BERKELEY LAB
Bringing Science Solutions to the World

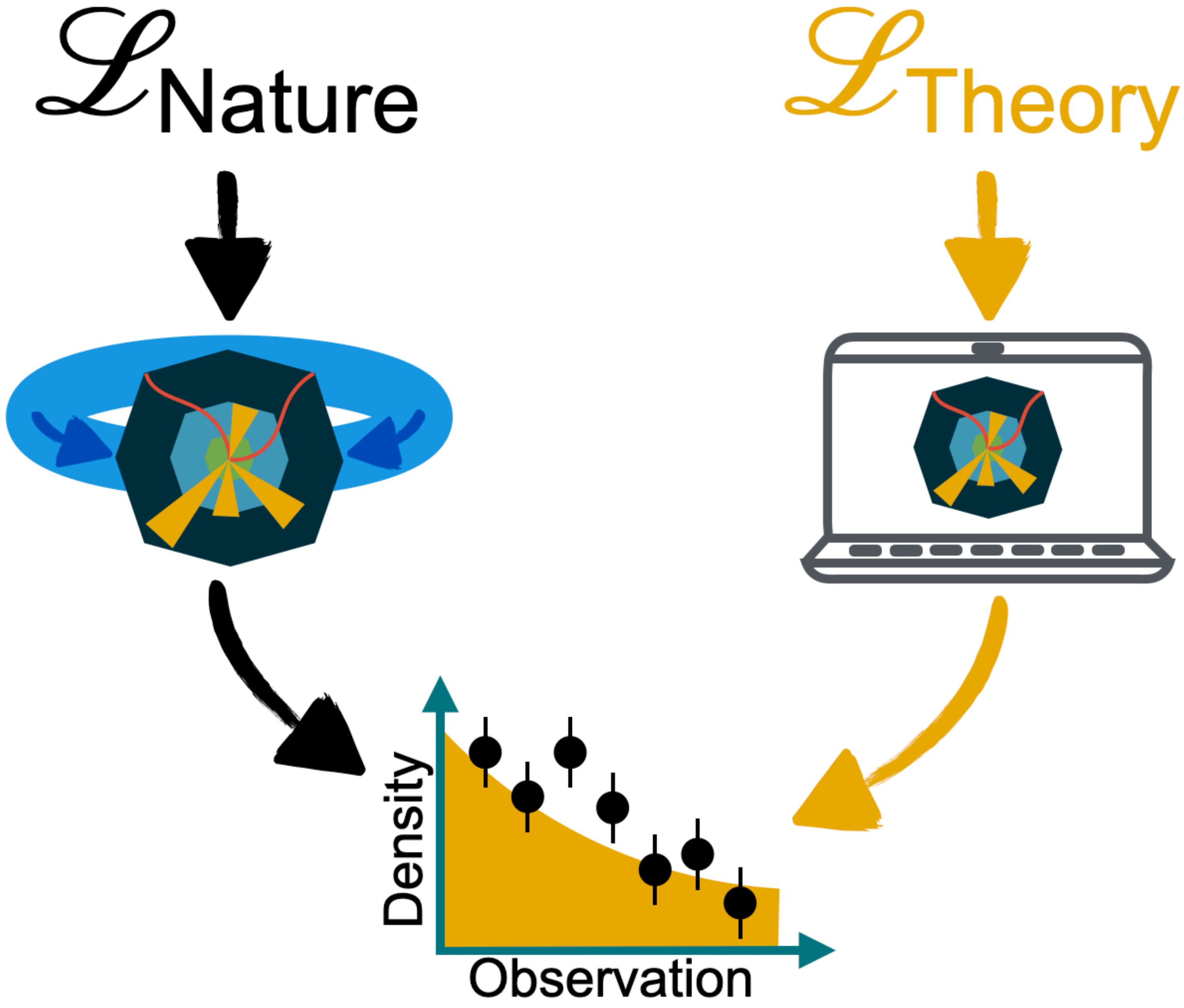
U.S. DEPARTMENT OF ENERGY
Office of Science

Machine Learning for HEP Data Analysis

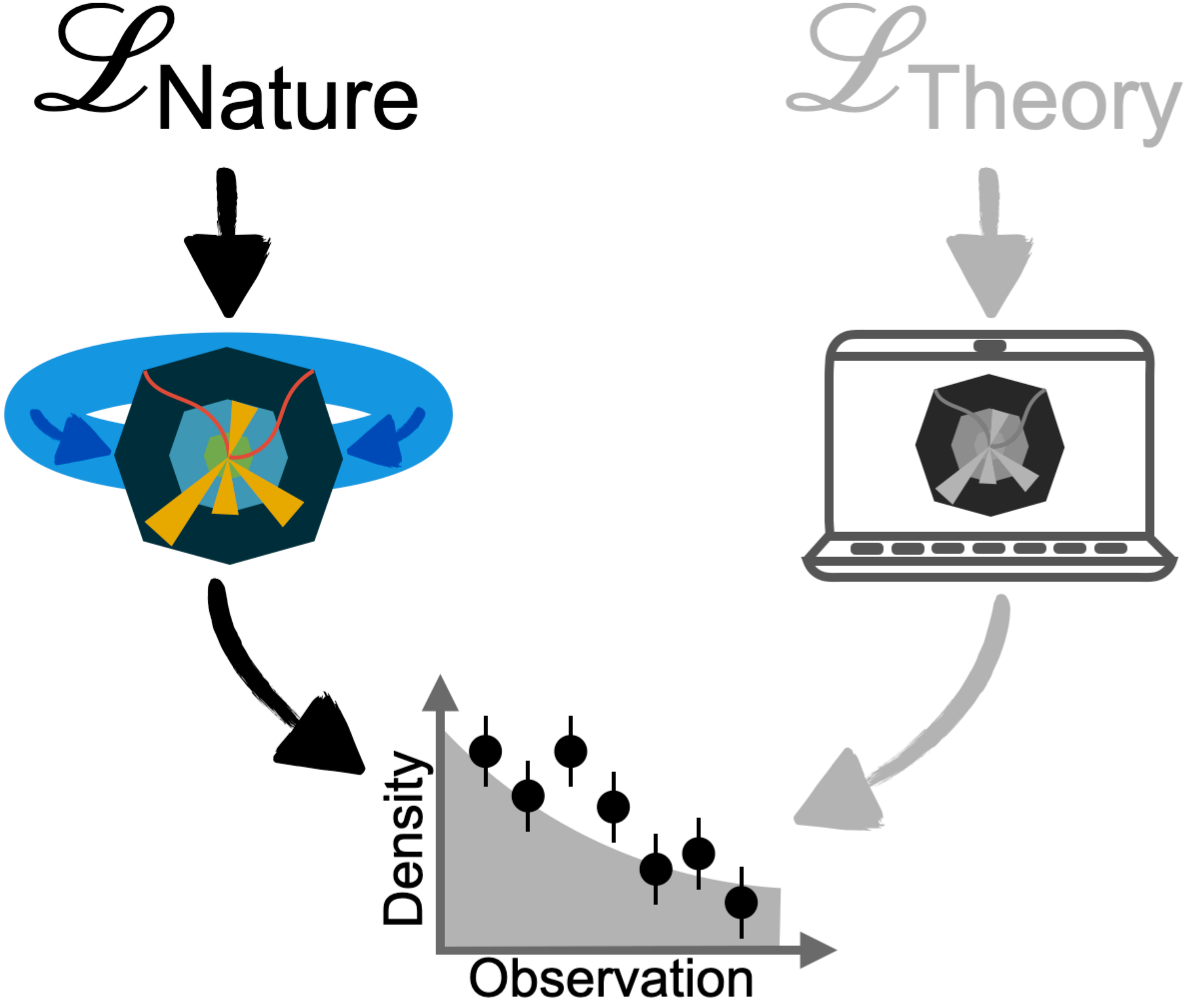
Dennis Noll

FCC Week 2024

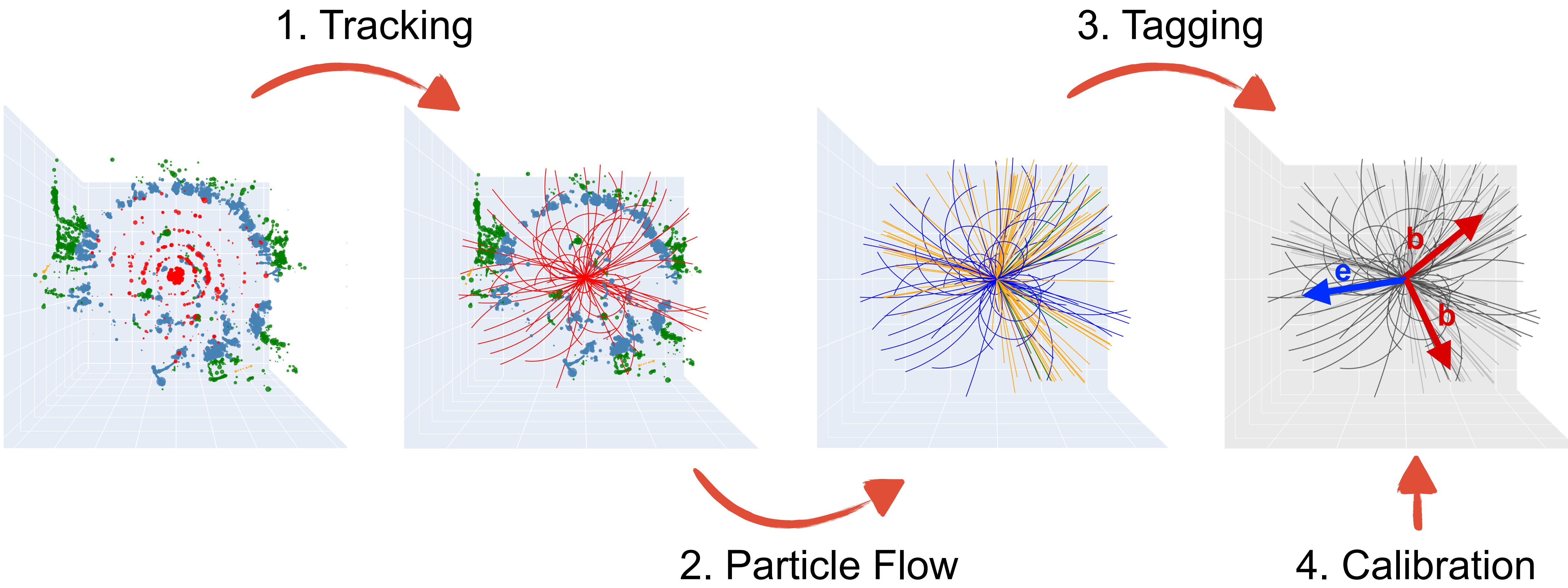
June 13, 2024



Inspired by [1]


$\mathcal{L}_{\text{Nature}}$
$\mathcal{L}_{\text{Theory}}$
Density
Observation

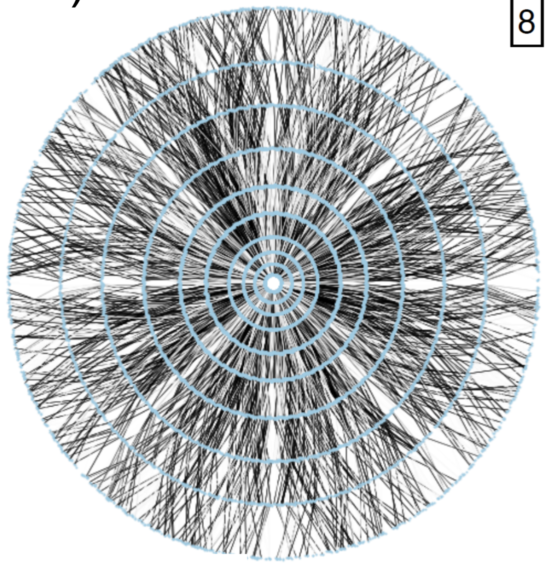
Introduction - Experiment [2]

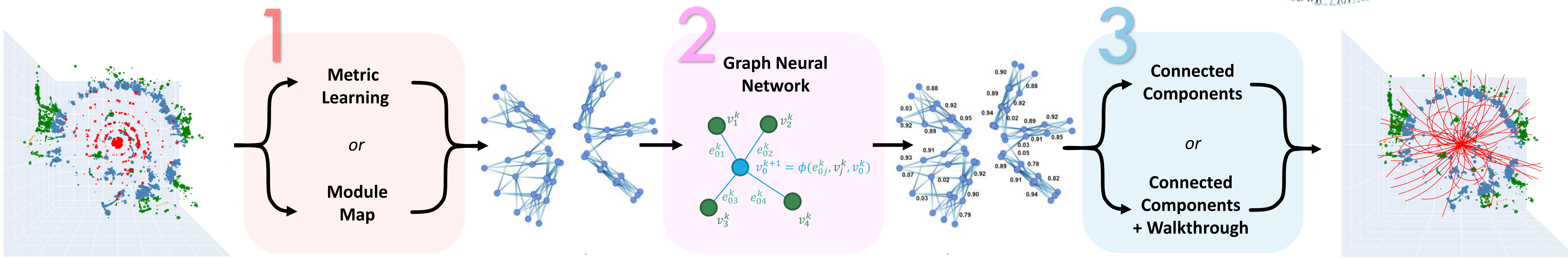


Experiment - 1. Tracking [3]

- Charged particles leave hist in tracker along their path (up to ~5000 per event)
- Turn tracker hits into tracks with graph-based ML
- Using ExaTrkX algorithm:
  1. Construct graph of hits
  2. Label graph edges
  3. Segment graph into tracks

[8]

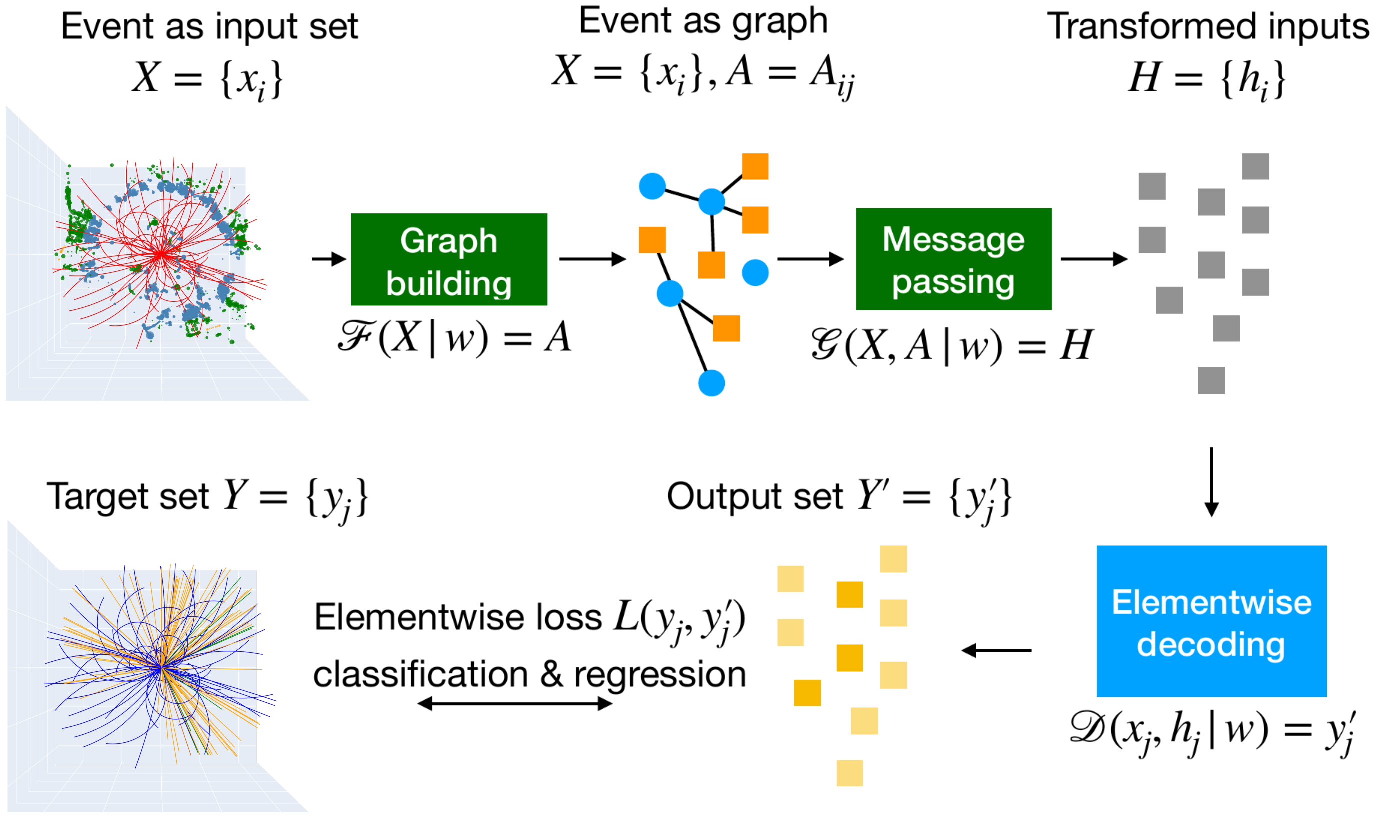




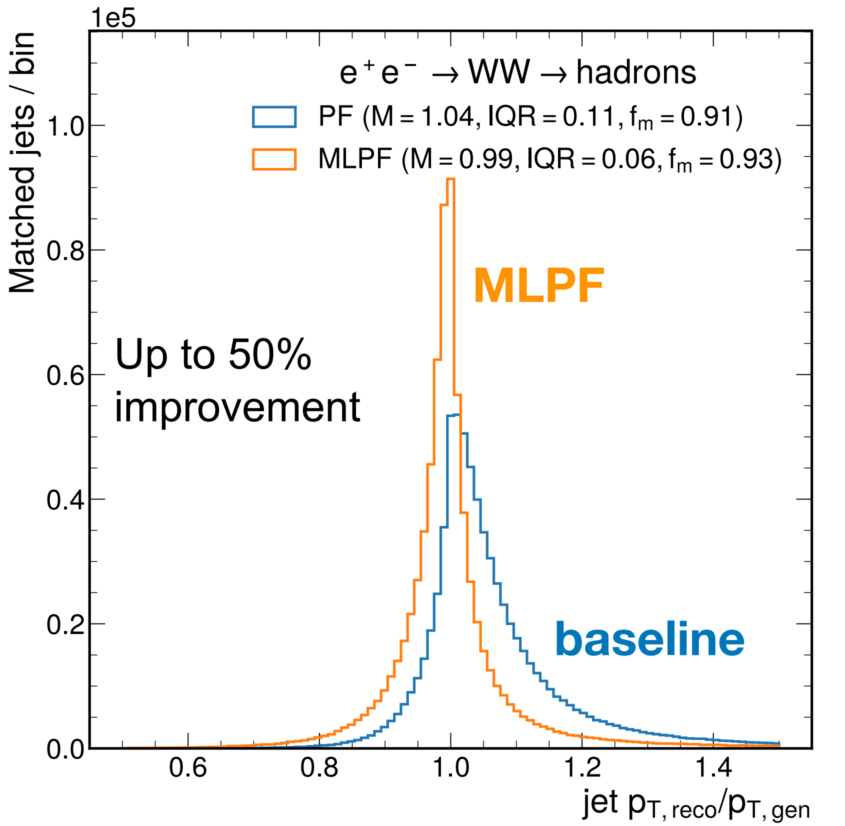
Experiment - 2. Particle Flow [2,4]

- Turn tracks and calorimenter clusters into particles
- Use granular detector layout optimally
- Different graph-based approaches ML approaches exist (MLPF, HGPflow, …)

MLPF Method

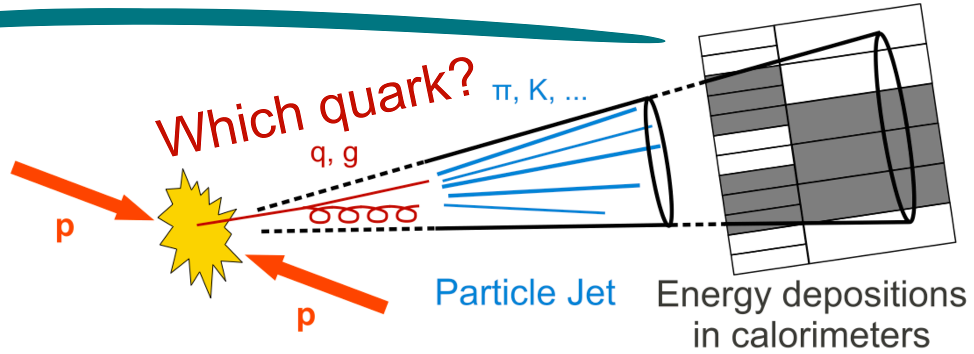


MLPF Result

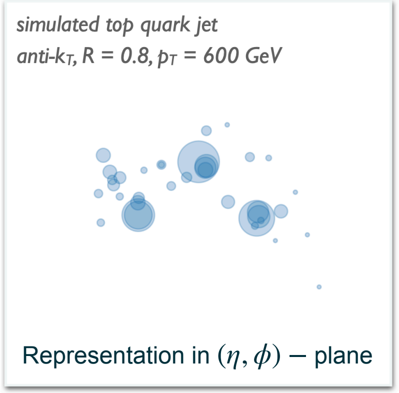


Experiment - 3. Tagging [5]

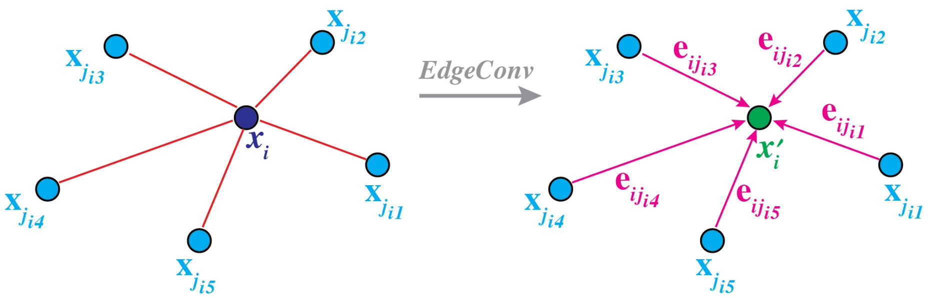
- Quarks from the hard interaction initiate jets in the detector
- Determine which type of quark initiated the jet (tagging)

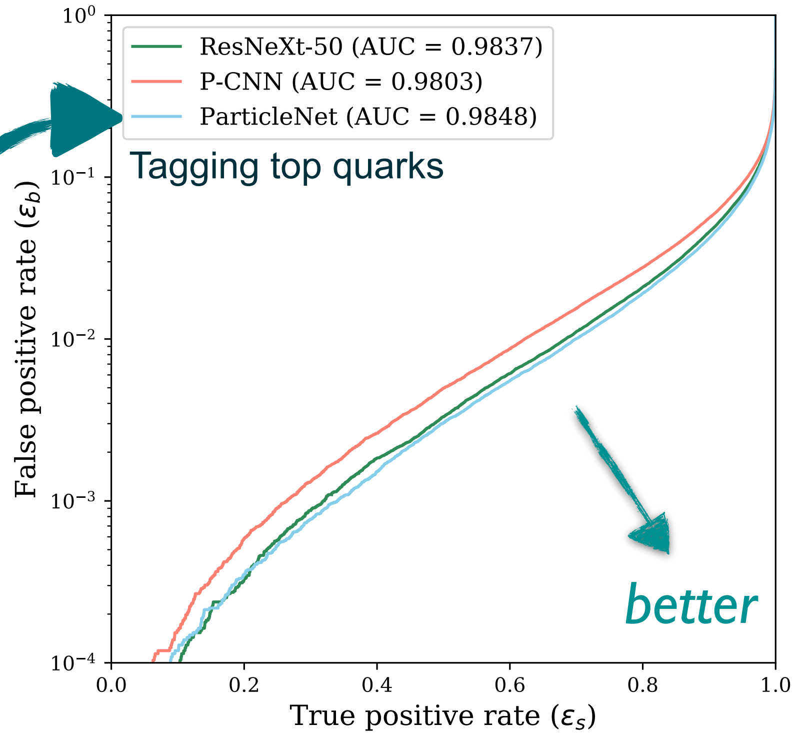


Jet as Point Cloud



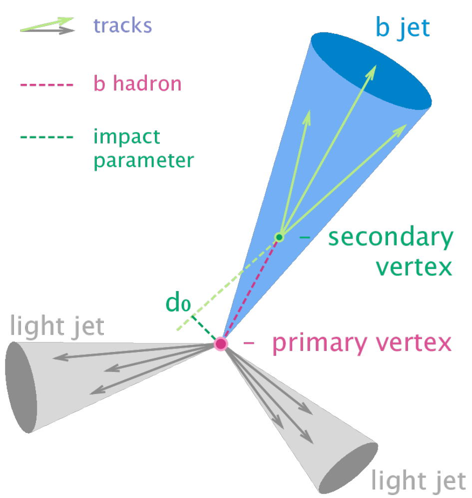
ParticleNet



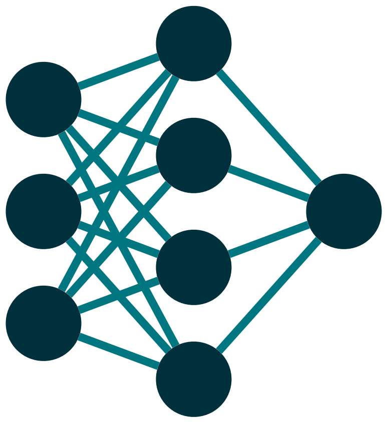


Experiment - 4. Calibration [6]

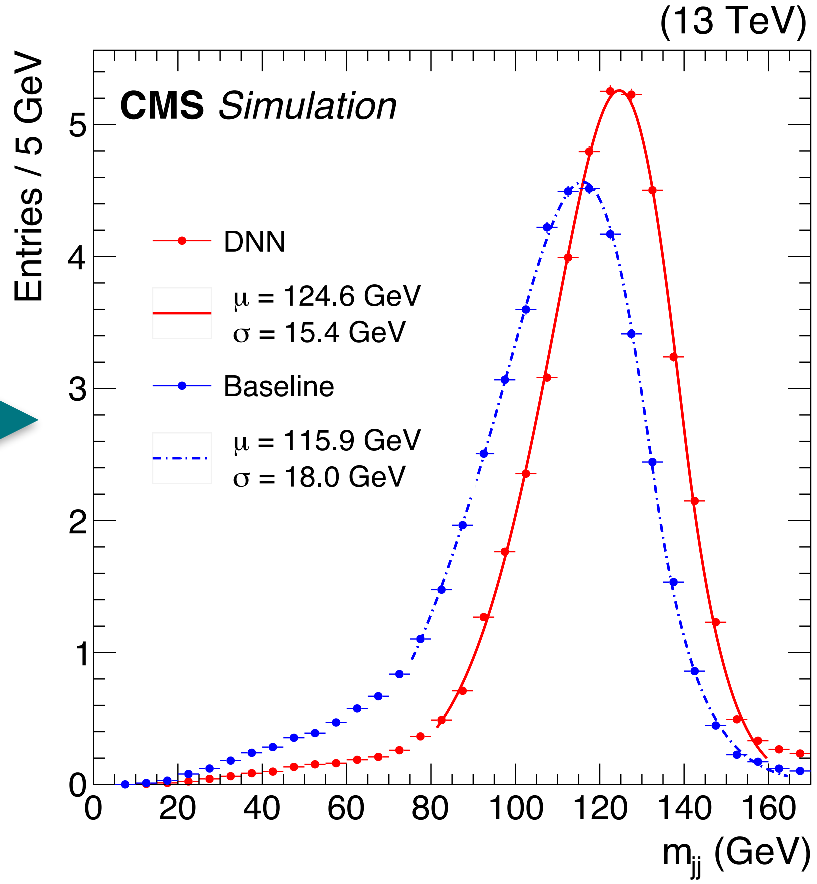
- Jets from b-quarks have large invisible (neutrino) contribution
- Calibrate the momentum of the jets with feed forward DNN regression
- Improved di-jet resolution by ~15% compared to baseline



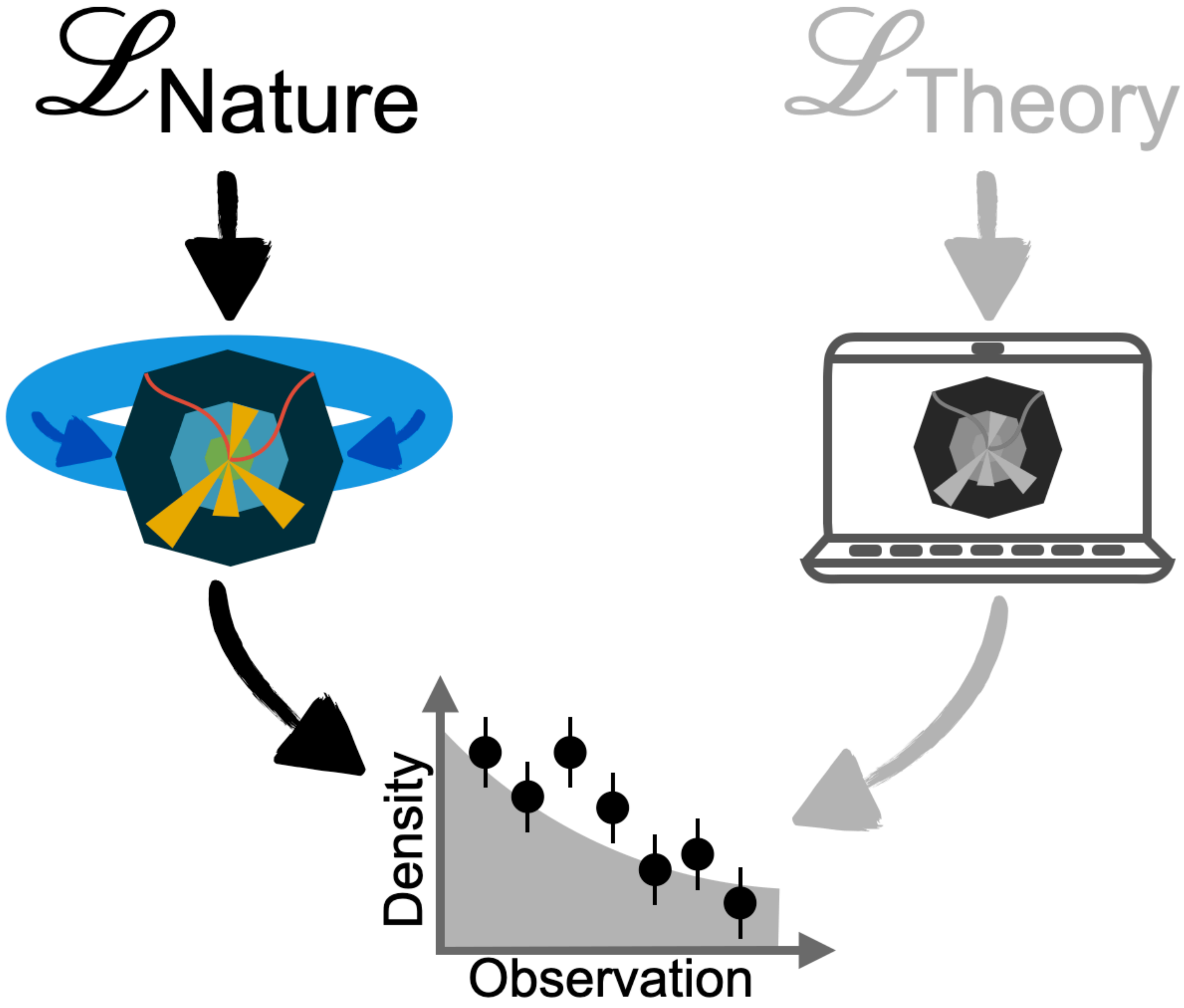
- Jet kinematics
- Jet composition
- PU info



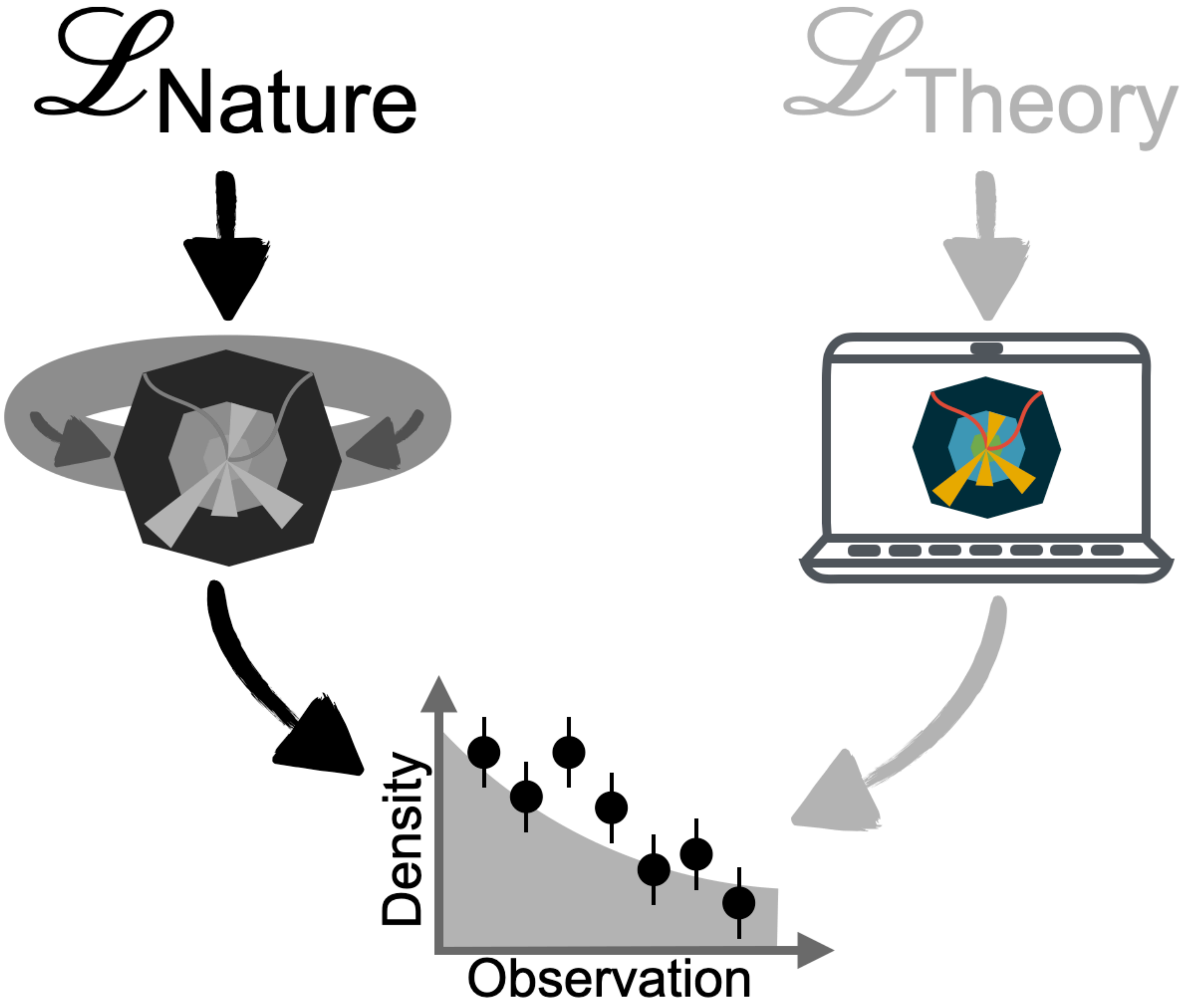
$$\frac{p_{T,true}}{p_{T,reco}}$$



Experiment:
- Tracking
- MLPF
- Tagging
- Calibration

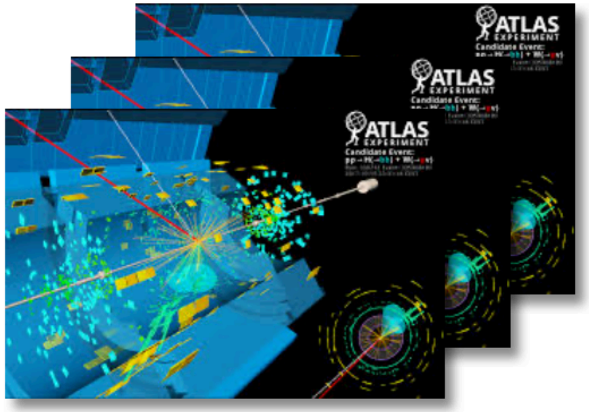


Experiment:
- Tracking
- MLPF
- Tagging
- Calibration



Simulation - Introduction [1]

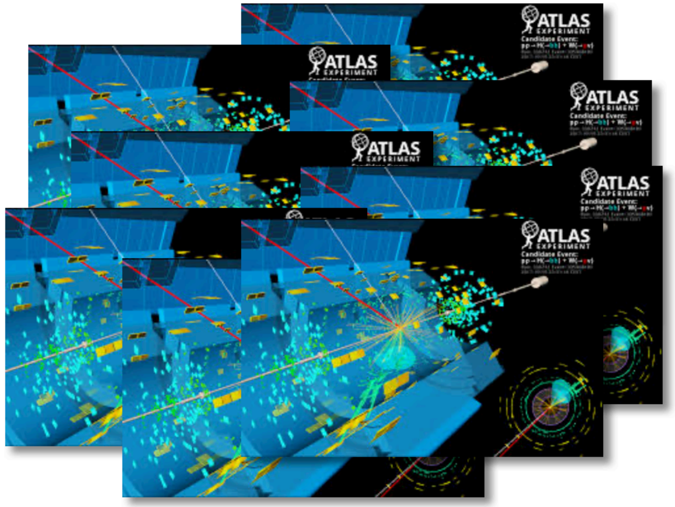
- Experiments spend significant computing budget on simulations
- Can make simulations much more efficient using ML
- **Simulations have different levels (full detector vs. particles), data types, complexities, …**



Simulation /
Recorded Data

Generative
Model

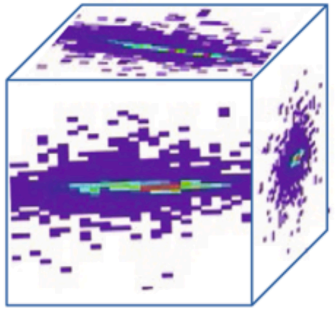
- GANs
- Variational AEs
- Normalizing Flows
- Diffusion Models
- …

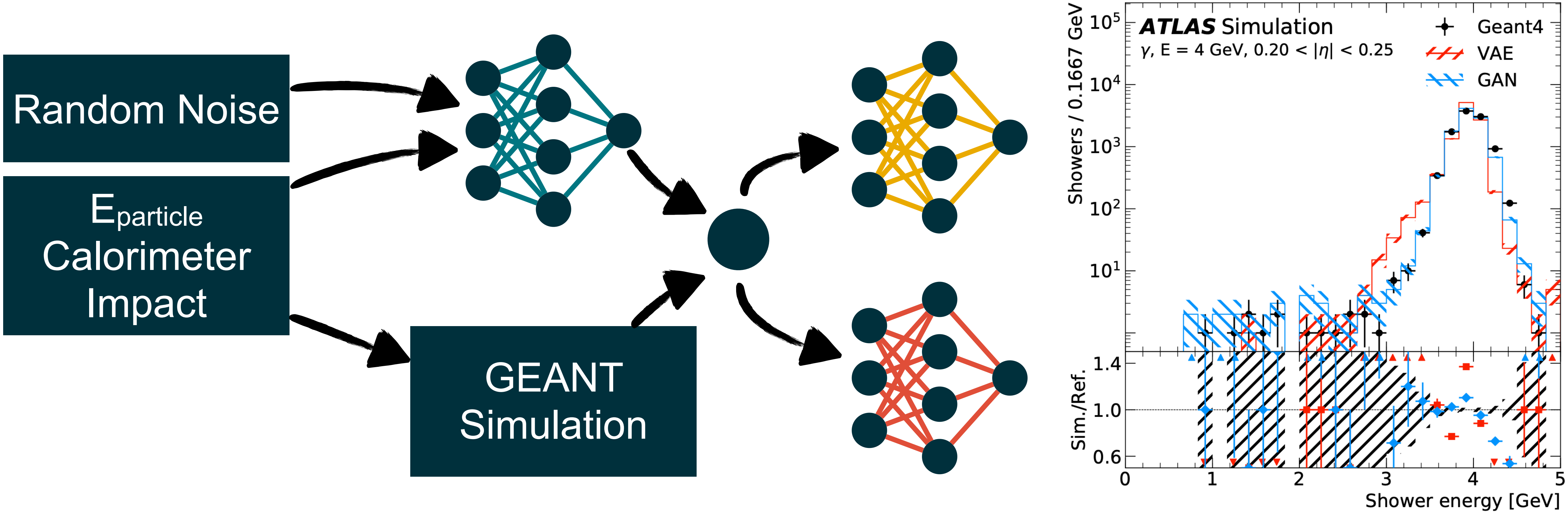


Oversampled

Simulation - Detector Level (Showers) [7]

- Simulate regular spatial shower profiles with ML
- Using generative adversarial networks (GANs)
- Parametrized by particle energy, calorimeter configuration, impact point
- Three networks: Generator, Critic, Energy Critic

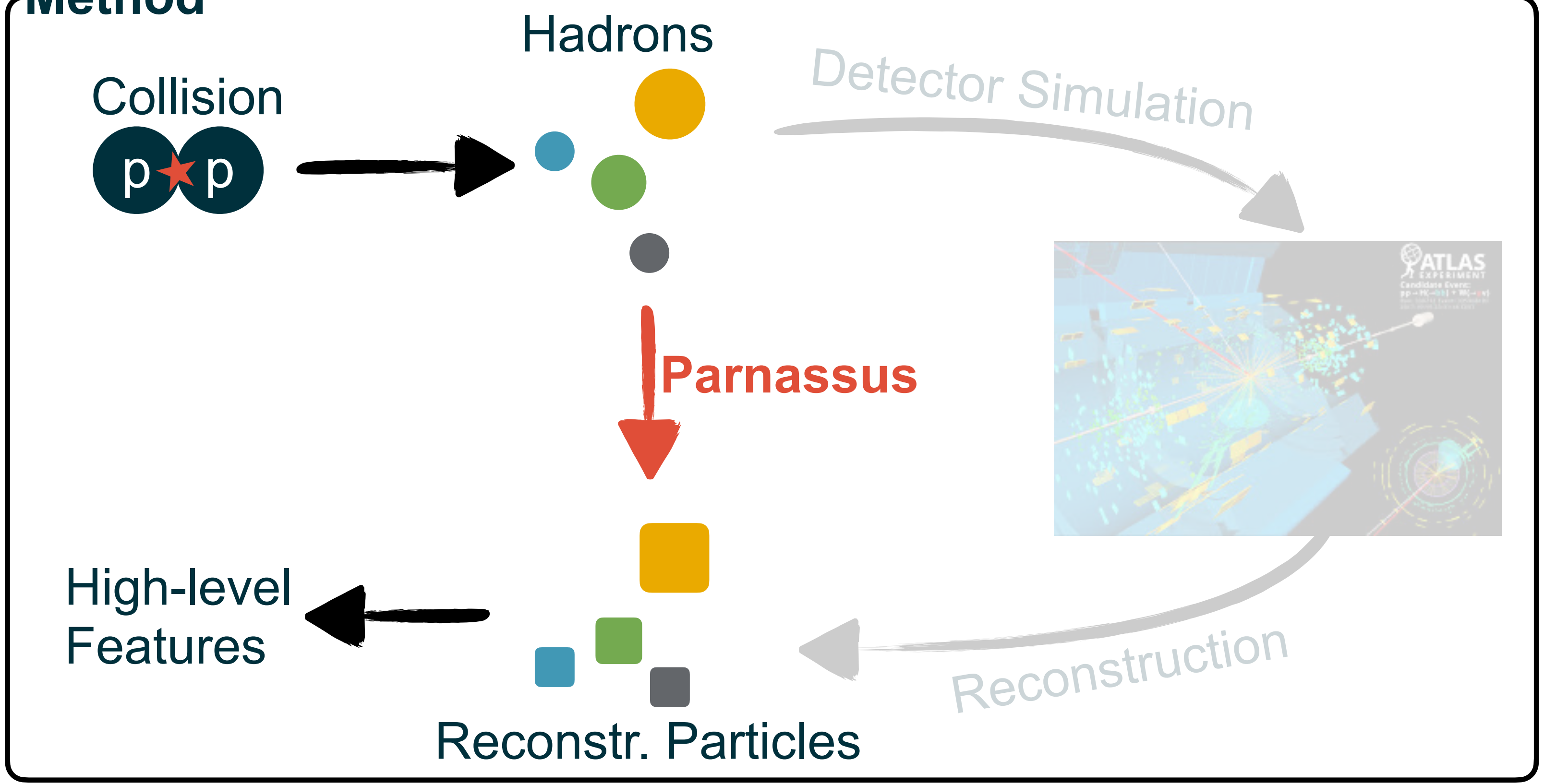




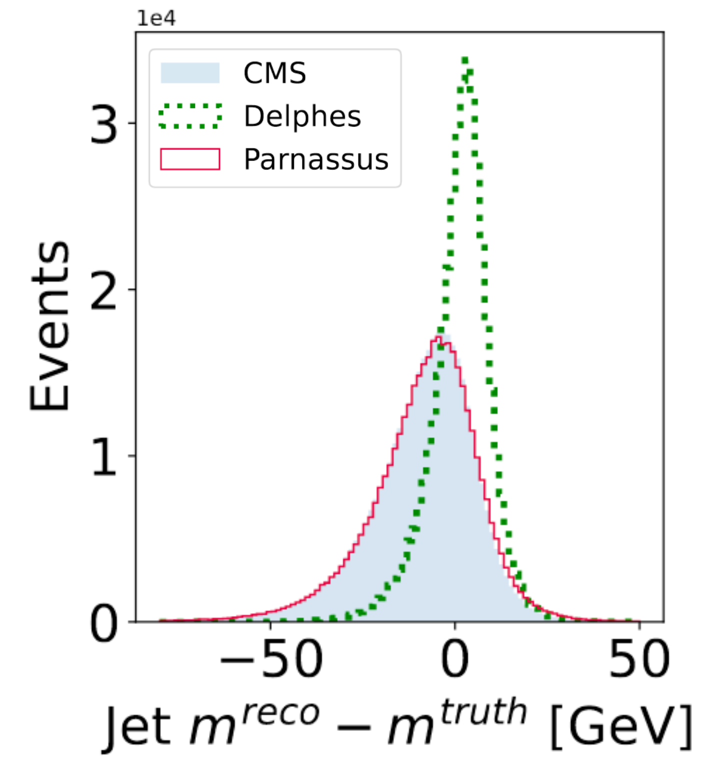
Simulation - Particle Level [8]

- Skip detector simulation and directly model reconstructed particles
- New ML method **Parnassus** [8]:
  - Normalizing Flow with Neural Ordinary Differential Equations
  - Transformer architecture (particle relations)

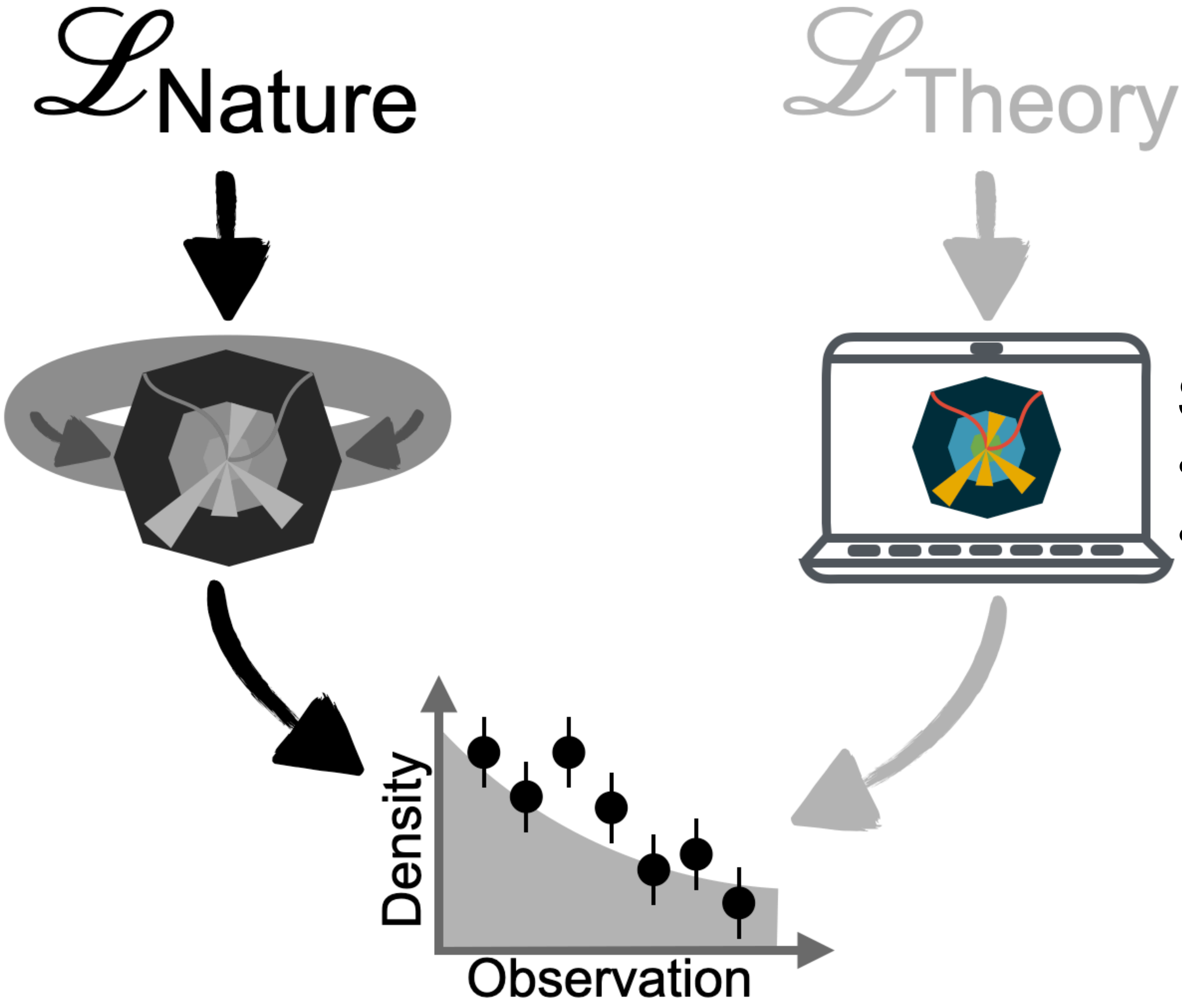
Method



Results

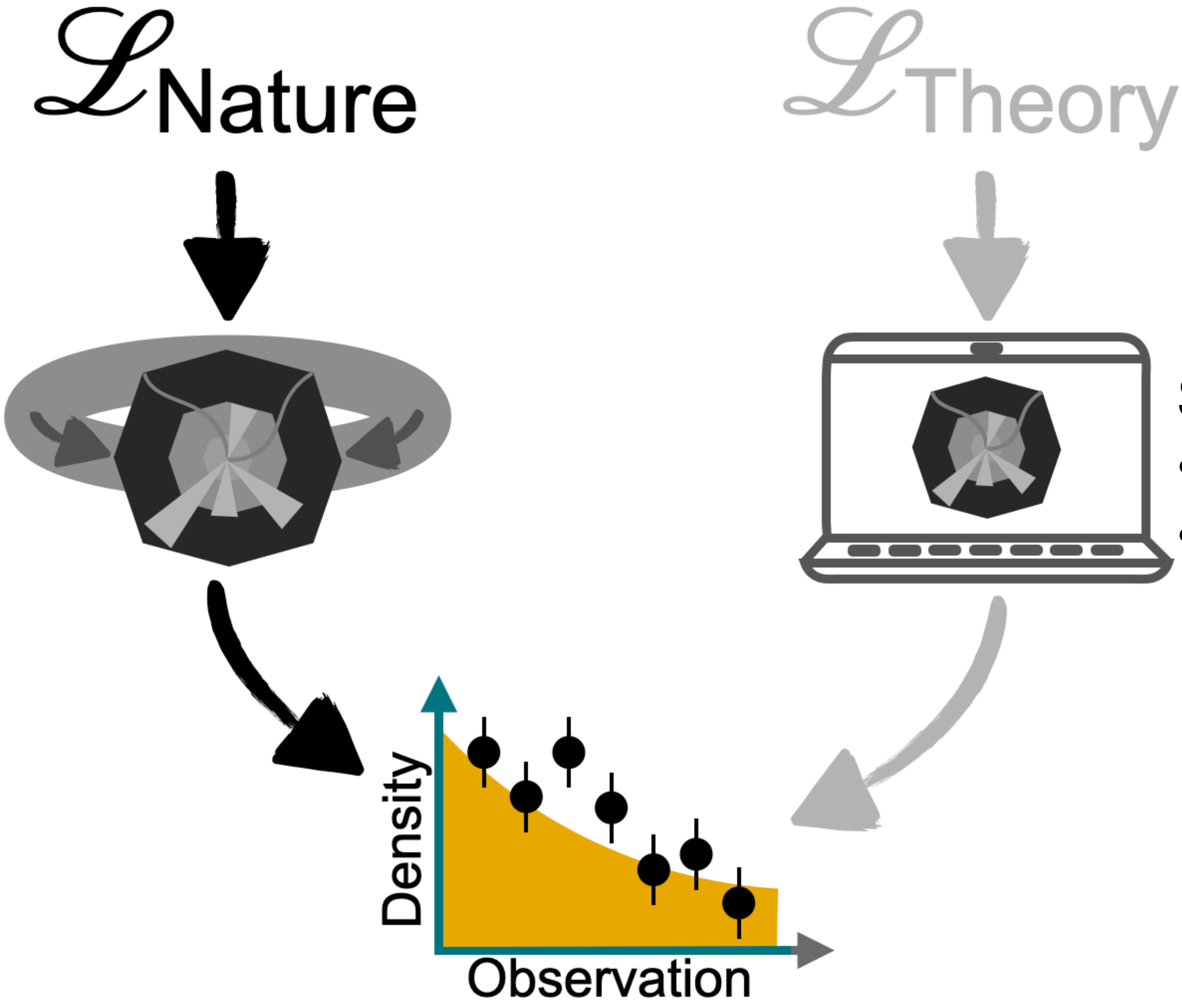


Experiment:
- Tracking
- MLPF
- Tagging
- Calibration



Simulation:
- Shower-level
- Particle-level

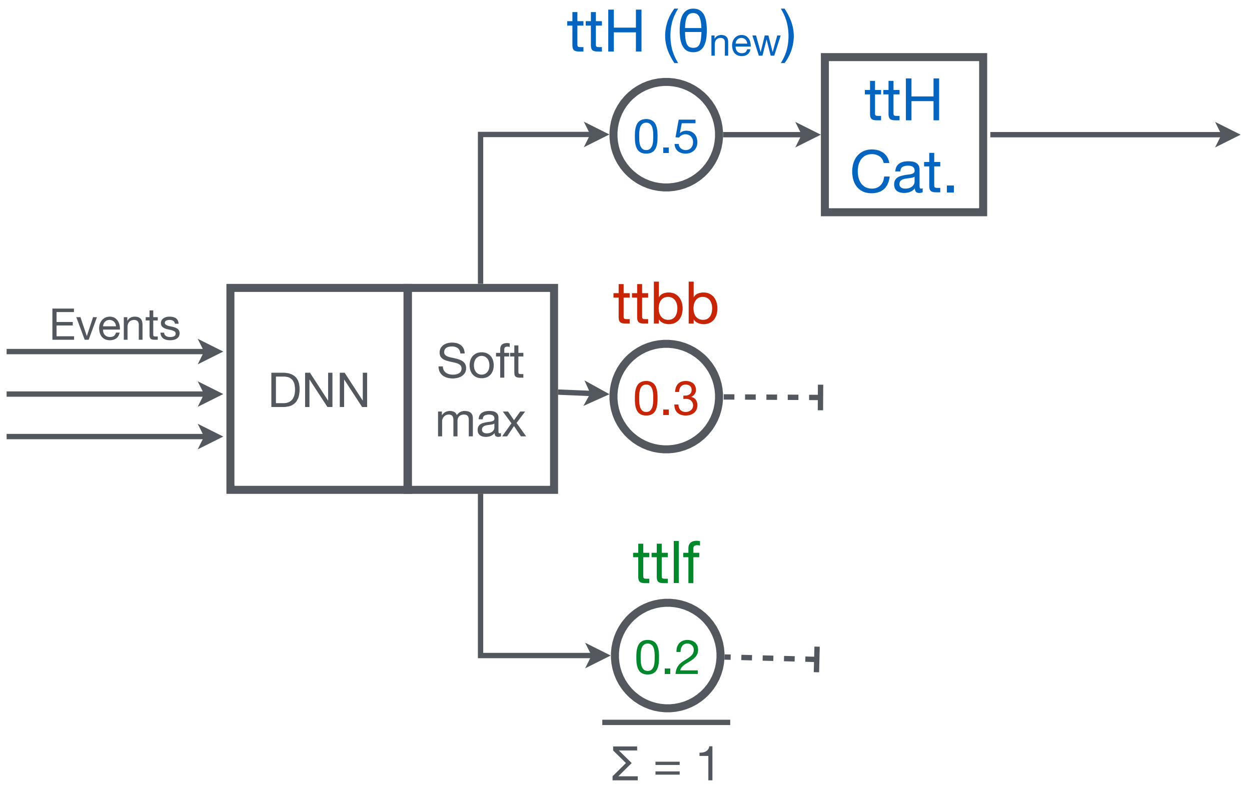
Experiment:
- Tracking
- MLPF
- Tagging
- Calibration

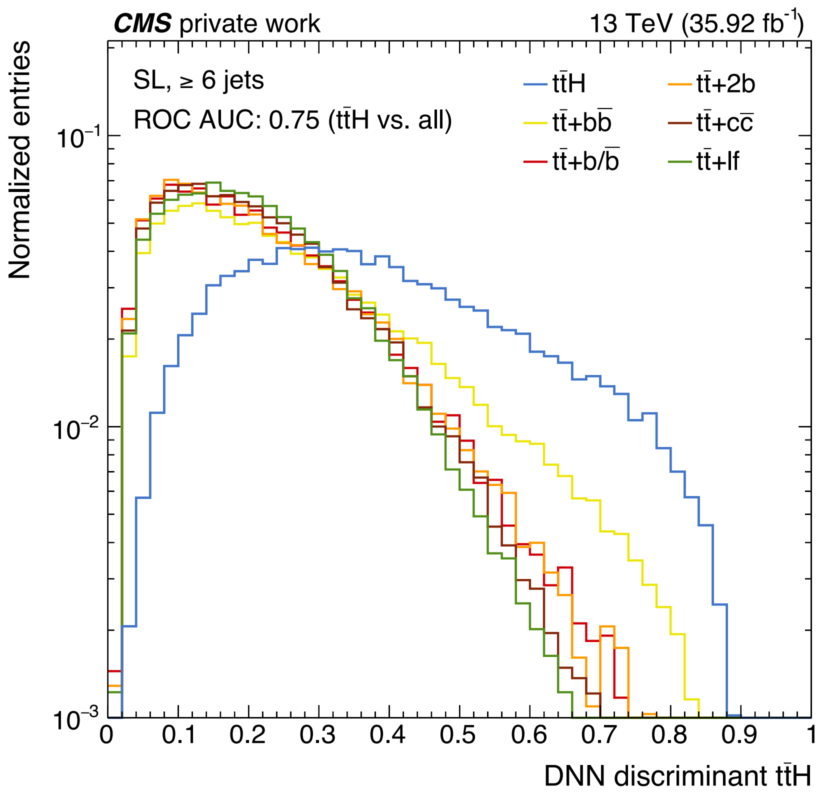


Simulation:
- Shower-level
- Particle-level

Regular Analysis [9]

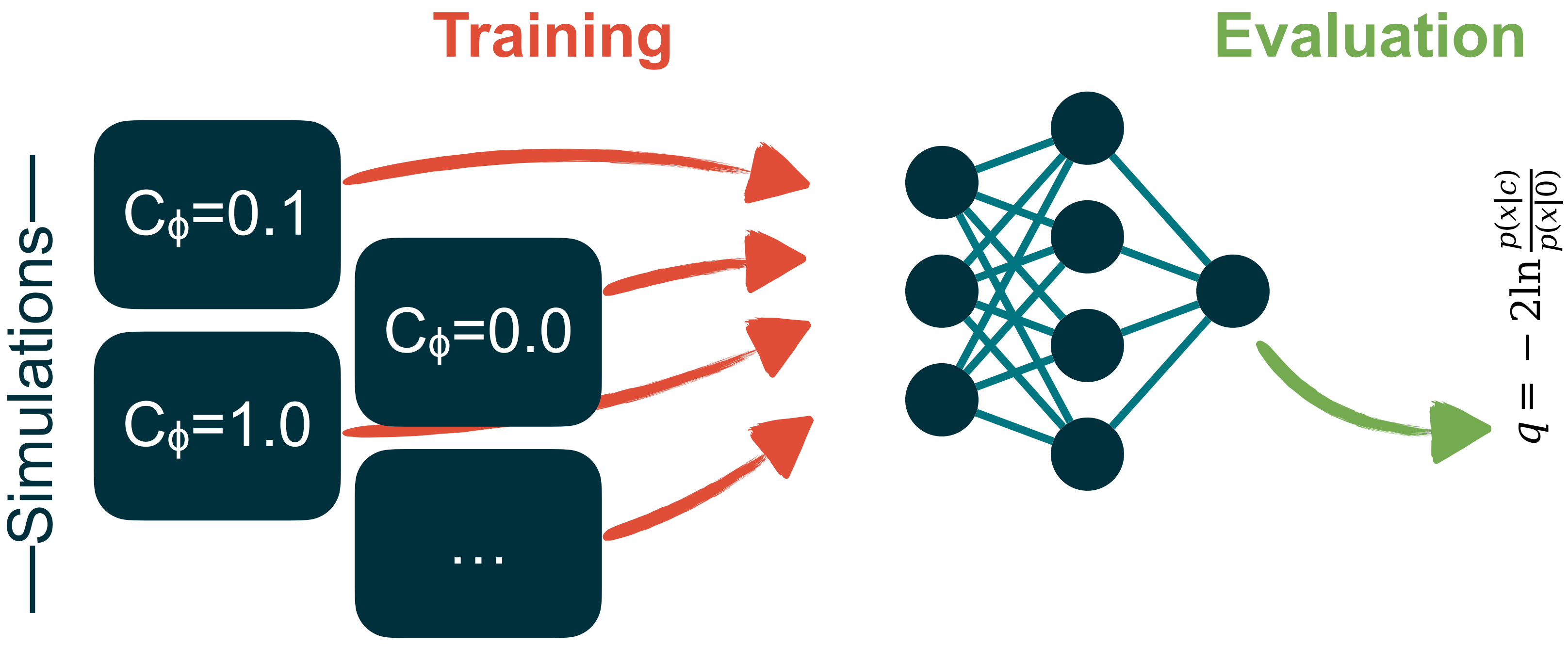
- Challenge: Decrease dimensionality of data (x) but keep physics information
- Optimal feature is probability of a new model $p(x|\theta_{new})$ vs $p(x|\theta_{old})$
- Multi-class classification trained with categorical cross-entropy (e.g. ttH measurement)

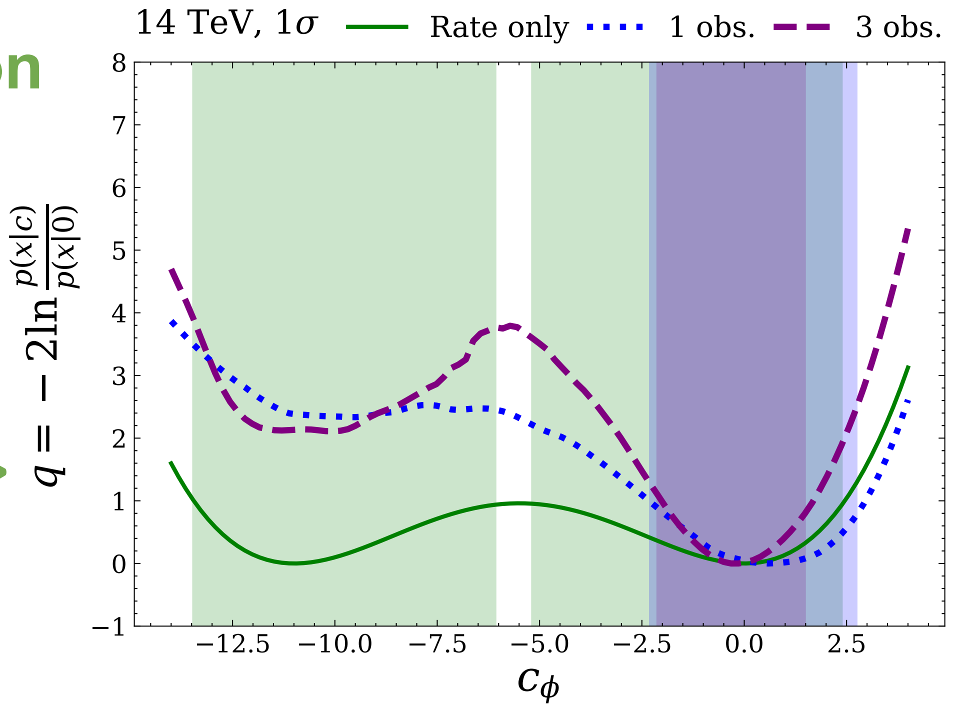




Simulation-Based Inference (SBI) [10,11]

- Techniques to directly infer p(θ|x) without using summary statistics / histograms
- Train networks to directly model likelihood ratio:
  - Trained via simple classification (e.g. $p(x|\theta_{BSM}) / p(x|\theta_{SM})$)
  - DNN can use low or high-dimensional data x





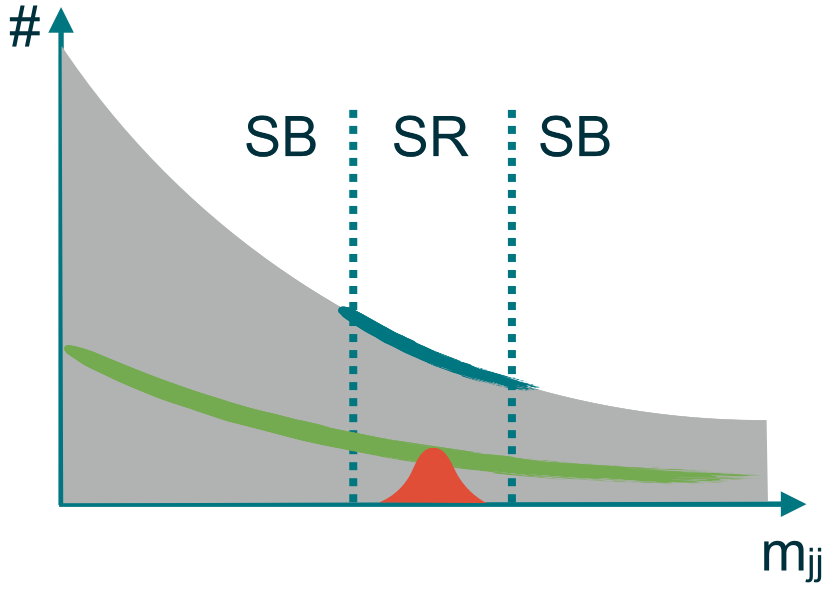
Anomaly Detection (AD) [12,13]

- Search for BSM in a model agnostic way
- Let machine figure out:
  - Interesting parts of phase space
  - How to look at them

2 Approaches	Assumption	Drawback
Unsupervised ML	Signal is rare	Not universal [14]
Weakly Supervised ML	Signal is peak	Need Bkg. Est.

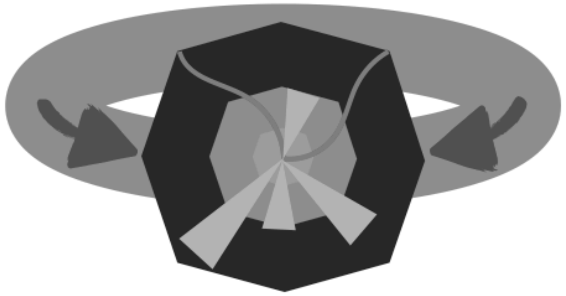
Weakly Supervised AD in a Nutshell:

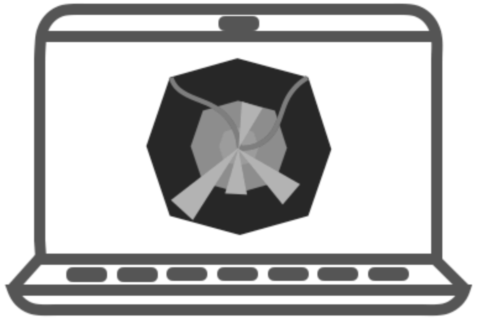
1. Define SR/SB
2. ML Bkg. Estimate
3. ML Class. ■ vs. ■
4. Bump Hunt



$\mathcal{L}_{\text{Nature}}$ $\mathcal{L}_{\text{Theory}}$

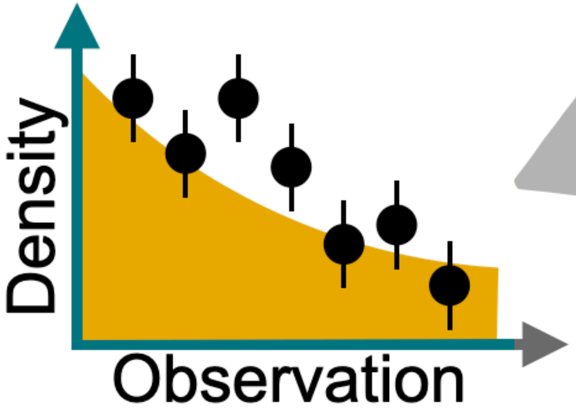
Experiment:
- Tracking
- MLPF
- Tagging
- Calibration





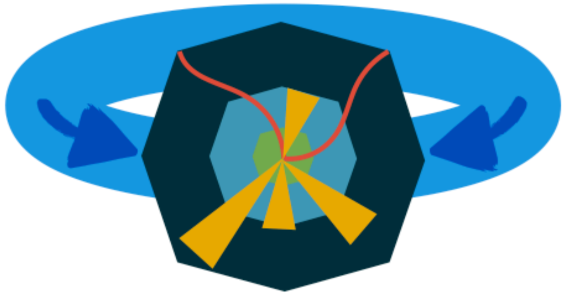
Simulation:
- Shower-level
- Particle-level

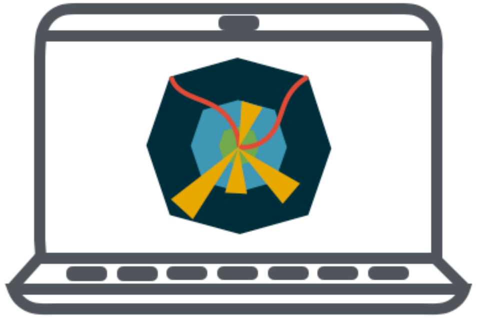
Comparison:
- Classification
- SBI
- AD



$\mathcal{L}_{\text{Nature}}$ $\mathcal{L}_{\text{Theory}}$

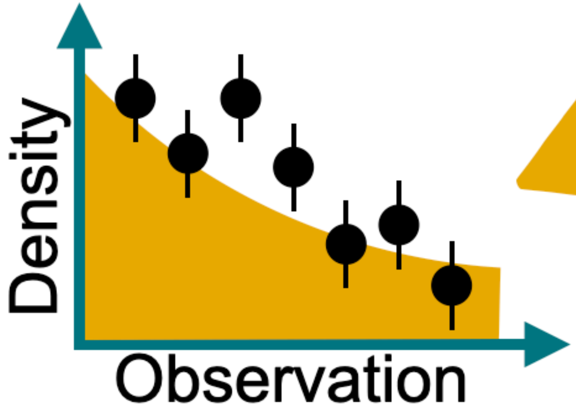
Experiment:
- Tracking
- MLPF
- Tagging
- Calibration



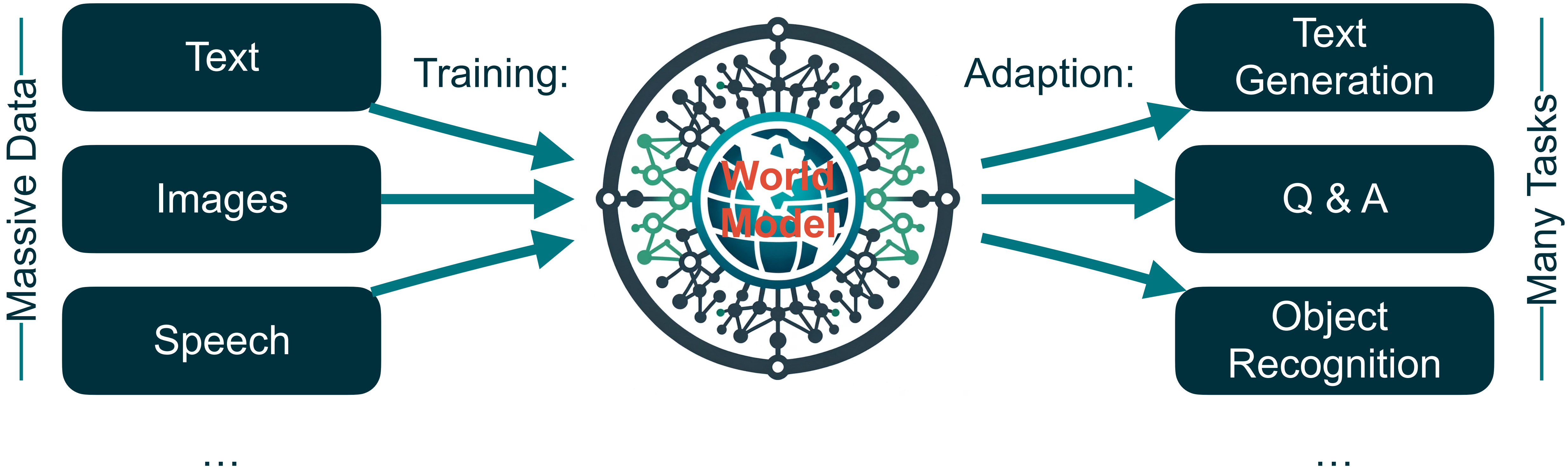


Simulation:
- Shower-level
- Particle-level

Comparison:
- Classification
- SBI
- AD

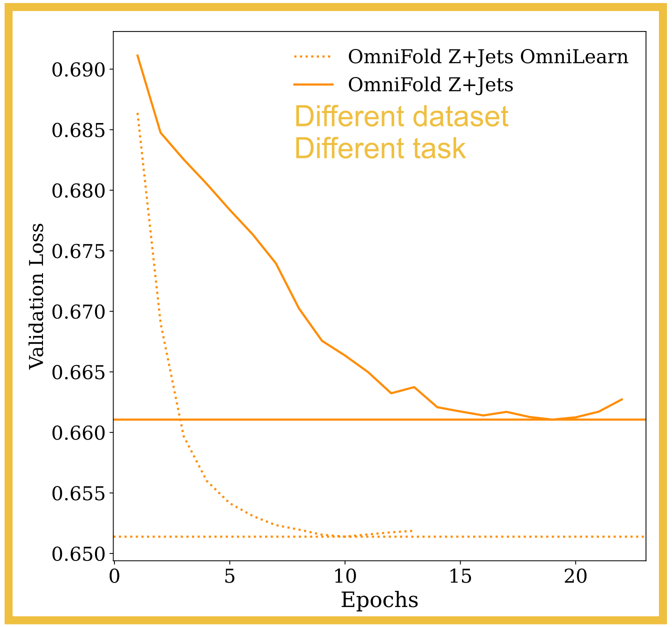


Foundation Models (Motivation) [15]

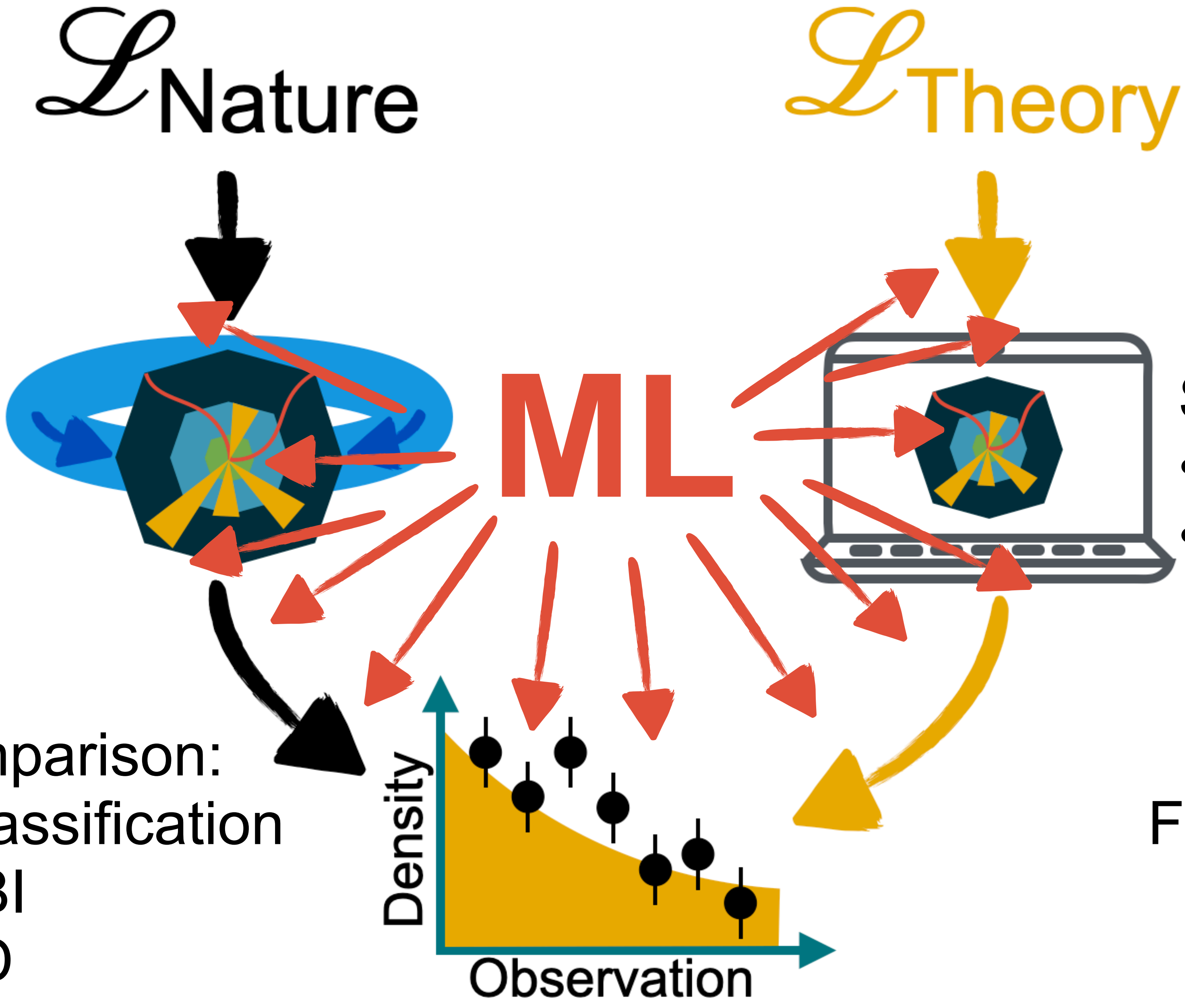


Foundation Models (Physics - OmniLearn) [16]

- OmniLearn:
  - Train foundation model for many jet-related tasks
  - Transformer model with Graph-attention networks
- Learns general World (Jet) Model
- Adaption better and more cost-efficient than training from scratch



Experiment:
- Tracking
- MLPF
- Tagging
- Calibration



Simulation:
- Shower-level
- Particle-level

Comparison:
- Classification
- SBI
- AD

Foundation Models

Contact

- First year postdoc at Berkeley Lab
- Since >7 years working in data analysis for CERN experiments

- **Physics**: Higgs, Anomaly Detection
- **Deep Learning**: Supervised, Unsupervised, Reinforcement
- **Computing**: Fast O(TB) Data Processing & Computing Pipelines

@ in

Citations

1. Gregor Kasieczka, Experimental particle physics and AI, Talk EUCAIFCon 2024 Amsterdam, https://indico.nikhef.nl/event/4875/contributions/21153/
2. Pata, J., Wulff, E., Mokhtar, F. et al. Improved particle-flow event reconstruction with scalable neural networks for current and future particle detectors. Commun Phys 7, 124 (2024). https://doi.org/10.1038/s42005-024-01599-5
3. Xiangyang Ju, HEP.TrkX Charged Particle Tracking Using Graph Neural Network, Connecting the Dots / Intelligens Trackers - Valencia Spain, April 3, 2019
4. Di Bello et al., Reconstructing particles in jets using set transformer and hypergraph prediction networks, Eur.Phys.J.C 83 (2023) 7, 596
5. Qu, H., & Gouskos, L. (2020). Jet tagging via particle clouds. Physical Review D, 101(5), 056019. DOI: 10.48550/arXiv.1902.08570
6. Nadezda Chernyavskaya, Deep Neural Network based b-jet energy correction and resolution, Presentation at SPS Annual Meeting, Lausanne (August 2018) Link
7. ATLAS Collaboration, Deep generative models for fast photon shower simulation in ATLAS, https://arxiv.org/pdf/2210.06204
8. Dreyer et al, Parnassus: An Automated Approach to Accurate, Precise, and Fast Detector Simulation and Reconstruction, https://arxiv.org/html/2406.01620v1
9. Marcel Rieger. 'Search for Higgs boson production in association with top quarks and decaying into bottom quarks using deep learning techniques with the CMS experiment'. PhD thesis. RWTH Aachen University, 2019. URL: http://publications.rwth-aachen.de/record/763526
10. Cranmer et al, Approximating Likelihood Ratios with Calibrated Discriminative Classifiers, https://arxiv.org/abs/1506.02169
11. Mastandrea et al, Constraining the Higgs Potential with Neural Simulation-based Inference for Di-Higgs Production, https://arxiv.org/pdf/2405.15847
12. Ben Nachman, Re-imagining the search for fundamental interactions with machine learning, Talk BIDMaP Seminar 2024, UC Berkeley
13. K. Benkendorfer, L. Le Pottier, Ben Nachman, PRD 104 (2021) 035003 and many more
14. G. Kasieczka, R. Mastandrea, V. Mikuni, BN, et. al, PRD 107 (2023) 015009
15. Rick Merrit (NVIDIA), What Are Foundation Models? https://blogs.nvidia.com/blog/what-are-foundation-models/
16. Vinicius Mikuni, Benjamin Nachman, OmniLearn: A Method to Simultaneously Facilitate All Jet Physics Tasks, https://arxiv.org/abs/2404.16091