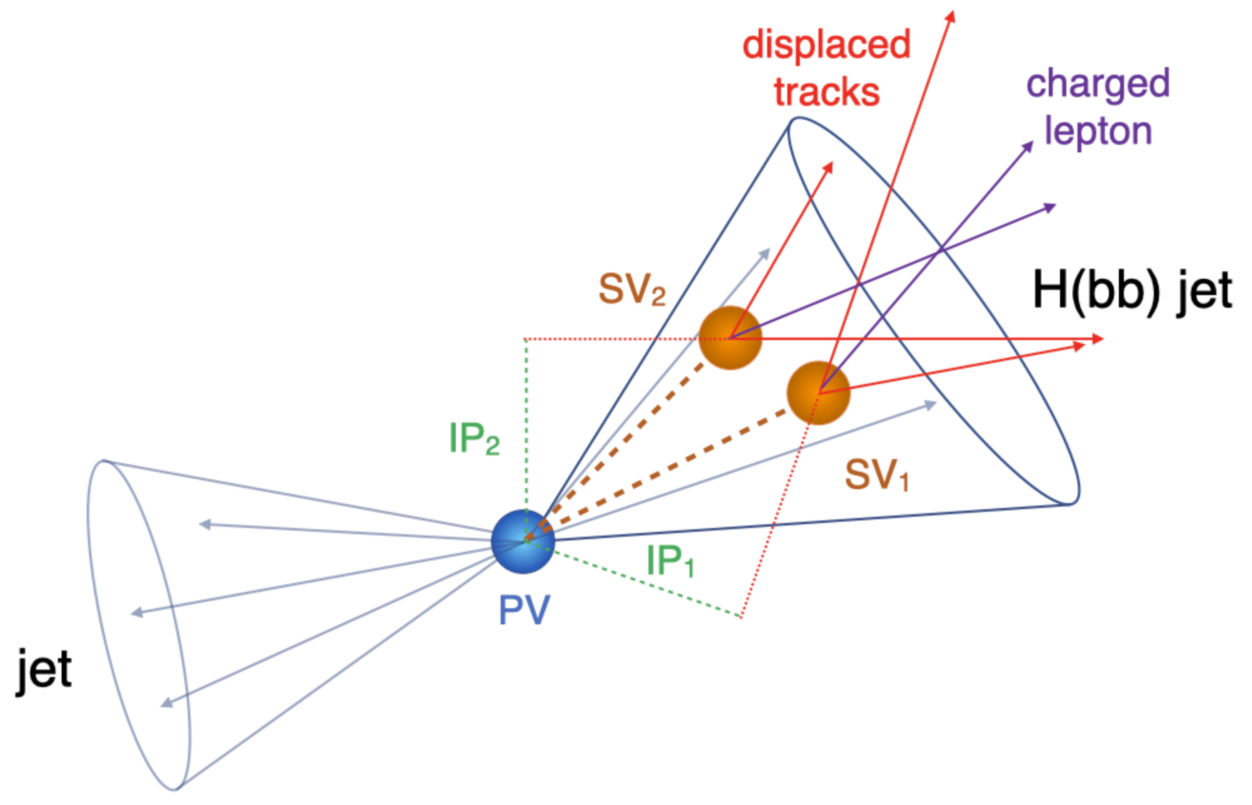
Lukas Heinrich, Nicole Hartman, Matthias Vigl

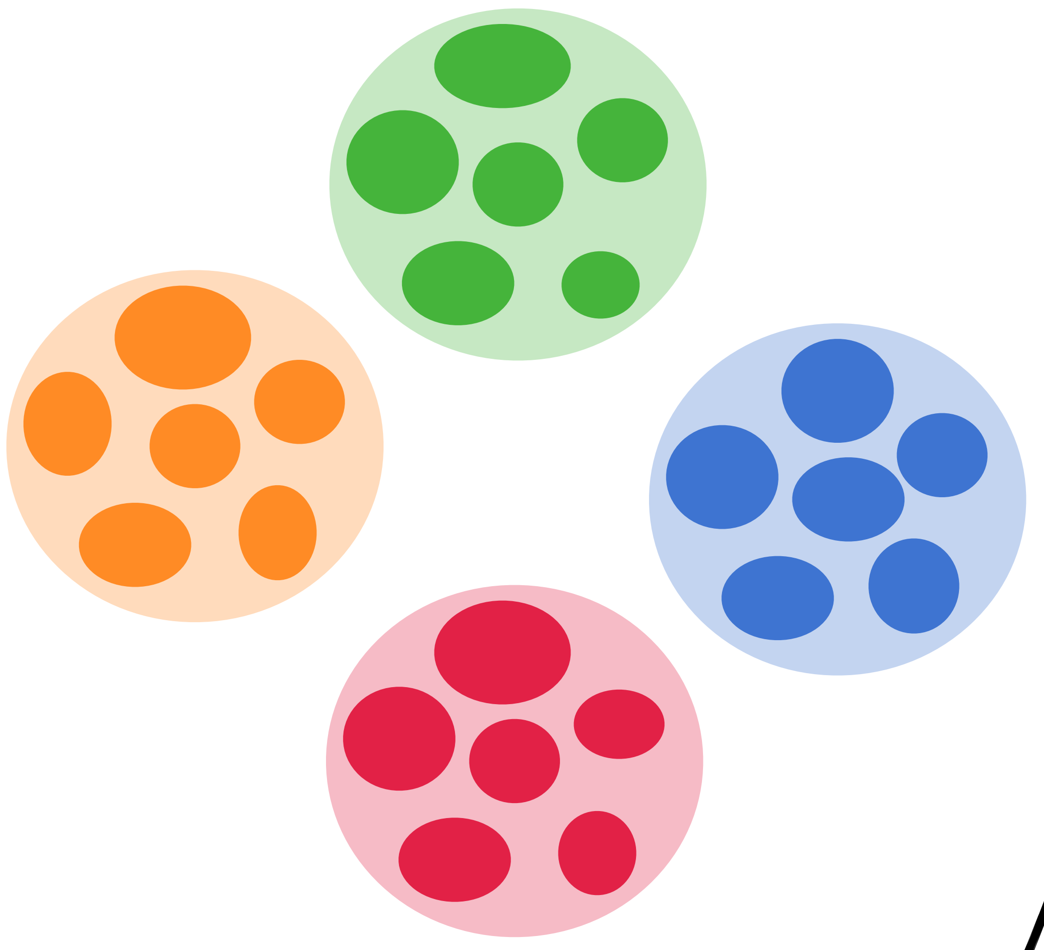
A toy end-to-end Analysis

$X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$[1]

Final state with Higgs/ QCD Jets







Jet constituents

Xbb flavour tagging

Backbone FM

Transformer (ParT)[2]

Xbb

Jet representation

Jet representation

Jet representation

Jet representation

Analysis head

DeepSets

S/B

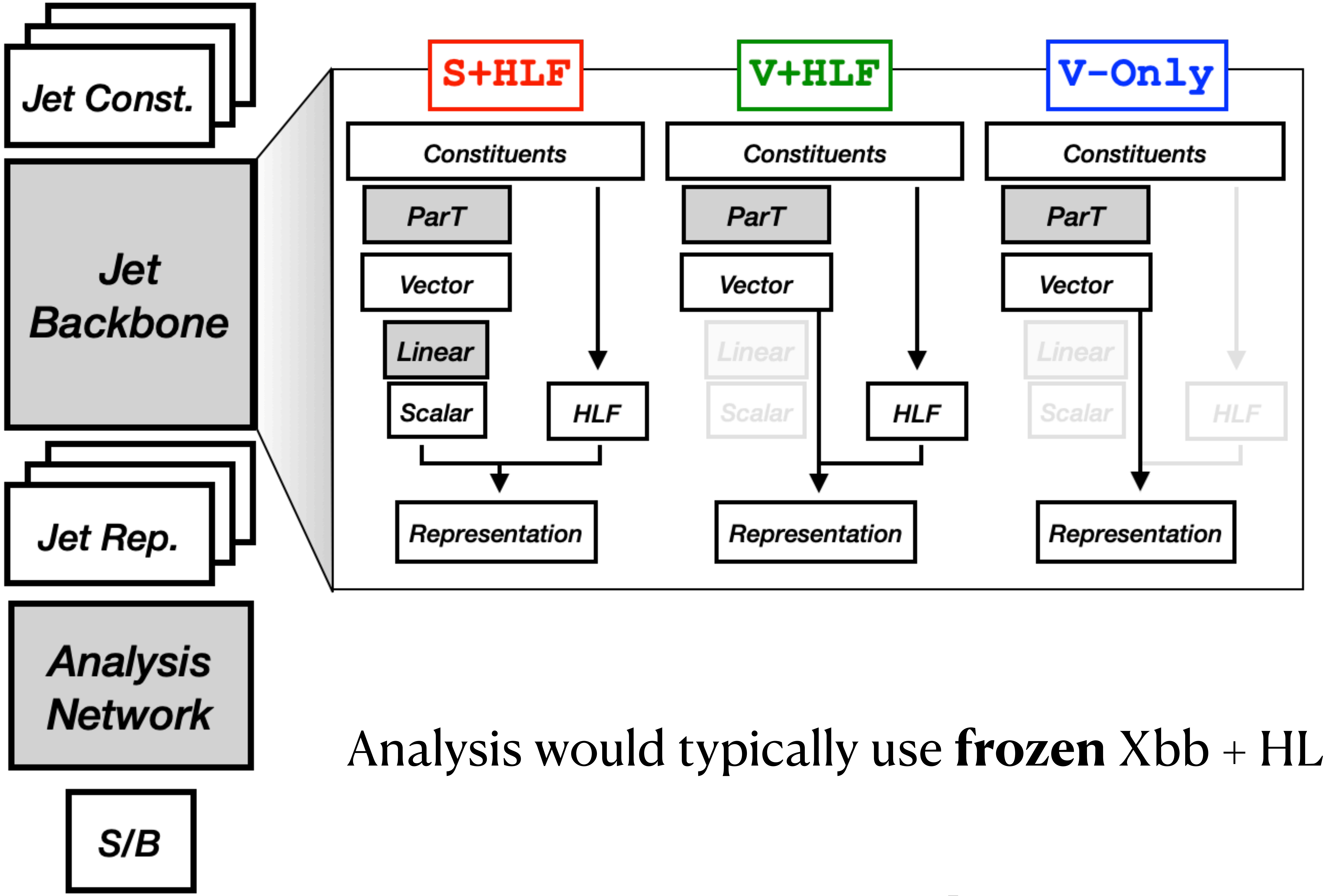
Result

[1]: Huilin Qu, Congqiao Li, and Sitian Qian, "Particle Transformer for Jet Tagging," (2022), arXiv:2202.03772

[2]: Duarte Javier, CMS open data [http://opendata.cern.ch/record/12102]

Lukas Heinrich, Nicole Hartman, Matthias Vigl

Backbone Jet representation



Q: Do high-dim embeddings hold more (useful) info than Xbb+HL Features?

Analysis would typically use **frozen** Xbb + HL Features (jet 4-momenta)

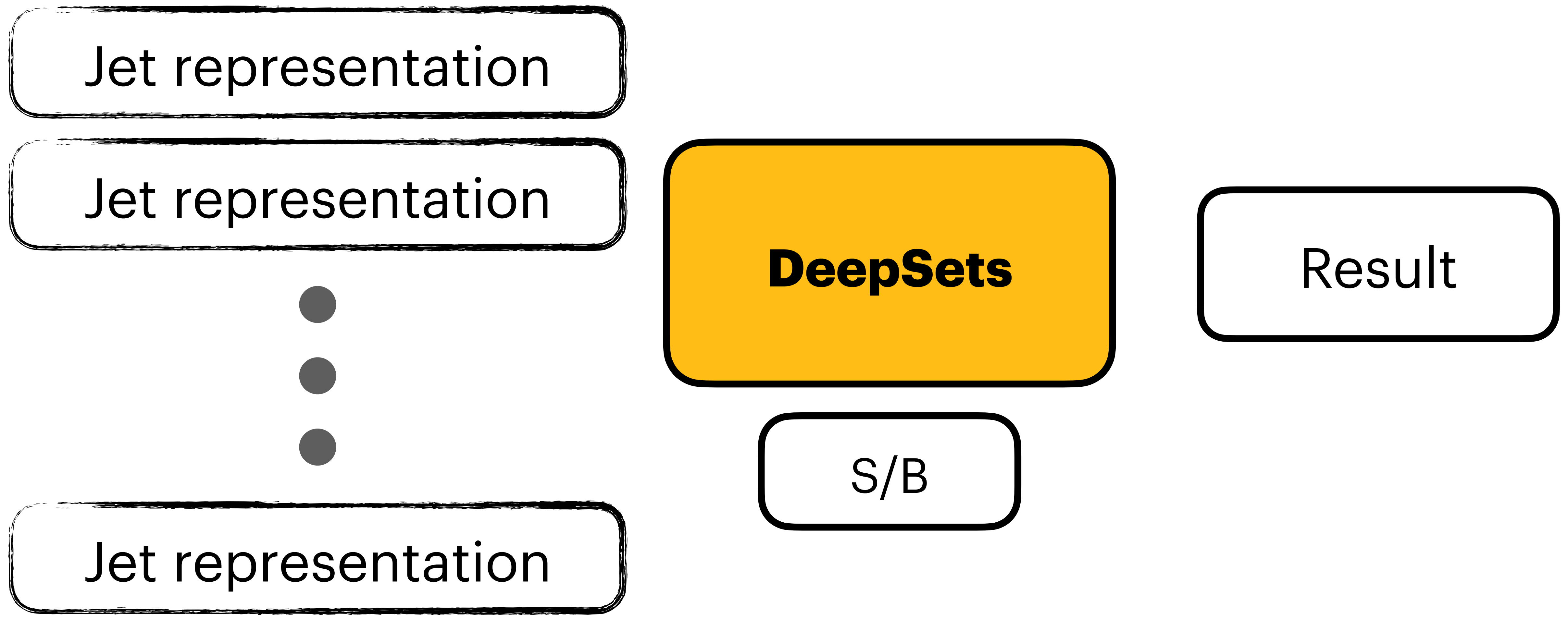
Lukas Heinrich, Nicole Hartman, Matthias Vigl

Analysis head

The head is trained for S/B discrimination with Jet representations from backbone as inputs

Variable number of jets per event + Permutation Invariance -> DeepSets

Q: Does fine-tuning the jet representation help?

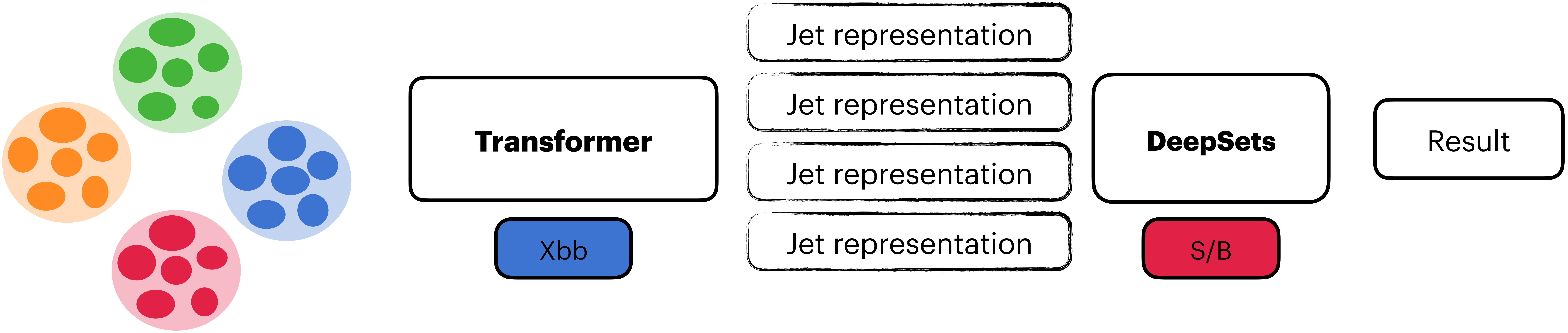


Lukas Heinrich, Nicole Hartman, Matthias Vigl

Frozen training

Backbone trained on **Xbb** task and then frozen

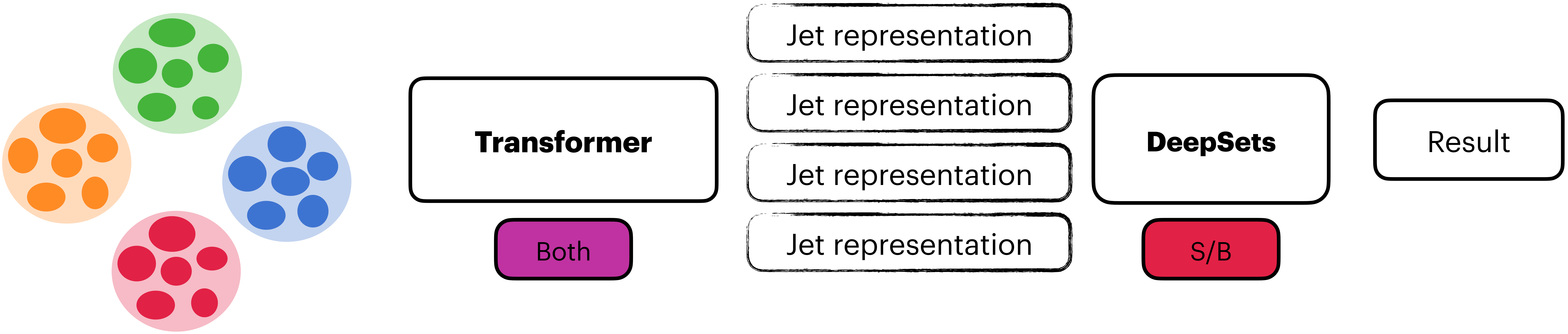
DeepSets + binary classification trained on **S/B**



Fine-tuned training

Backbone pre-trained on **Xbb** task

Then **fine-tuned** on **S/B**

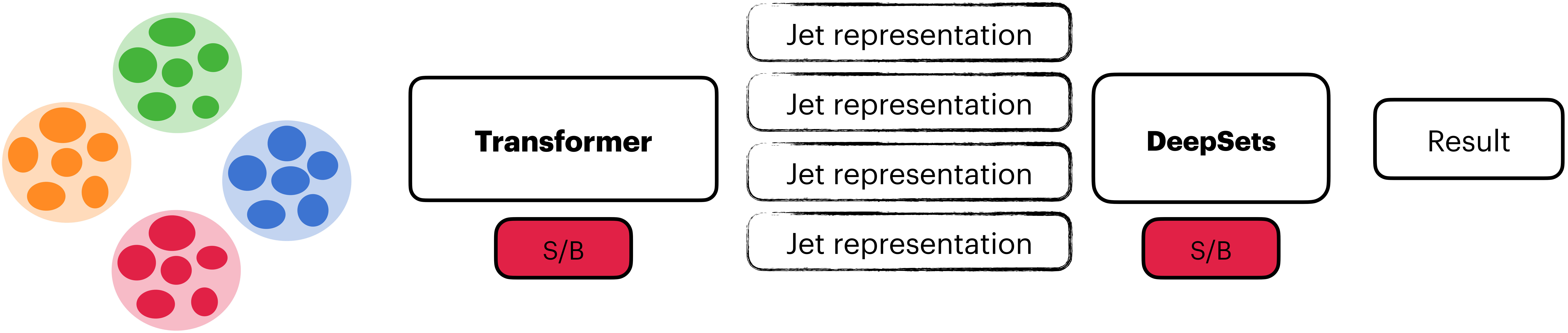


Lukas Heinrich, Nicole Hartman, Matthias Vigl

From scratch training

No backbone pre-training

Backbone + head trained from scratch on **S/B**



Lukas Heinrich, Nicole Hartman, Matthias Vigl

Architecture autonomy

Structural Autonomy

Representational Autonomy

Training \ Architecture	Scalar + HL	Vector + HL	Vector
Frozen	Standard HEP		Hope for sufficient stat
Fine-tuned		ML-assisted HEP	
From scratch	Inductive Bias is all you need		'Hits to Higgs'

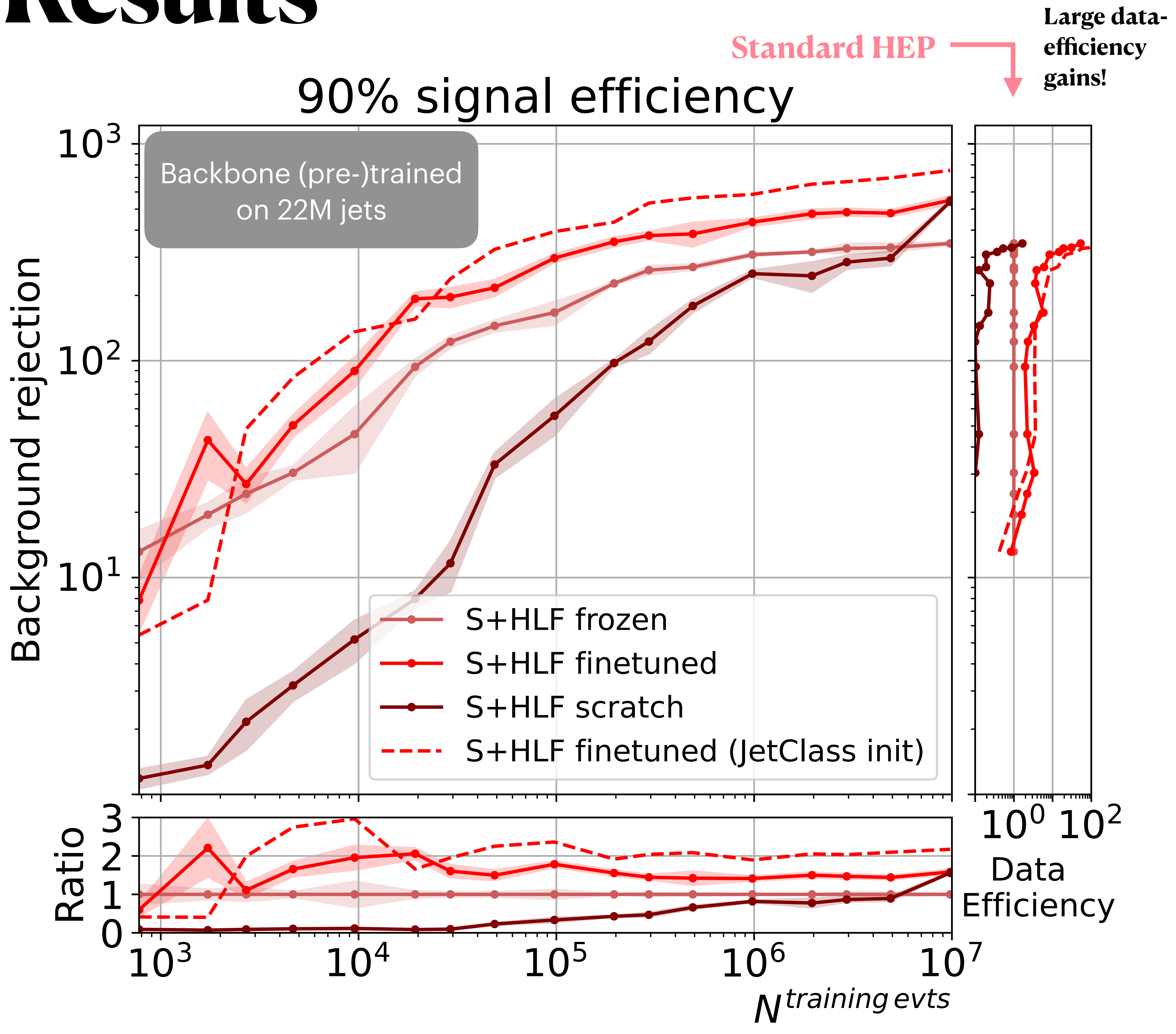
Lukas Heinrich, Nicole Hartman, Matthias Vigl

Results

Well-known patterns from ML seem to hold also in HEP

- Fine-tuning workflow improves both **performance** & **data efficiency** (10-100x wrt standard hep)

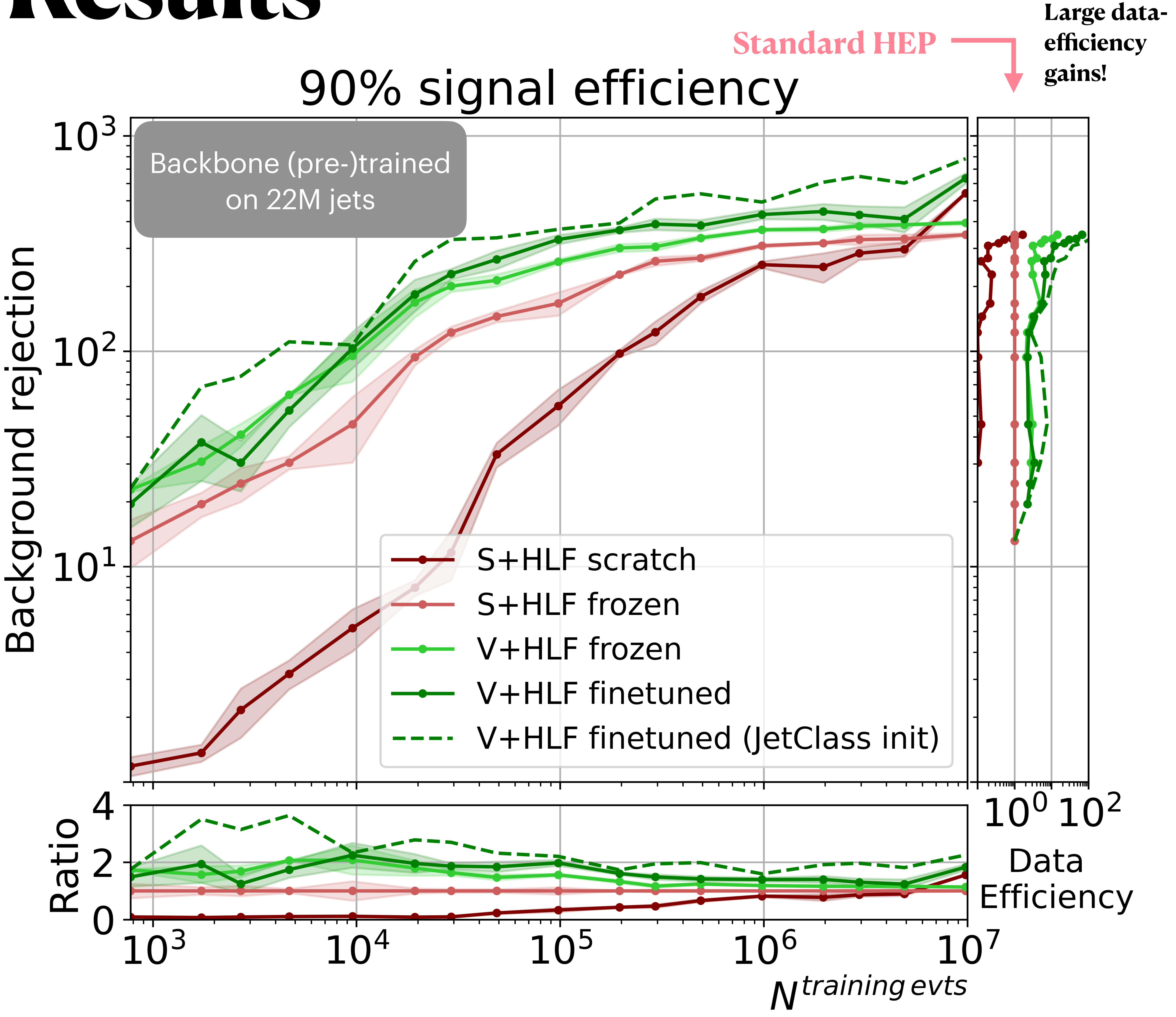
- **Domain adaptation**: Pre-training on a different dataset (JetClass[3]) helps



[3]: Huilin Qu, Congqiao Li, and Sitian Qian, "JetClass: A Large-Scale Dataset for Deep Learning in Jet Physics," (2022).

Lukas Heinrich, Nicole Hartman, Matthias Vigl

Results



Well-known patterns from ML seem to hold also in HEP

- Fine-tuning workflow improves both **performance** & **data efficiency** (10-100x wrt standard hep)
- High-dim embeddings also seem to be useful in the frozen case
- **Domain adaptation**: Pre-training on a different dataset (JetClass[3]) helps

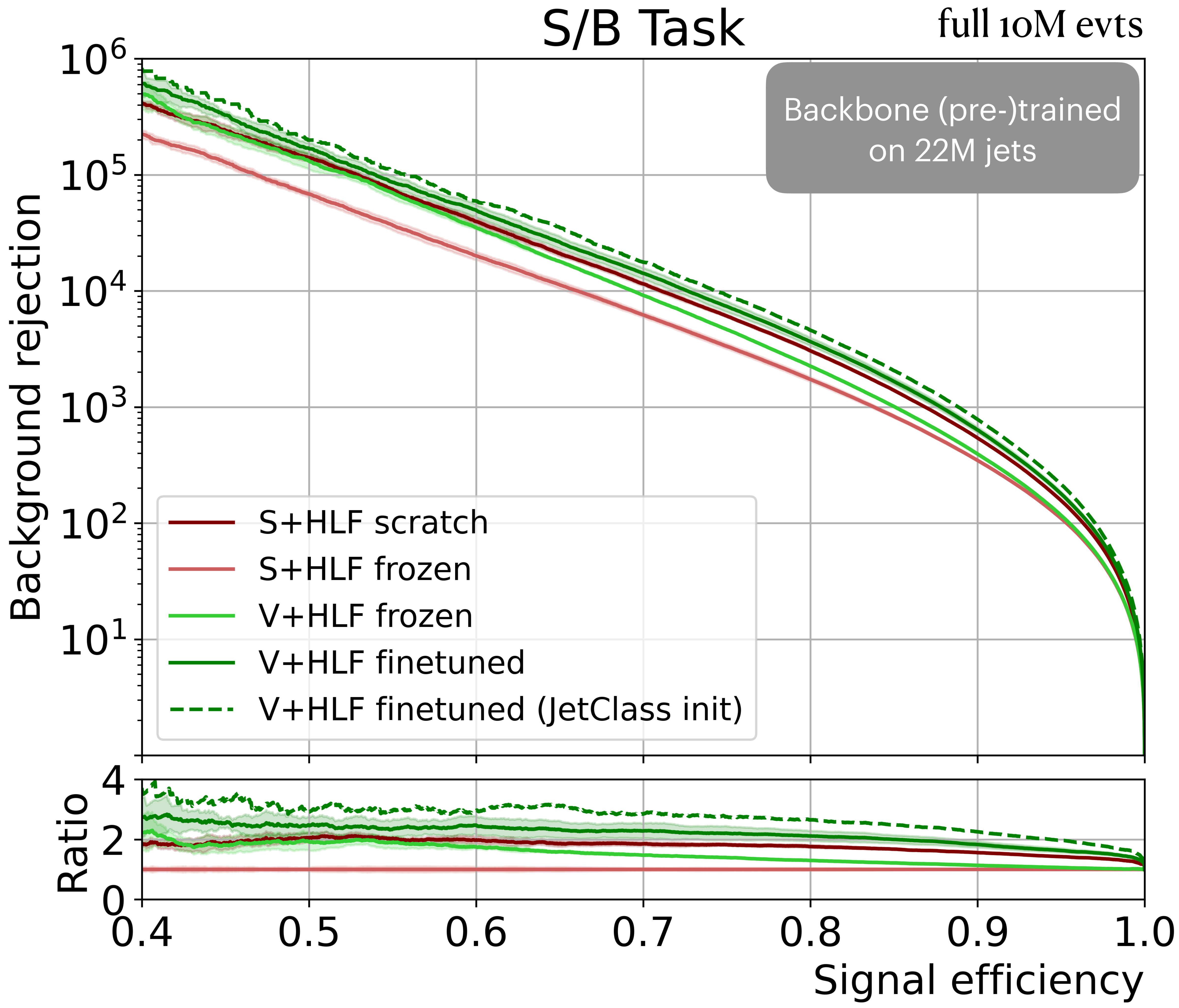
[3]: Huilin Qu, Congqiao Li, and Sitian Qian, "JetClass: A Large-Scale Dataset for Deep Learning in Jet Physics," (2022).

Conclusions

Fine-tuning workflow for end to end analysis works and is useful even for simple examples

Compared to standard HEP approach:

- **2x in background rejection**
- **10-100x in data efficiency**
- There might be more to gain in more complex topologies



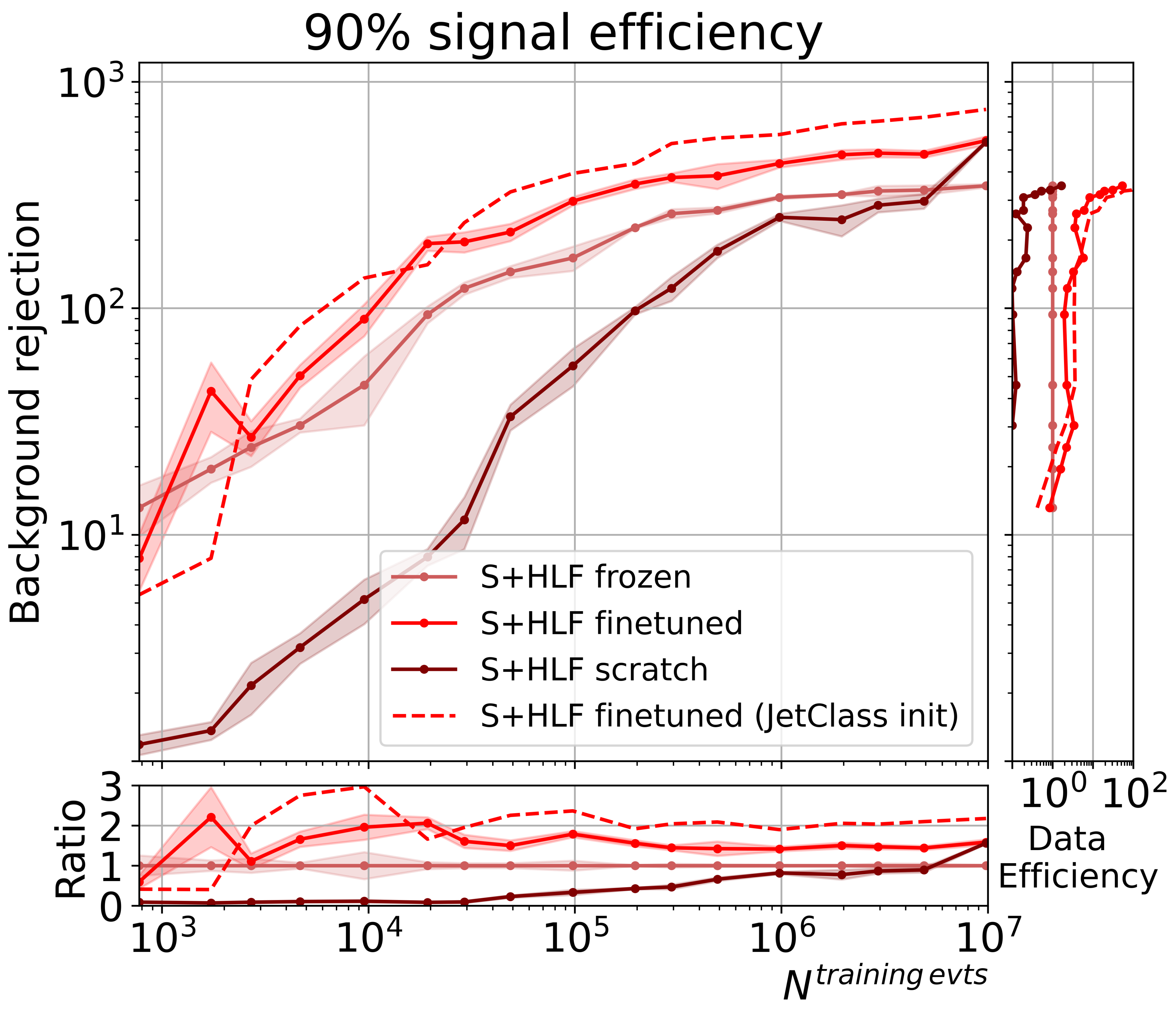
Link to the paper: **arXiv:2401.13536**

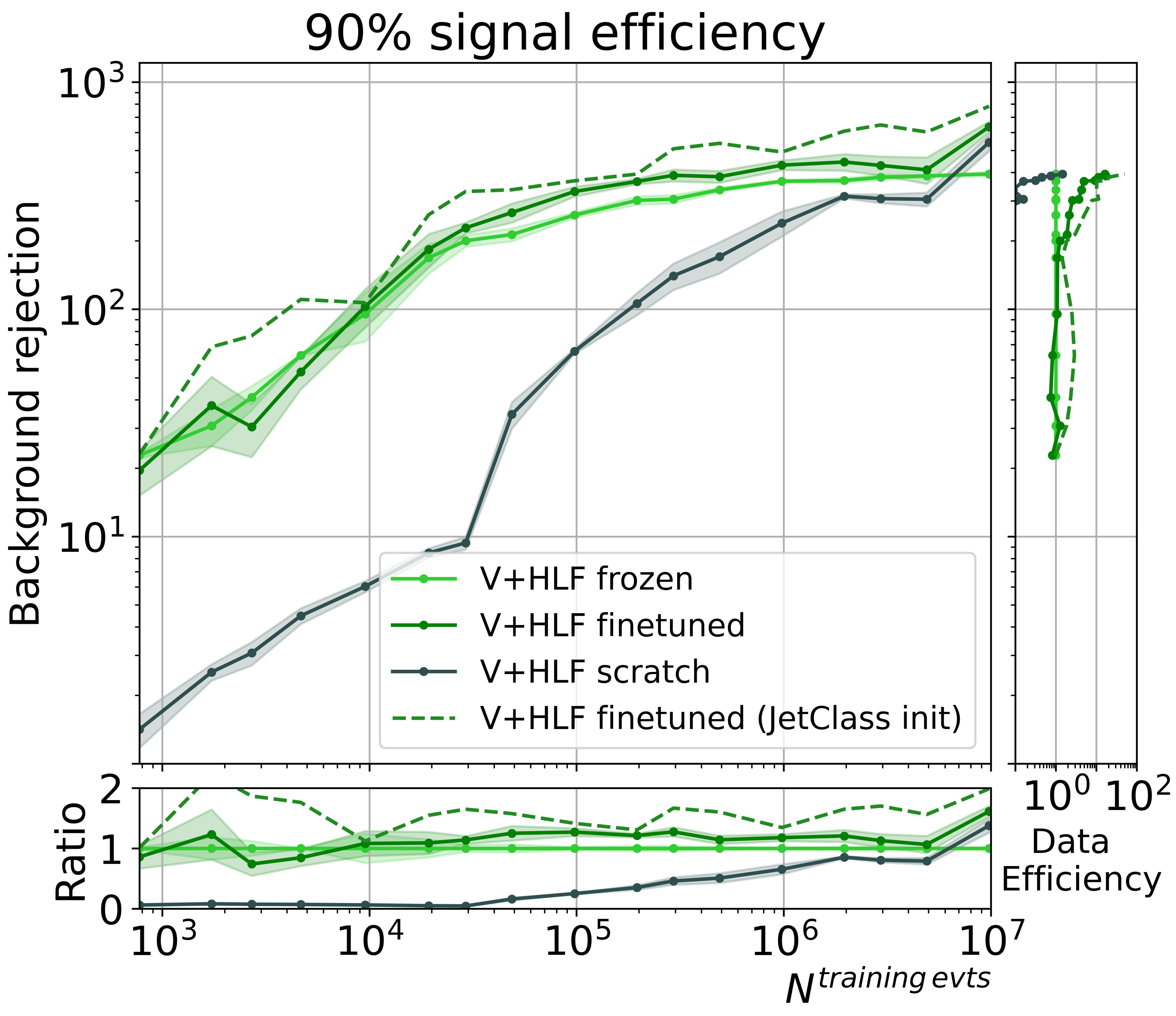
Unsupervised backbone? Previous talk on "**Masked Particle Modeling**" by Samuel Byrne Klein

Lukas Heinrich, Nicole Hartman, Matthias Vigl

Backup

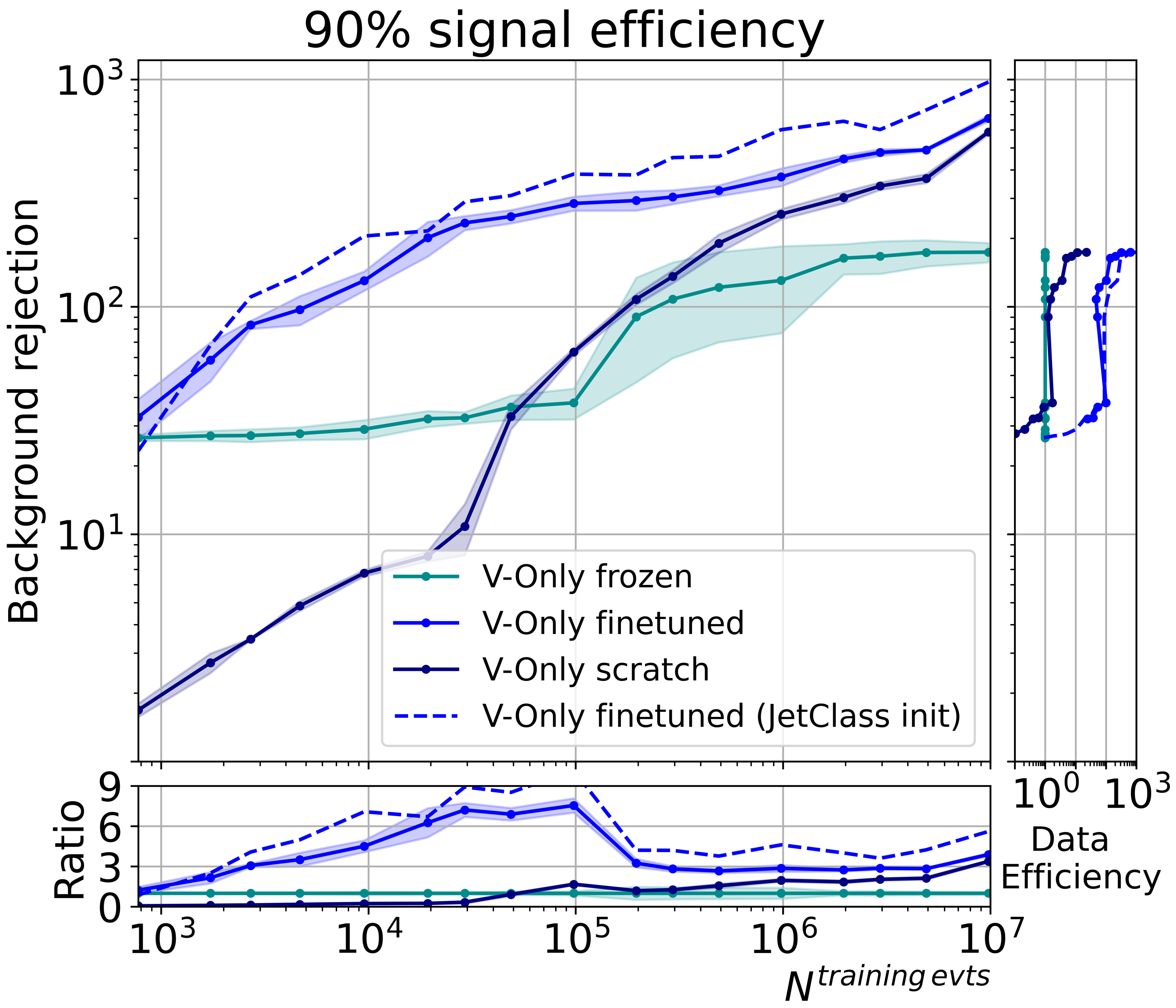
From scratch training eventually surpasses frozen models, it's just slow



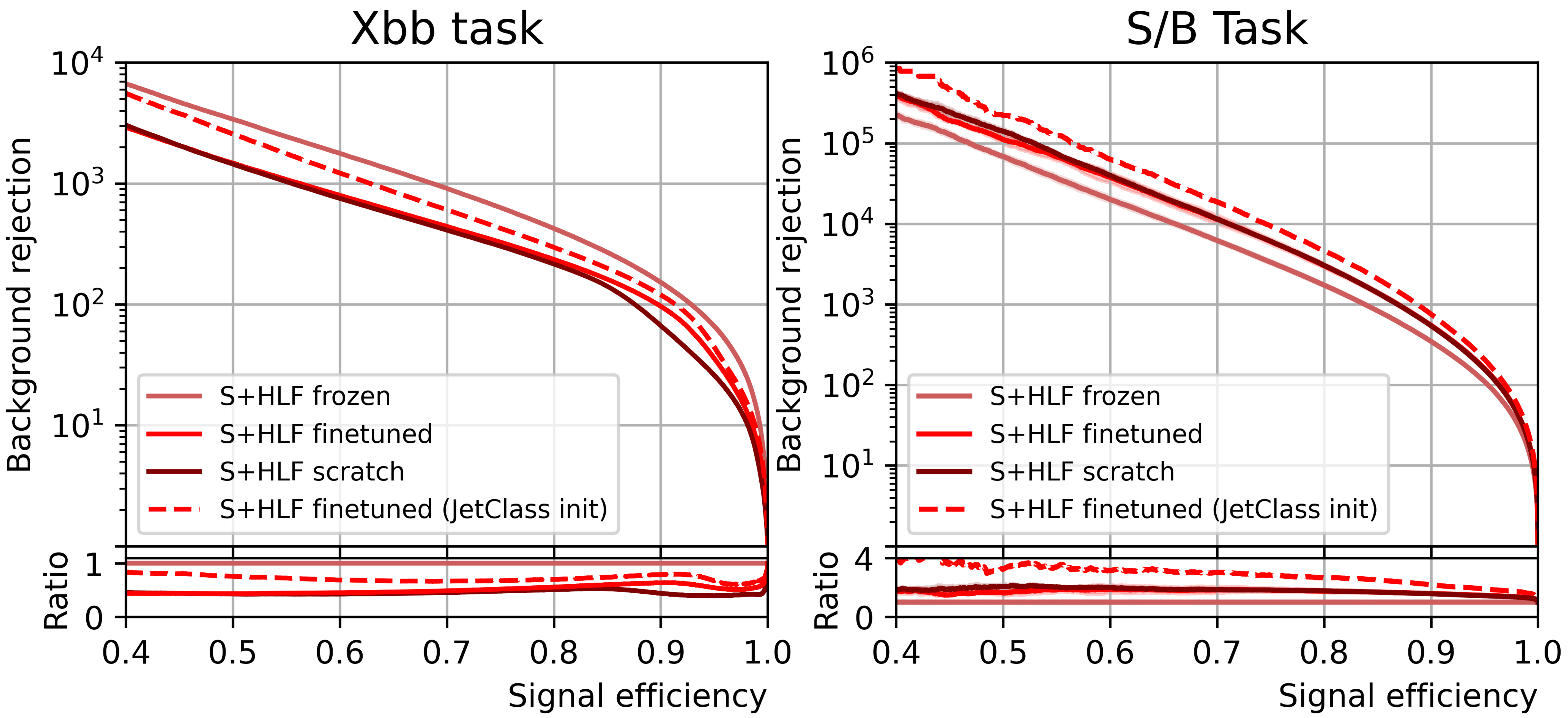


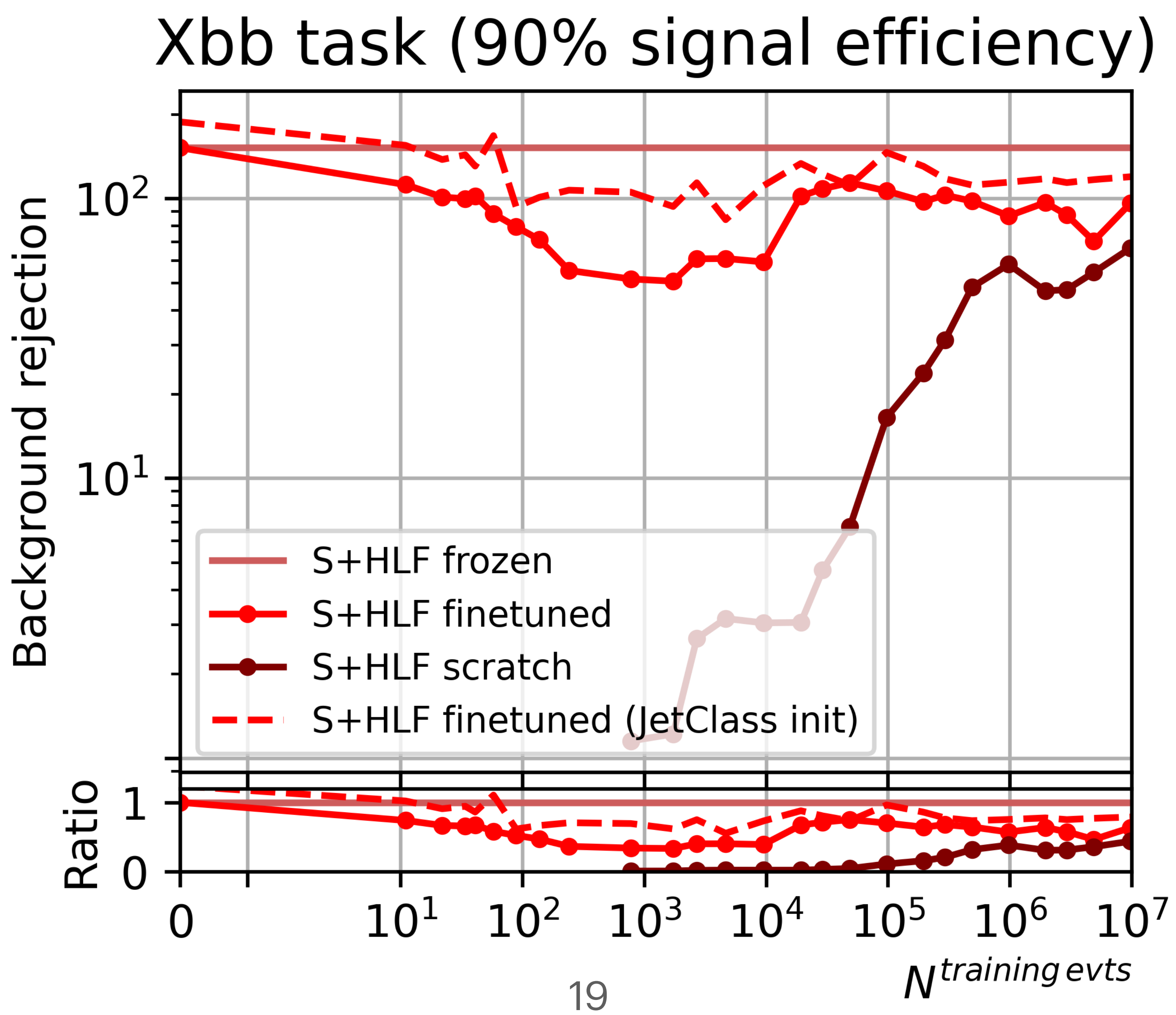
Lukas Heinrich, Nicole Hartman, Matthias Vigl

From scratch training eventually surpasses frozen models, it's just slow



Lukas Heinrich, Nicole Hartman, Matthias Vigl

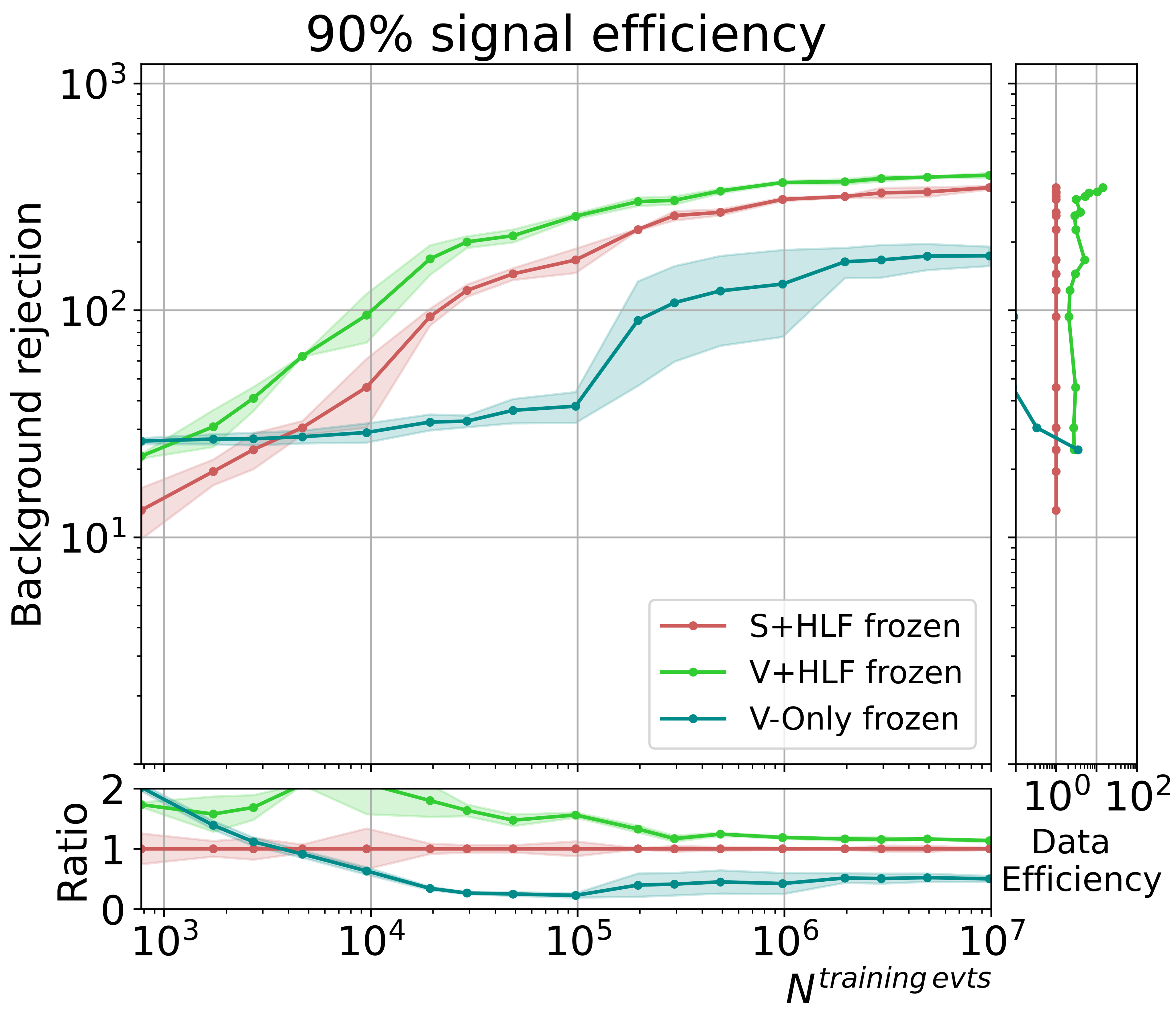




Xbb is learned when solving the downstream task even without actual jet labels

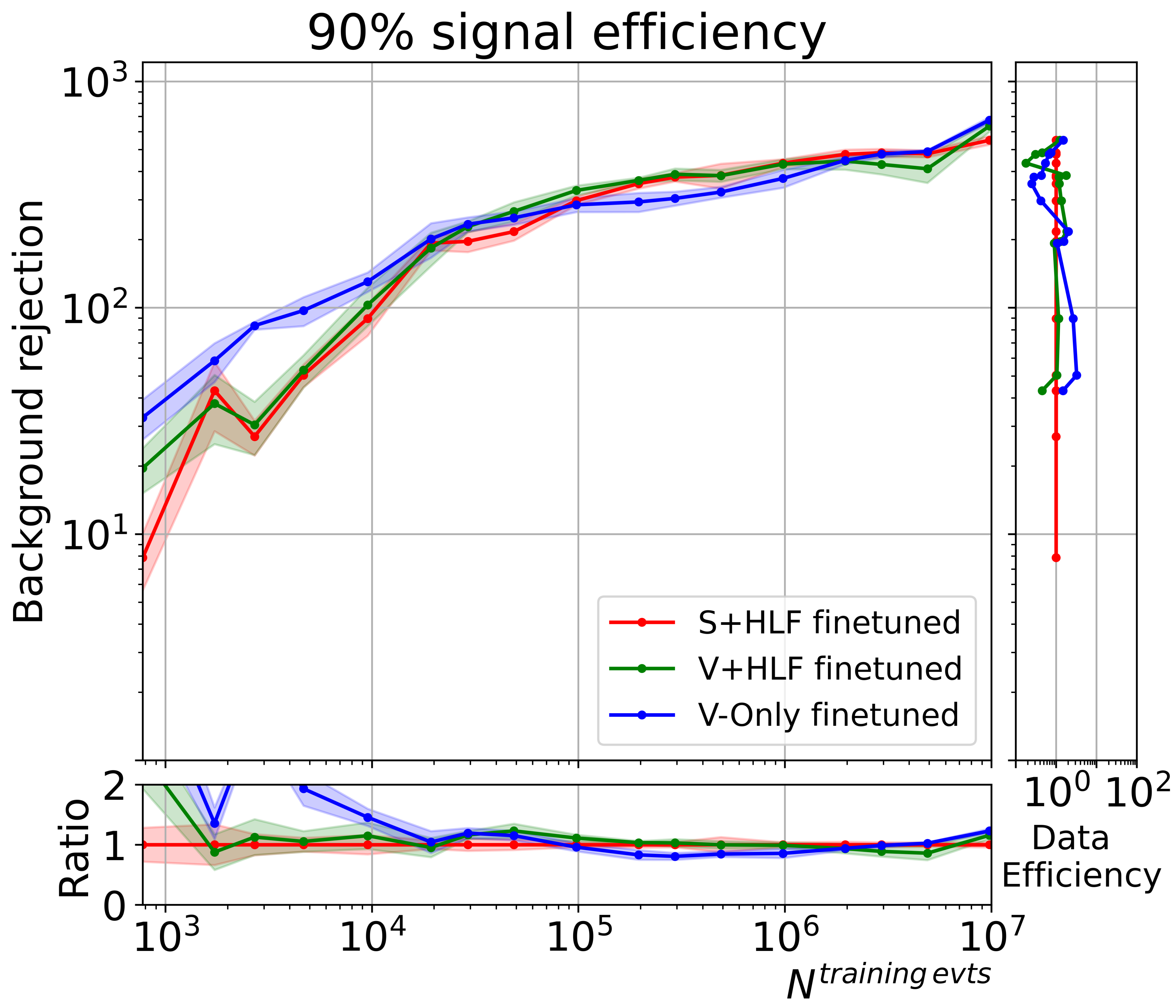
Lukas Heinrich, Nicole Hartman, Matthias Vigl

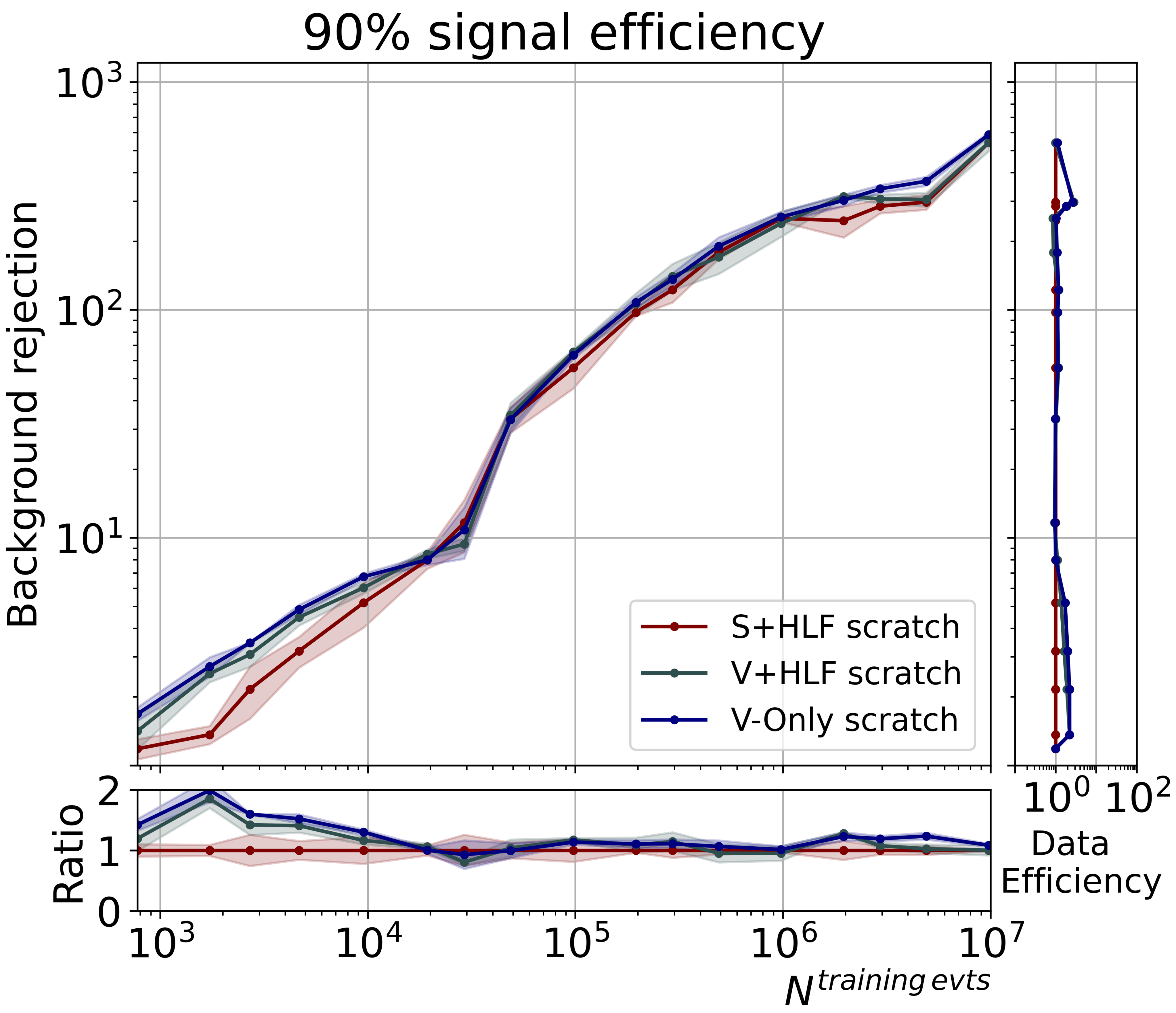
High dim embeddings help for frozen jet representations



Lukas Heinrich, Nicole Hartman, Matthias Vigl

Dimensionality becomes less important when training end-to-end





Lukas Heinrich, Nicole Hartman, Matthias Vigl

Setup: CMS open data and ParT

CMS open data: Duarte Javier, [http://opendata.cern.ch/record/12102]

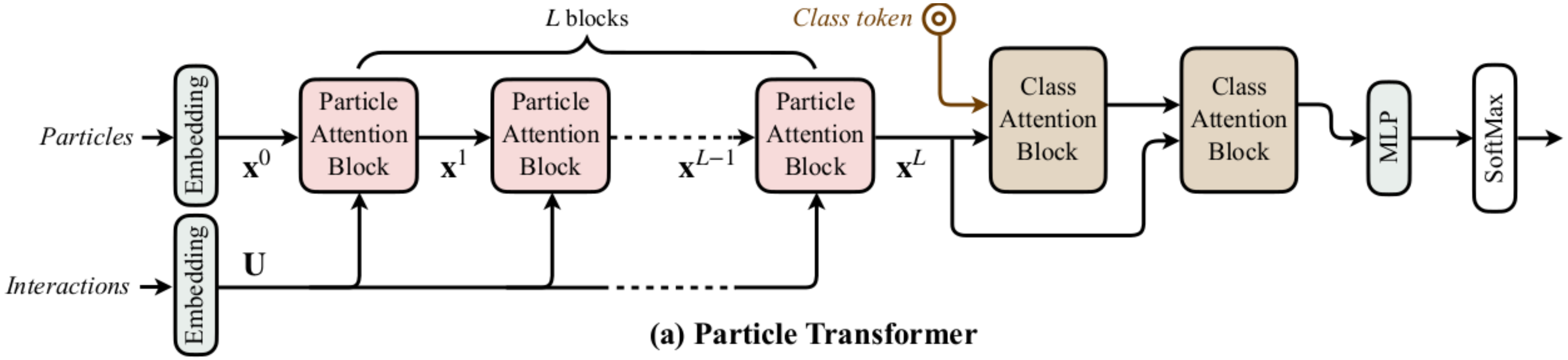
Jets are clustered using the anti-kT algorithm with R=0.8 from particle flow (PF) candidates

Constituents features:

- up to 100 PF per jet
- 17 features per PF

High-level features:

- Jet 4-momenta
- Xbb scores from ParT



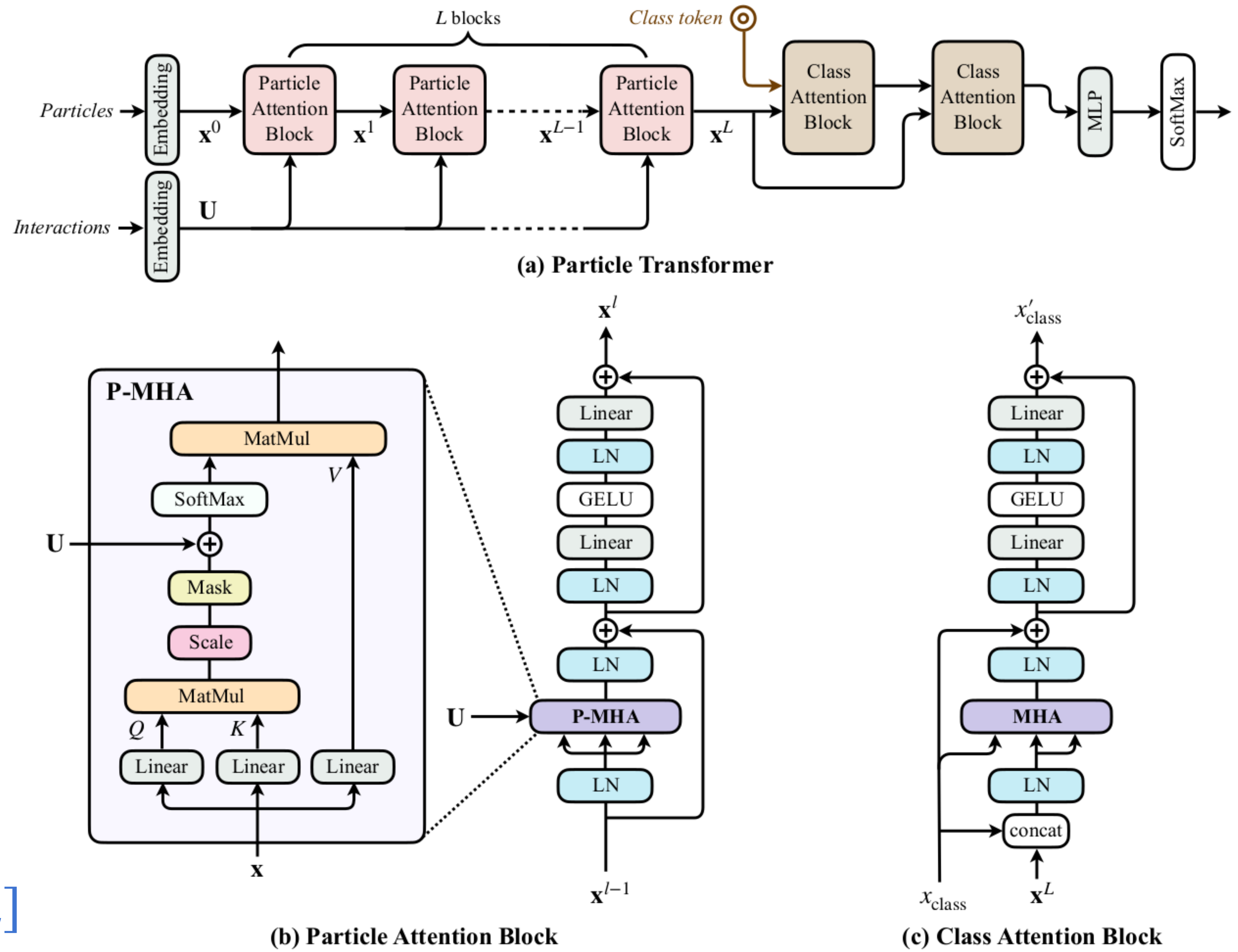
(a) Particle Transformer

Particle transformer for FTAG [arXiv:2202.03772]

Training: QCD vs Higgs jets

10M events / 22M jets

ParT



(a) Particle Transformer

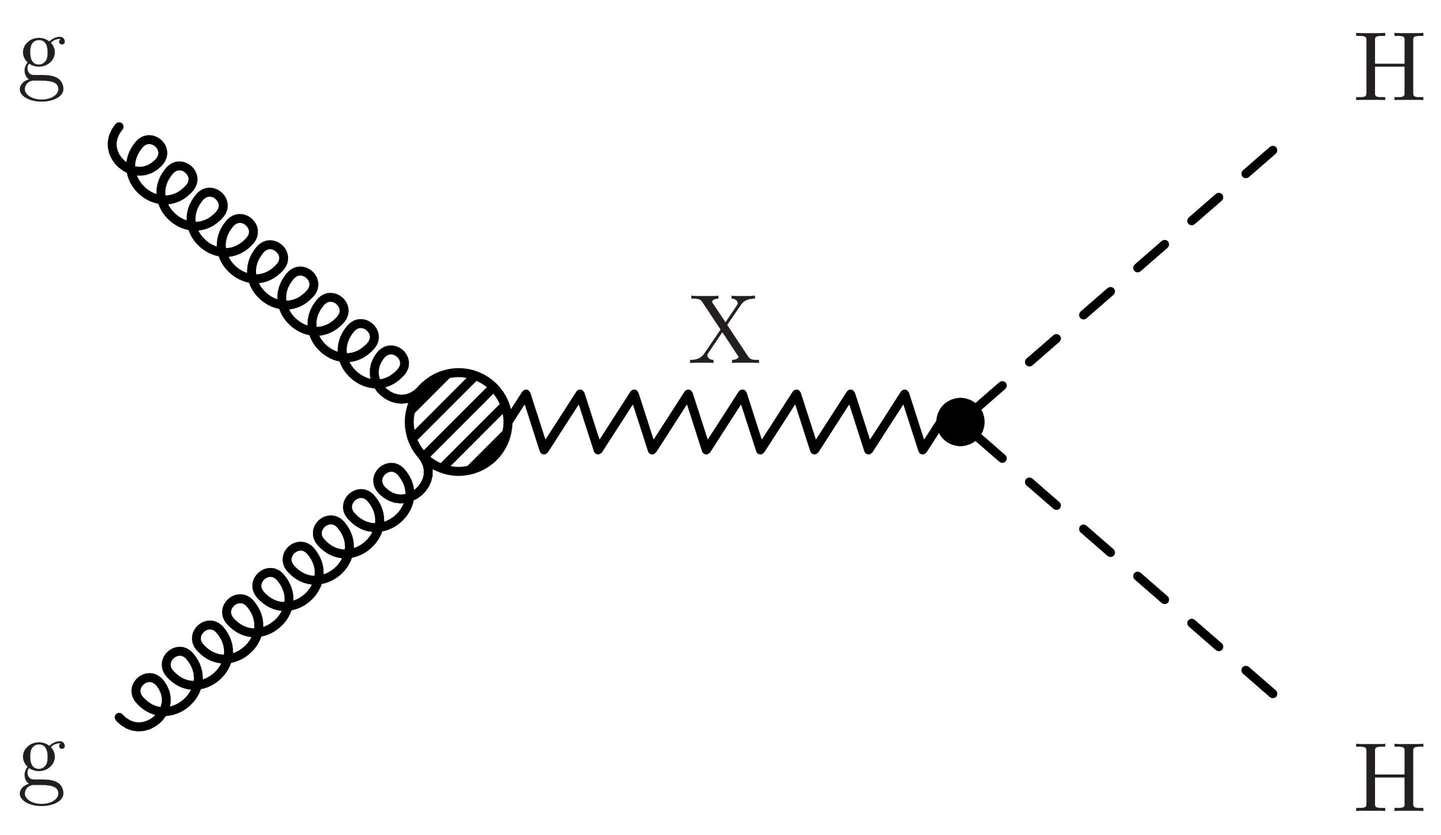
(b) Particle Attention Block

(c) Class Attention Block

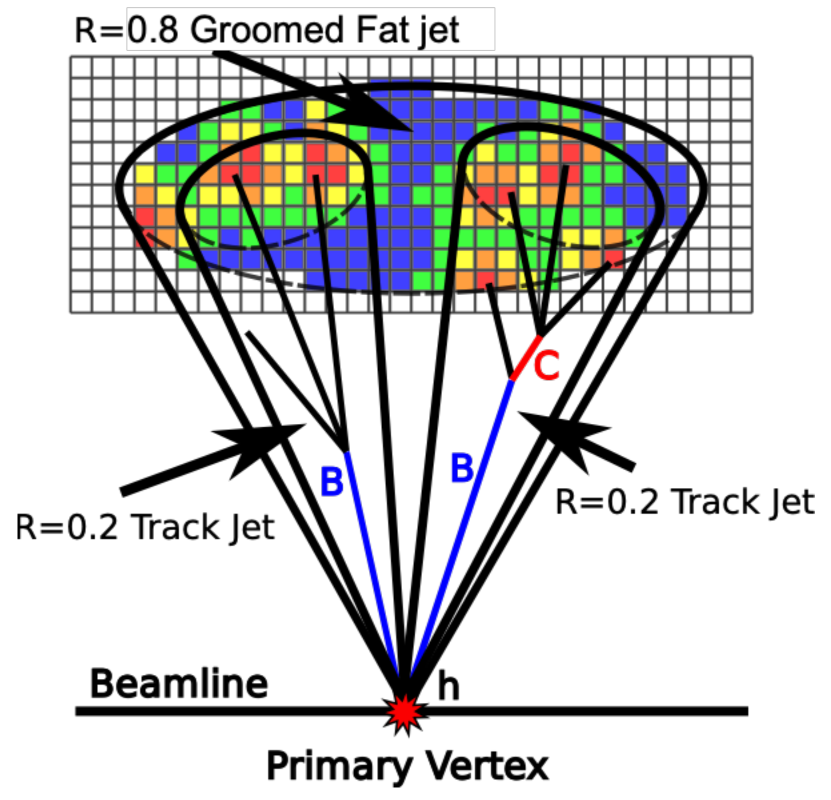
[arXiv:2202.03772]

CMS open data

- CMS simulated dataset:
- Sample with jet, track and secondary vertex properties for H(bb) tagging (http://opendata.cern.ch/record/12102)
- meant for jet tagging, up to 100 pf cand per jet - 17 feats each
- signal samples: 11 mass points - M_x from 600 GeV to 4500 GeV, bkg: QCD multijet
- 'fat jets' (fj) 4-momenta and (old) Xbb score



10M events / 22M jets




```
  - ['pfcand_pt_log', null]
  - ['pfcand_e_log', null]
  - ['pfcand_logptrel', null]
  - ['pfcand_logerel', null]
  - ['pfcand_deltaR', null]
  - ['pfcand_charge', null]
  - ['pfcand_isChargedHad', null]
  - ['pfcand_isNeutralHad', null]
  - ['pfcand_isGamma', null]
  - ['pfcand_isEl', null]
  - ['pfcand_isMu', null]
  - ['pfcand_dz', null]
  - ['pfcand_dzerr', null]
  - ['pfcand_dz', null]
  - ['pfcand_dzerr', null]
  - ['pfcand_deta', null]
  - ['pfcand_dphi', null]
pf_vectors:
  length: 110
  pad_mode: wrap
  vars:
  - [pfcand_px, null]
  - [pfcand_py, null]
  - [pfcand_pz, null]
  - [pfcand_energy, null]
```

[http://cms-results.web.cern.ch/cms-results/public-results/publications/BTV-16-002/]