Parnassus for Detector Simulation with ~~Legacy~~ e^+e^- Collider Data

Ya-Feng Celia Lo for the Parnassus team

ML4Jets 2026

Why legacy e^+e^- collider data?

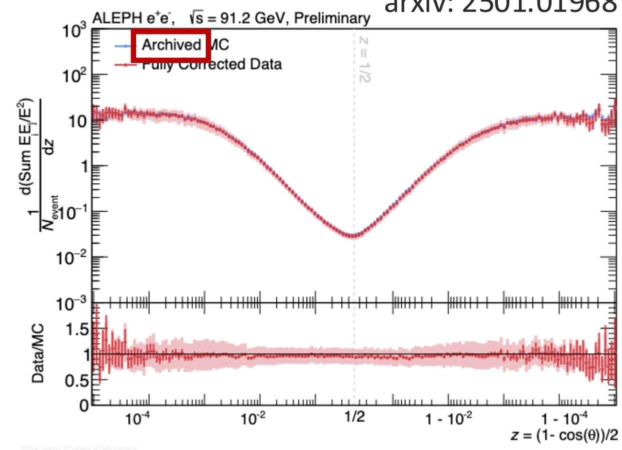
Archived e^+e^- data are being reanalyzed

- Cleanest QCD environment from Z pole: jet spectra and substructure, energy correlators, advanced flavour tagging and ML-assisted unbinned unfolding
- Growing interest from theory community with modern interpretation and up-to-date simulation

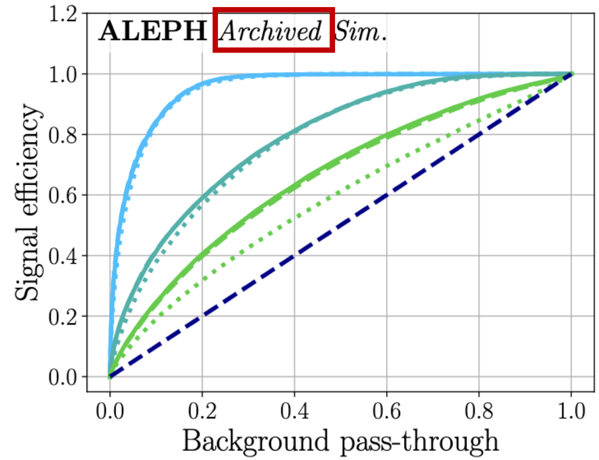
…with no modern simulation to support them

- Archives are fixed; the original Geant3 simulation and reconstruction chains are hard or impossible to run today
- No modern full simulation exists for these detectors

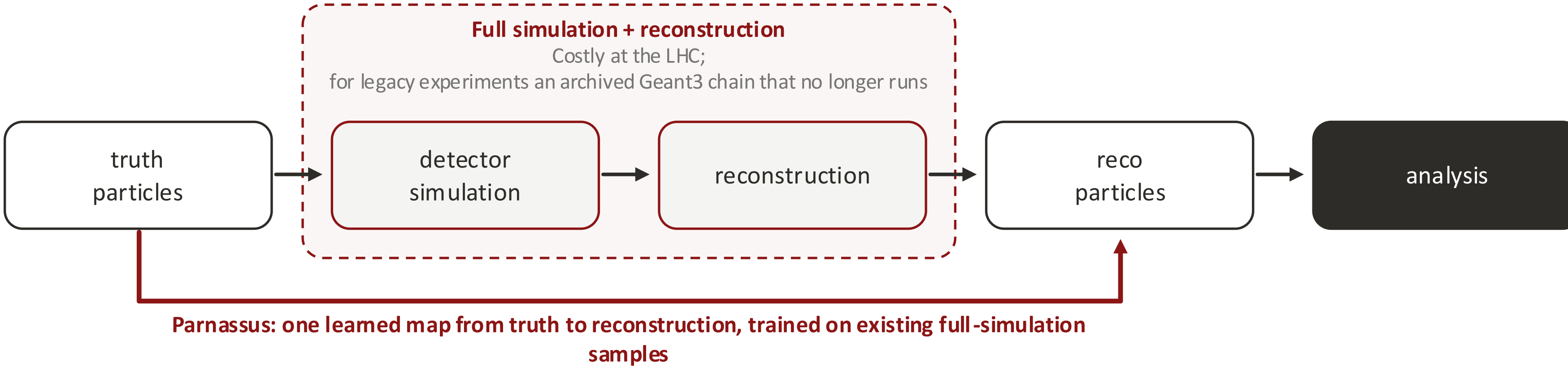
arxiv: 2501.01968



arxiv: 2603.06524

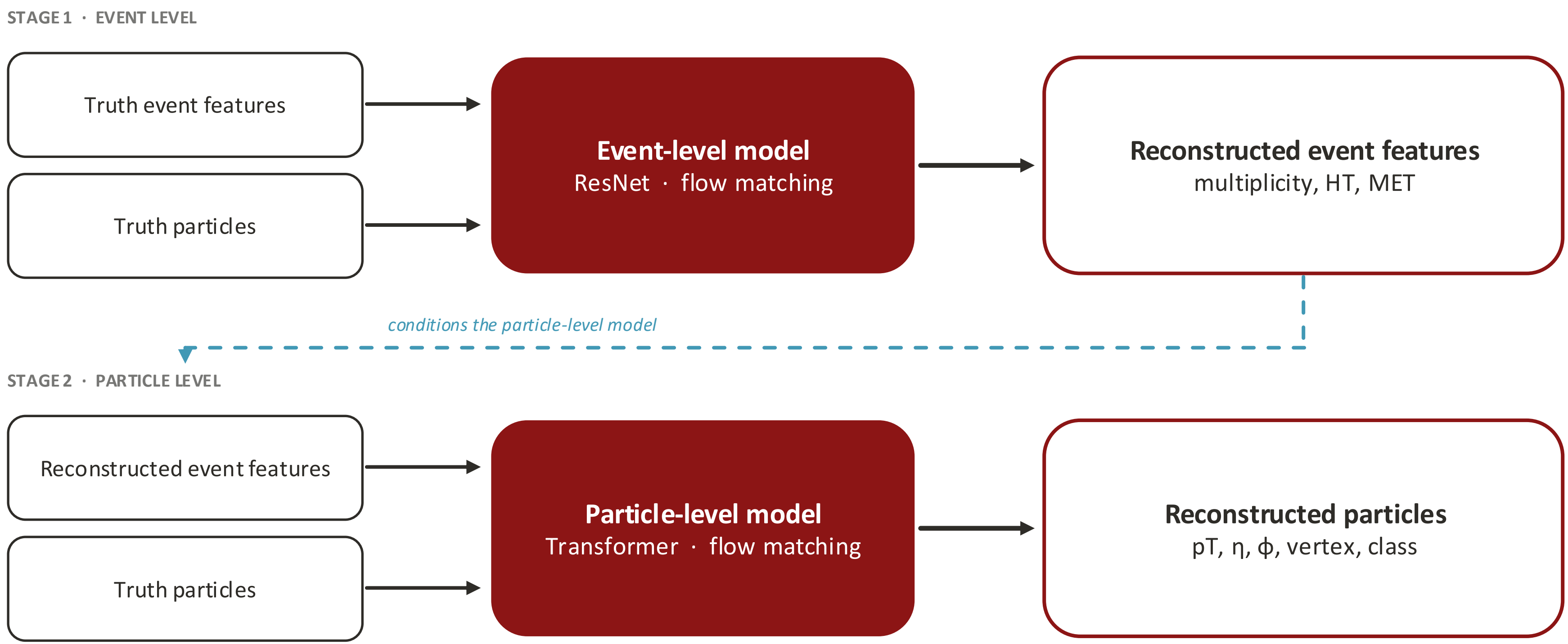


Parnassus detector simulation with ML



- Full detector simulation and reconstruction are computationally expensive for current experiments and may no longer be runnable for preserved experiments
- Parnassus learns the mapping from truth-level particles directly to reconstructed particles using existing full-simulation samples
  - Particle-level output allows a legacy reanalysis to compute its own observables, like new jet definitions
- The same software framework can be used across experiments; the neural network automatically learns each detector model

Parnassus: two generative flow-matching networks



Why legacy e^+e^- collider data with Parnassus?

Archived e^+e^- data are being reanalyzed

- Cleanest QCD environment from Z pole: jet spectra and substructure, energy correlators, advanced flavour tagging and ML-assisted unbinned unfolding;
- Growing interest from theory community with modern interpretation and up-to-date simulation

…with no modern simulation to support them

- Archives are fixed; the original Geant3 simulation and reconstruction chains are hard or impossible to run today
- No modern full simulation exists for these detectors

What Parnassus can unlock here

- Using the existing archived samples, we only need to extract gen/reco-level information to build the fast-simulation samples
- Generative Model with neural network preserves the multidimensional correlation faithfully
- Able to change the generator or tune behind the archived detector response
  - e.g. run Pythia 8, Herwig, or Sherpa through ALEPH, which the archived samples cannot do
  - Produce larger samples with more statistics
  - Simulate processes and signal hypotheses that were never part of the original production for BSM

Parnassus deployments so far

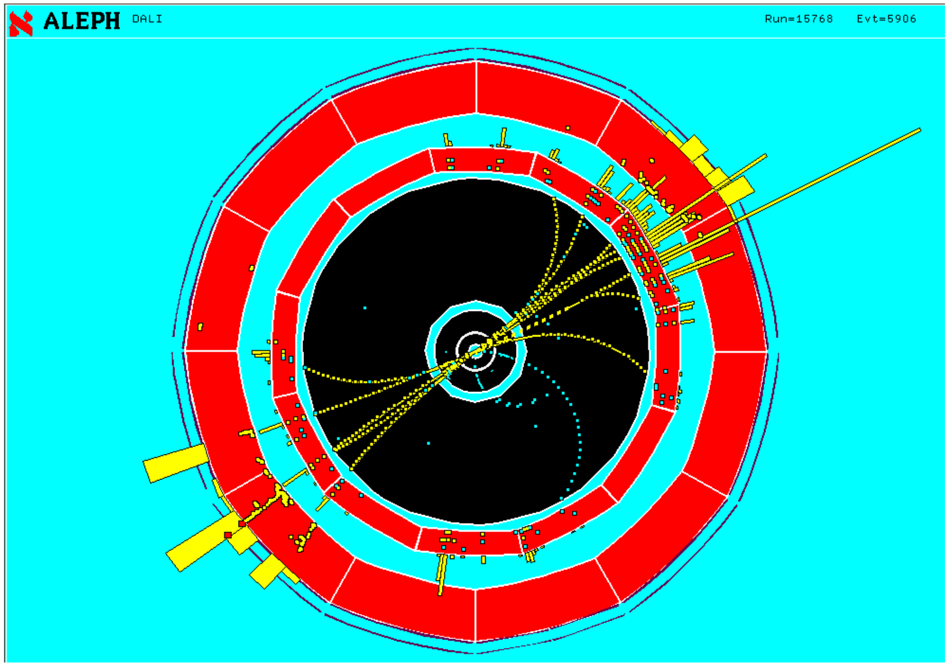
Experiment	Sample	Training events	Reference
CMS	QCD, ttbar H→4l (test-only)	3M	arXiv:2406.01620 arXiv:2503.19981
ATLAS	dijet	4M	CHEP 2026
ALEPH	Z→qqbar	1M	arXiv:2604.11834
SLD	Z→qqbar	**0.2M**	Preprint to appear

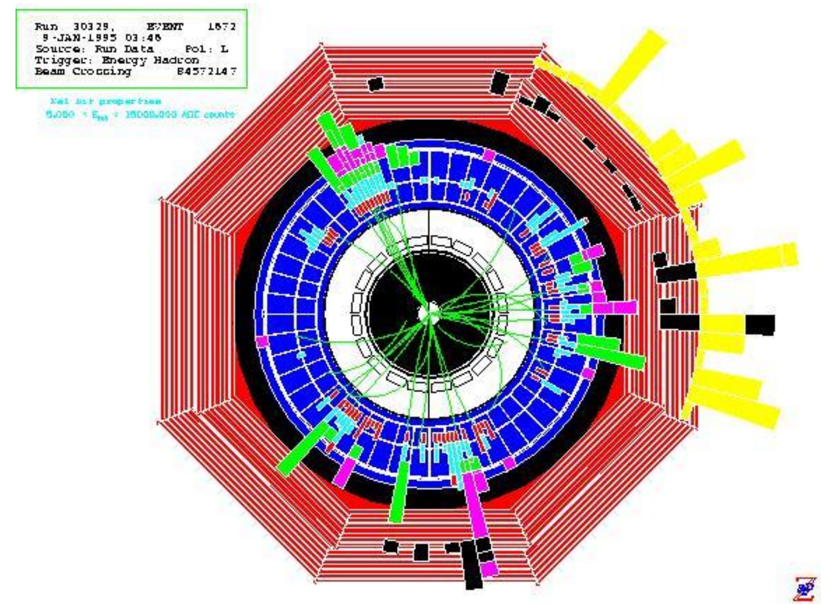
This talk

- One common Parnassus software package and architecture
- Separate pretrained models for different experiments and detector configurations

Issues here?

Today's focus: ALEPH and SLD

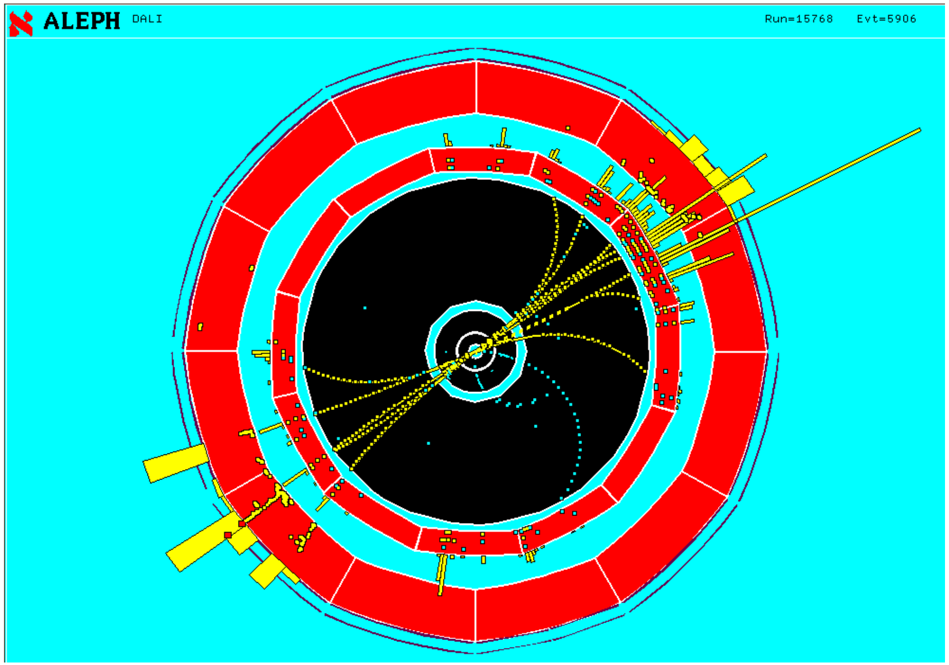




- ALEPH: one of the 4 experiments at LEP *(this talk focus on LEP-I era)*

- SLD: unique polarized e+ e- machine at Z pole

Part I — ALEPH model



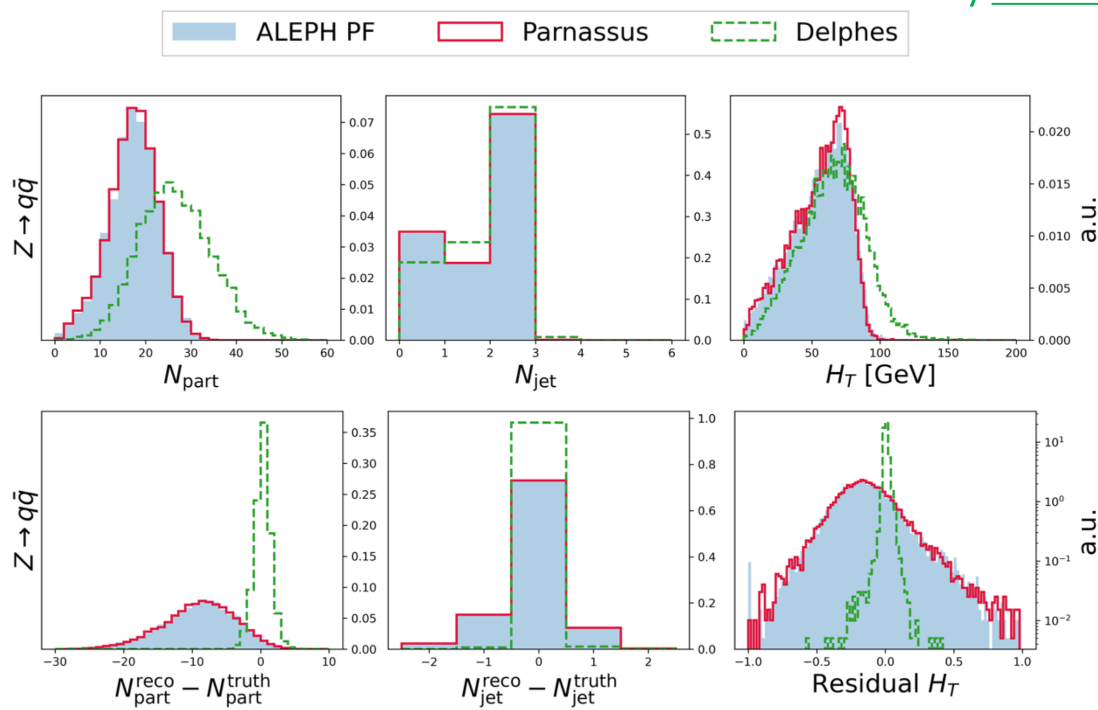
Sample

- Archived Geant3 simulation of $e^+e^- \to Z \to q\bar{q}$ at $\sqrt{s} \approx 91.2$ GeV, in EDM4hep (Thanks Sanghyun Ko for providing samples information!)
- 1,005,986 events: 805k train / 144k validation / 57k test

Parnassus reproduces the ALEPH reference accurately across a range of event-, jet-, and particle-level observables

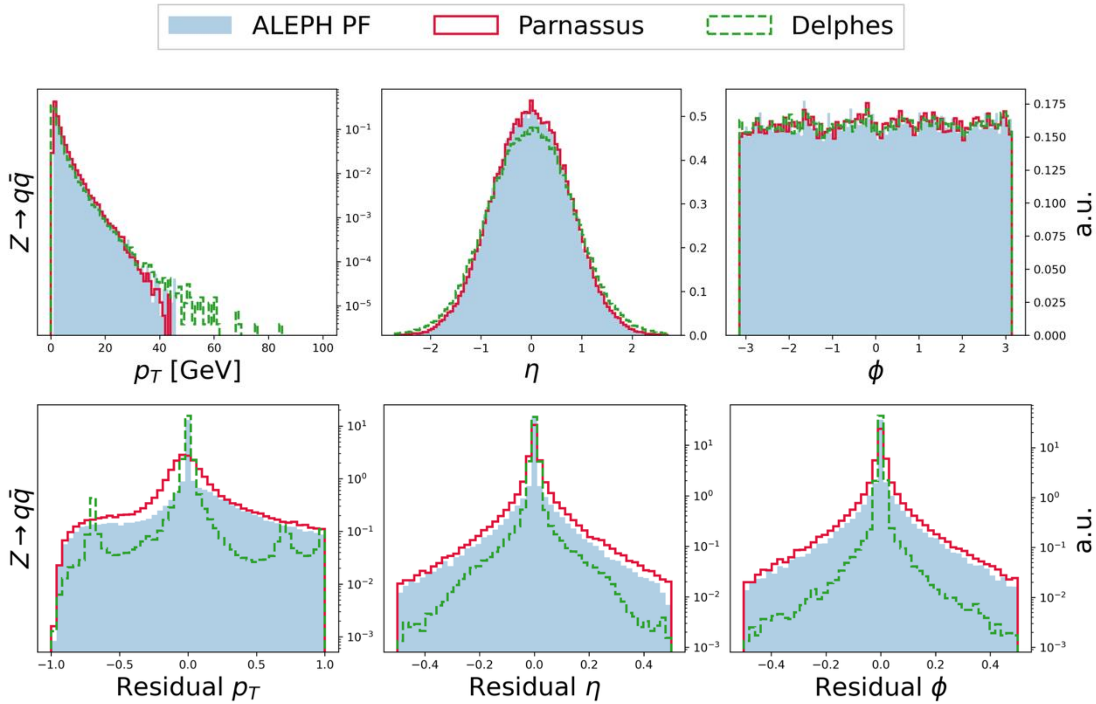
ALEPH — event-level observables

Produced by official Delphes ALEPH card



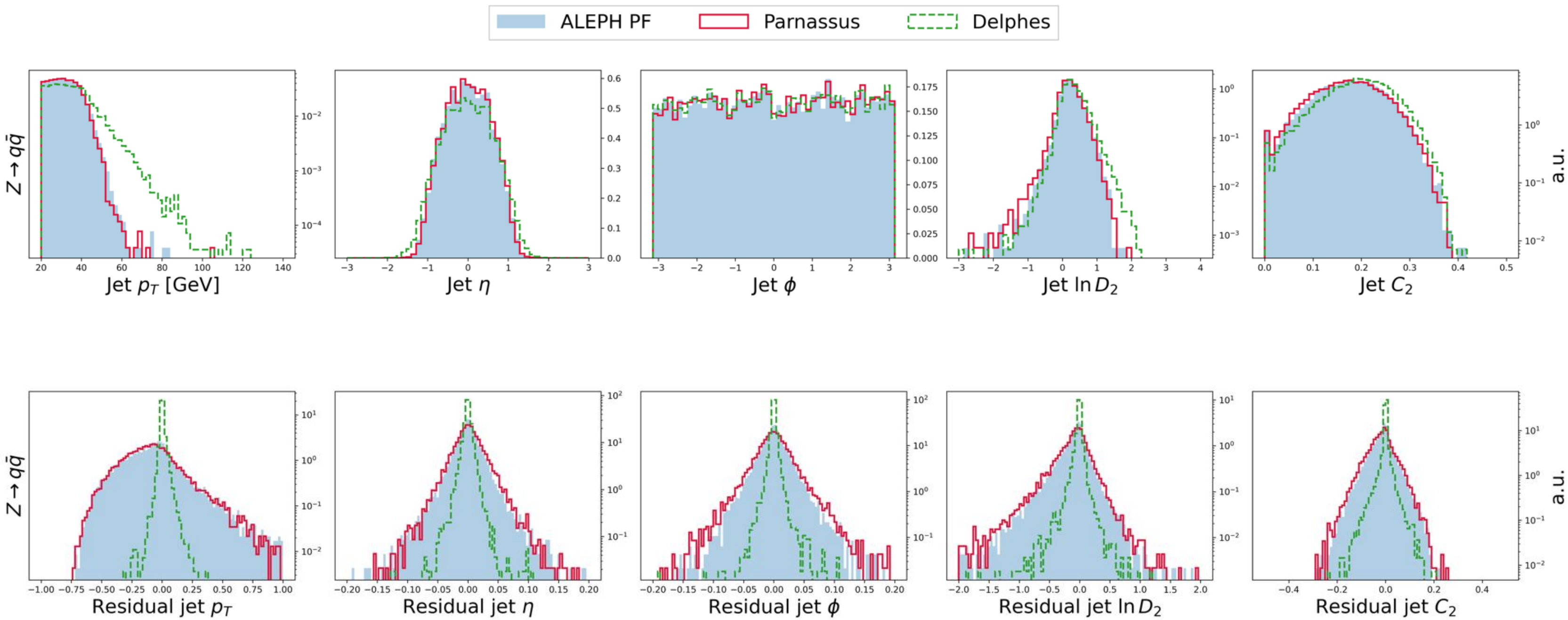
Multiplicity, jet count and H_T in both the marginals and the reco–truth residuals match well with reference

ALEPH — particle-level response



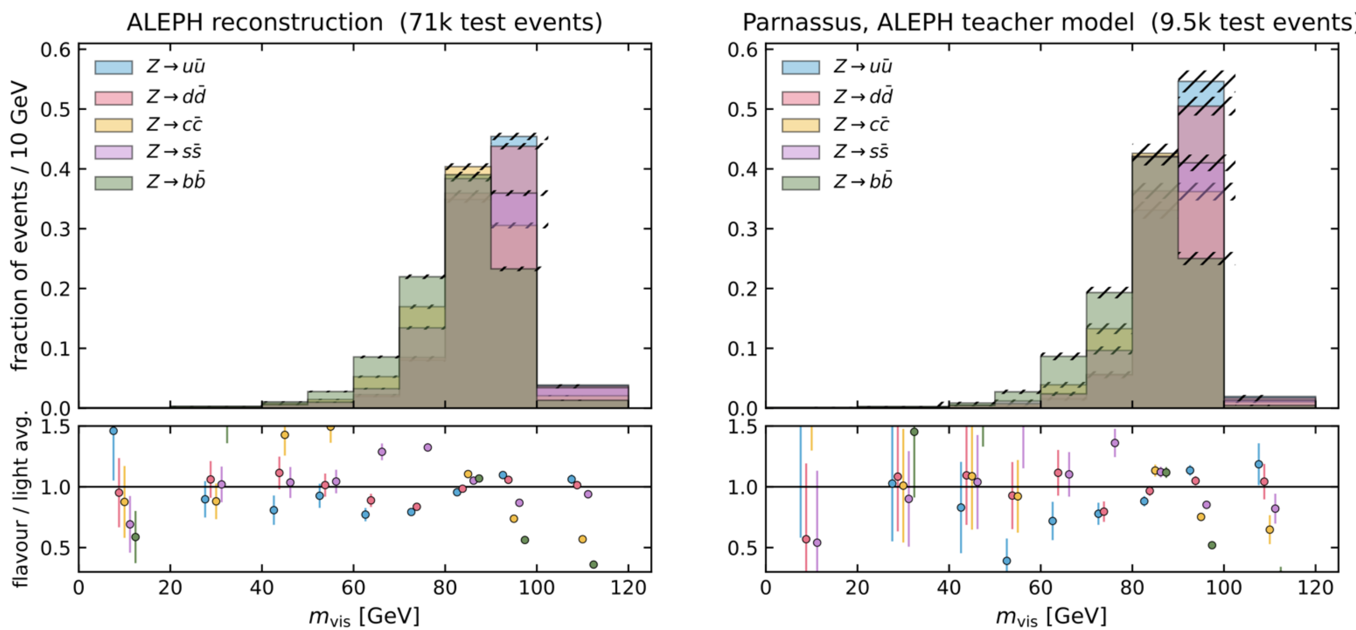
Particle pT, η, ϕ marginals agree with the reference

ALEPH — jet kinematics and substructure



Jet pт, η, ϕ and the substructure variables ln D_2 and C_2 are reproduced well
Low-level particles propagate information well to higher-level objects

ALEPH — flavour information preserve



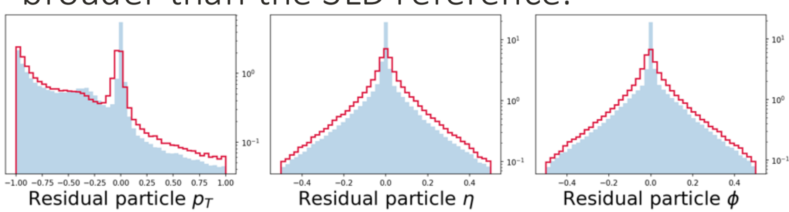
Visible mass distribution and the flavour composition follow the ALEPH reconstruction within the statistical uncertainties

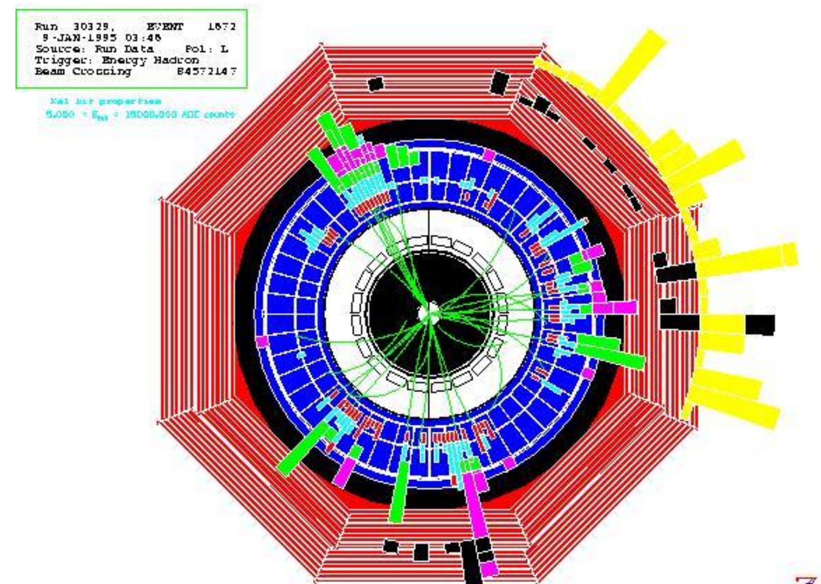
Part II — SLD model

Sample

- Archived SLD Monte Carlo simulation (provided from Alkaid Cheng) of $e^+e^- \to Z \to q\bar{q}$ at $\sqrt{s} \approx 91.2$ GeV
- 260,266 events: 212.5k train / 37.8k validation / 10k test

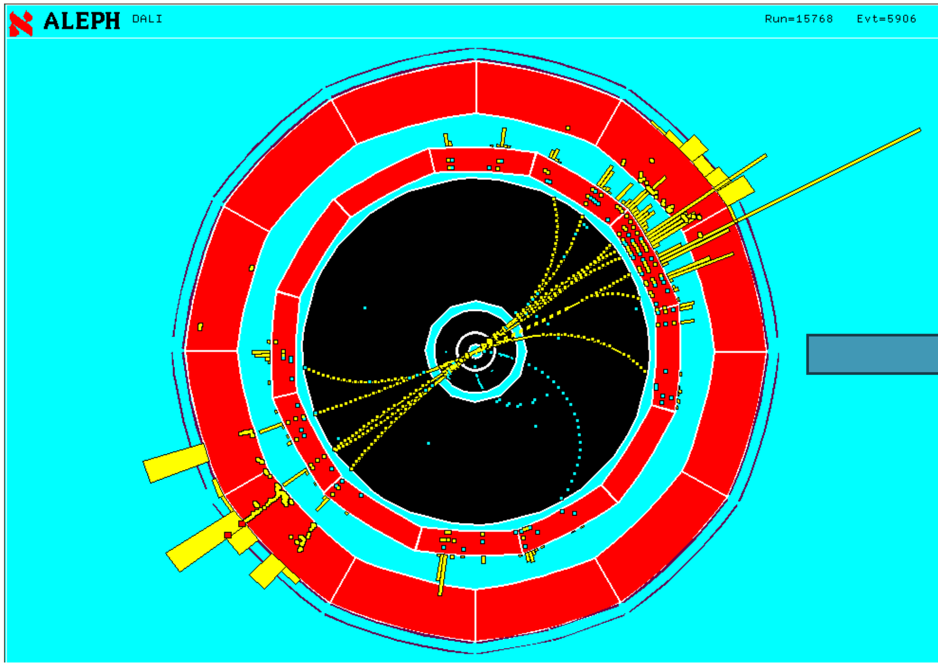
Direct training reproduces the reconstructed distributions, but the per-particle response remains substantially broader than the SLD reference.

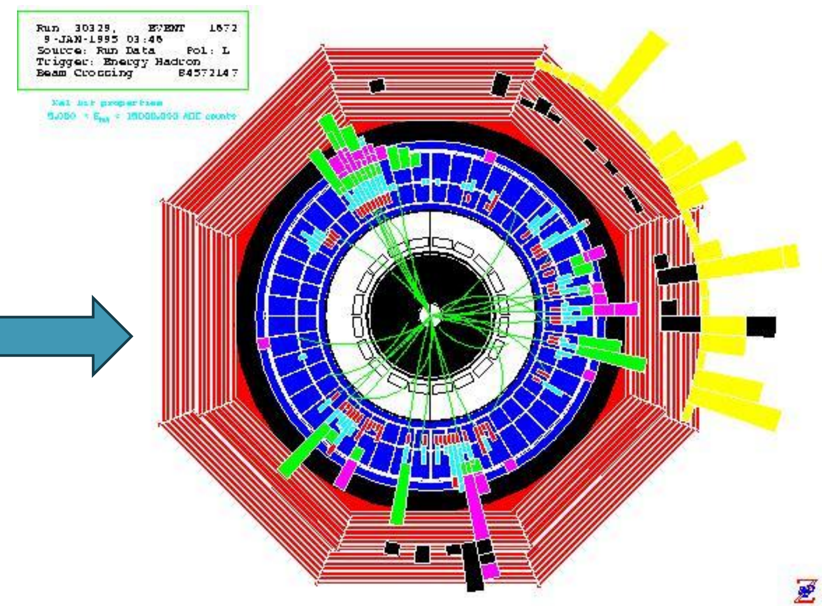




Direct SLD train struggled from low statistics, how to improve?

Part III — fine-tune SLD model from ALEPH





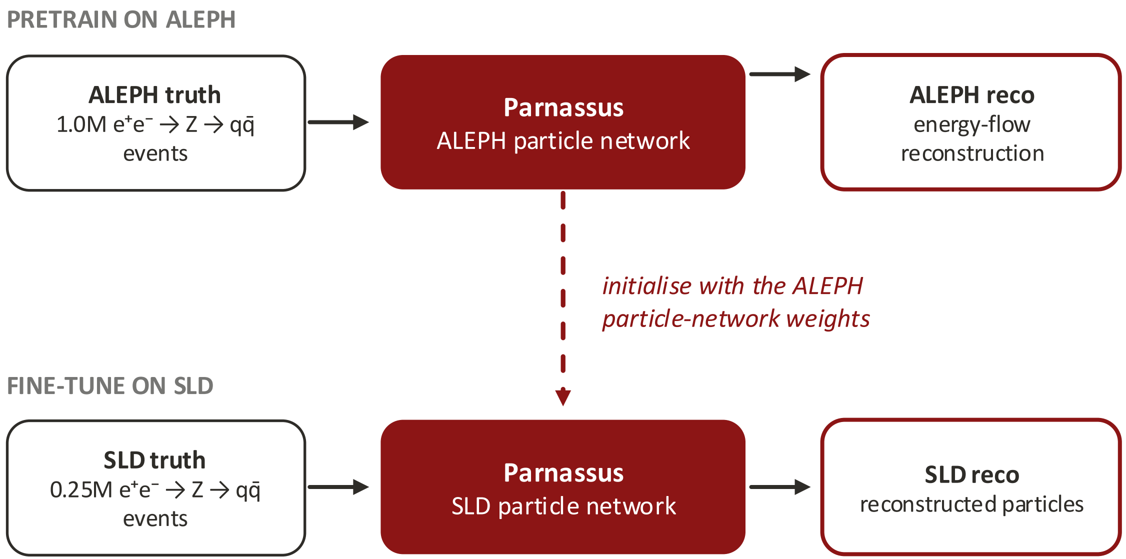
1M events available

0.2M events available

Direct SLD train struggled from low statistics, how to improve?

Transfer learning from another e+ e- model?!

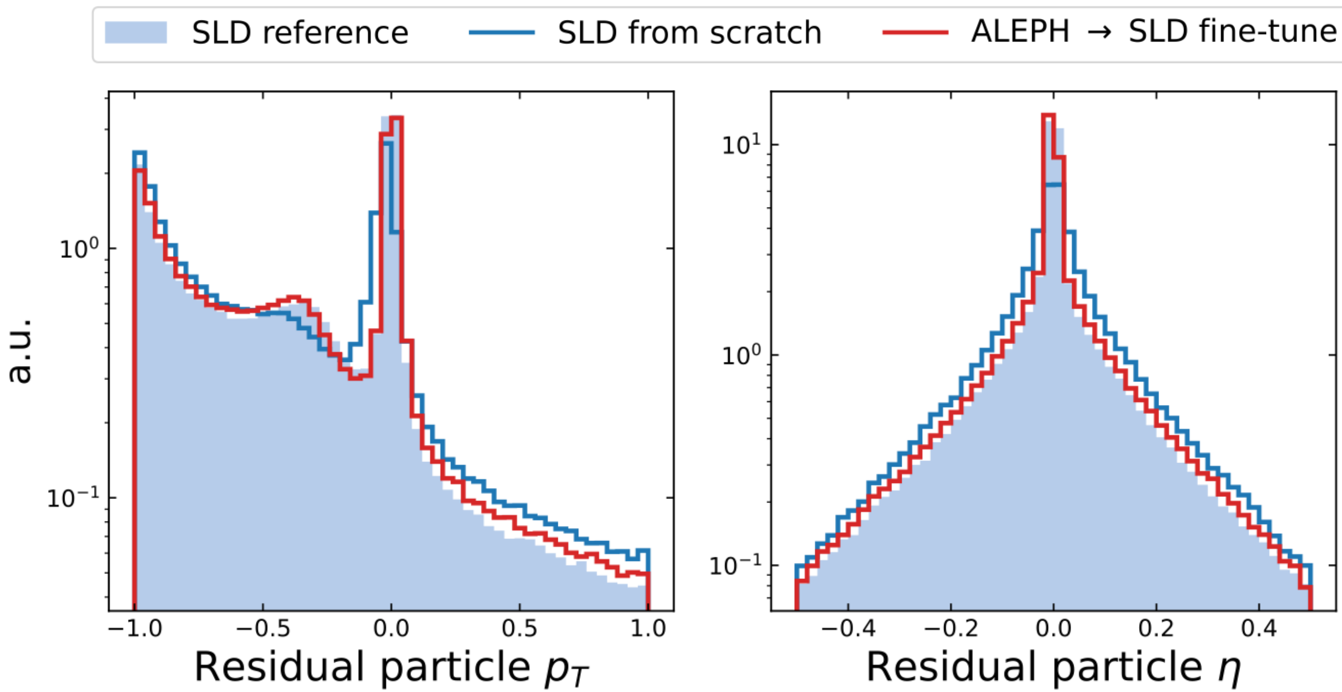
Part III — fine-tune SLD model from ALEPH



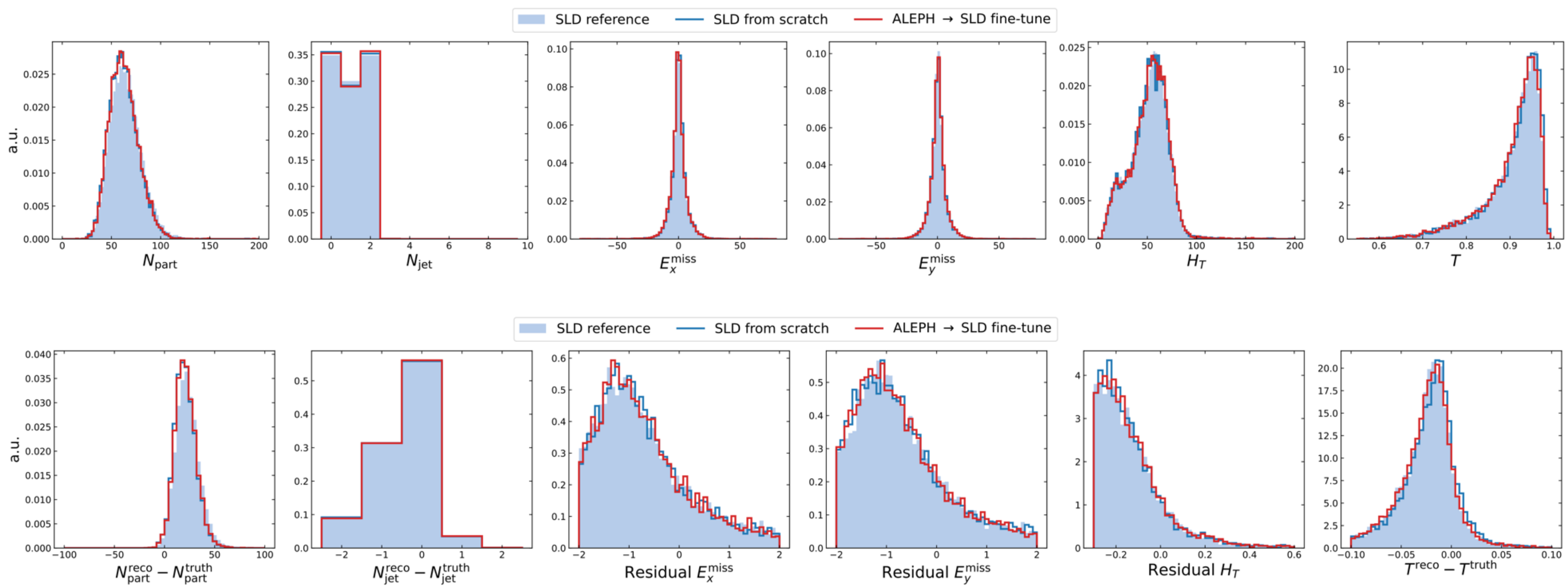
- Same physics domain: hadronic Z decays at $\sqrt{s} \approx 91.2$ GeV
- Different detector, reconstruction, archive representation, and sample size
- Initialize the SLD particle network with the parameters learned on ALEPH

Results — Fine-tuning recovers the SLD particle response

- Same SLD dataset, *same* pipeline across different strategies, only the initialization changes
- Δη width and particle pT response improve

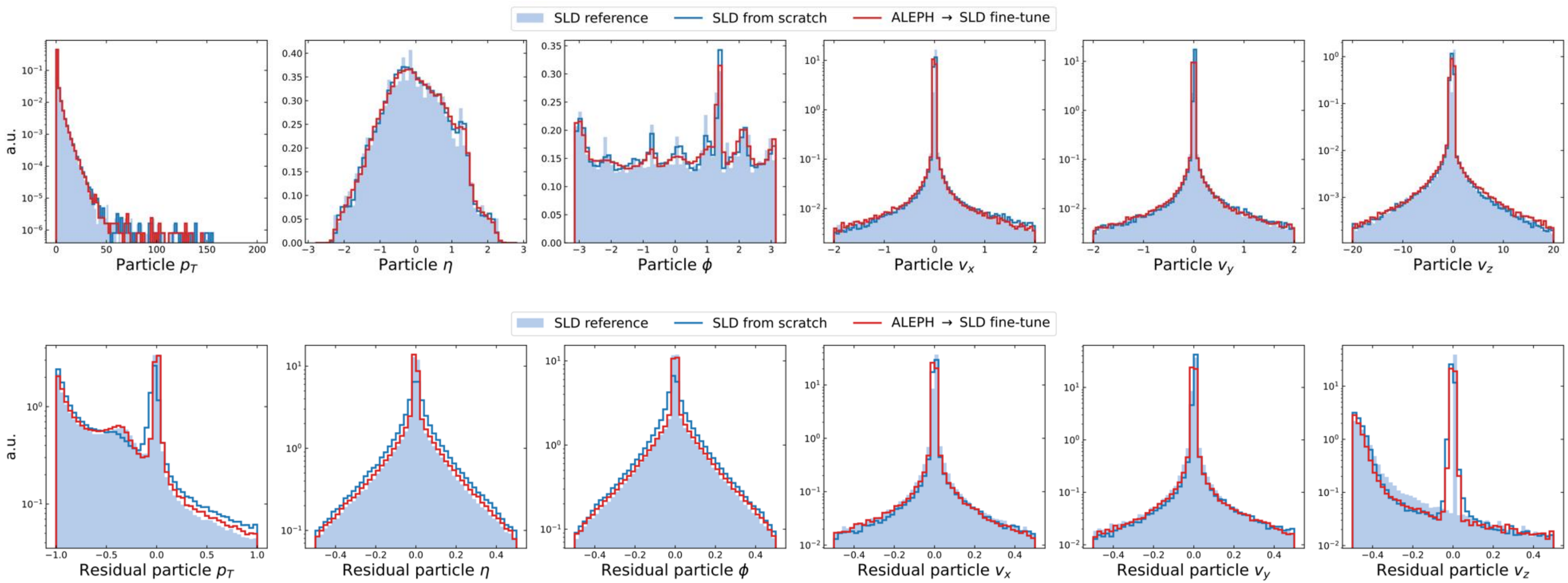


Results — Event-level performance



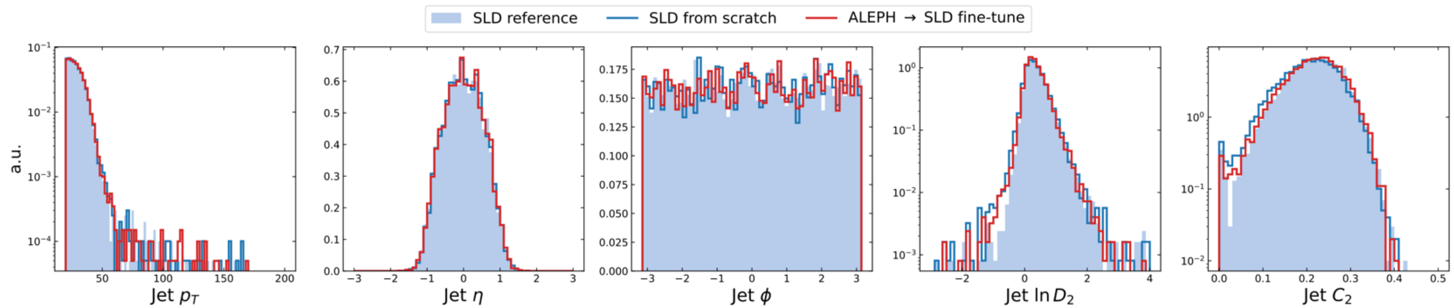
Event-level quantities, multiplicity, energy sums, Hᴛ, and their residuals remain unchanged.

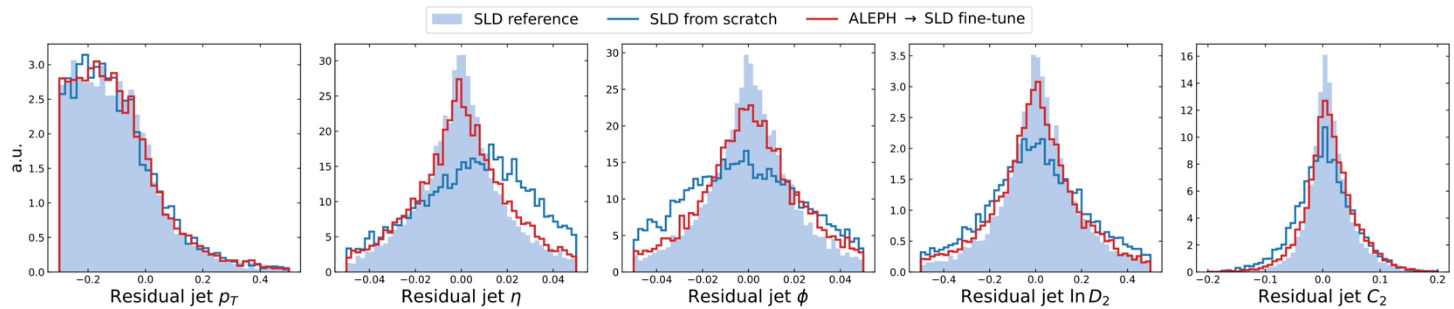
Results — particle-level response



Fine-tuning brings the residuals back to the reference.

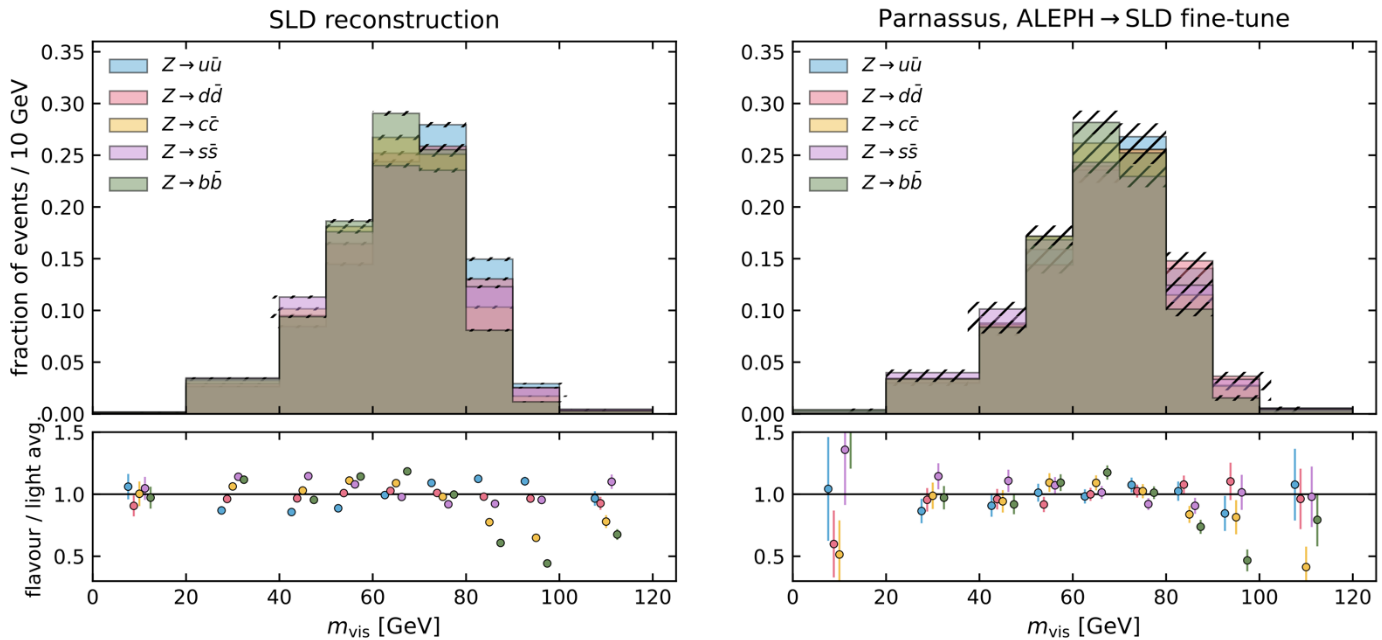
Results — jet kinematics and substructure





Jet residuals of the fine-tuned model are visibly closer to the SLD reference:
the better particle response propagates to jets.

SLD — flavour information preserve

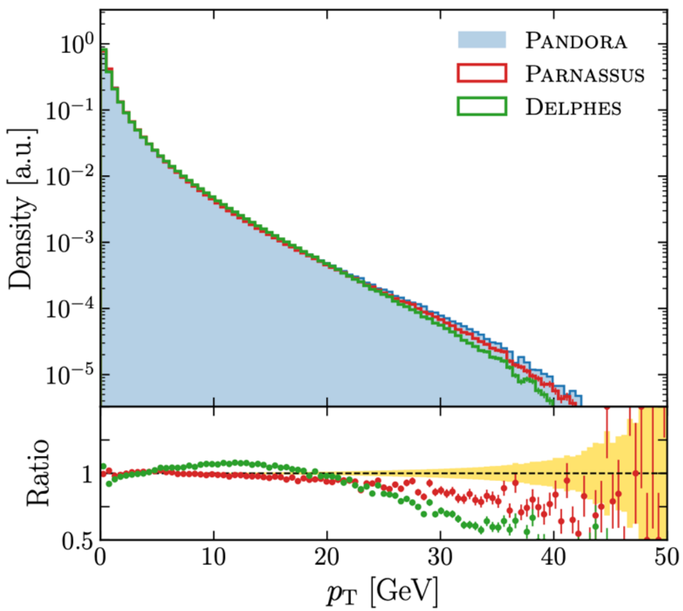


Visible mass distribution and the flavour composition follow the SLD reconstruction within the statistical uncertainties

Part 3.5: Can we transfer to other detectors?

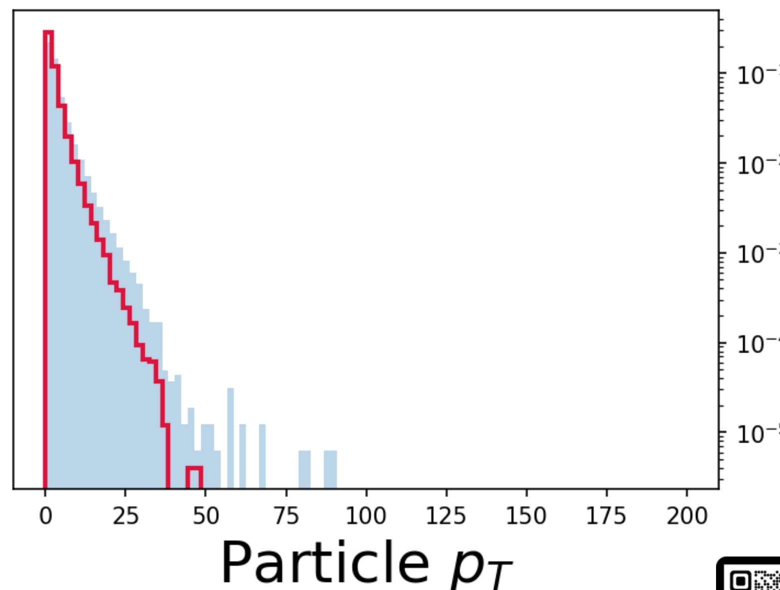
Preliminary direct training results from other experiments

FCC - CLD



From Umar Sohail Qureshi

OPAL



With help from Jingyu Zhang and Sitian Qian

These results motivate a foundation model for e^+e^- detectors.

Conclusion and outlook

- ALEPH: Parnassus reproduces a legacy e^+e^- detector at event, jet, and particle level from archived simulation alone
- SLD: with only 250k archived events, training from scratch misses the per-particle response. However, initializing from the ALEPH weights recovers it.
- The agreement holds for higher-level observables so we are confident Parnassus can support physics analyses

Outlook

- These results demonstrate that Parnassus could transfer across detectors
- Build a foundation model for e^+e^- detectors that transfers across experiments

Parnassus framework: https://parnassus-hep.github.io/parnassus/

- Etienne Dreyer (Weizmann)
- Abdel Elabd (U. Washington)
- Dan Godi (Weizmann)
- Hamsa Hanif (Simon Fraser)
- Eilam Gross (Weizmann)
- Tyler Hellstern (Duke/Stanford)
- Dmitrii Kobylianskii (Weizmann)
- Runze Li (Yale/Stanford)
- Ya-Feng (Celia) Lo (Stanford/SLAC)
- Vinicius Mikuni (KMI)
- Benjamin Nachman (Stanford/SLAC)

Thank You