48,000 UC academic workers on strike for fair compensation, support for working parents, international scholar rights, and more.
To find out how you can support: fairucnow.org

PROFESSIONAL RESEARCHERS
STUDENT RESEARCHERS
PROJECT SCIENTISTS
SPECIALISTS
POSTDOCS
READERS
TUTORS
CPPs
TAs

48,000 STRONG

RECENT ADVANCES IN

Machine learning
in
particle physics

JAVIER DUARTE

DARK INTERACTIONS WORKSHOP

NOVEMBER 16, 2022

a3d3

UCSD

RECENT ADVANCES IN

Machine learning
in
particle physics

JAVIER DUARTE

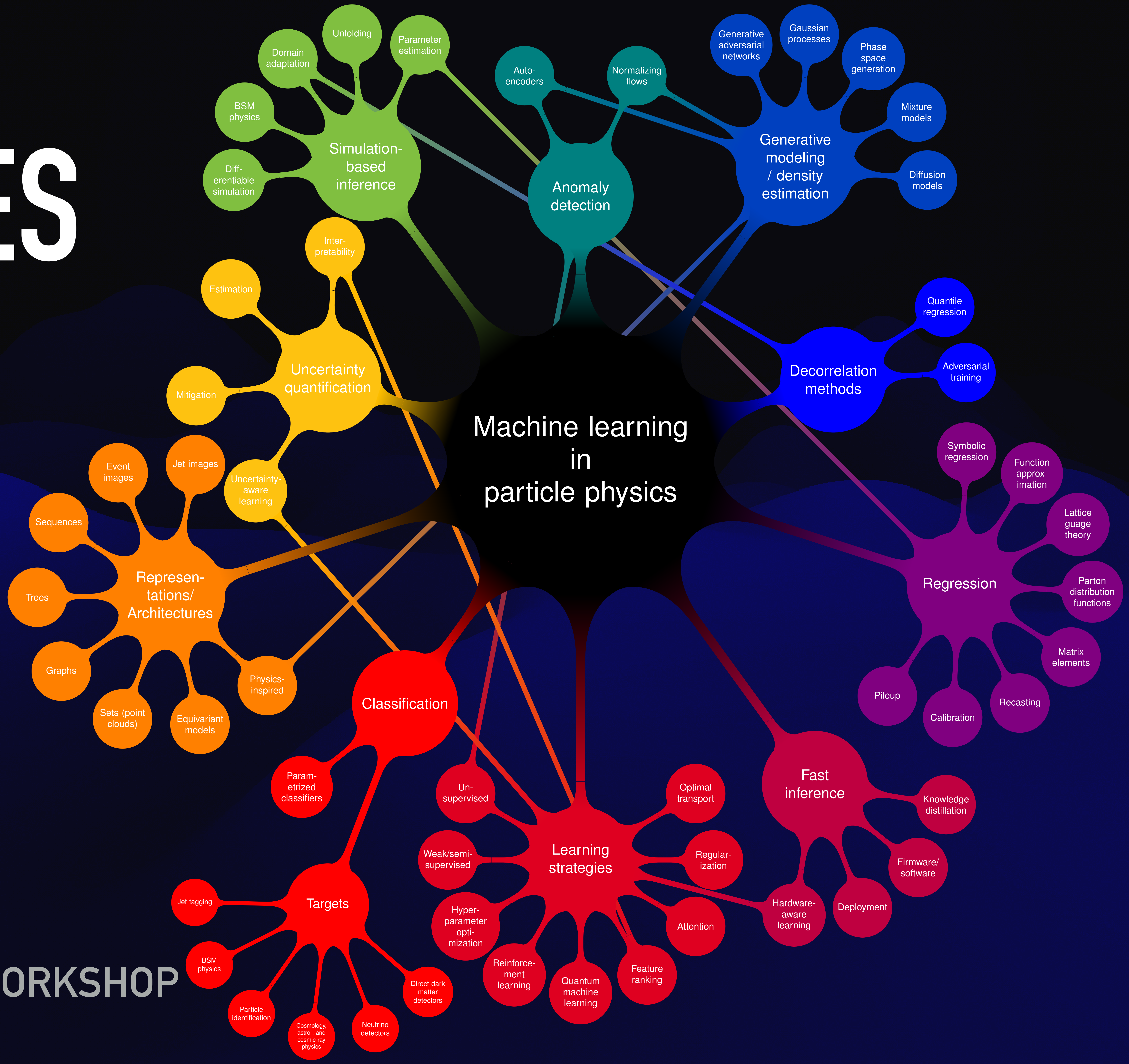
DARK INTERACTIONS WORKSHOP

NOVEMBER 16, 2022

a3d3

UCSD

RECENT ADVANCES IN



JAVIER DUARTE

DARK INTERACTIONS WORKSHOP

NOVEMBER 16, 2022

iml-wg.github.io/HEPML-LivingReview

github.com/jmduarte/Nomological_Net_ML_Particle_Physics

I. DATA REPRESENTATIONS & SYMMETRIES
II. ANOMALY DETECTION
III. GENERATIVE MODELING
III. FAST INFERENCE
VI. SUMMARY & OUTLOOK

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

- High-level (expert) variables

- Shallow neural network, boosted decision tree, …

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

- High-level (expert) variables
- Shallow neural network, boosted decision tree, …
- Ordered list of particles
- 1D convolutional neural network, recurrent neural network

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

- High-level (expert) variables
- Ordered list of particles
- Images

- Shallow neural network, boosted decision tree, …
- 1D convolutional neural network, recurrent neural network
- 2D convolutional neural network

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

- High-level (expert) variables
- Ordered list of particles
- Images
- Set of particles

- Shallow neural network, boosted decision tree, …
- 1D convolutional neural network, recurrent neural network
- 2D convolutional neural network
- Deep set (energy flow network)

REPRESENTATIONS ⟷ INDUCTIVE BIAS ⟷ ALGORITHMS

- High-level (expert) variables
- Ordered list of particles
- Images
- Set of particles
- Graph of particles

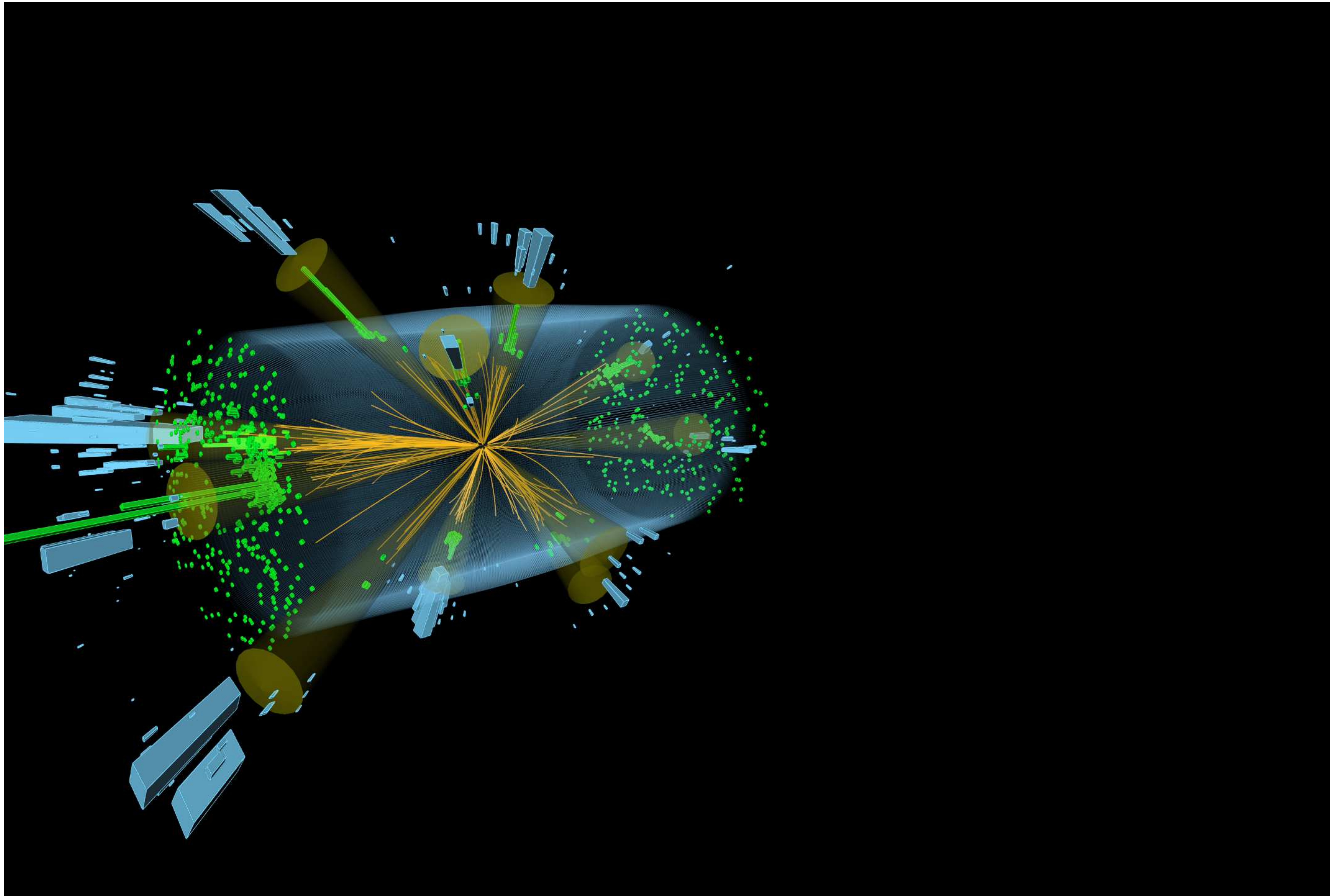
- Shallow neural network, boosted decision tree, …
- 1D convolutional neural network, recurrent neural network
- 2D convolutional neural network
- Deep set (energy flow network)
- Graph neural network

REPRESENTATIONS $\longleftrightarrow$ INDUCTIVE BIAS $\longleftrightarrow$ ALGORITHMS

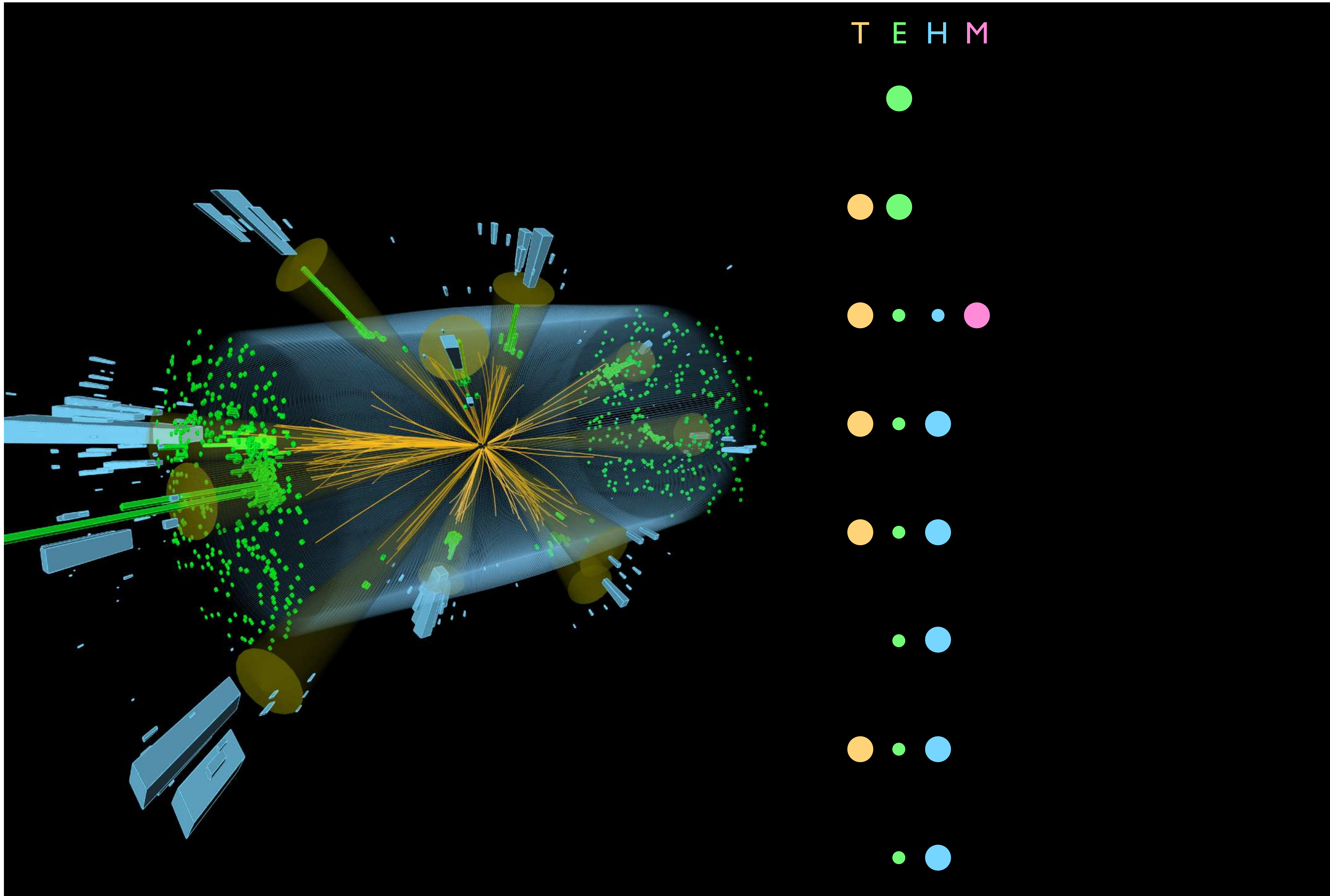
- High-level (expert) variables
- Ordered list of particles
- Images
- Set of particles
- Graph of particles
- Lorentz scalars/vectors

- Shallow neural network, boosted decision tree, …
- 1D convolutional neural network, recurrent neural network
- 2D convolutional neural network
- Deep set (energy flow network)
- Graph neural network
- Lorentz-equivariant network

COLLISION EVENT

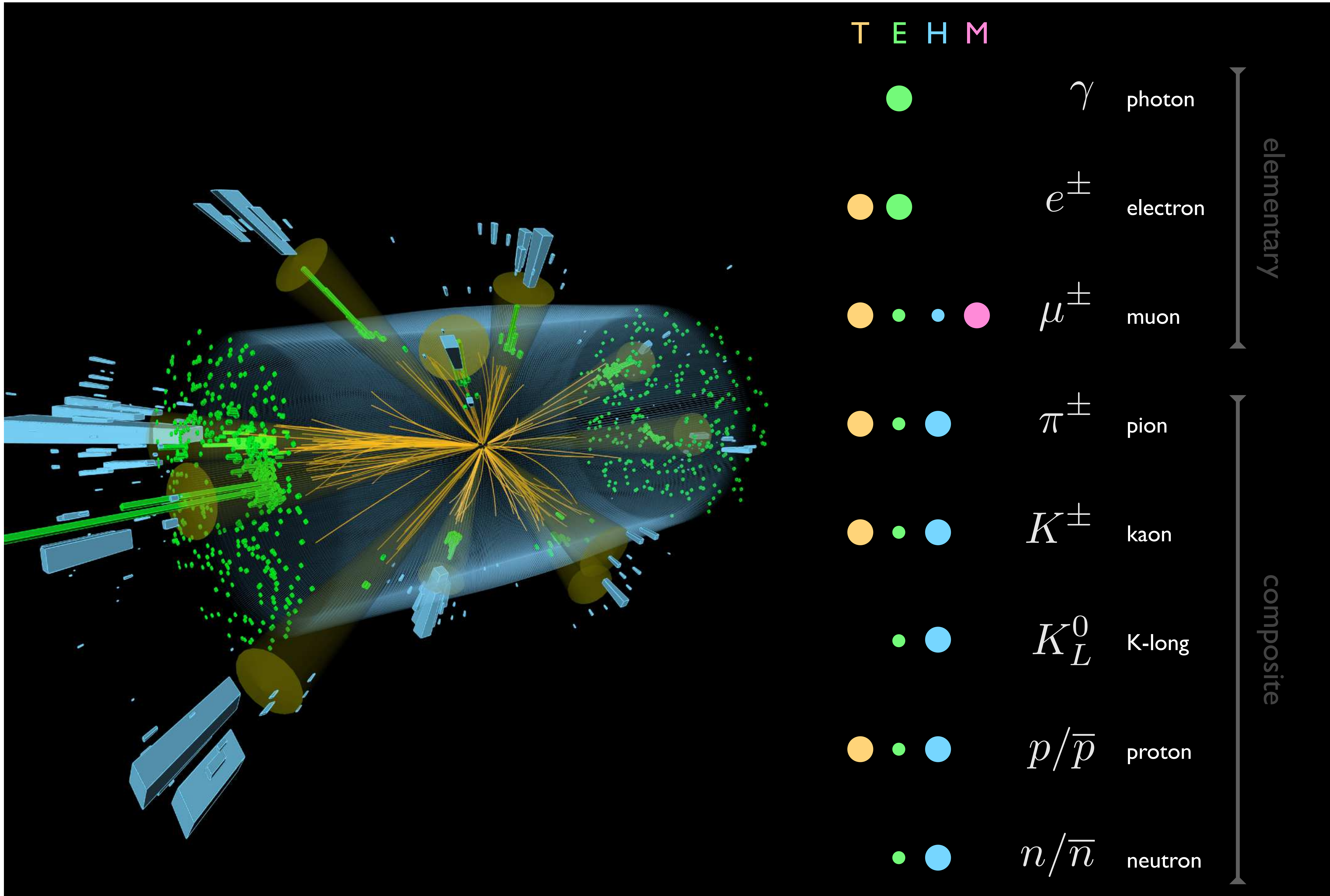


COLLISION EVENT



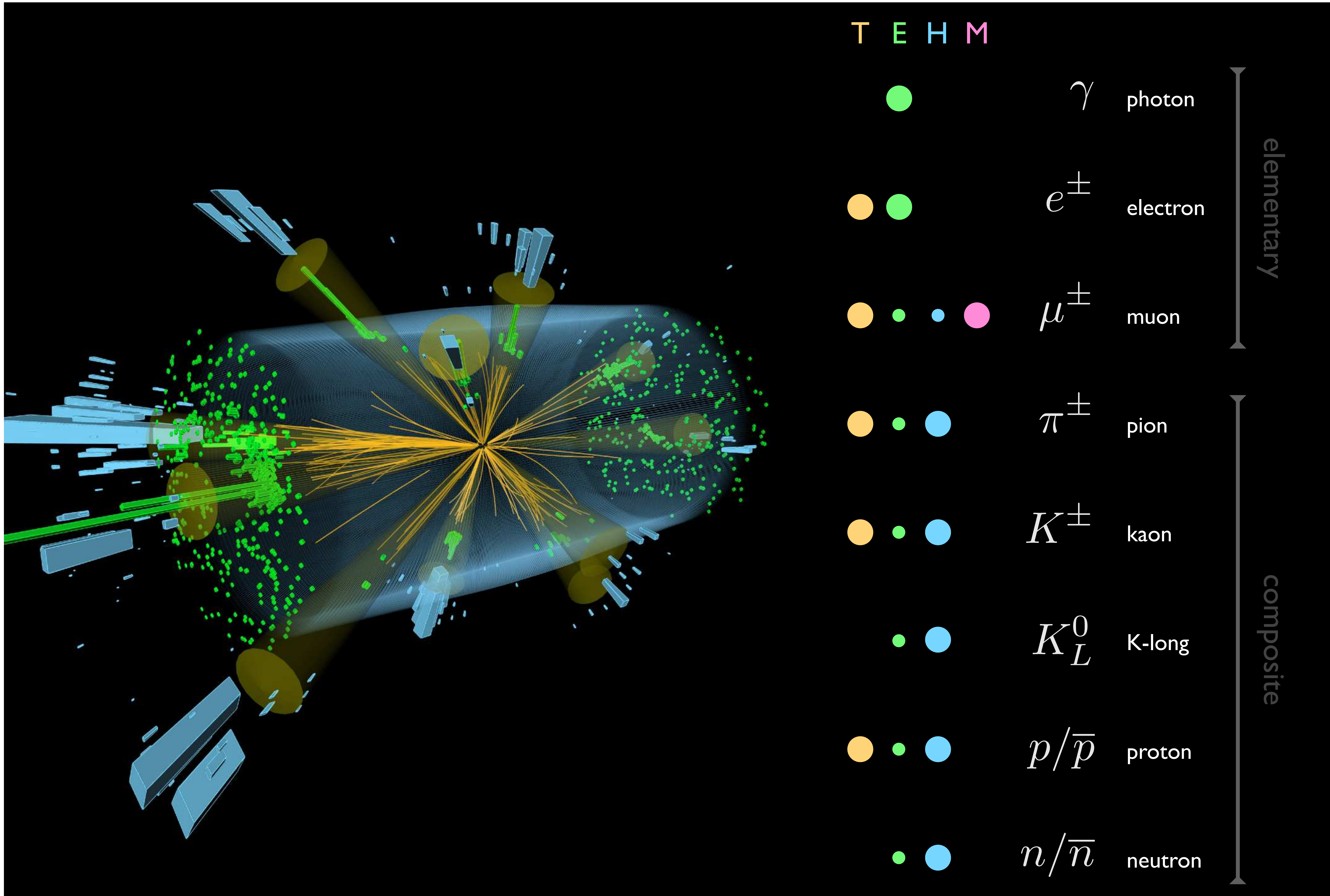
COLLISION EVENT

- After "particle-flow reconstruction," can think of event as a collection of points in momentum space



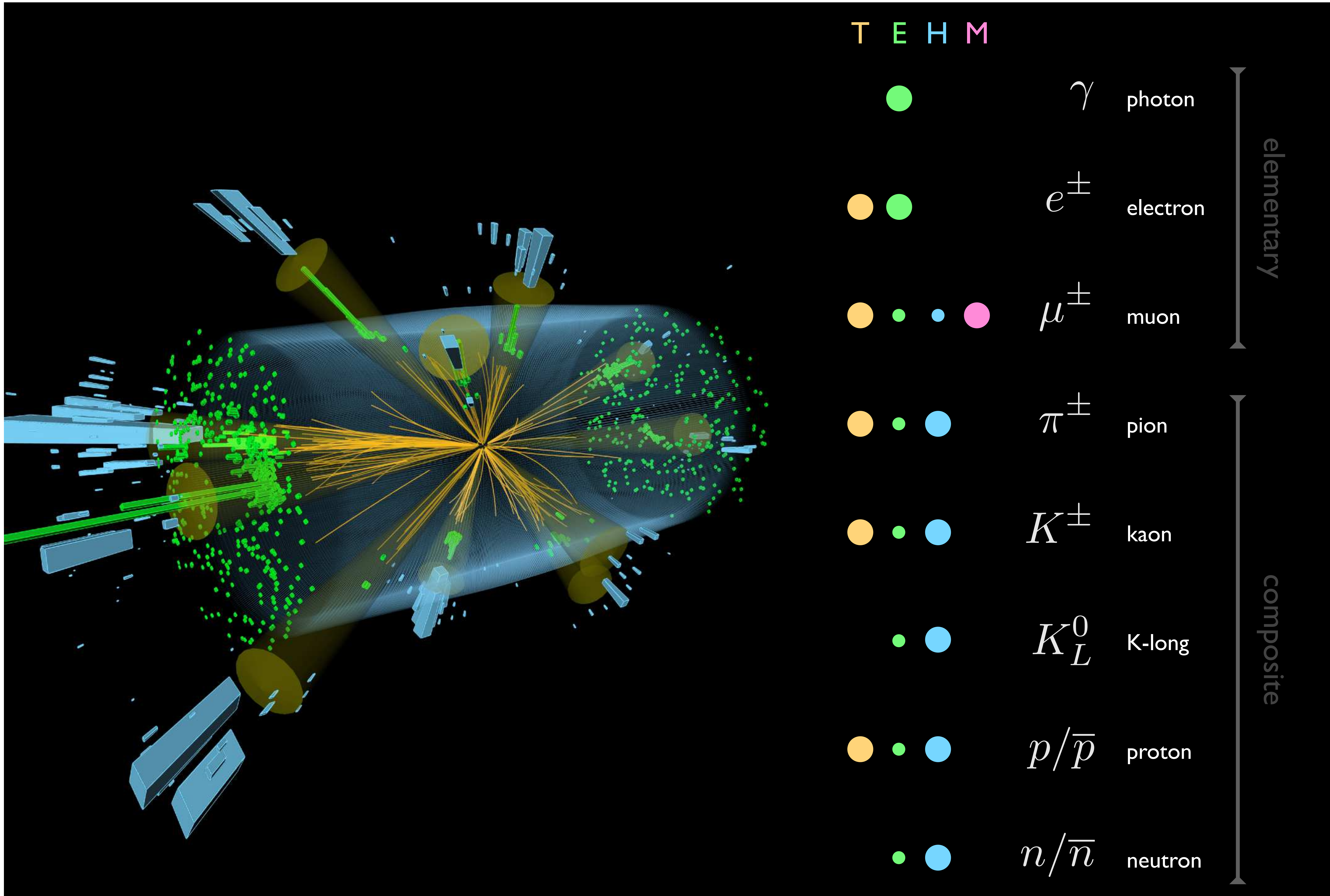
COLLISION EVENT

- After "particle-flow reconstruction," can think of event as a collection of points in momentum space
- For jets (localized clusters of particles), dimensionality $(N_{\text{particles}} \sim 100,\ 4 + M)$

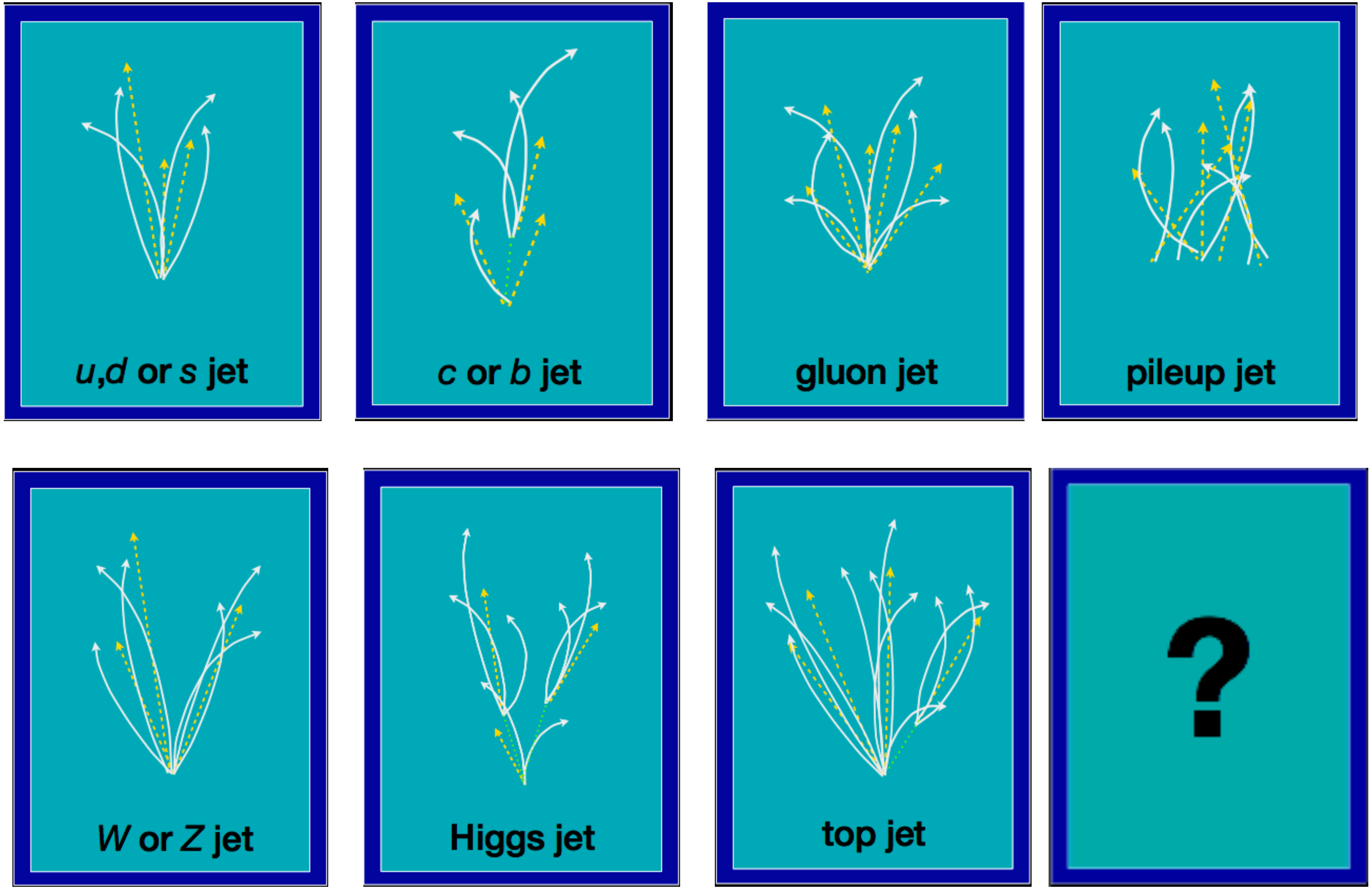


COLLISION EVENT

- After "particle-flow reconstruction," can think of event as a collection of points in momentum space
- For jets (localized clusters of particles), dimensionality ($N_{\text{particles}} \sim 100$, $4 + M$)
- Variable jet length requires:
  - Preprocessing into another rep. (tab. data, jet images, …)
  - Truncation to fixed size
  - Graph NN



TASK: JET CLASSIFICATION

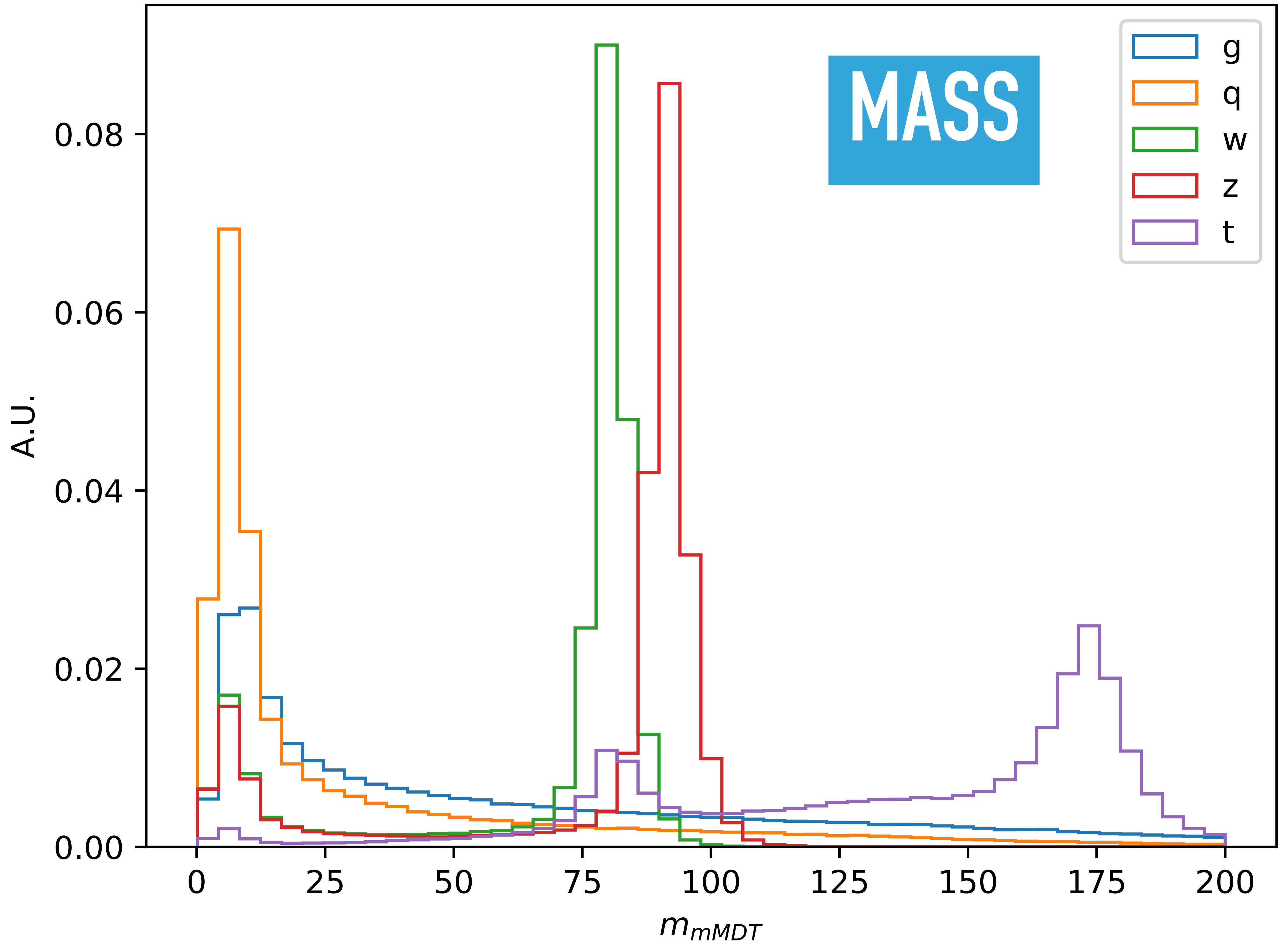


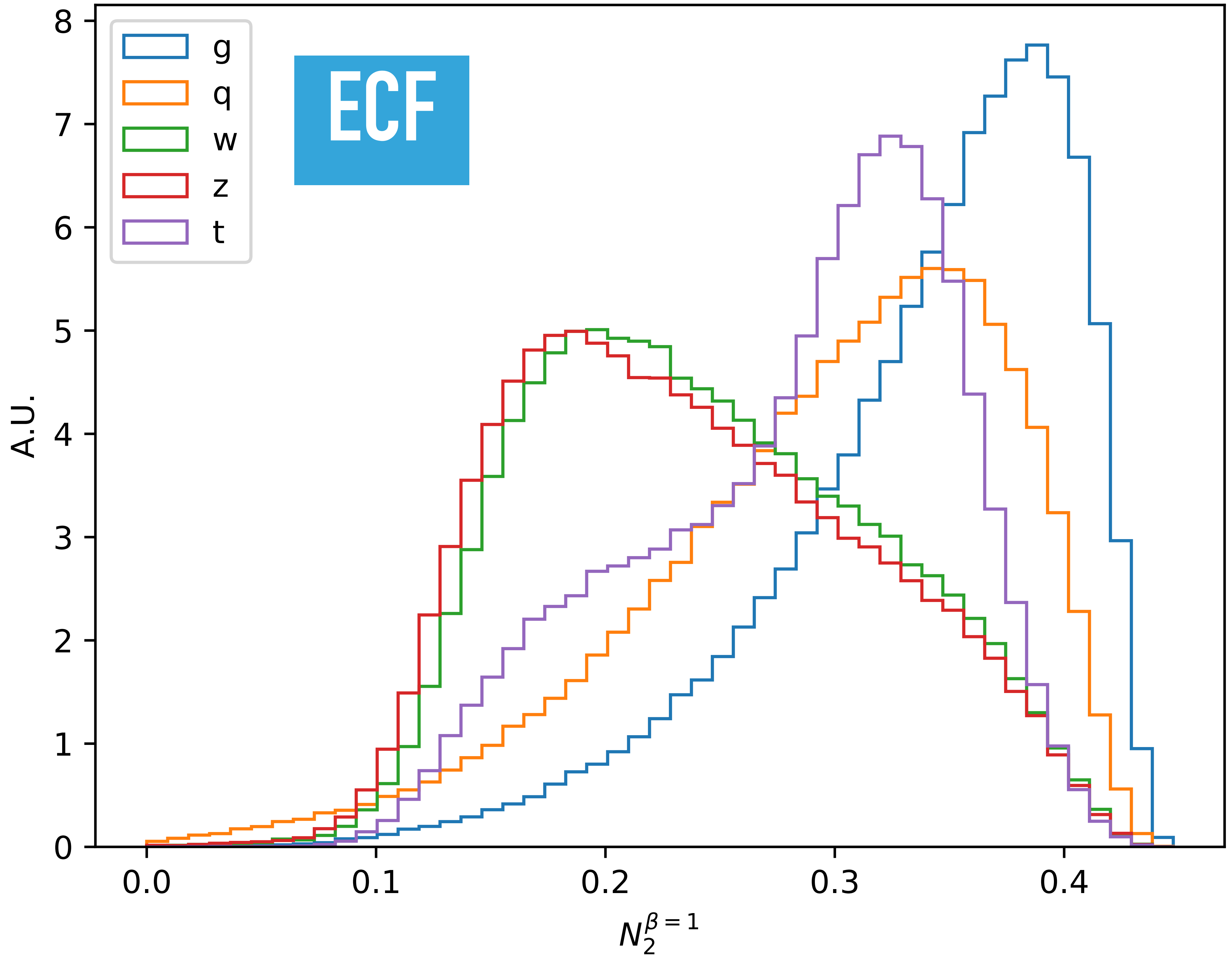
TABULAR DATA: JET SUBSTRUCTURE VARIABLES

- Tabular data: use physics knowledge to preprocess jet information into a set of high-level features
- Substructure variable:
    - jet mass
    - energy correlation functions, e.g. $N_2^{\beta=1} = {}_2e_3^{\beta=1}/({}_1e_3^{\beta=1})^2$

$${}_1e_3^{\beta} = \sum_{1\leq i<j<k\leq n_J} z_i z_j z_k \min\{\Delta R_{ij}^{\beta}, \Delta R_{ik}^{\beta}, \Delta R_{jk}^{\beta}\}$$

$${}_2e_3^{\beta} = \sum_{1\leq i<j<k\leq n_J} z_i z_j z_k \min\{\Delta R_{ij}^{\beta}\Delta R_{ik}^{\beta}, \Delta R_{ij}^{\beta}\Delta R_{jk}^{\beta}, \Delta R_{ik}^{\beta}\Delta R_{jk}^{\beta}\}$$



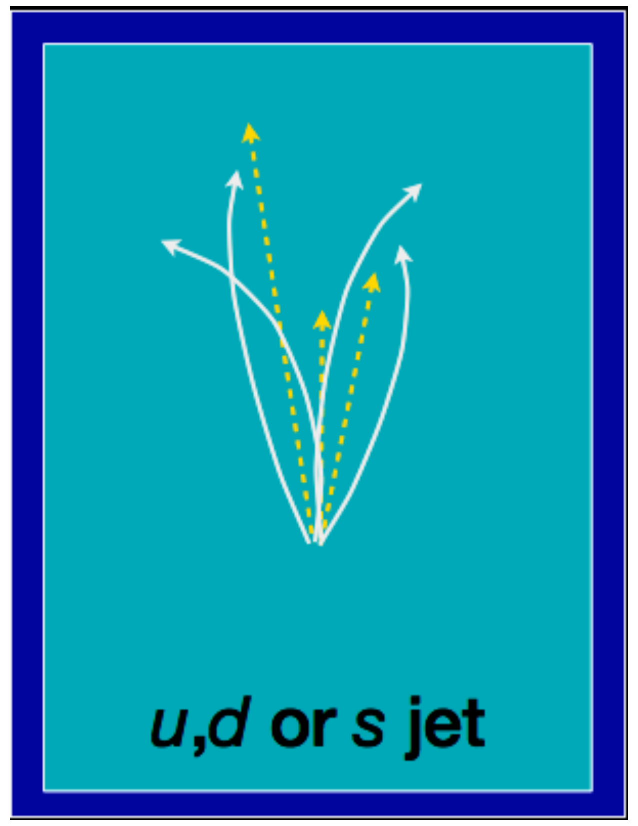


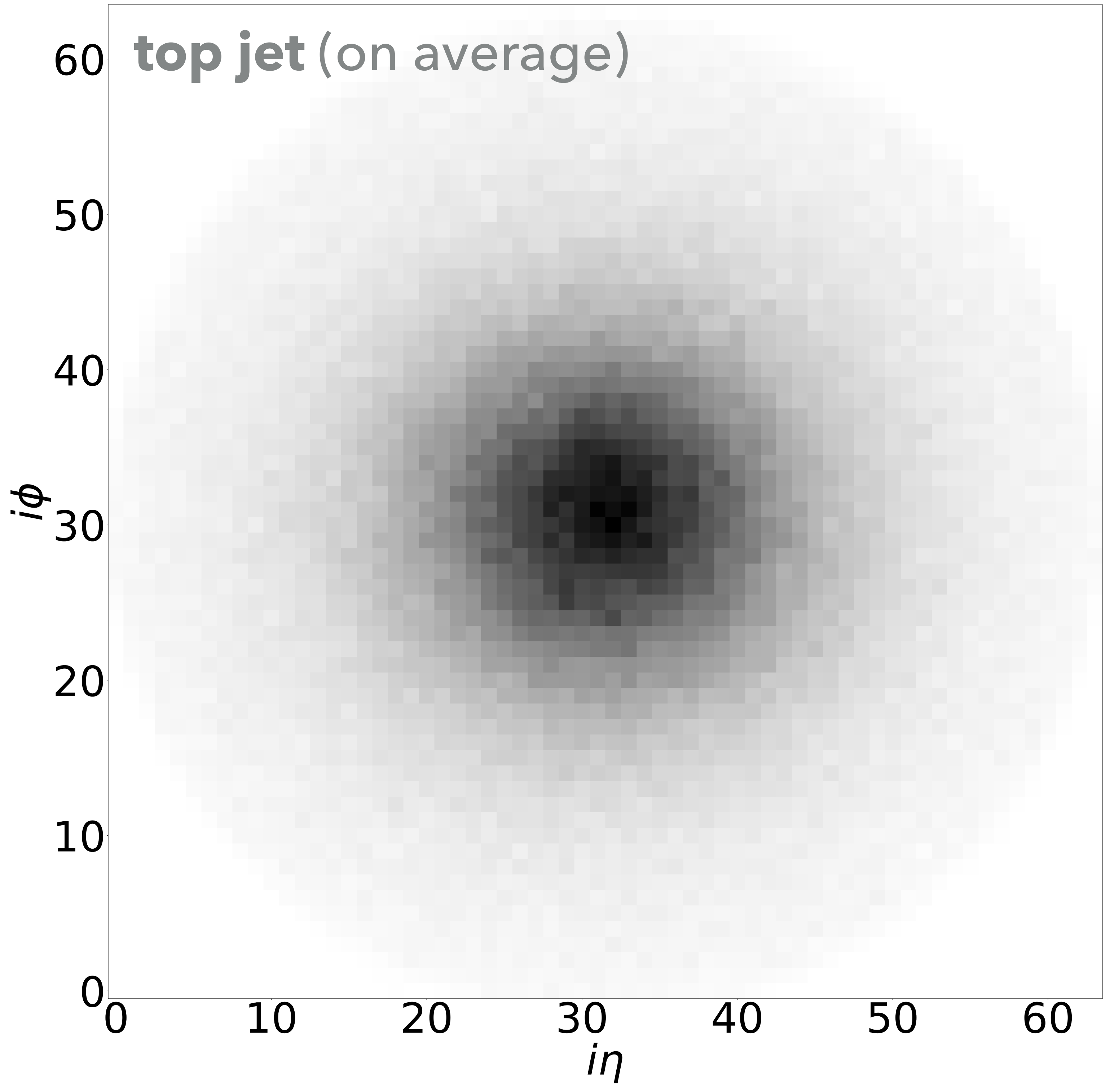
JET IMAGES

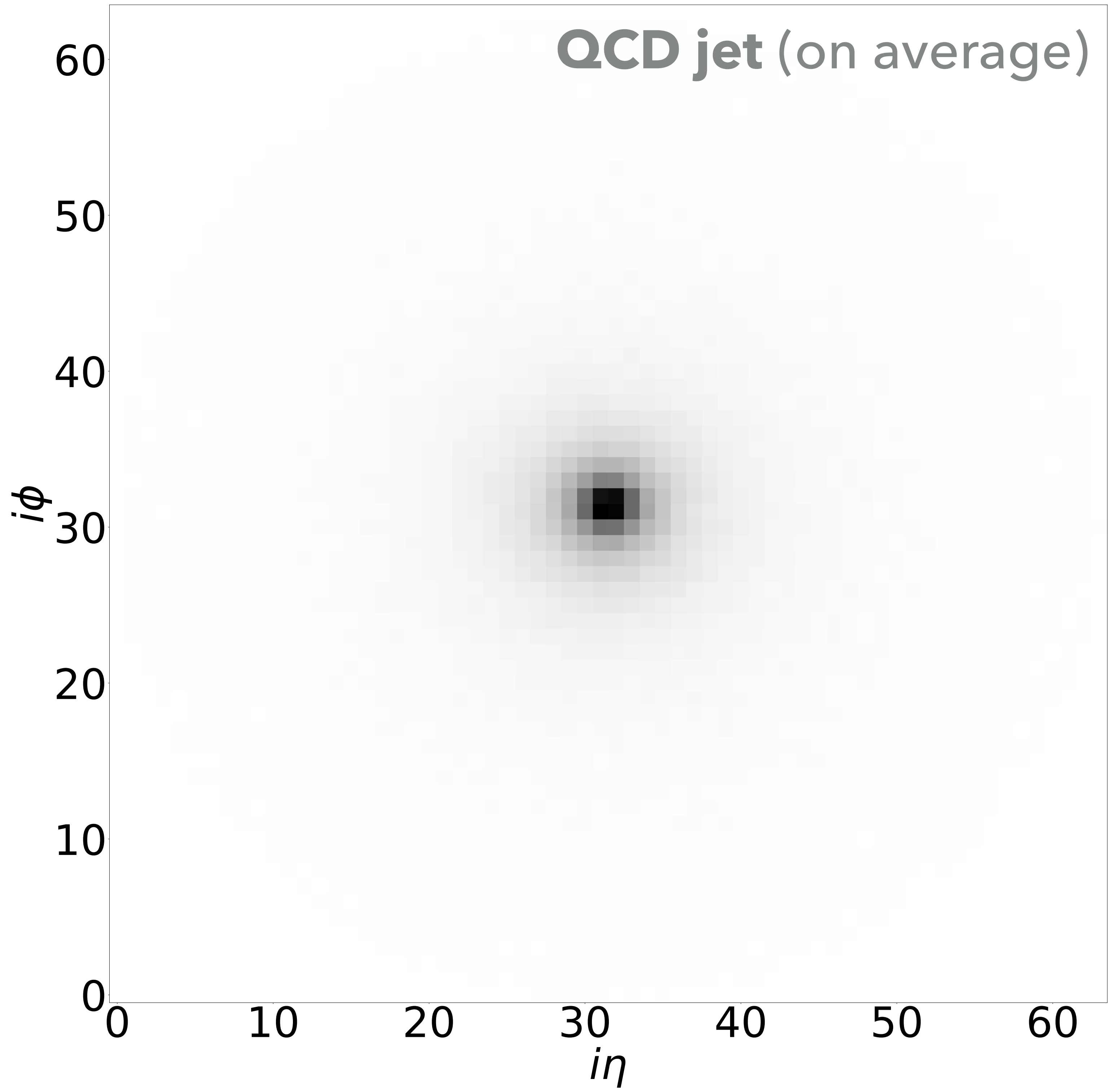
- Jet images = pixelated versions of calorimeter hits in 2D (η, ϕ)

- Much lower level

vs.

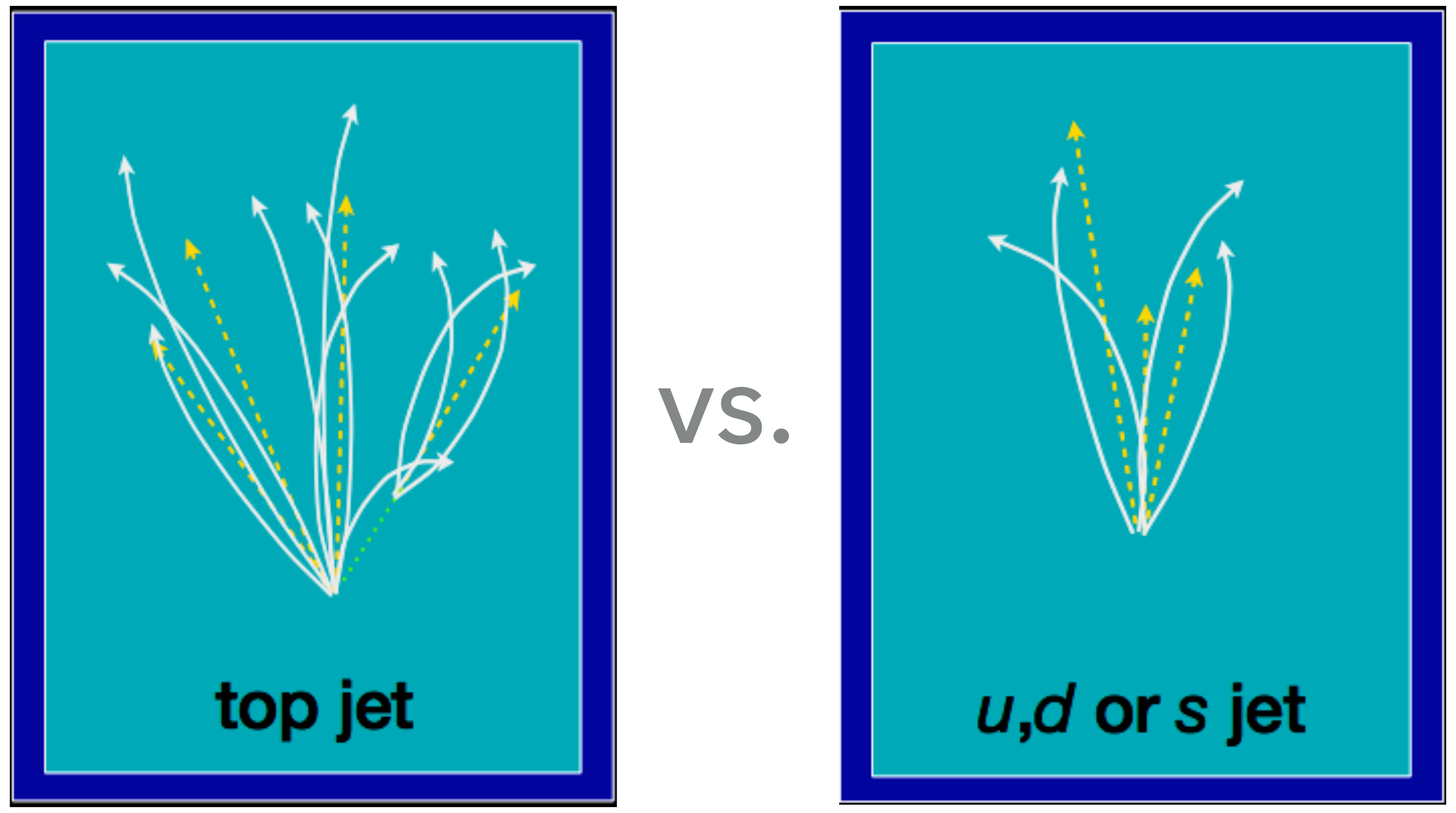


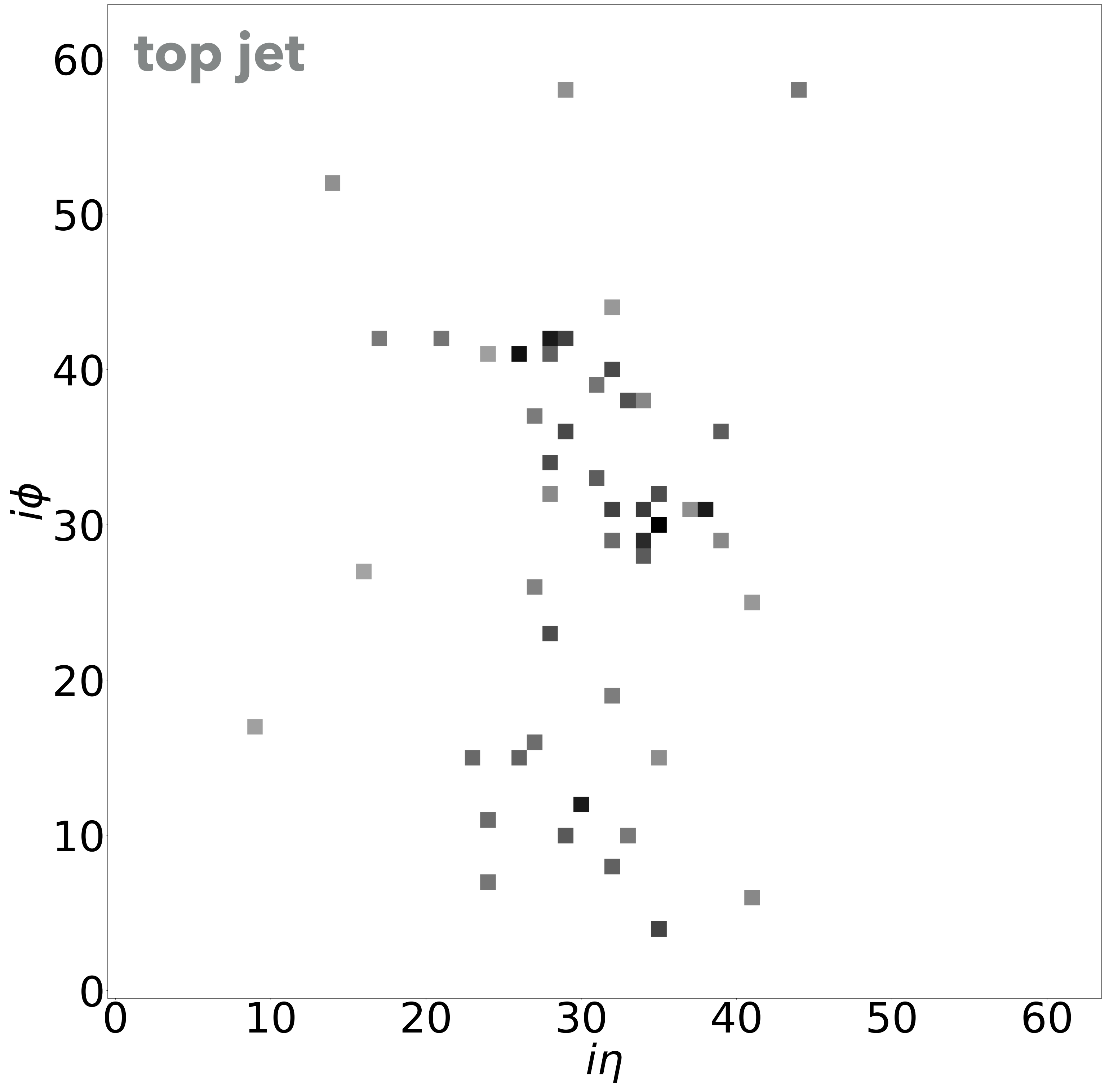


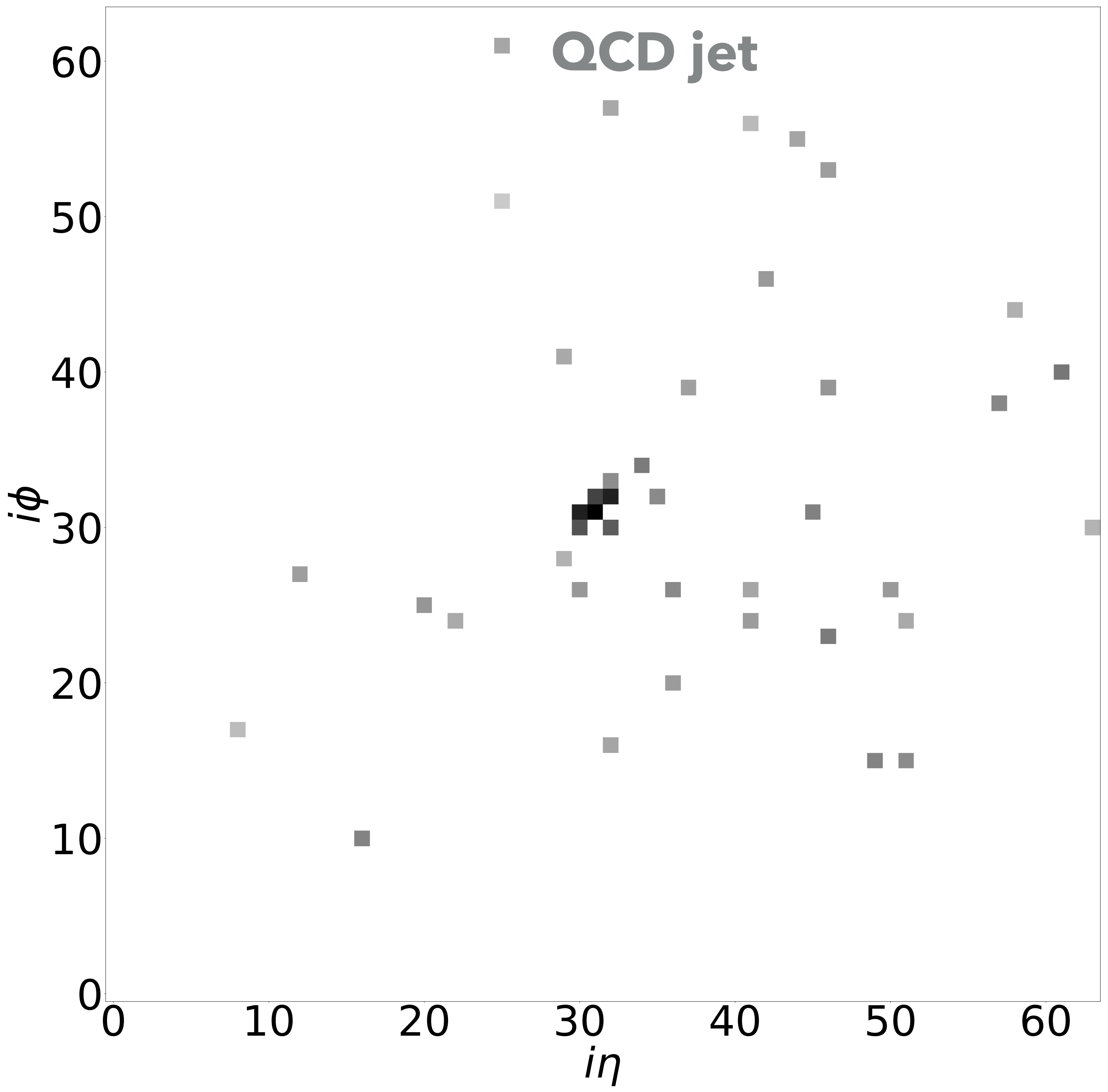


JET IMAGES

- Jet images = pixelated versions of calorimeter hits in 2D (η, ϕ)
- Much lower level

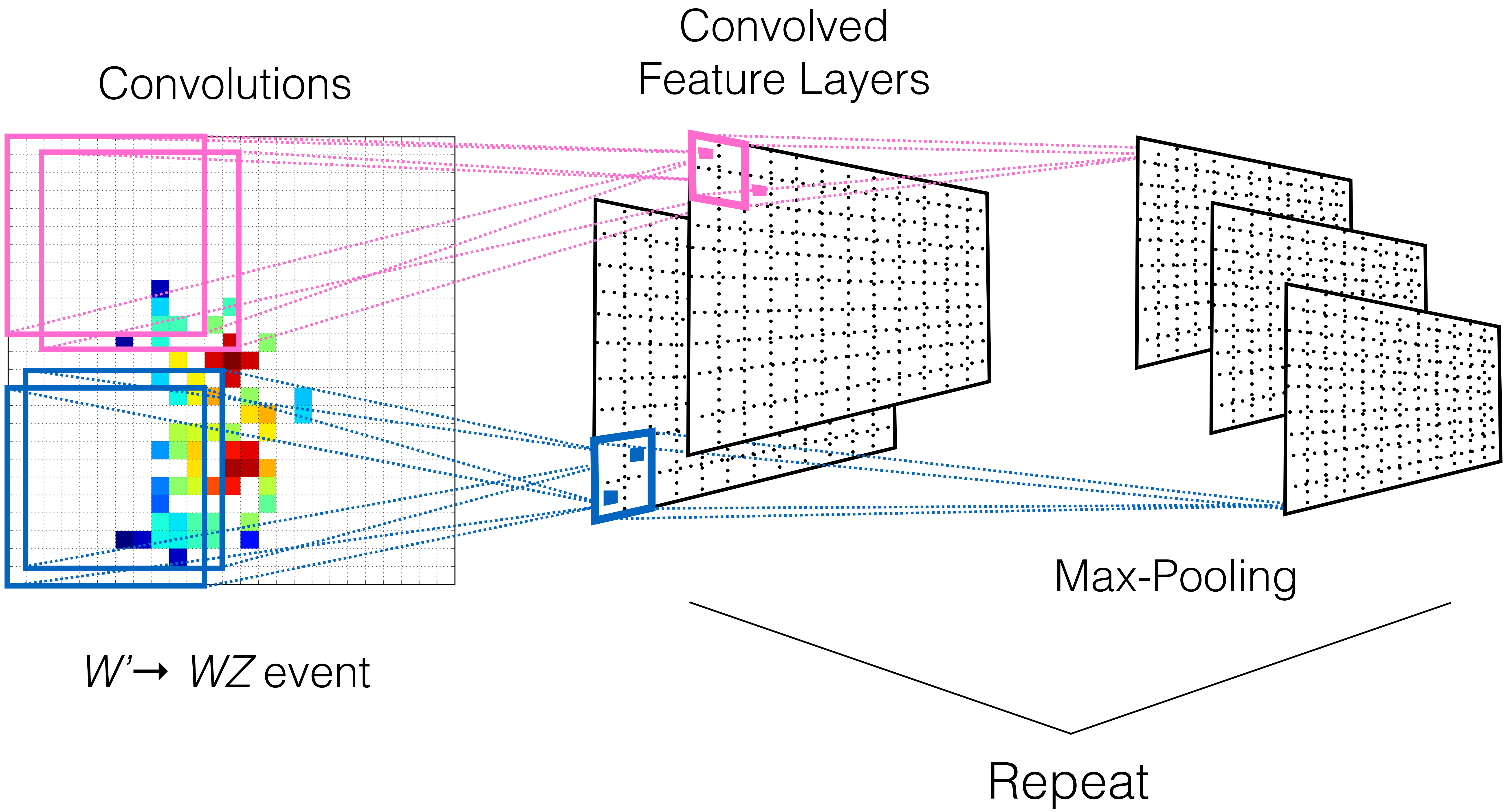






CONVOLUTIONAL NEURAL NETWORKS

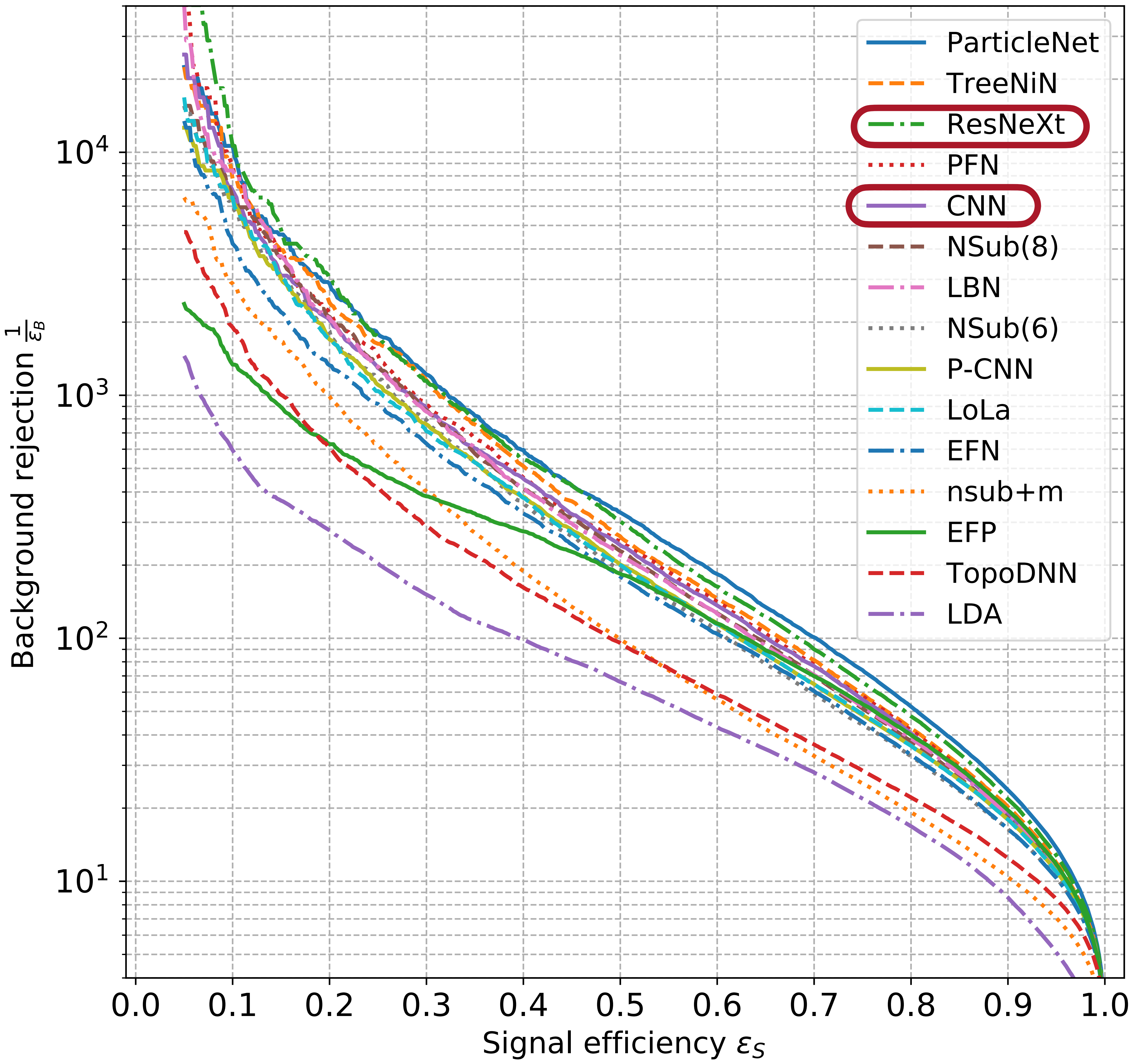
- Natural to apply 2D CNN



CNN PERFORMANCE

- CNNs among the best performing algorithms

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k



INNOVATING WITH NEW REPRESENTATIONS

arXiv:2007.13681
arXiv:2012.01249

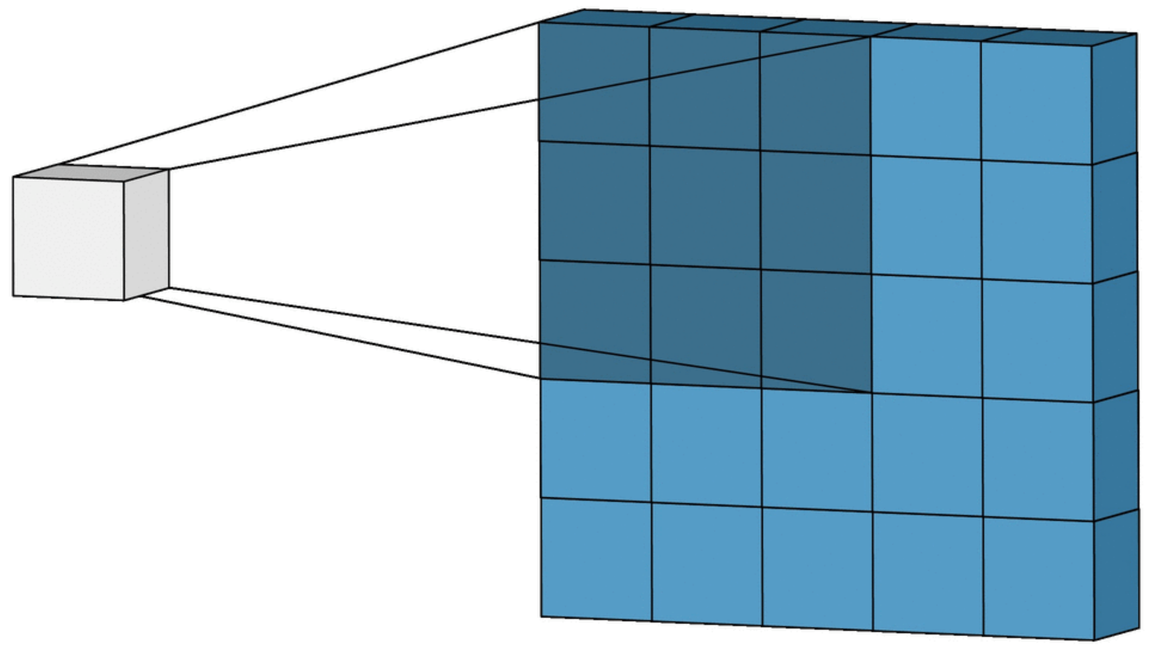
INNOVATING WITH NEW REPRESENTATIONS

arXiv:2007.13681
arXiv:2012.01249

- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance

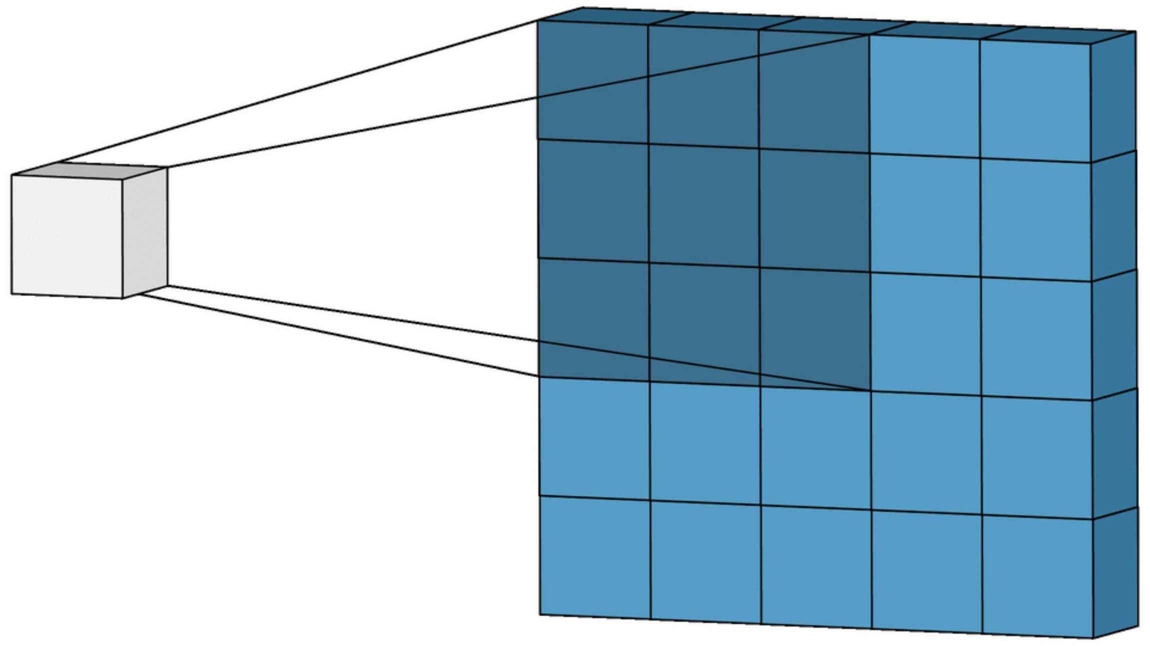
INNOVATING WITH NEW REPRESENTATIONS

- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance
  - CNNs for images

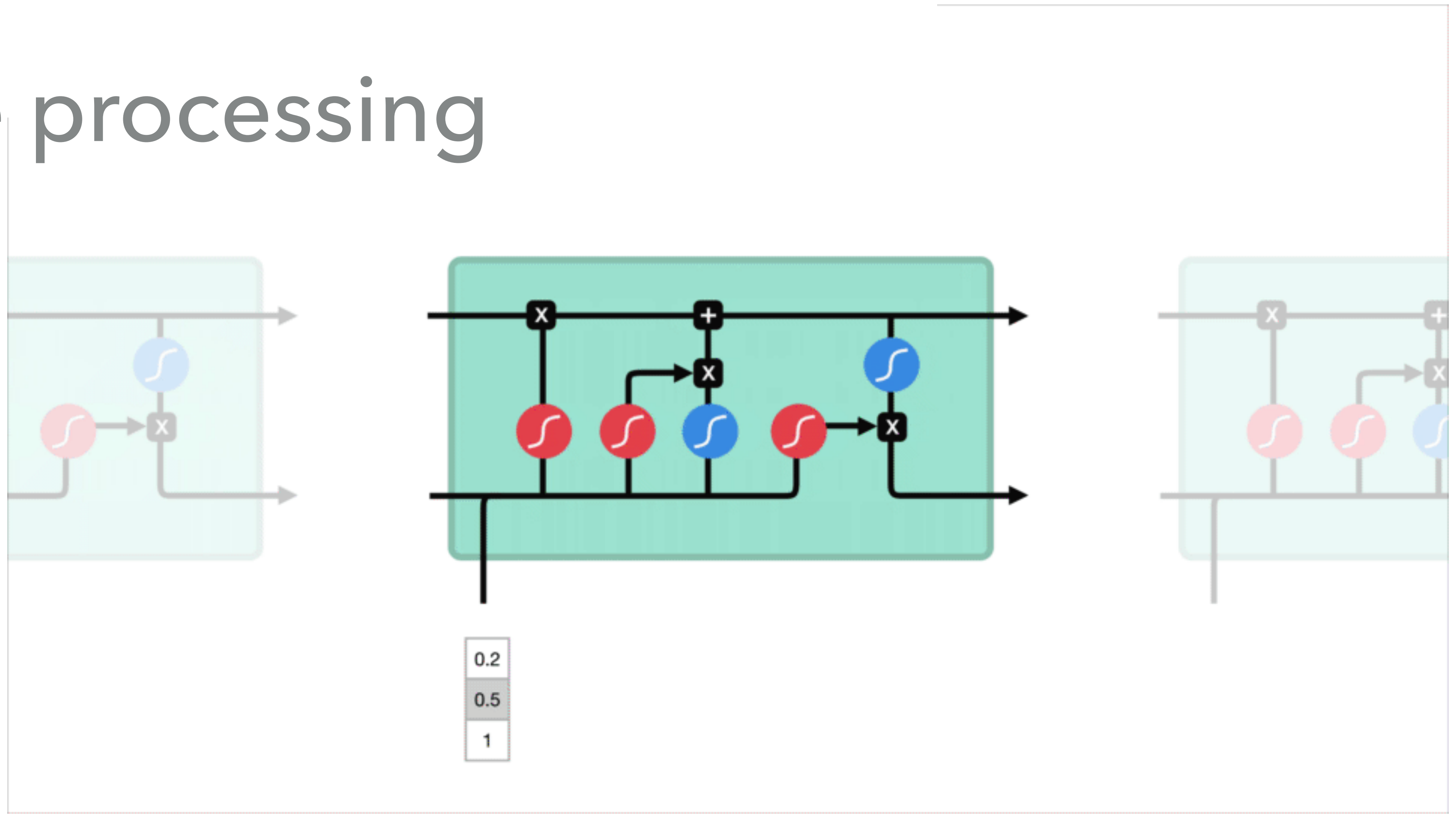


INNOVATING WITH NEW REPRESENTATIONS

- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance
  - CNNs for images

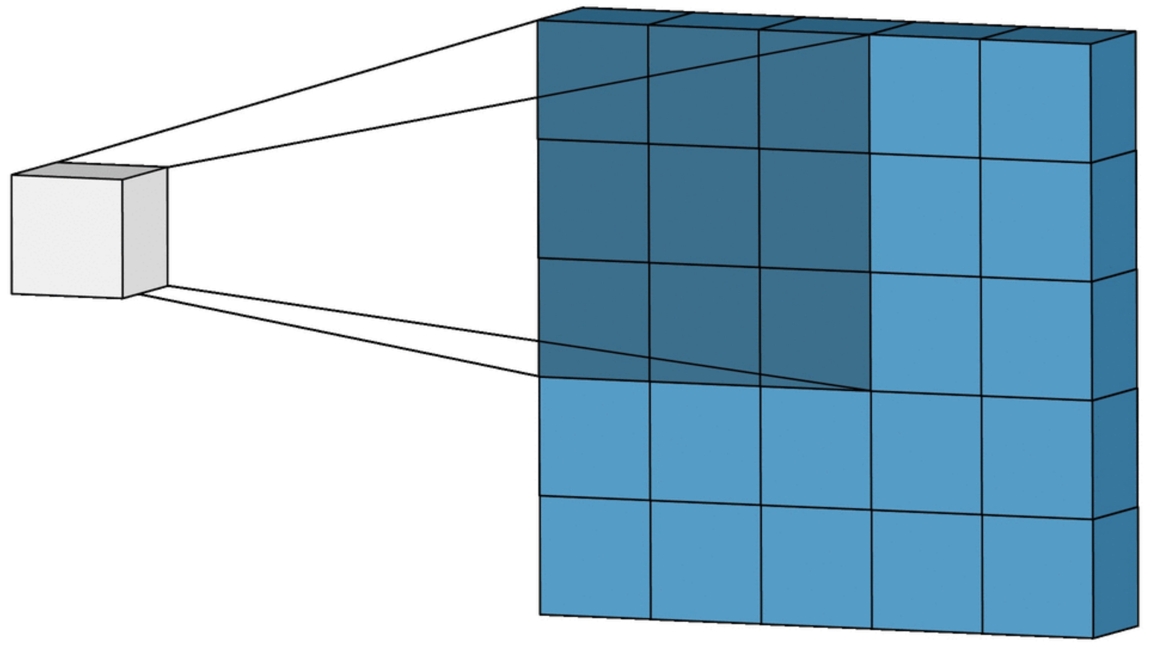


  - RNNs for language processing

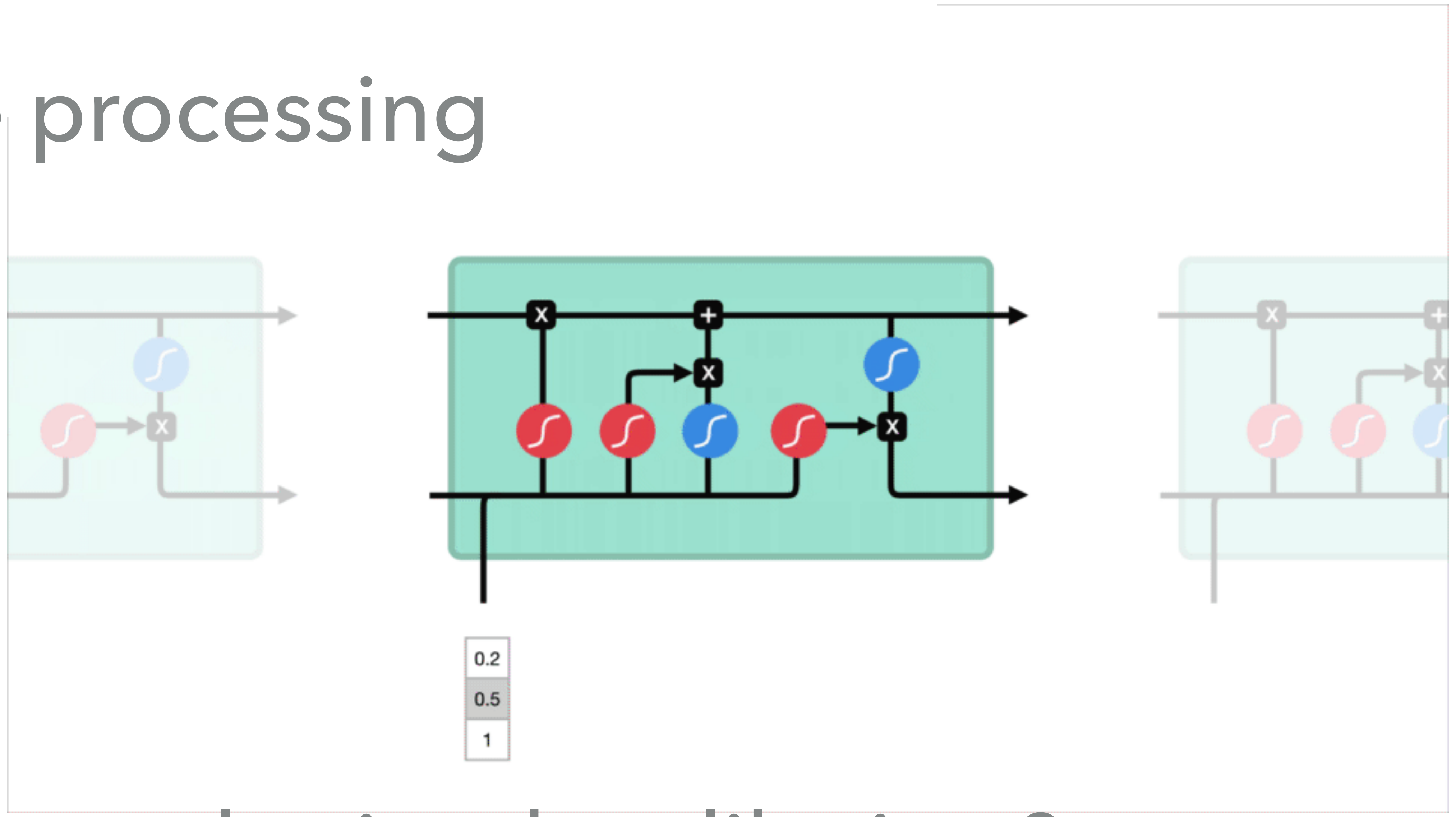


INNOVATING WITH NEW REPRESENTATIONS

- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance
  - CNNs for images



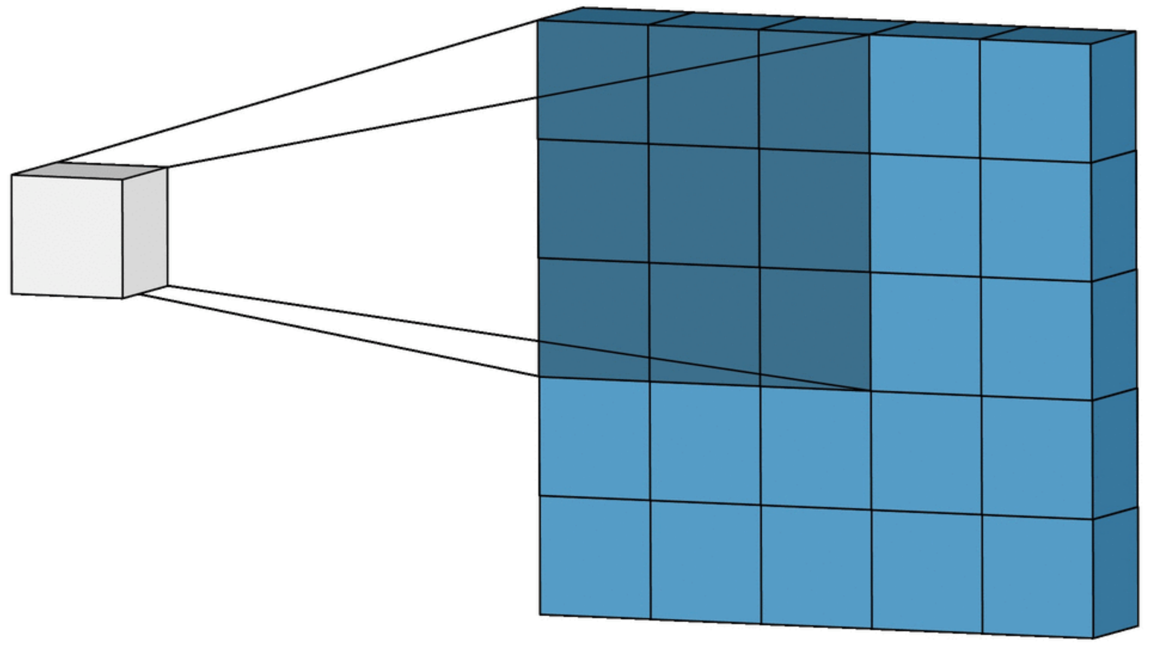
  - RNNs for language processing



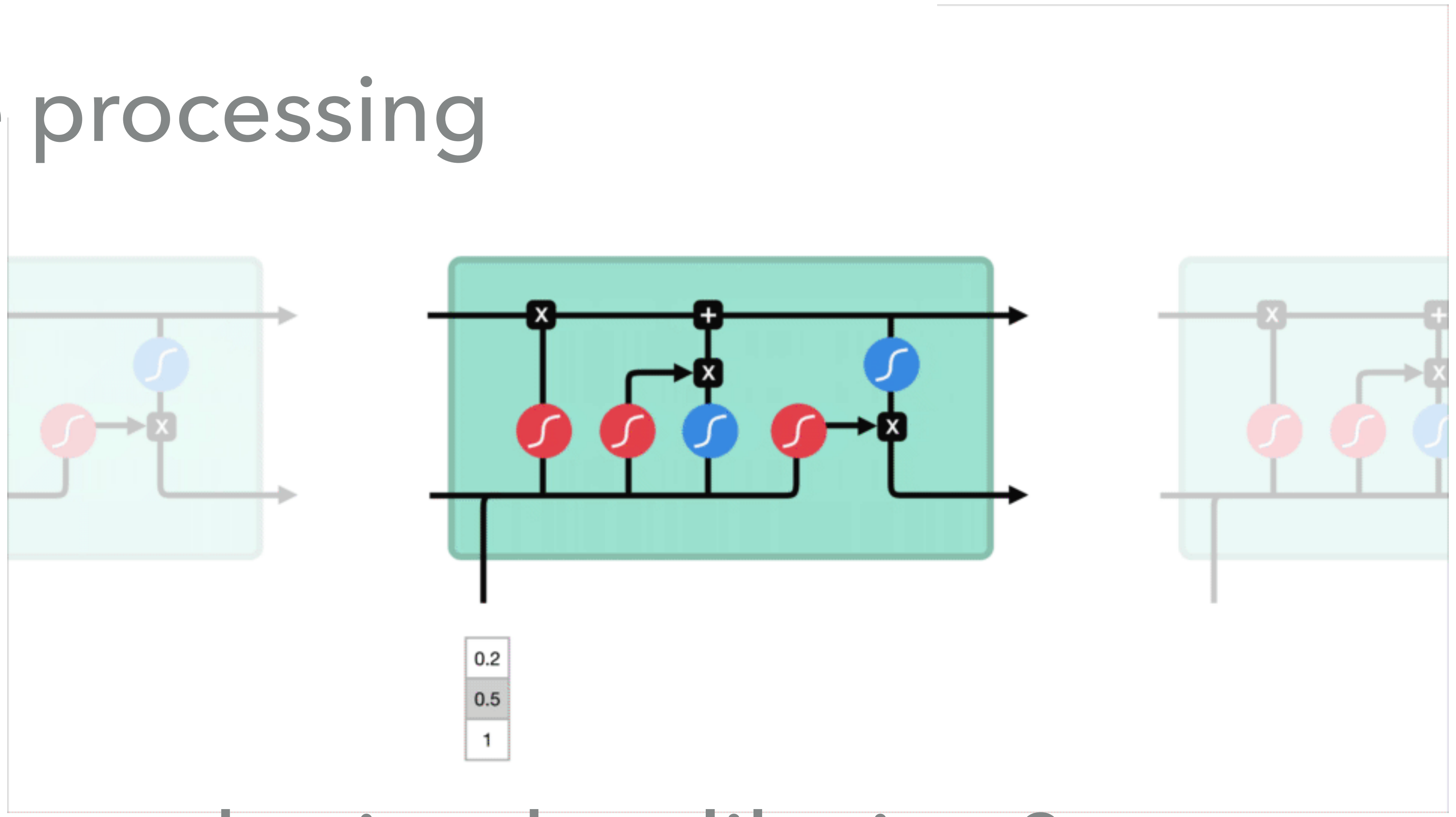
- What about high energy physics data like jets?

INNOVATING WITH NEW REPRESENTATIONS

- In deep learning, tailoring algorithms to the structure (and symmetries) of the data has led to groundbreaking performance
  - CNNs for images

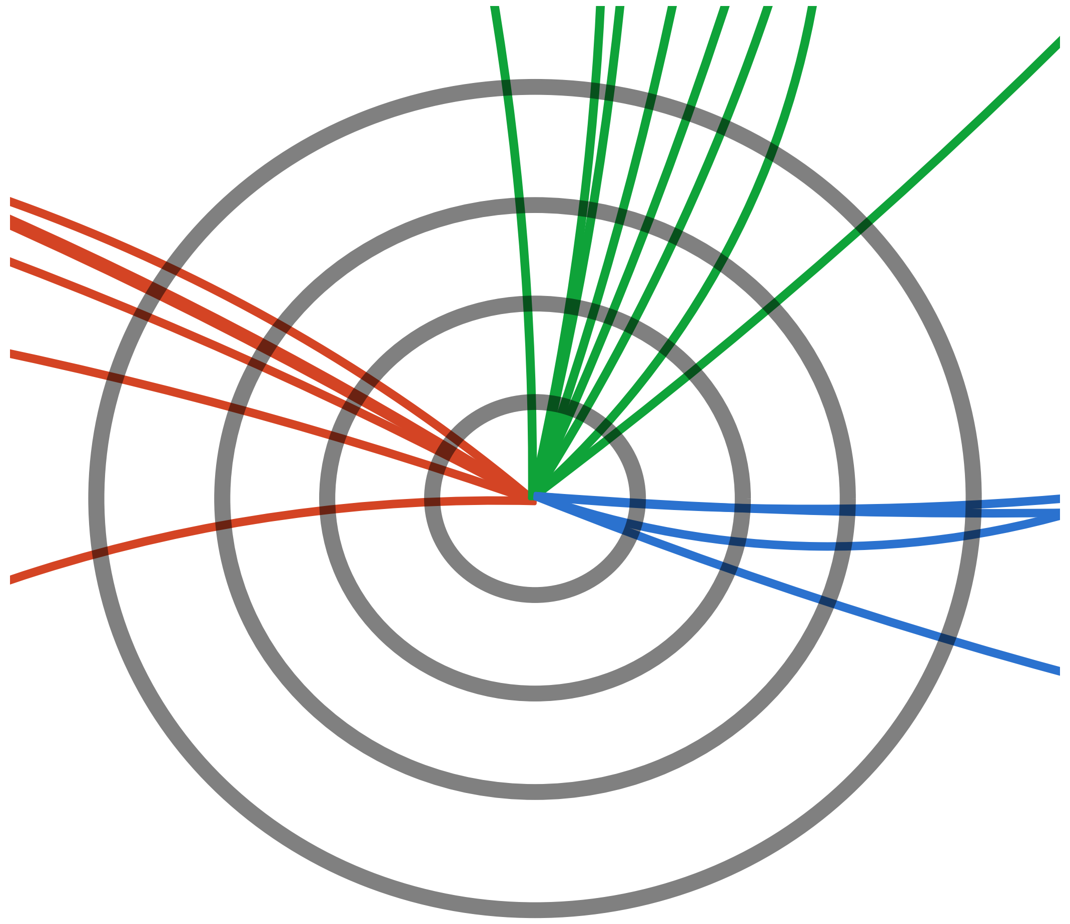


  - RNNs for language processing

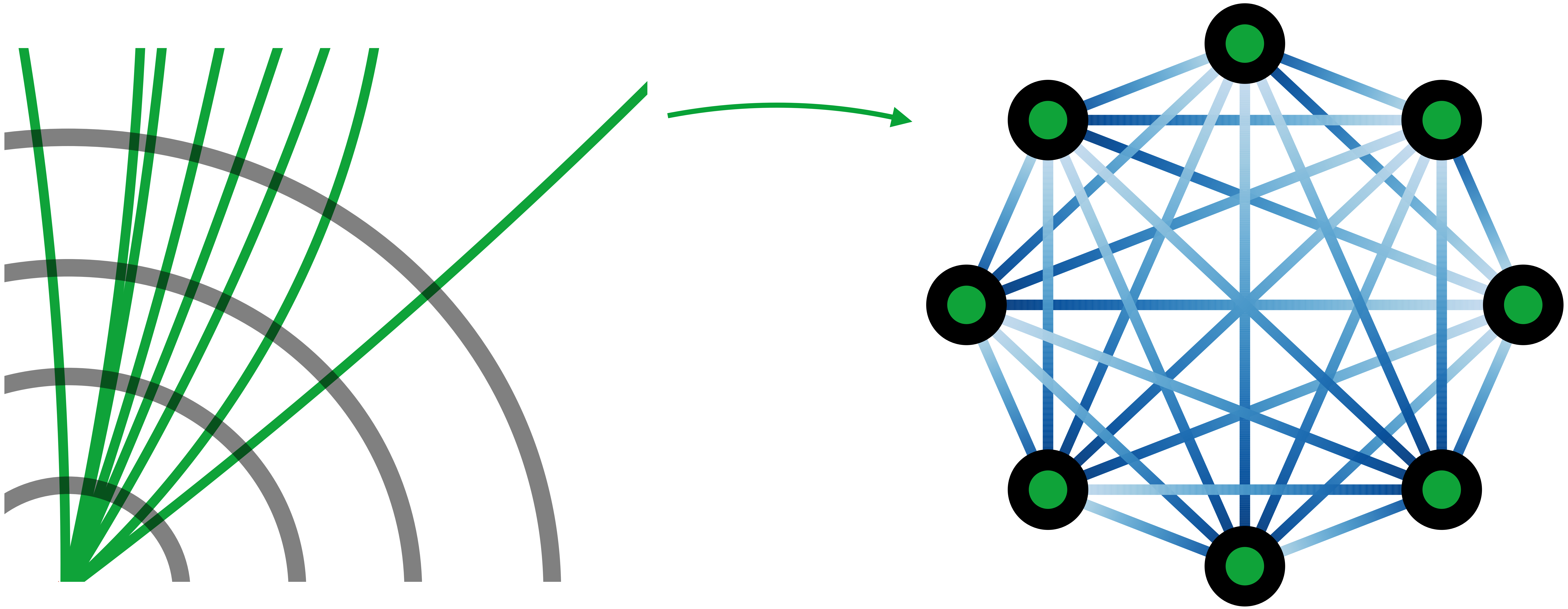


- What about high energy physics data like jets?
  - Distributed unevenly in space
  - Sparse
  - Variable size
  - No defined order
  - Interconnections

    → Graphs



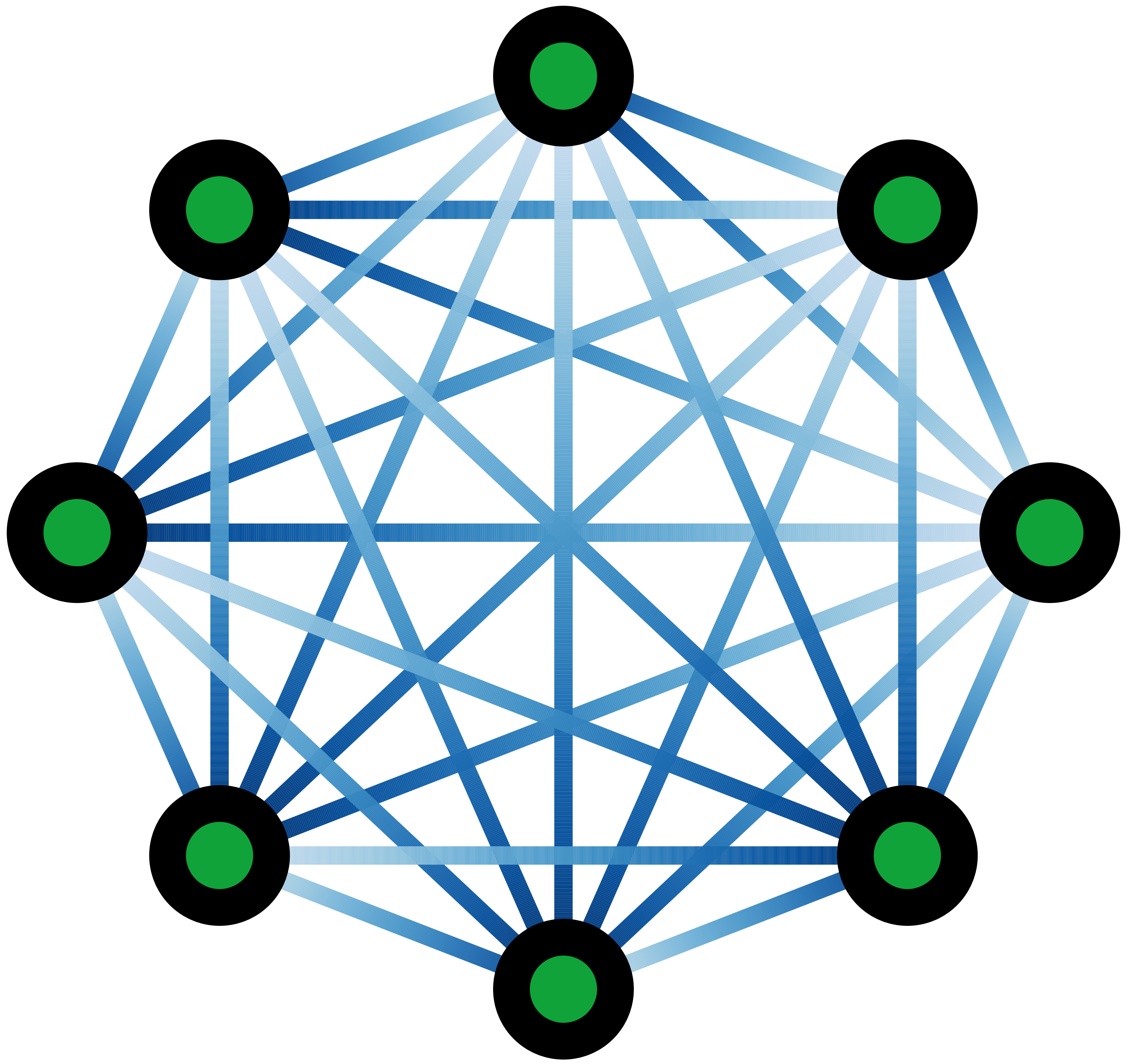
NODE, EDGE, GRAPH FEATURES IN HEP (E.G. JET)



NODE, EDGE, GRAPH FEATURES IN HEP (E.G. JET)

- Node features $\mathbf{v}_i$: particle 4-momentum

$$p = [E, p_x, p_y, p_z] \equiv [p_\mathrm{T}, \eta, \phi, m]$$



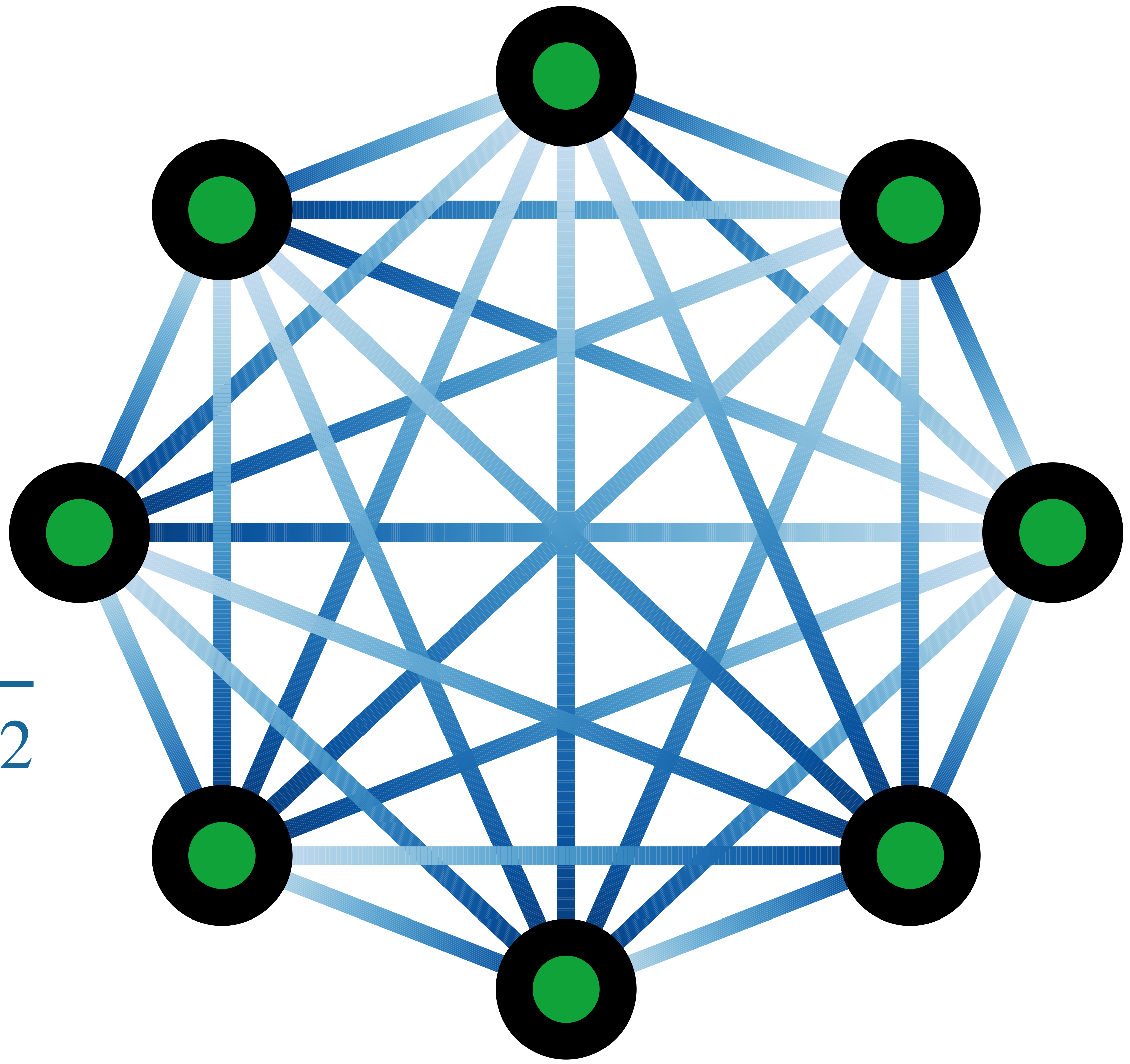
NODE, EDGE, GRAPH FEATURES IN HEP (E.G. JET)

- Node features $\mathbf{v}_i$: particle 4-momentum

$$p = [E, p_x, p_y, p_z] \equiv [p_\mathrm{T}, \eta, \phi, m]$$

- Edge features $\mathbf{e}_k$: pseudoangular distance between particles

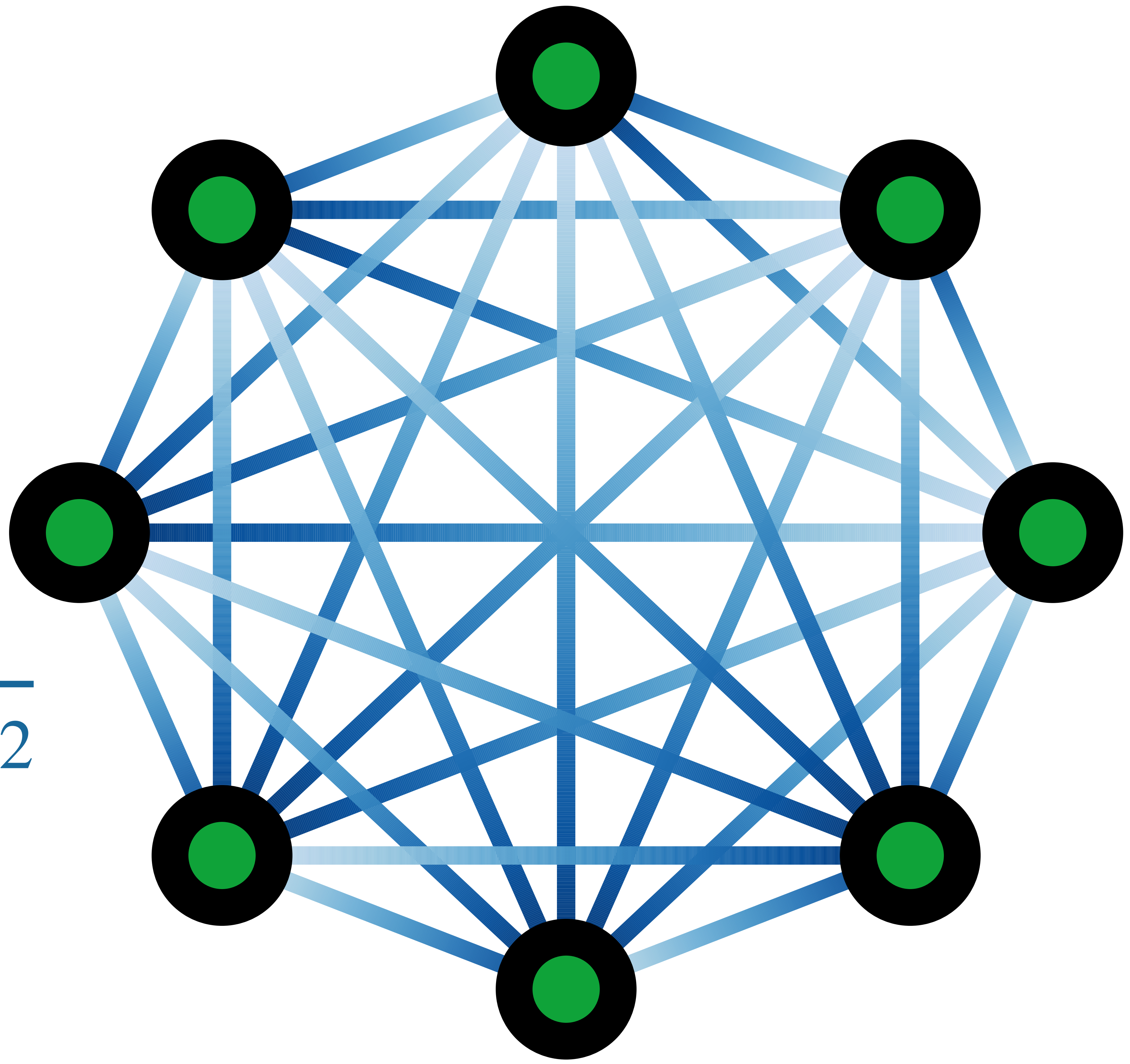
$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$



NODE, EDGE, GRAPH FEATURES IN HEP (E.G. JET)

- Node features $\mathbf{v}_i$: particle 4-momentum

$$p = [E, p_x, p_y, p_z] \equiv [p_\mathrm{T}, \eta, \phi, m]$$



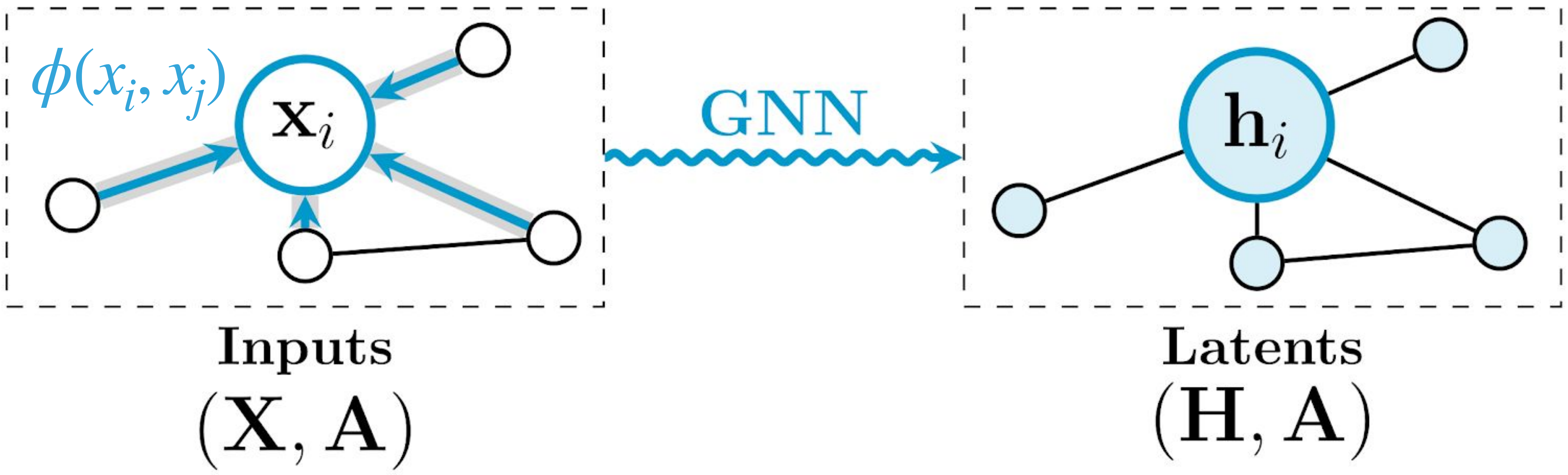
- Edge features $\mathbf{e}_k$: pseudoangular distance between particles

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

- Graph (global) features $\mathbf{u}$: jet mass

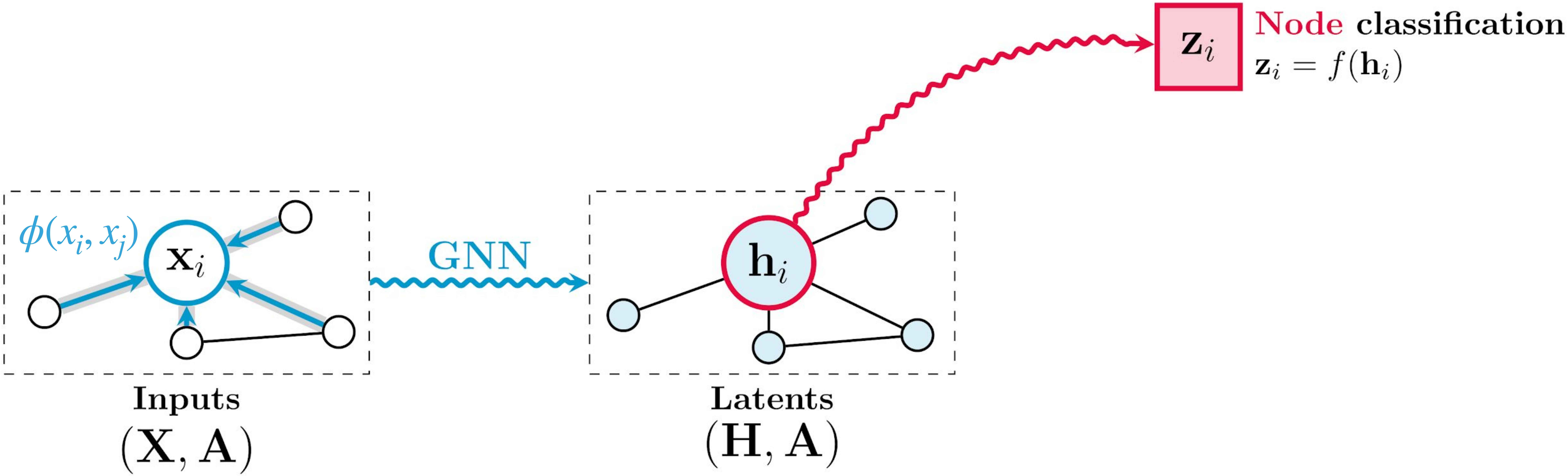
$$m = \sqrt{\sum_{i \in \mathrm{jet}} E_i^2 - p_{x,i}^2 - p_{y,i}^2 - p_{z,i}^2}$$

HOW TO USE GNNS IN HEP



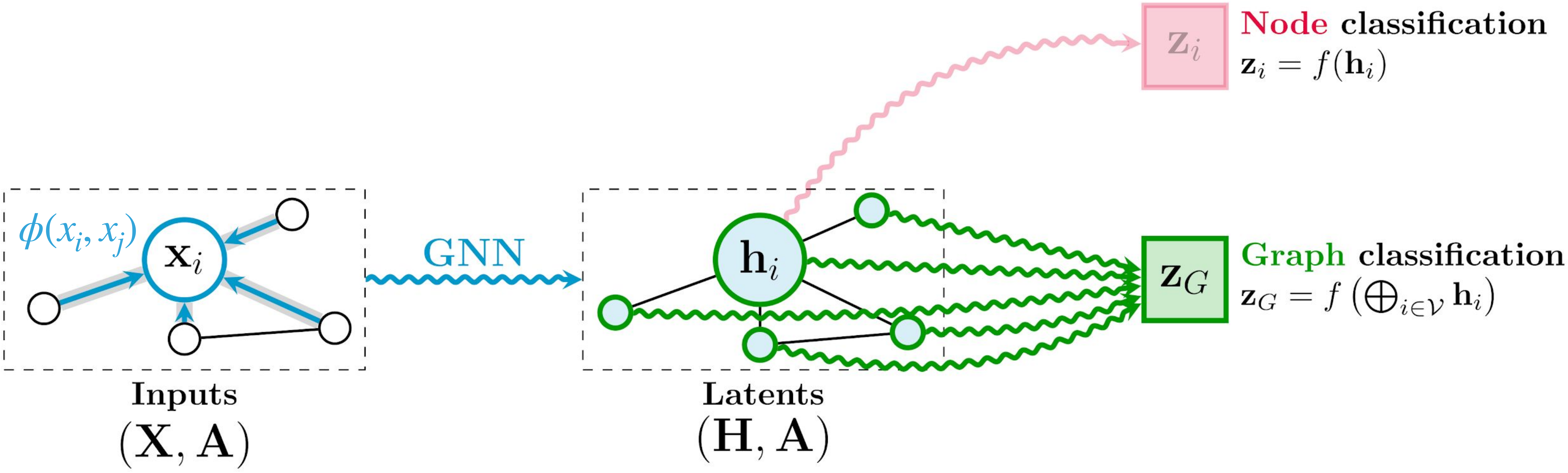
HOW TO USE GNNS IN HEP

- Node-level tasks
  - Identify "pileup" particles



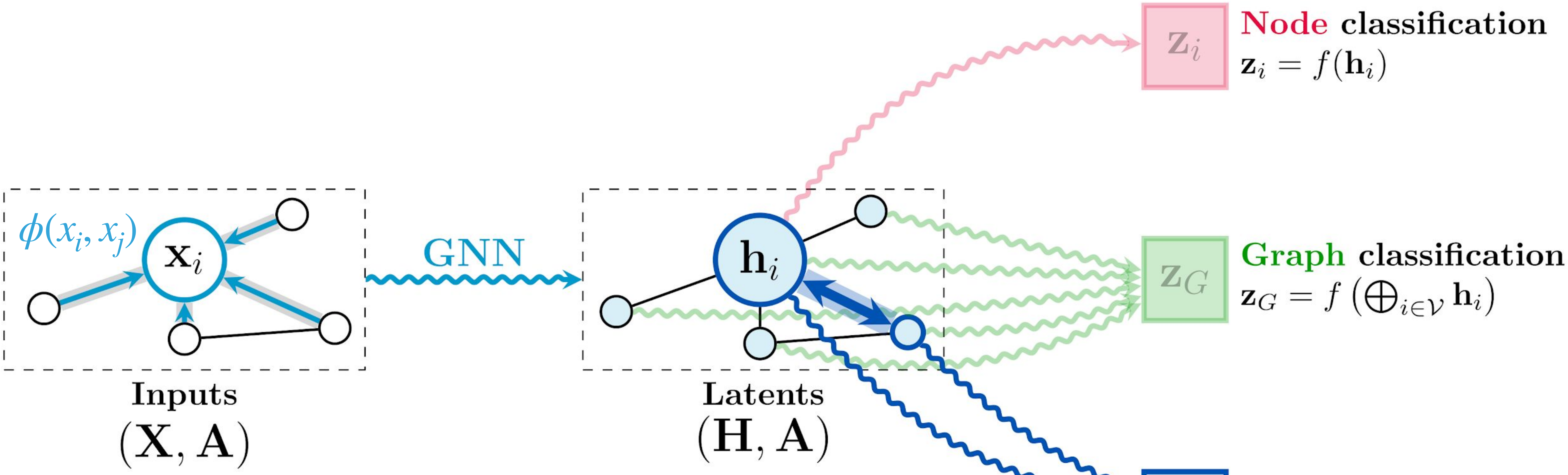
Source: https://youtu.be/uF53xsT7mjc

- Node-level tasks
  - Identify "pileup" particles
- Graph-level tasks
  - **Jet tagging**



HOW TO USE GNNS IN HEP

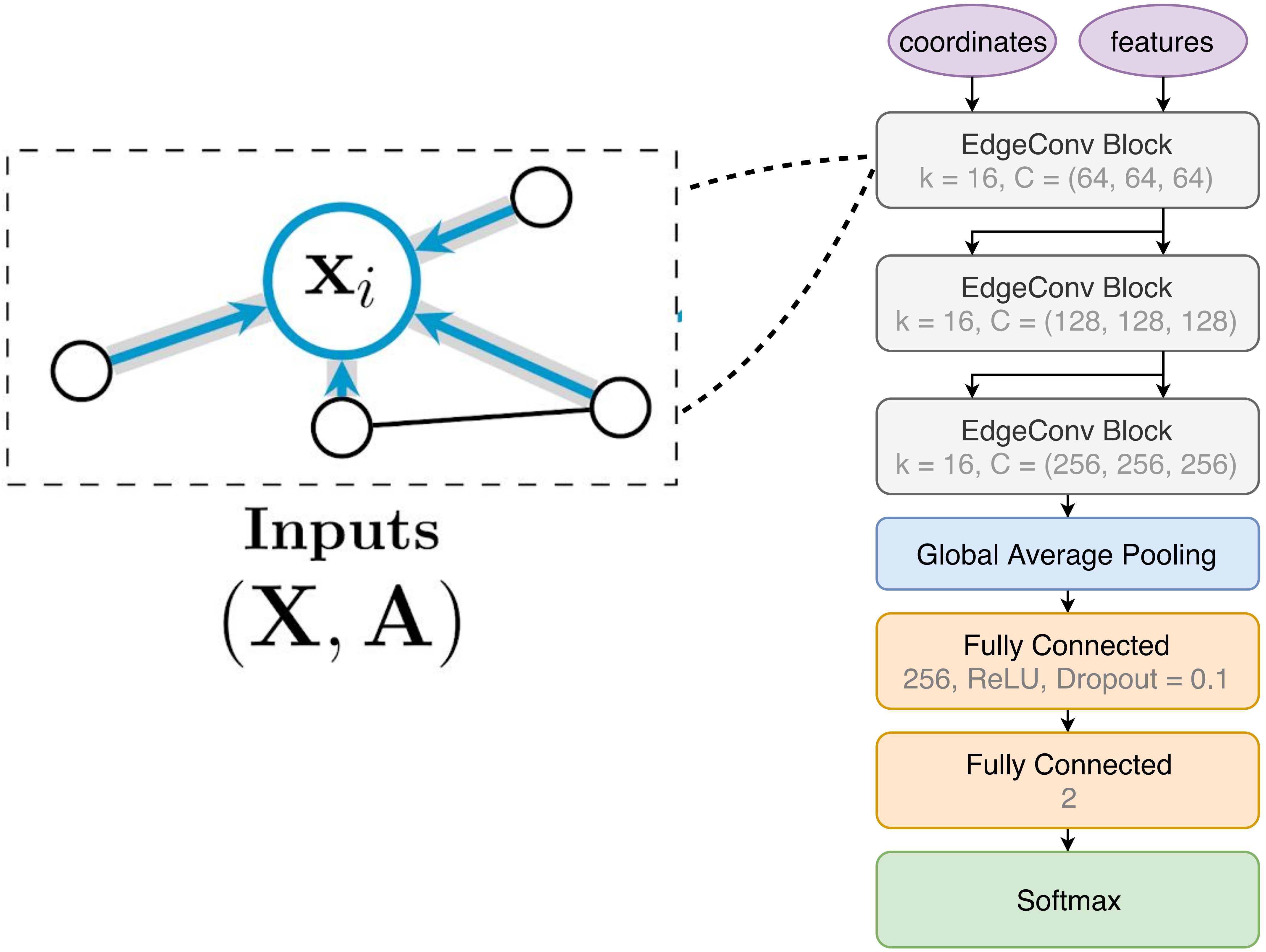
- Node-level tasks
  - Identify "pileup" particles
- Graph-level tasks
  - **Jet tagging**



- Edge-level tasks
  - Identify good track doublets

PARTICLENET: GNN FOR TAGGING H(BB) IN CMS

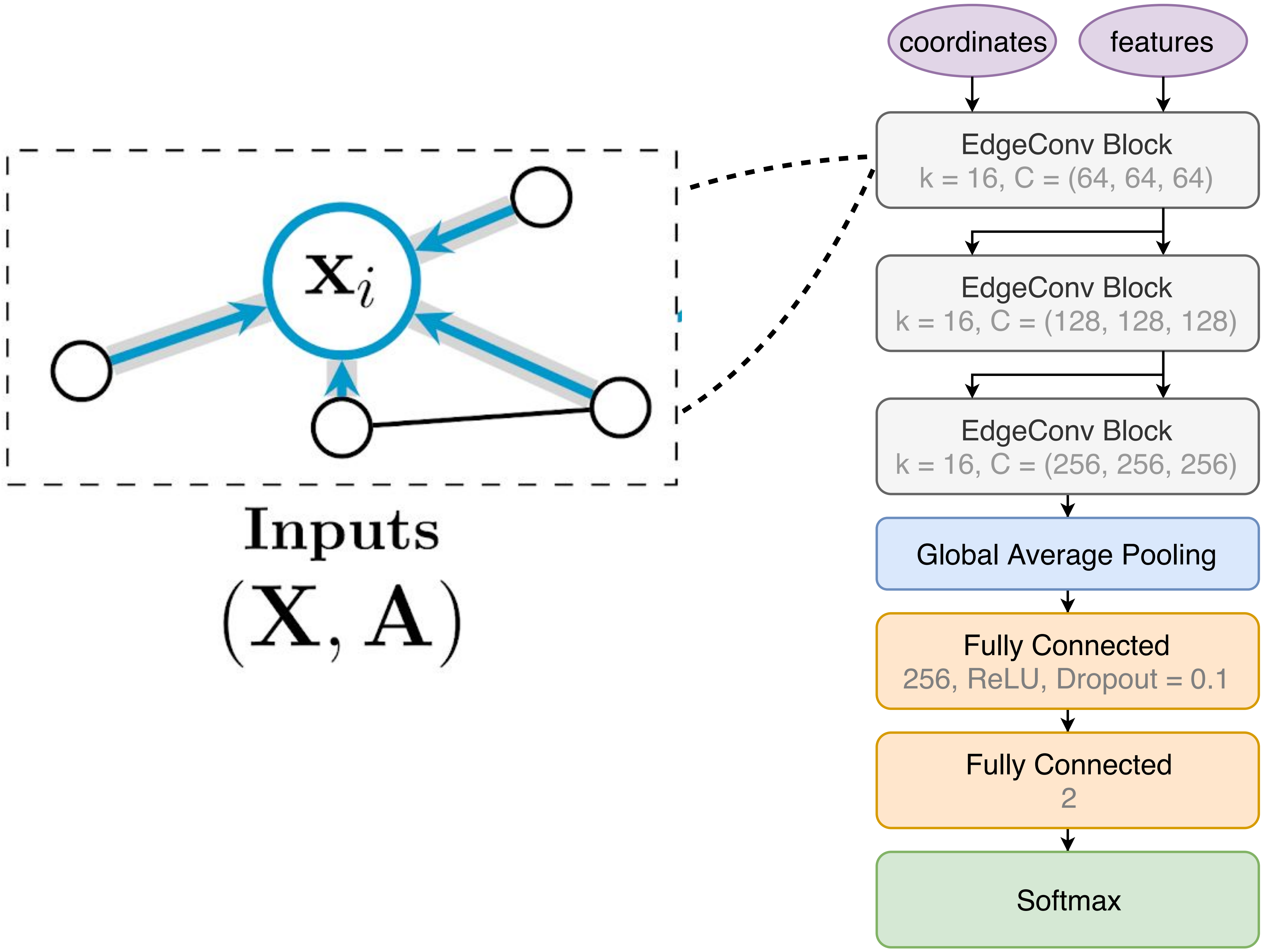
arXiv:1902.08570
CMS-DP-2020-002



PARTICLENET: GNN FOR TAGGING H(BB) IN CMS

arXiv:1902.08570
CMS-DP-2020-002

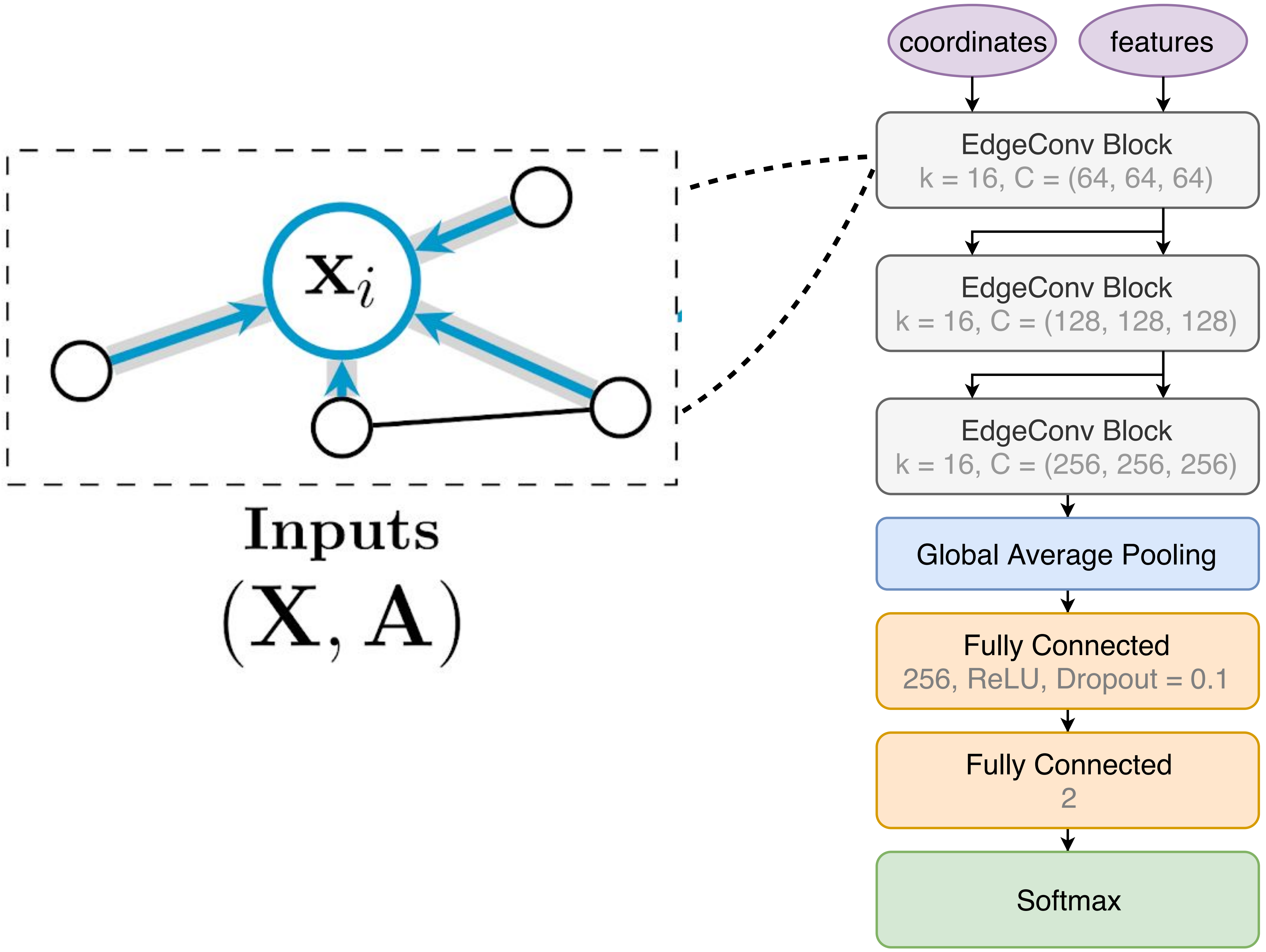
- ParticleNet, using "dynamic edge convolutions:" graph is constructed based on "closeness" in an abstract "latent" space

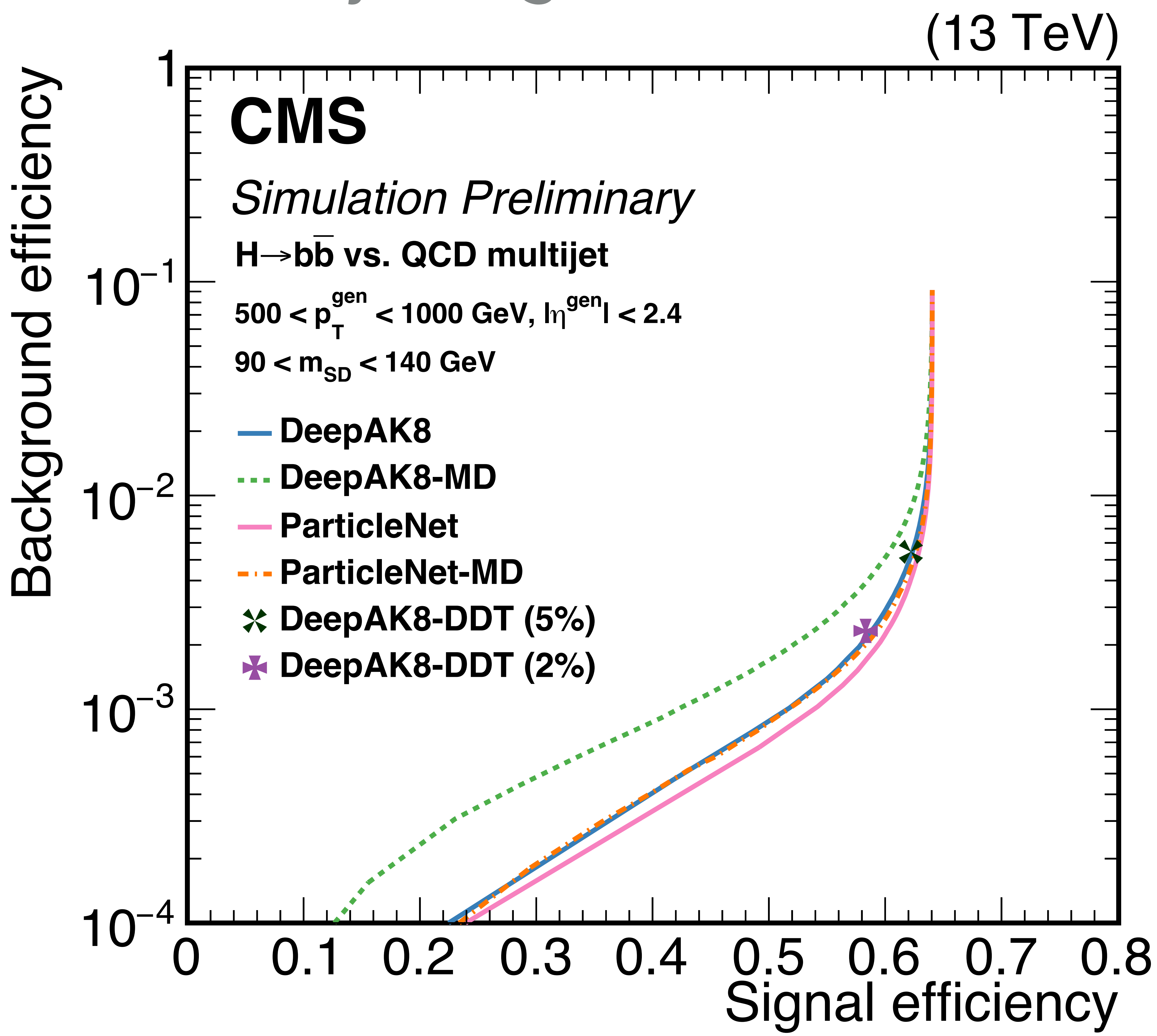


PARTICLENET: GNN FOR TAGGING H(BB) IN CMS

arXiv:1902.08570
CMS-DP-2020-002

- ParticleNet, using "dynamic edge convolutions:" graph is constructed based on "closeness" in an abstract "latent" space
- Identifies H(bb) with an efficiency of ~50% while rejecting 99.9% of background

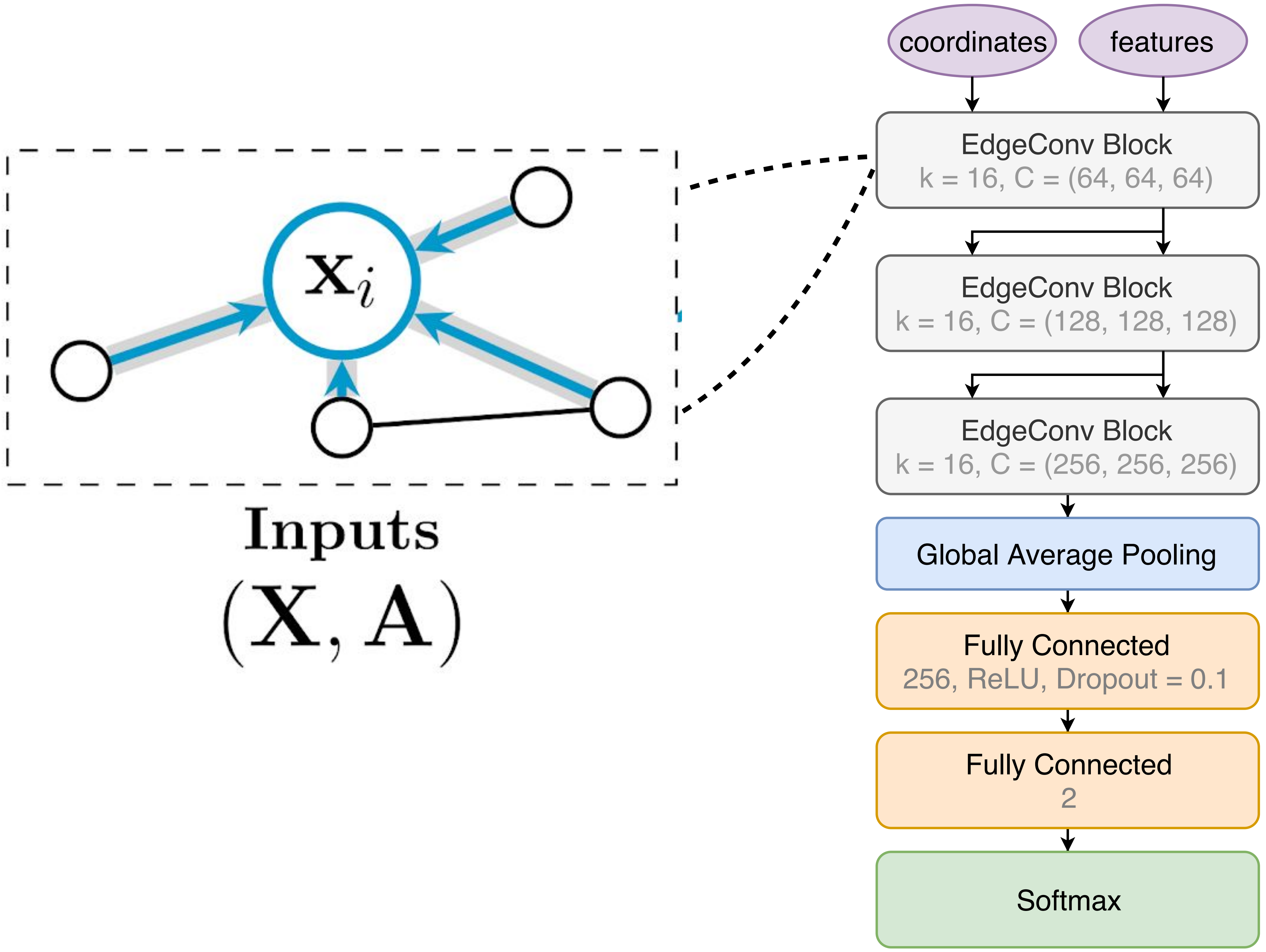


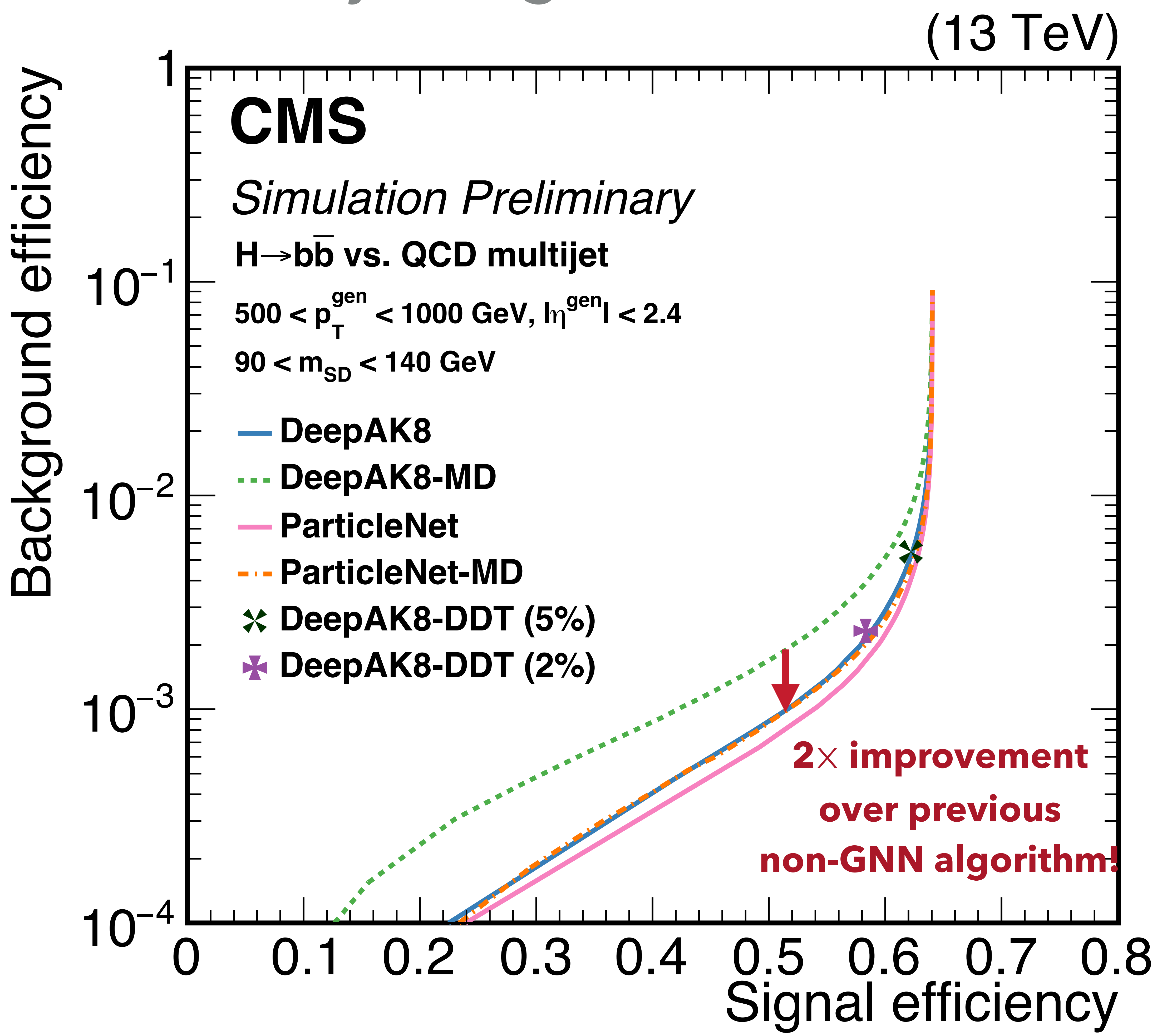


PARTICLENET: GNN FOR TAGGING H(BB) IN CMS

arXiv:1902.08570
CMS-DP-2020-002

- ParticleNet, using "dynamic edge convolutions:" graph is constructed based on "closeness" in an abstract "latent" space
- Identifies H(bb) with an efficiency of ~50% while rejecting 99.9% of background





INDUCTIVE BIAS & EQUIVARIANCE

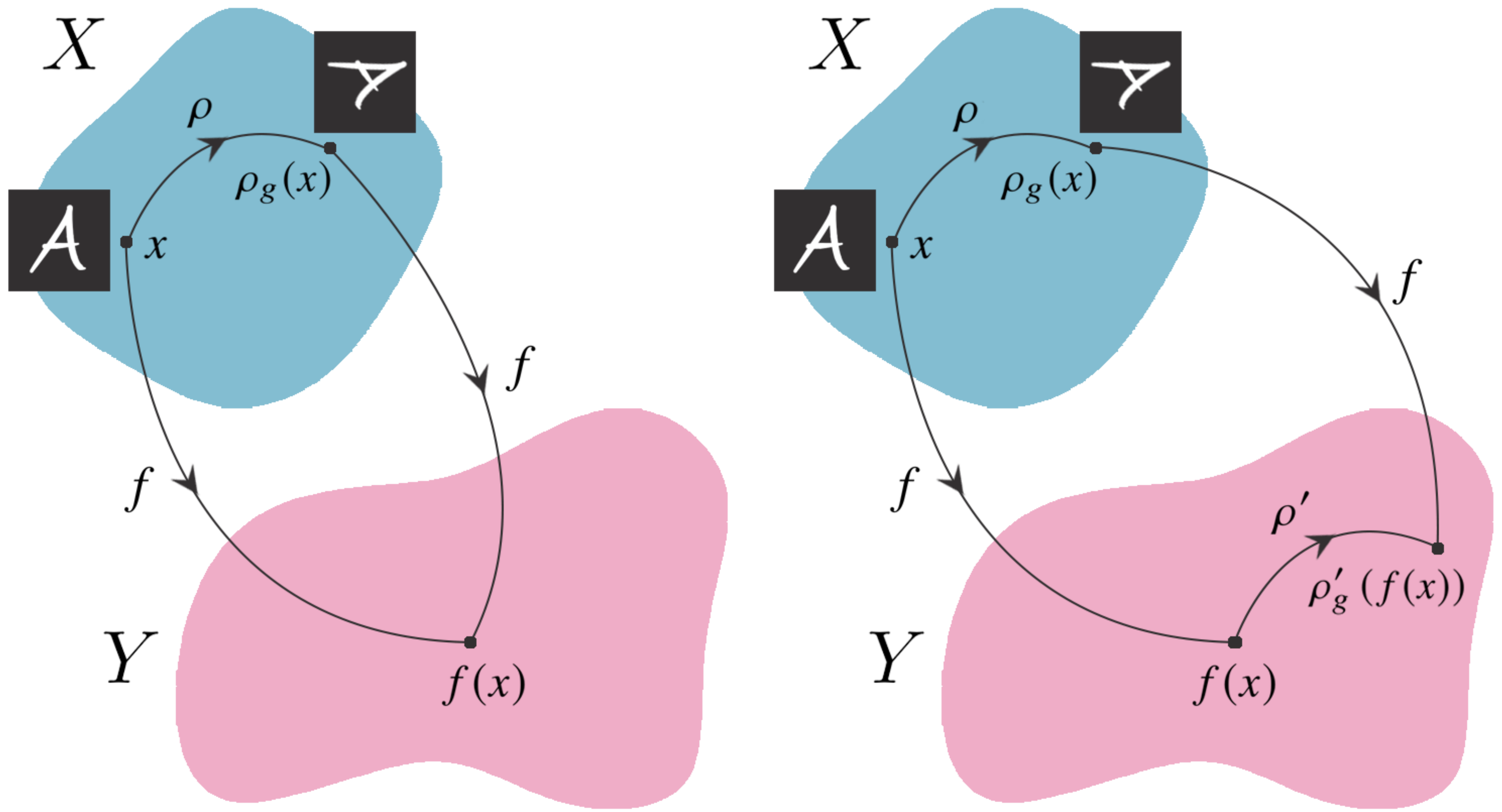
- Symmetry-equivariant networks
  - More economical (fewer, but more expressive parameters), interpretable, and trainable

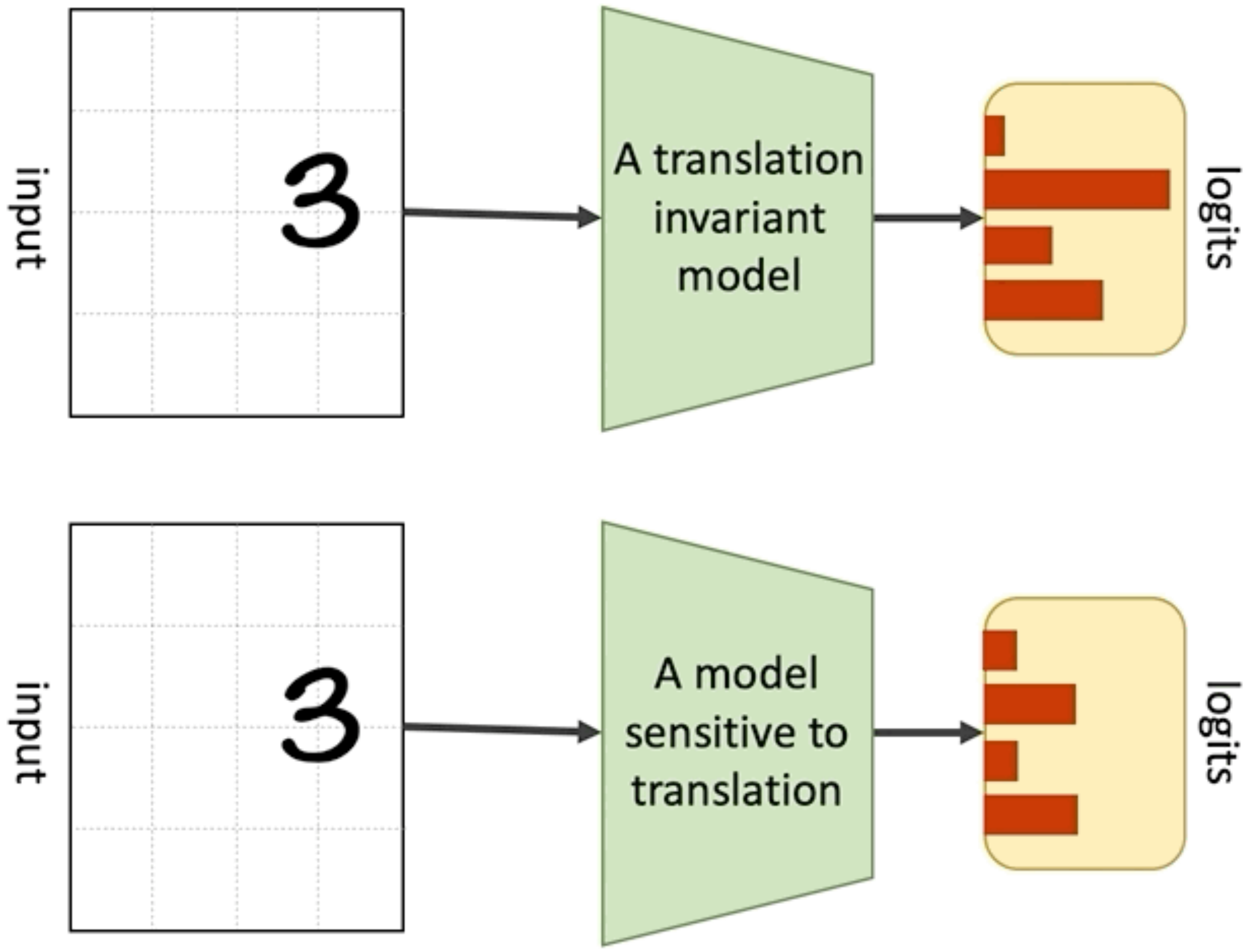
Invariance

$$f(\rho_g(x)) = f(x)$$

Equivariance

$$f(\rho_g(x)) = \rho'_g\,(f(x))$$





INDUCTIVE BIAS & EQUIVARIANCE

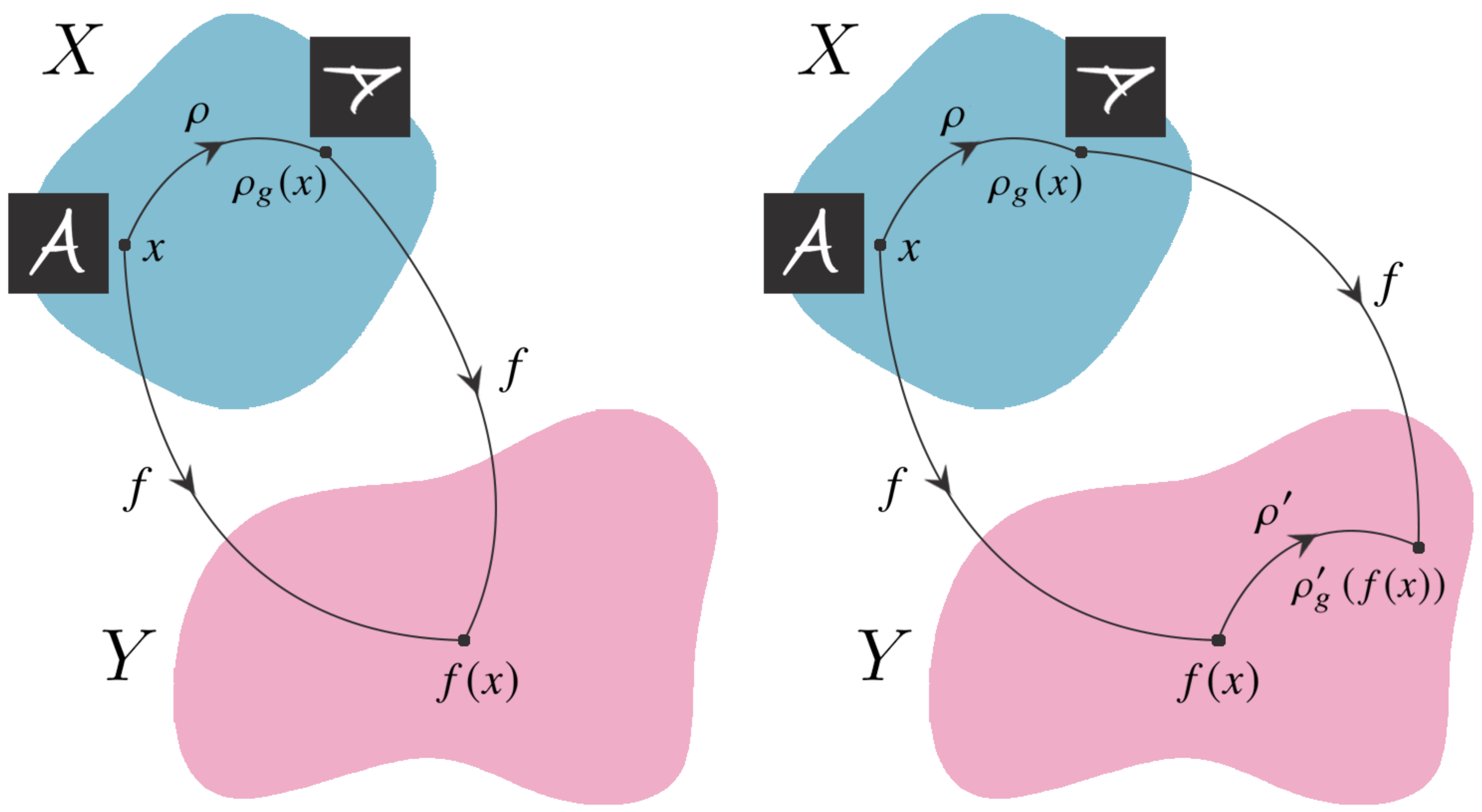
- Symmetry-equivariant networks
  - More economical (fewer, but more expressive parameters), interpretable, and trainable

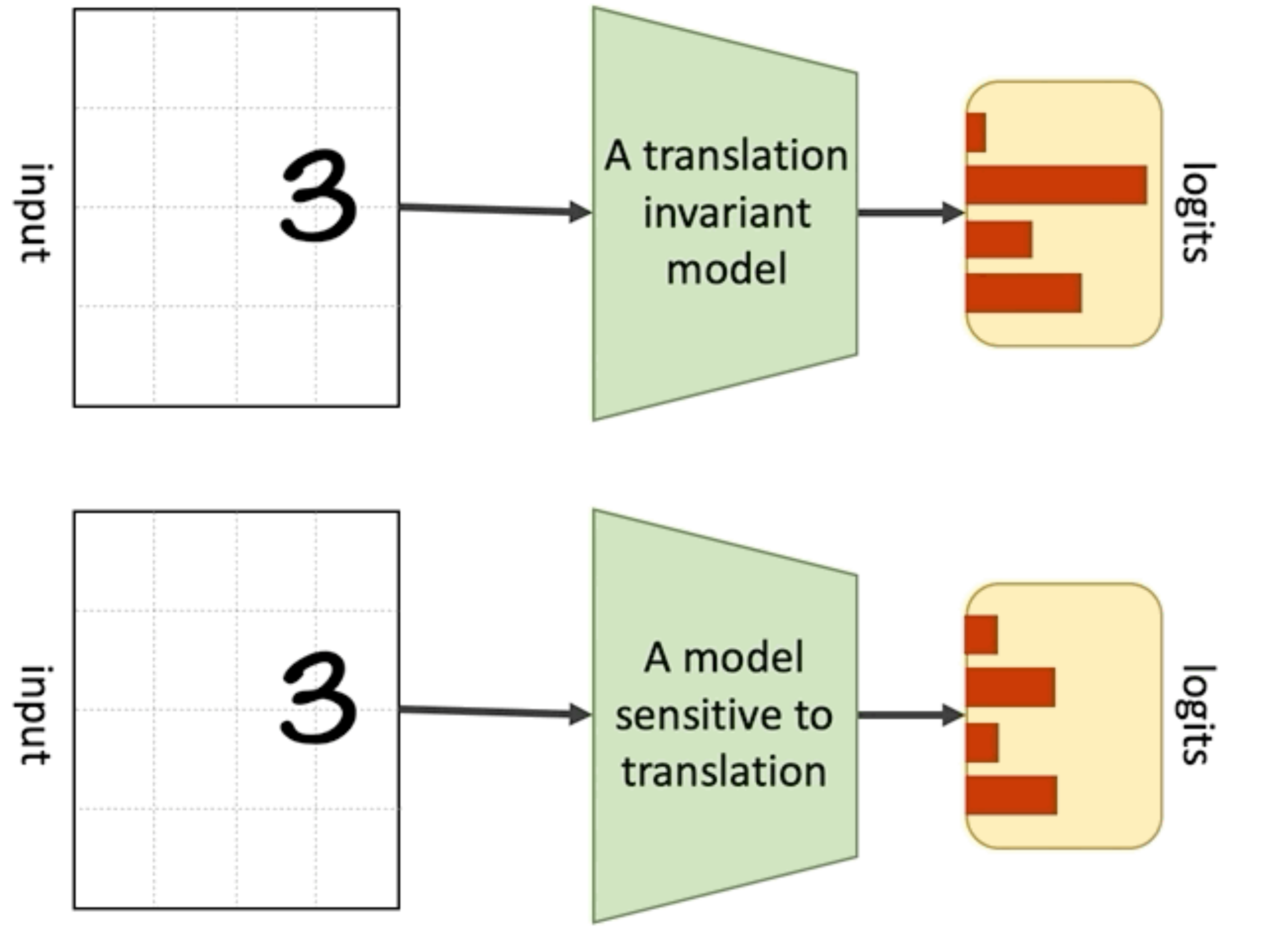
Invariance

$$f(\rho_g(x)) = f(x)$$

Equivariance

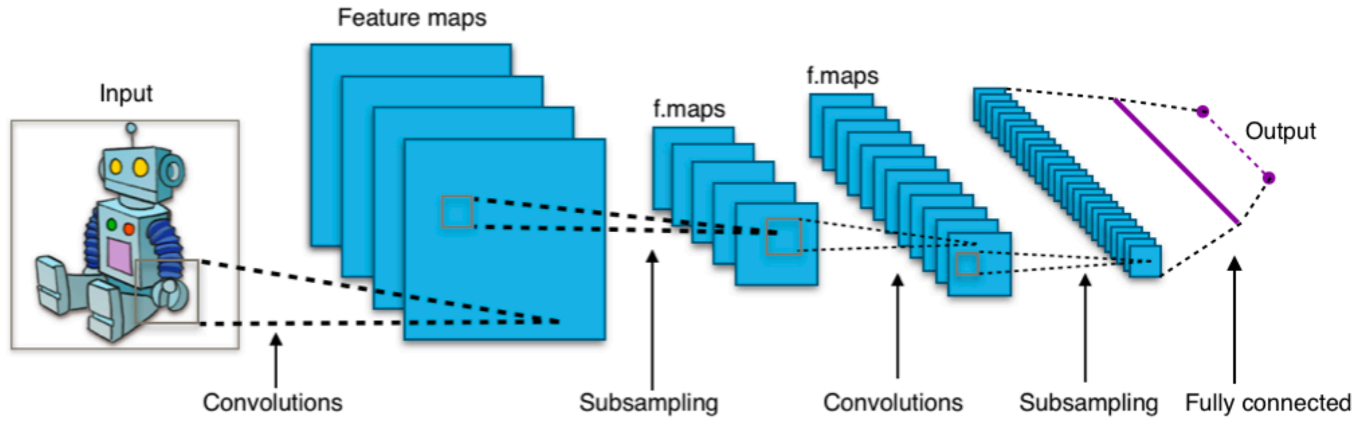
$$f(\rho_g(x)) = \rho'_g\,(f(x))$$





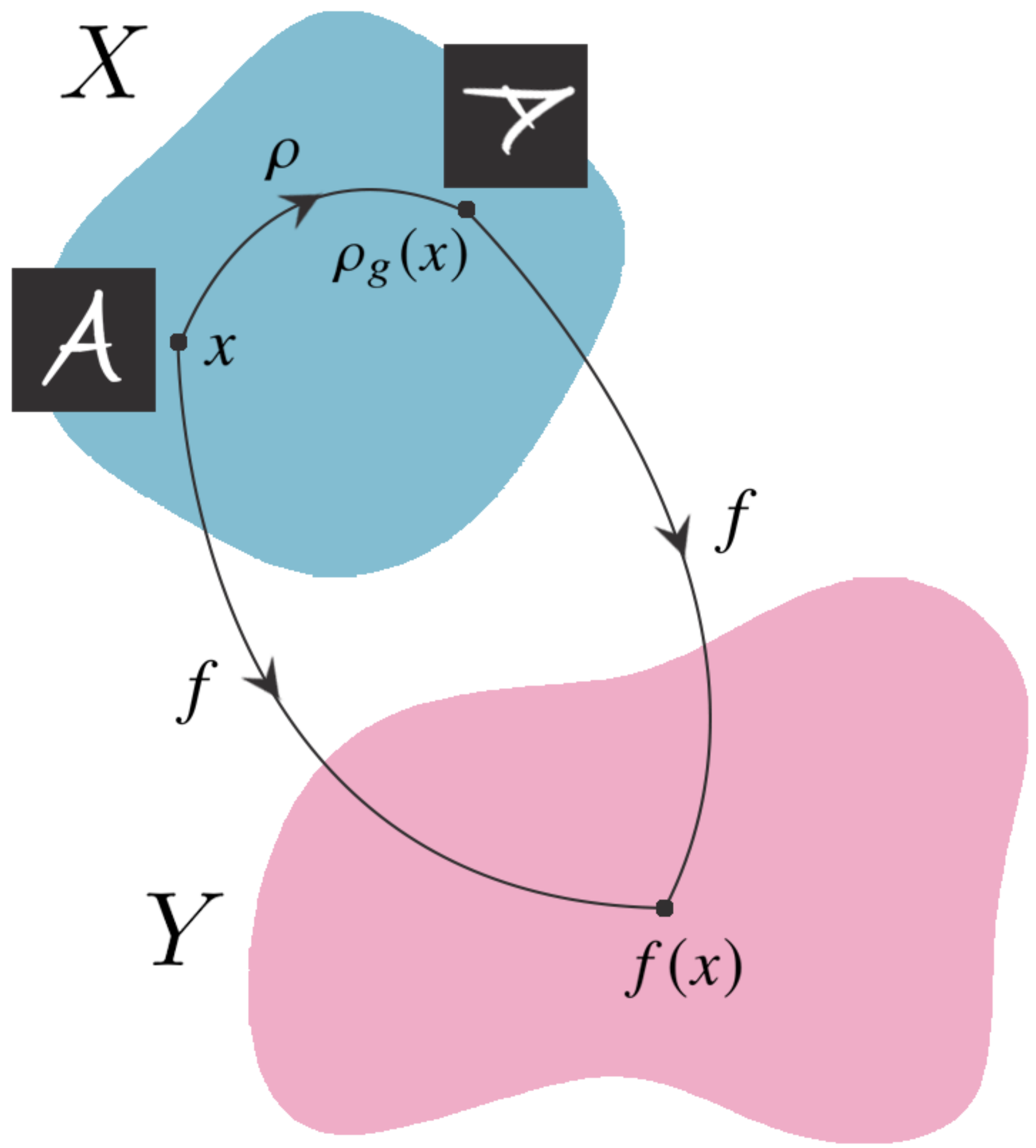
INDUCTIVE BIAS & EQUIVARIANCE

- Symmetry-equivariant networks
  - More economical (fewer, but more expressive parameters), interpretable, and trainable



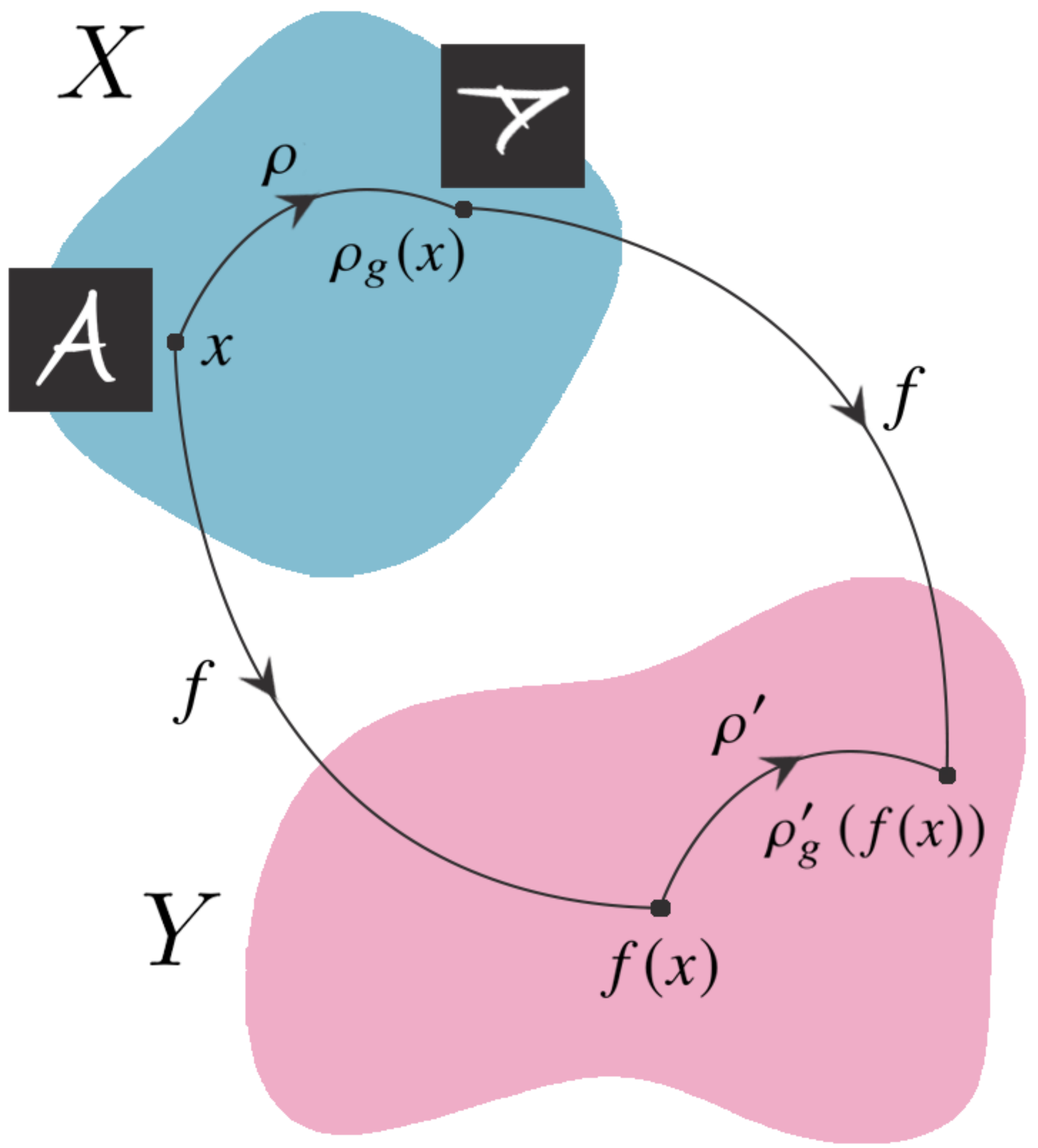
Invariance

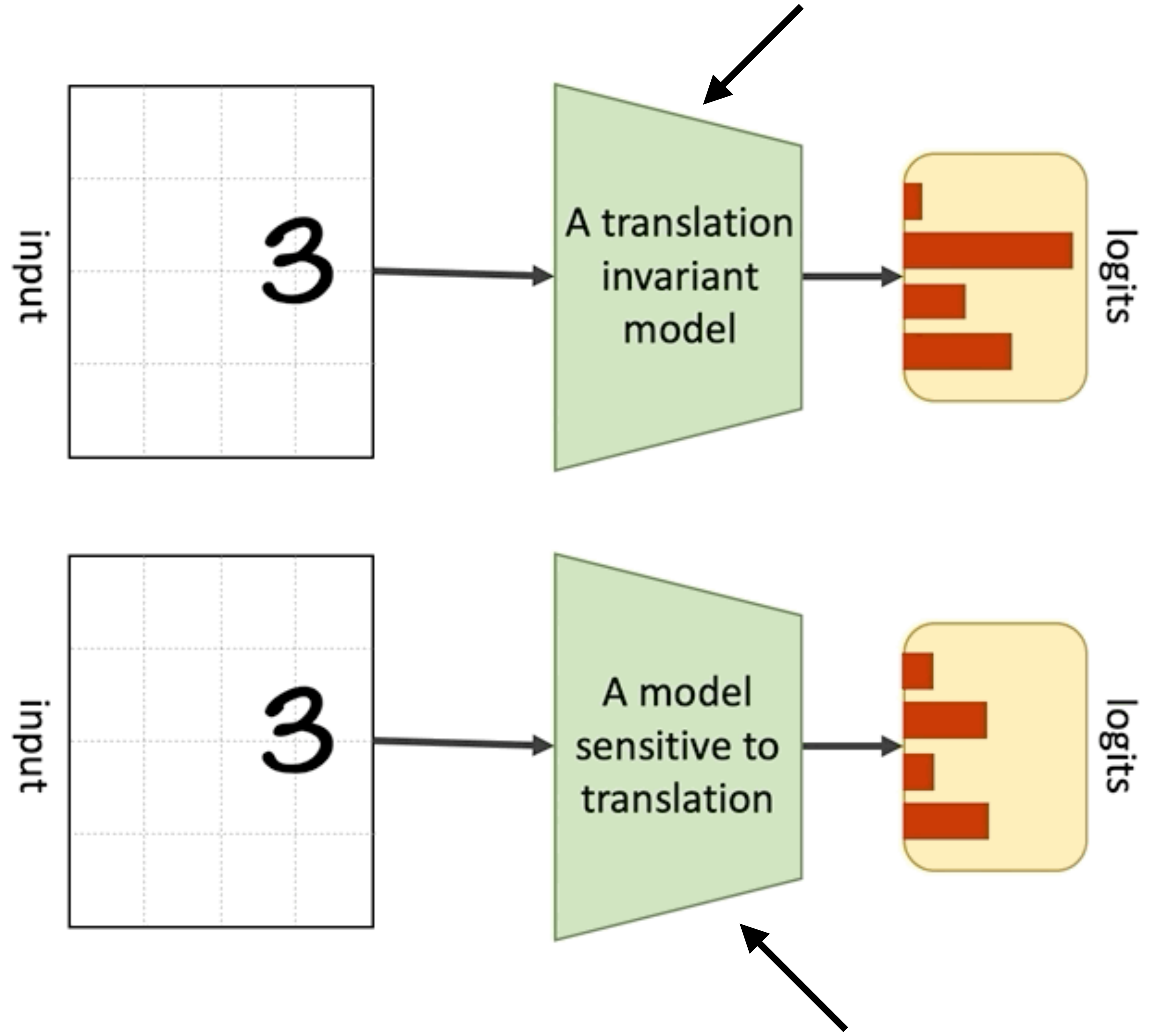
$$f(\rho_g(x)) = f(x)$$

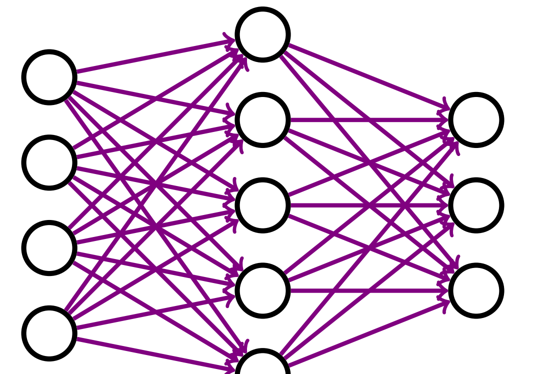


Equivariance

$$f(\rho_g(x)) = \rho'_g(f(x))$$

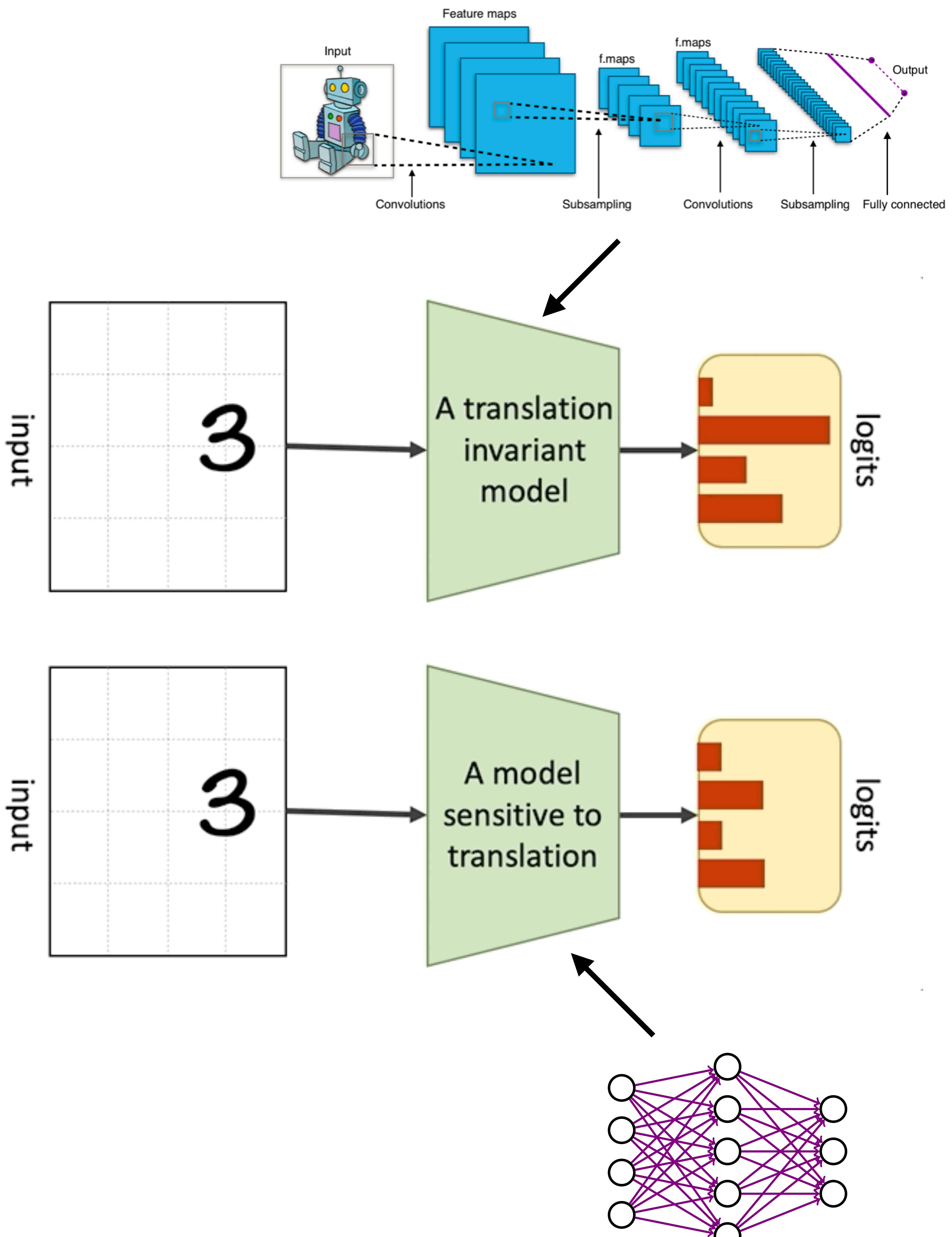






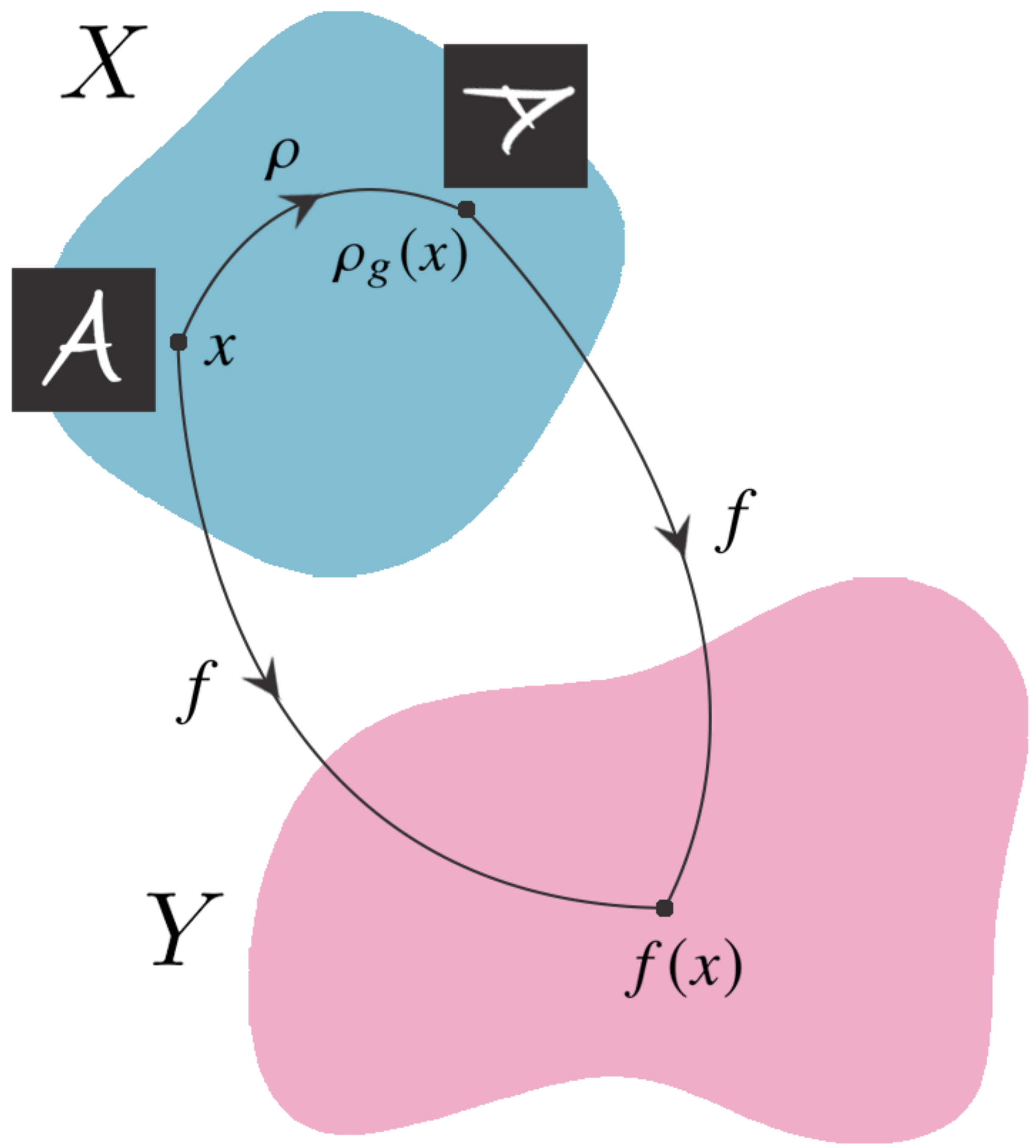
INDUCTIVE BIAS & EQUIVARIANCE

- Symmetry-equivariant networks
  - More economical (fewer, but more expressive parameters), interpretable, and trainable



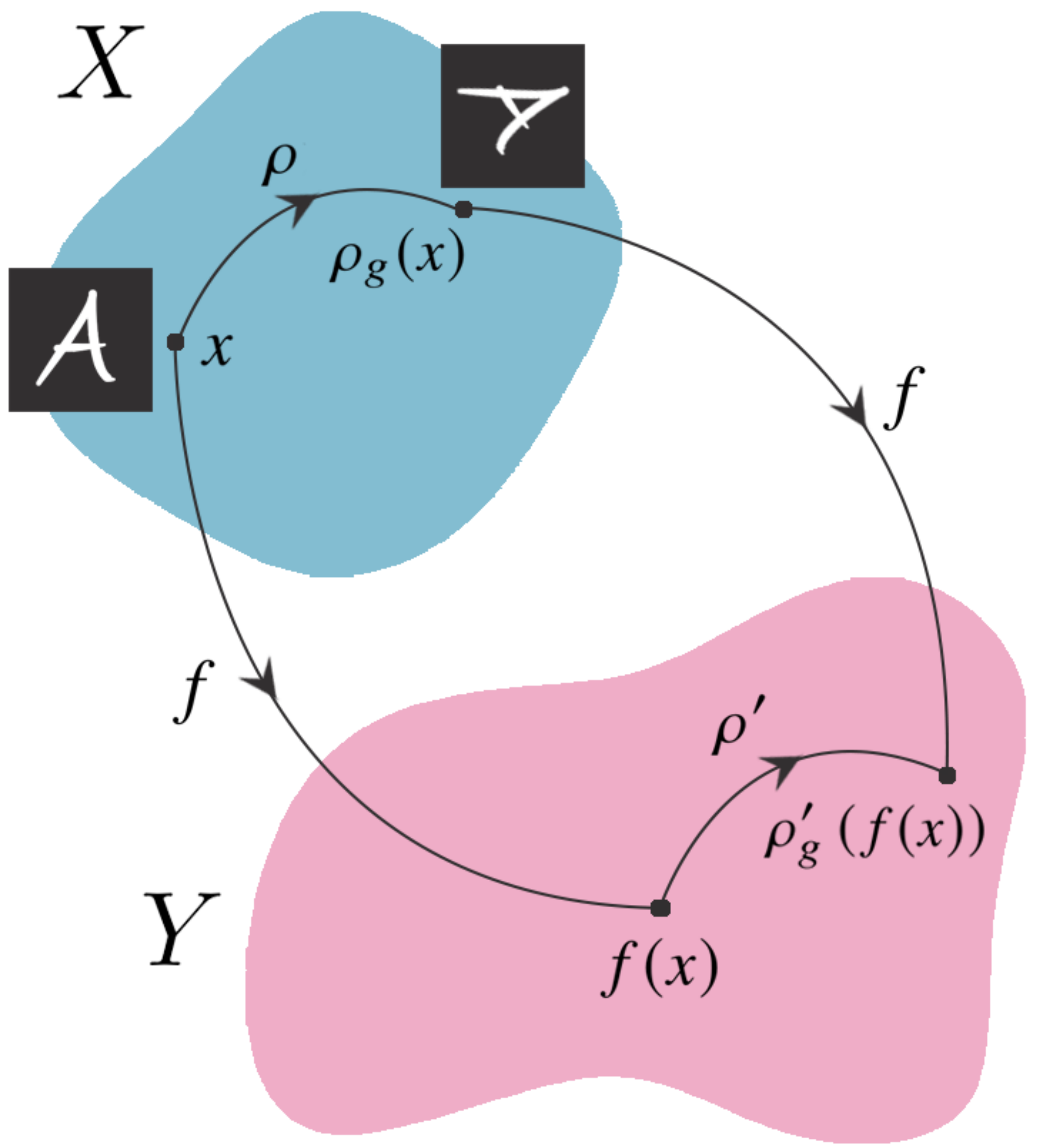
Invariance

$$f(\rho_g(x)) = f(x)$$



Equivariance

$$f(\rho_g(x)) = \rho'_g\,(f(x))$$

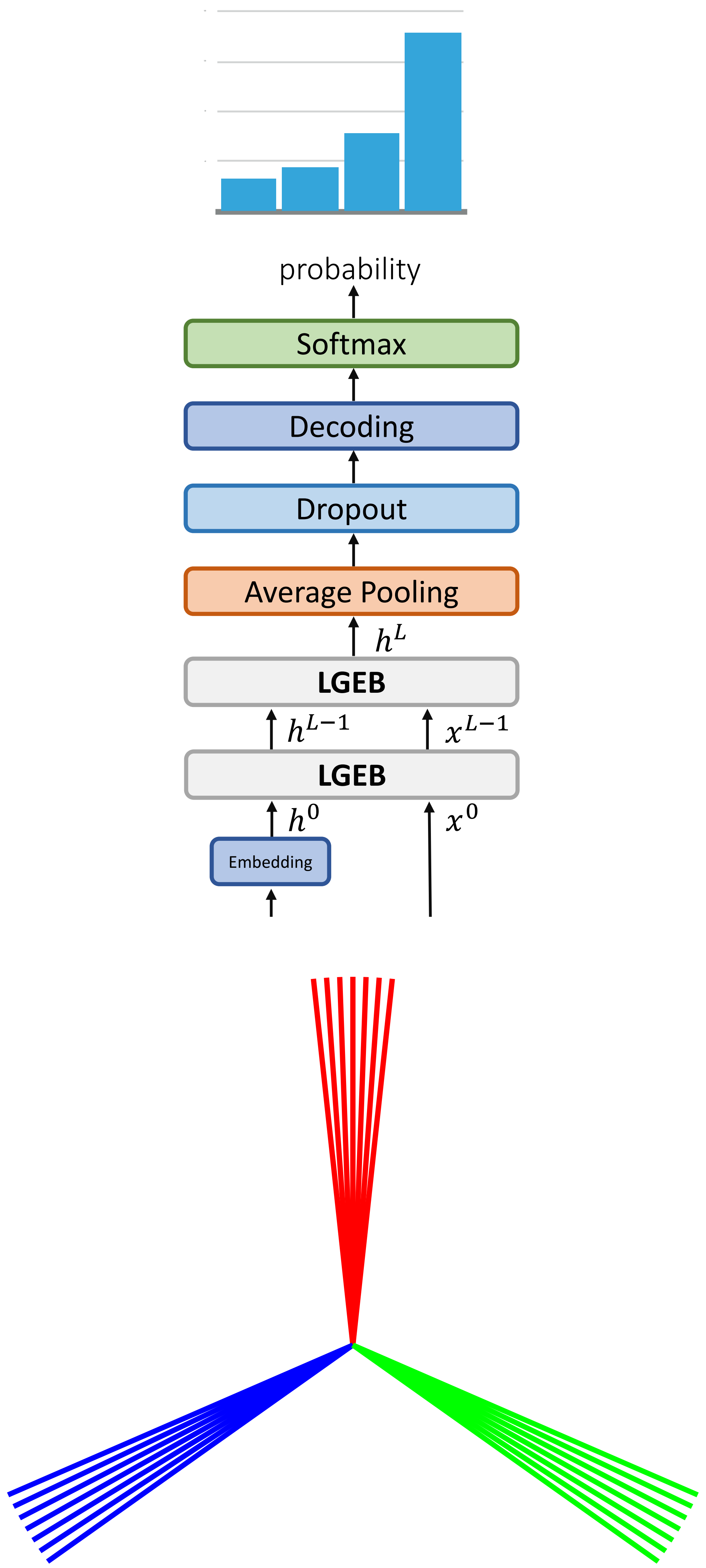


HOW DO WE ENFORCE LORENTZ SYMMETRY?

- Lorentz-invariant networks:
  - Boosting all particles into a new frame should give the same result
- Lorentz-equivariant networks:
  - Boosting all particles into a new frame should give an output that transforms the same way

HOW DO WE ENFORCE LORENTZ SYMMETRY?

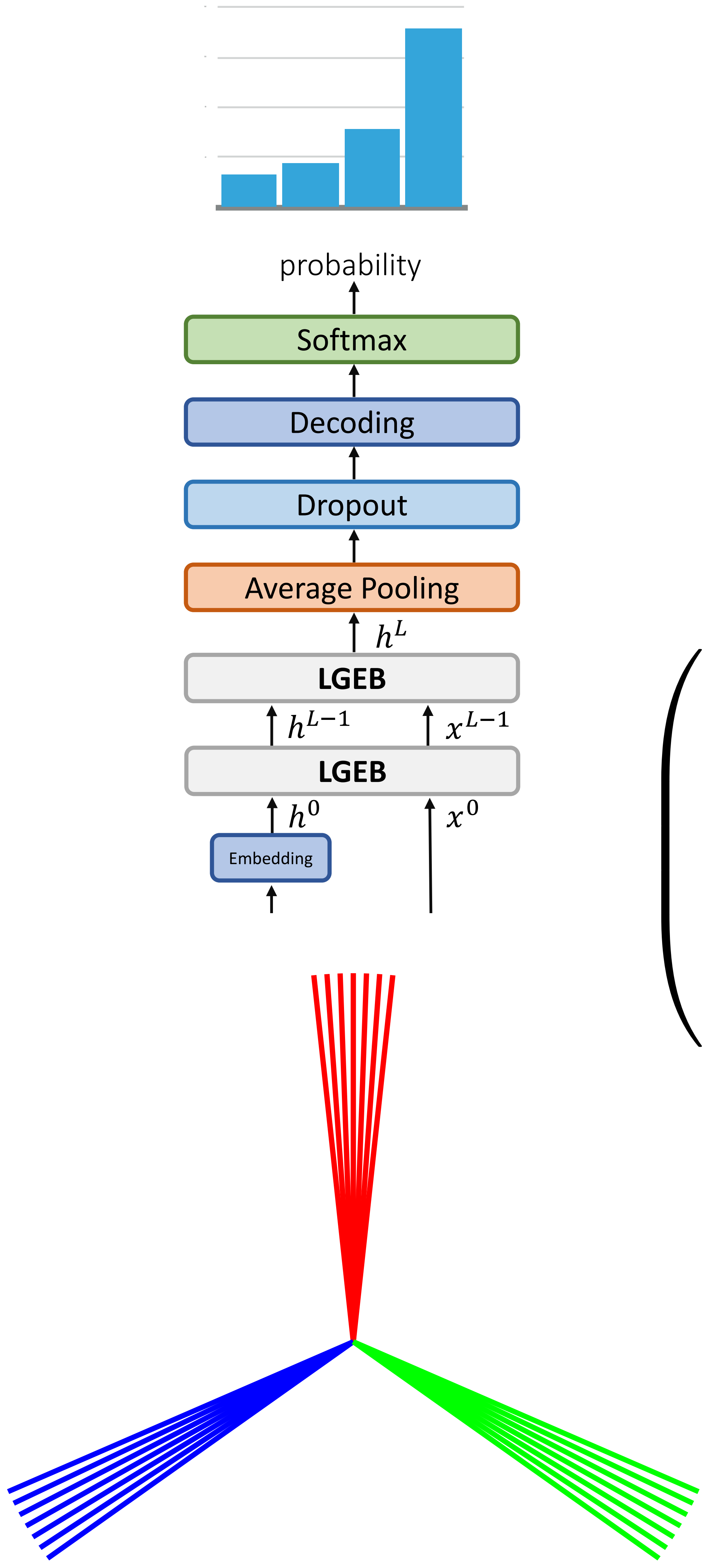
- Lorentz-invariant networks:
  - Boosting all particles into a new frame should give the same result
- Lorentz-equivariant networks:
  - Boosting all particles into a new frame should give an output that transforms the same way



arXiv:2201.08187

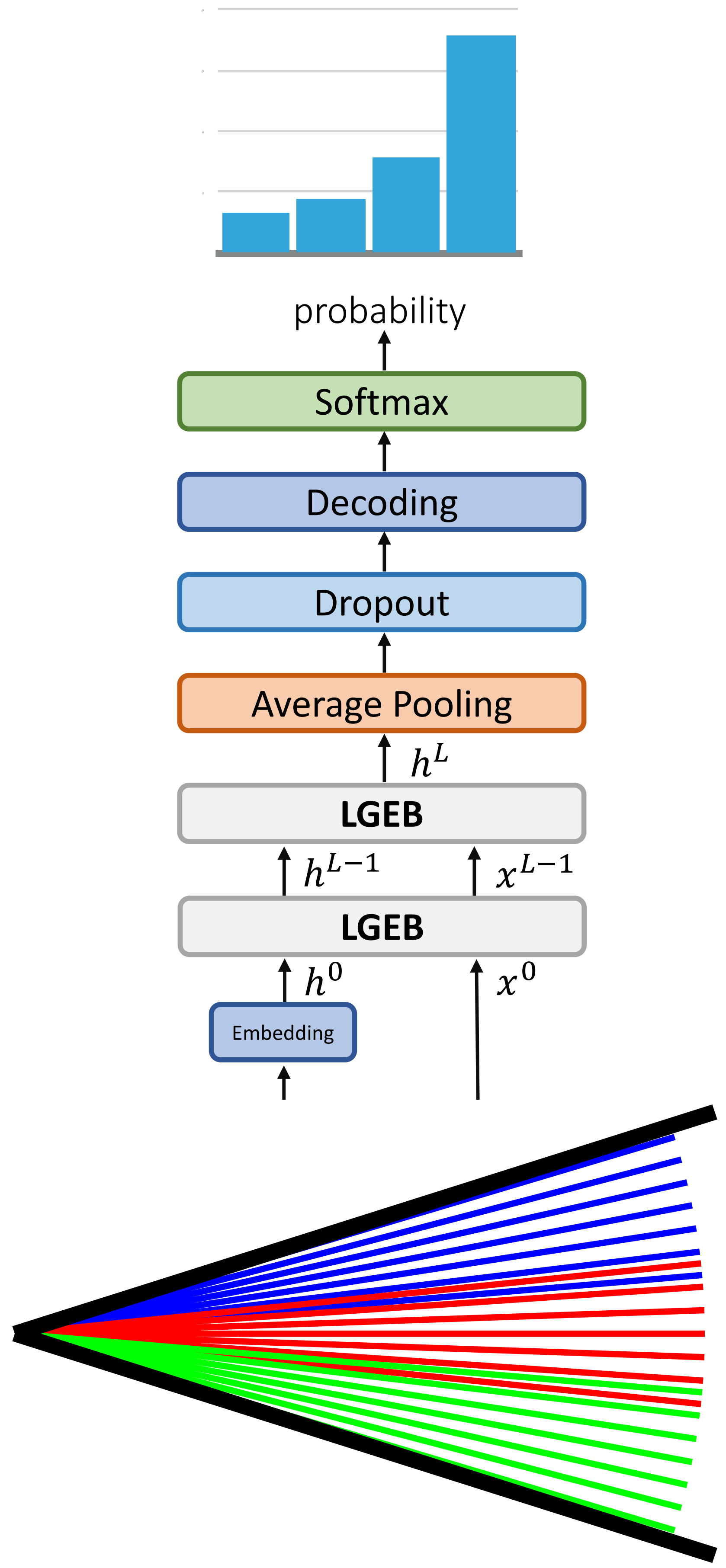
HOW DO WE ENFORCE LORENTZ SYMMETRY?

- Lorentz-invariant networks:
  - Boosting all particles into a new frame should give the same result
- Lorentz-equivariant networks:
  - Boosting all particles into a new frame should give an output that transforms the same way



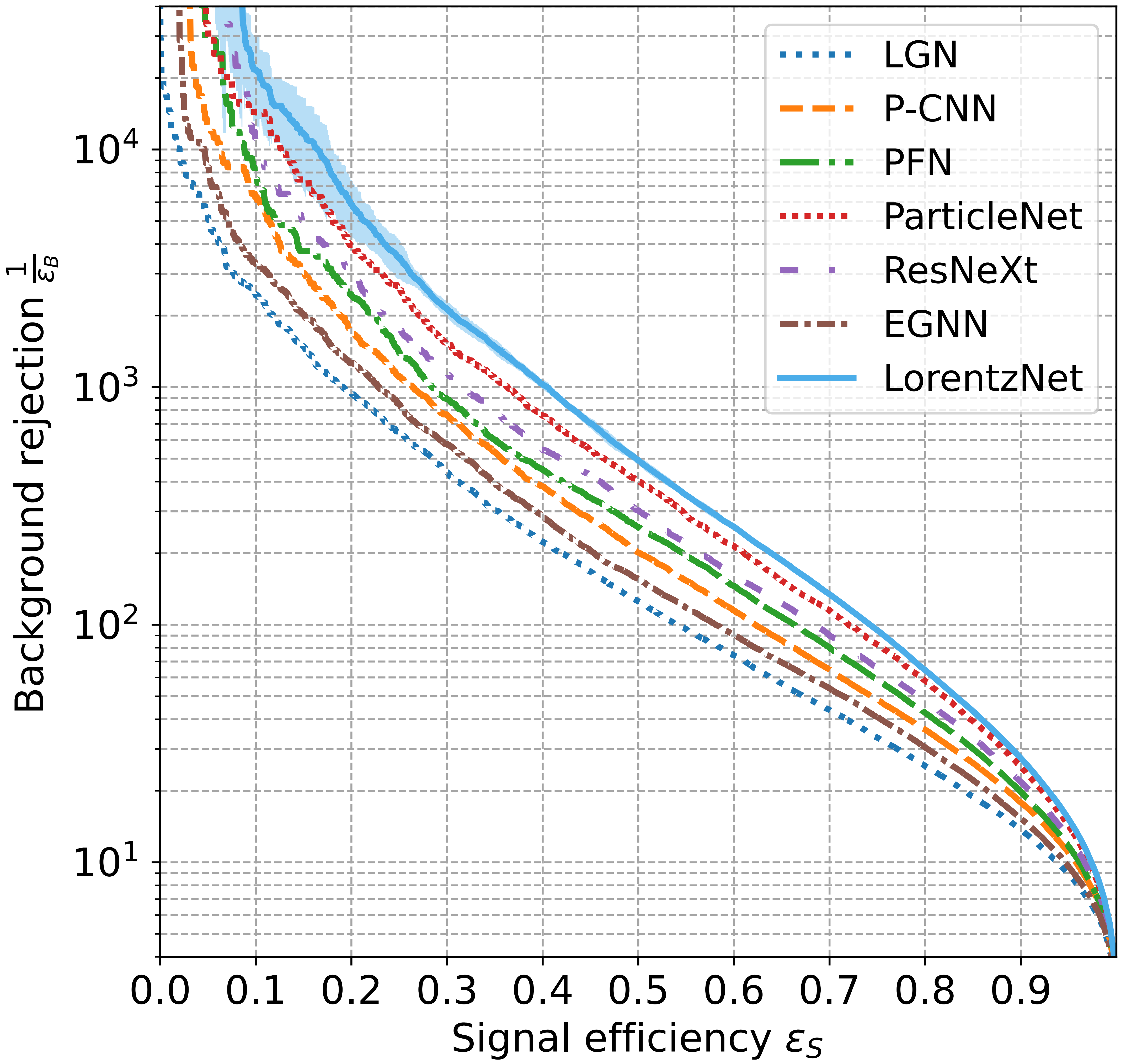
$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Lorentz boost



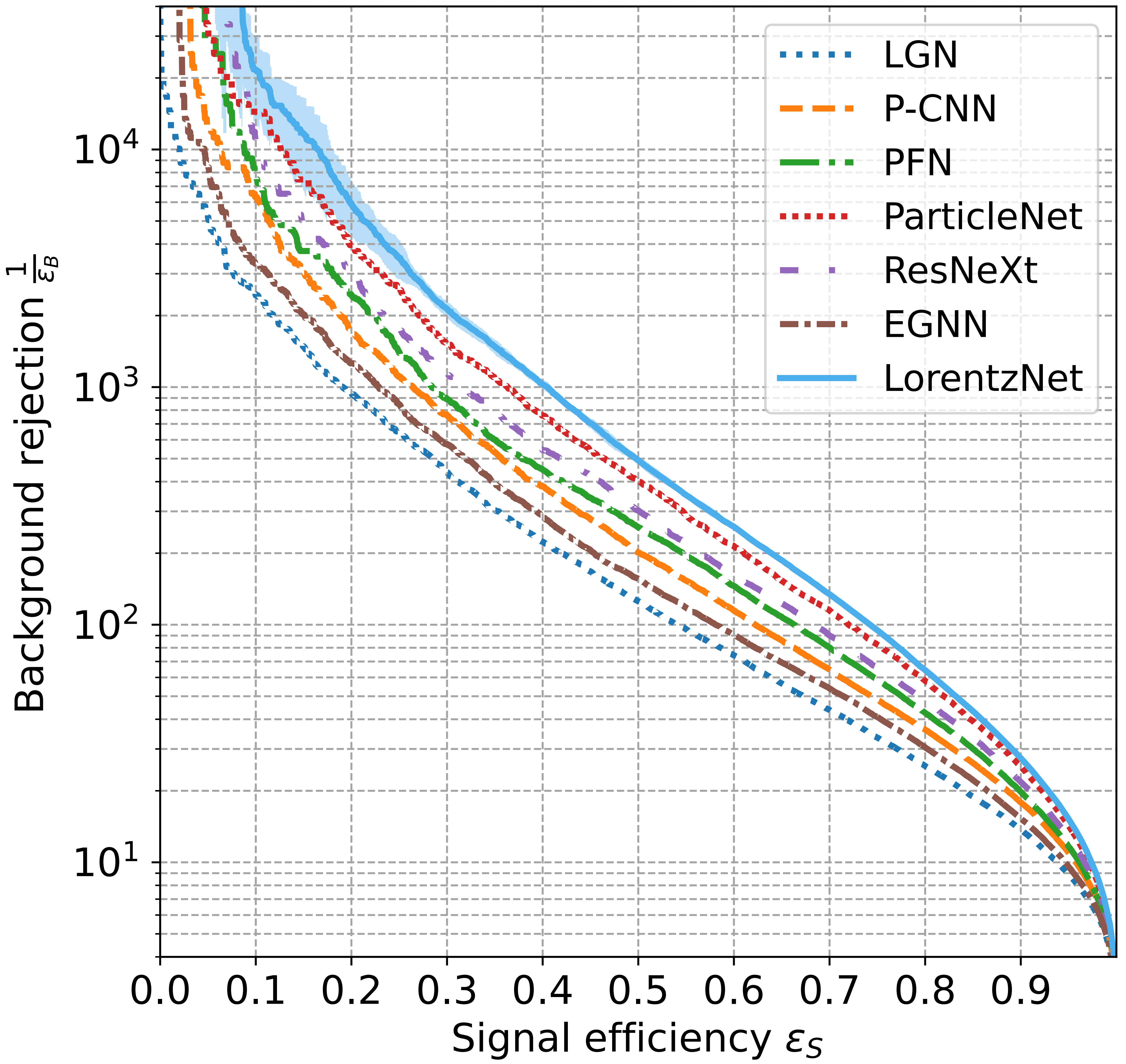
arXiv:2201.08187

LORENTZNET PERFORMANCE



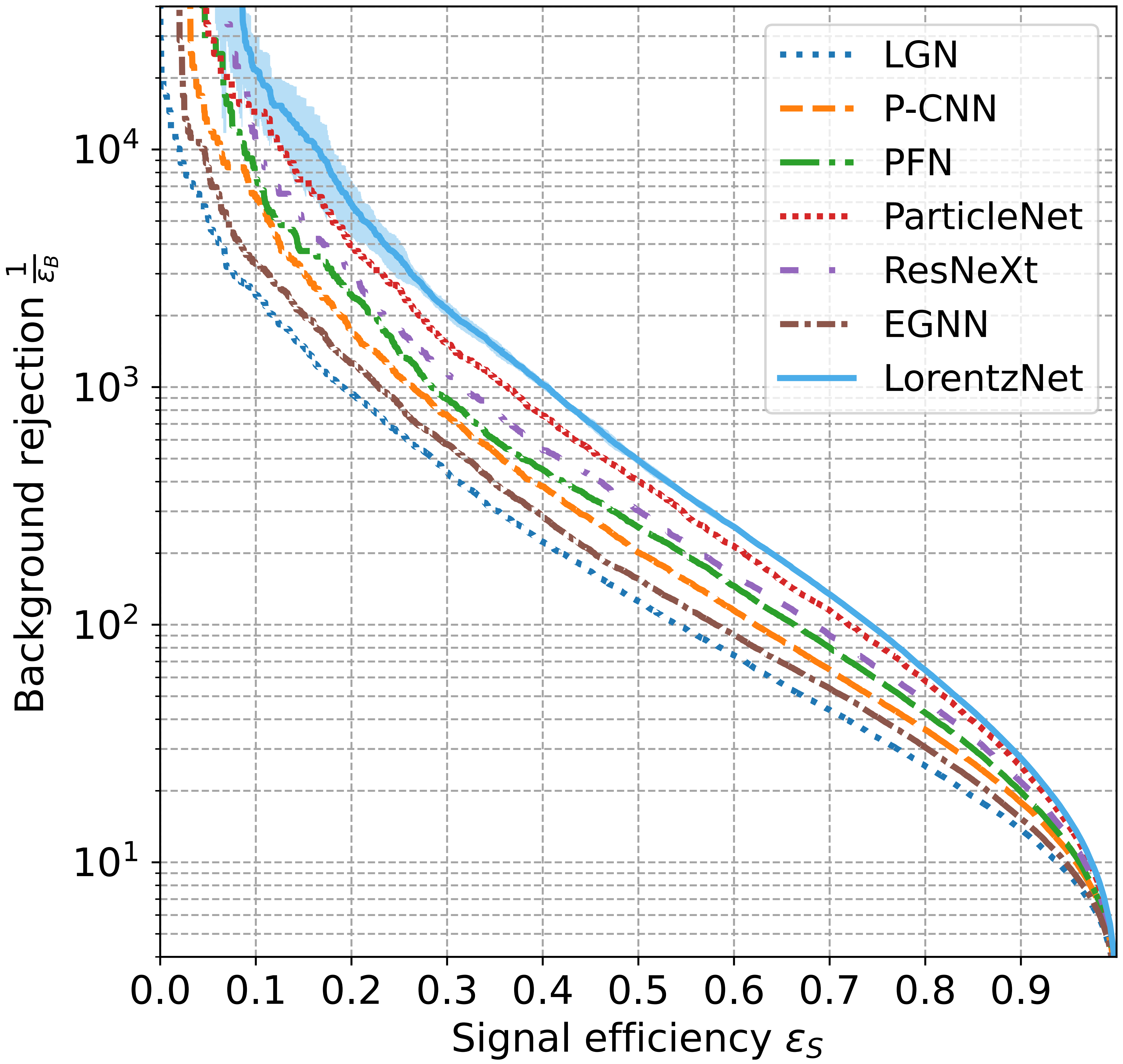
LORENTZNET PERFORMANCE

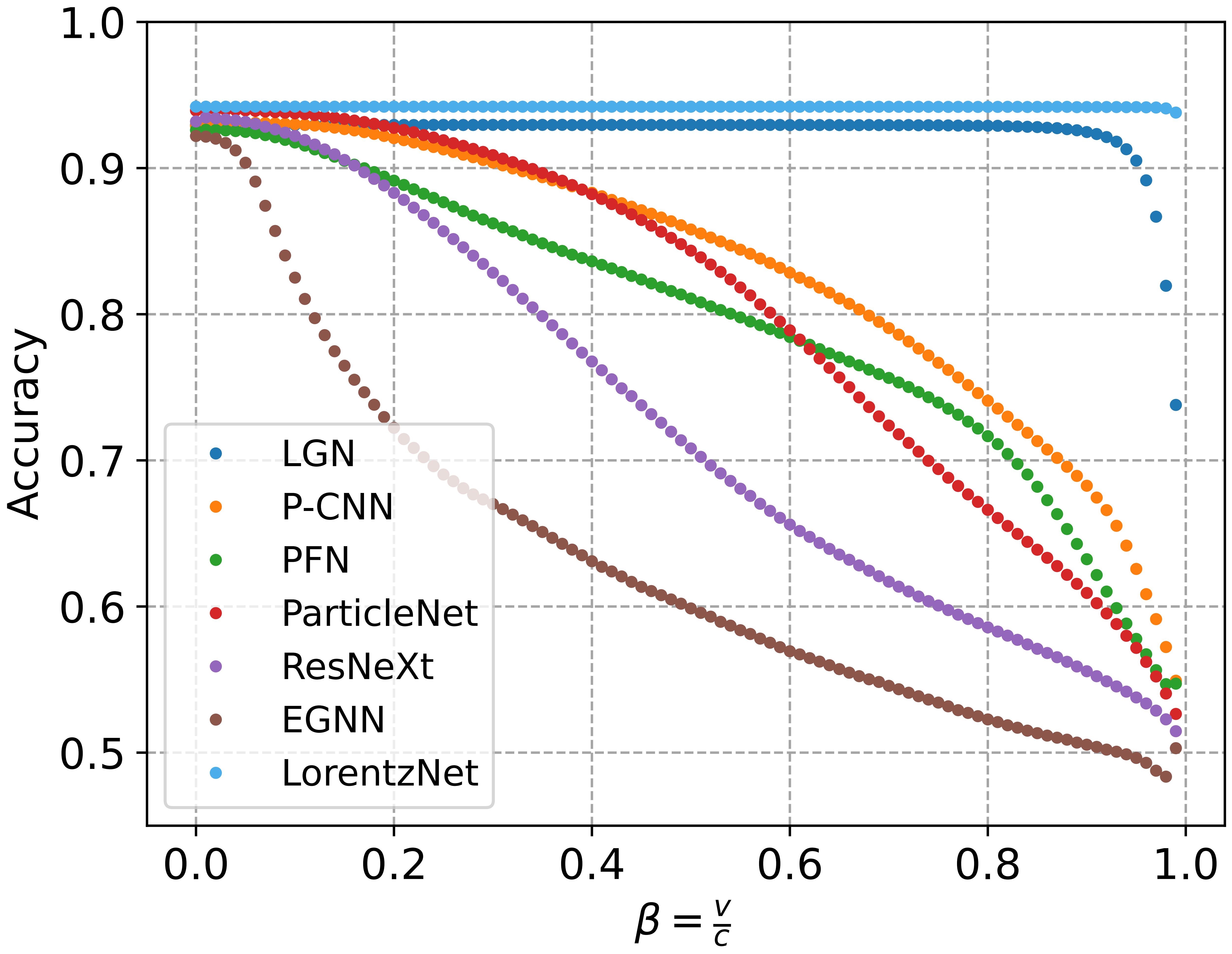
- State-of-the-art performance for top quark tagging



LORENTZNET PERFORMANCE

- State-of-the-art performance for top quark tagging
- Lorentz group invariance confirmed

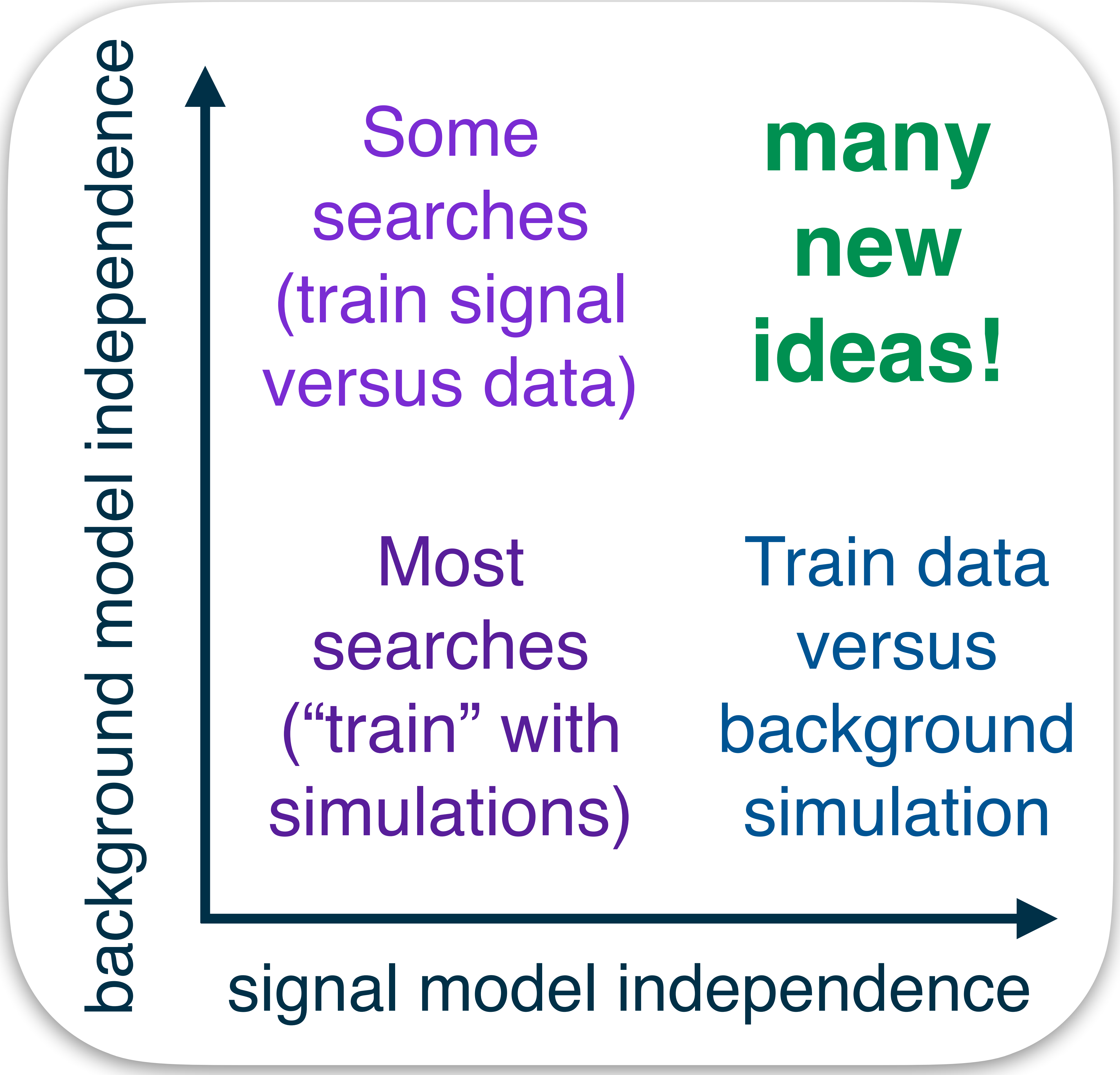




I. DATA REPRESENTATIONS & SYMMETRIES
II. ANOMALY DETECTION
III. GENERATIVE MODELING
III. FAST INFERENCE
VI. SUMMARY & OUTLOOK

LANDSCAPE OF BSM SEARCHES

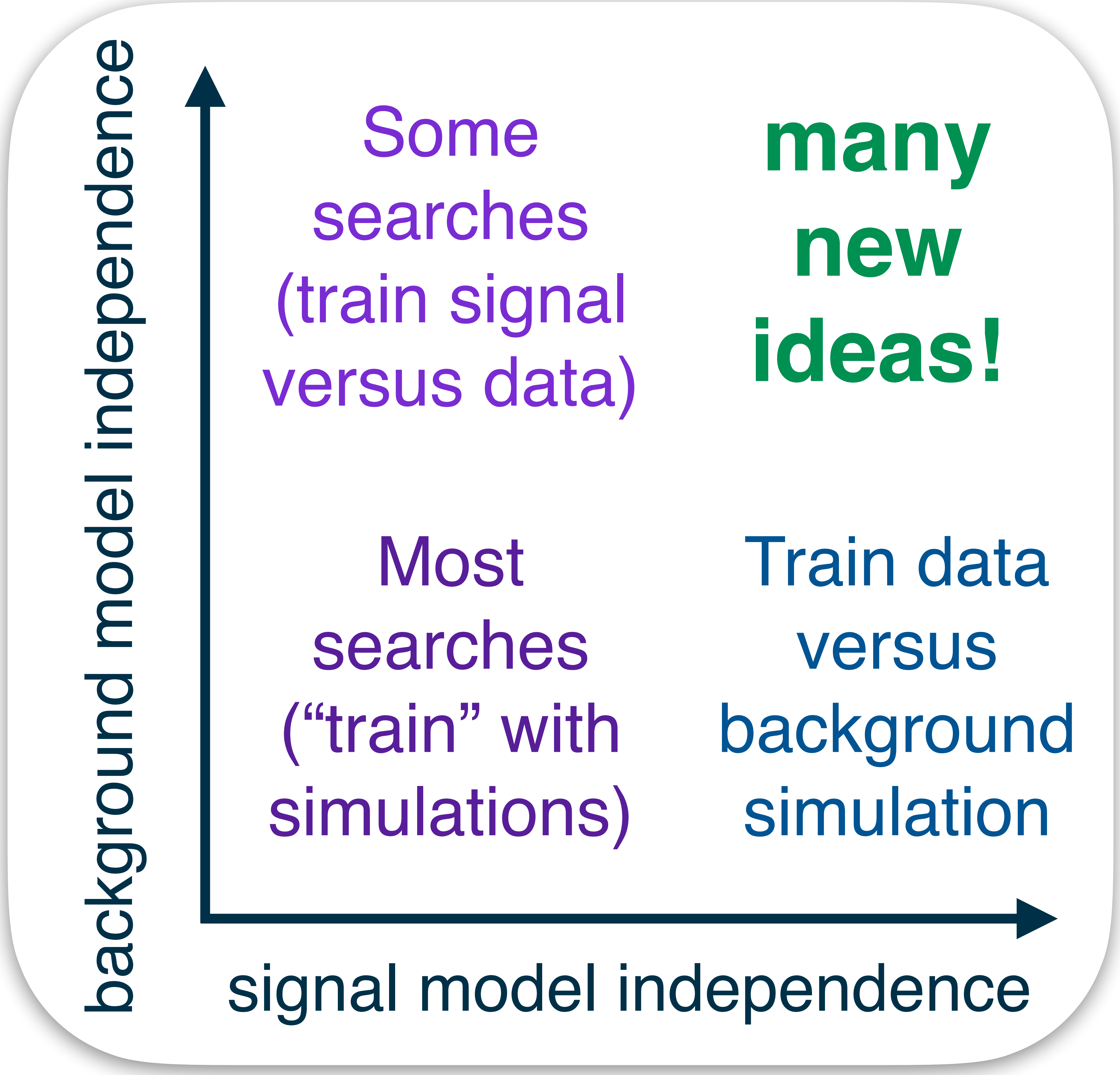
Credit: B. Nachman
https://indico.cern.ch/event/1188153/



LANDSCAPE OF BSM SEARCHES

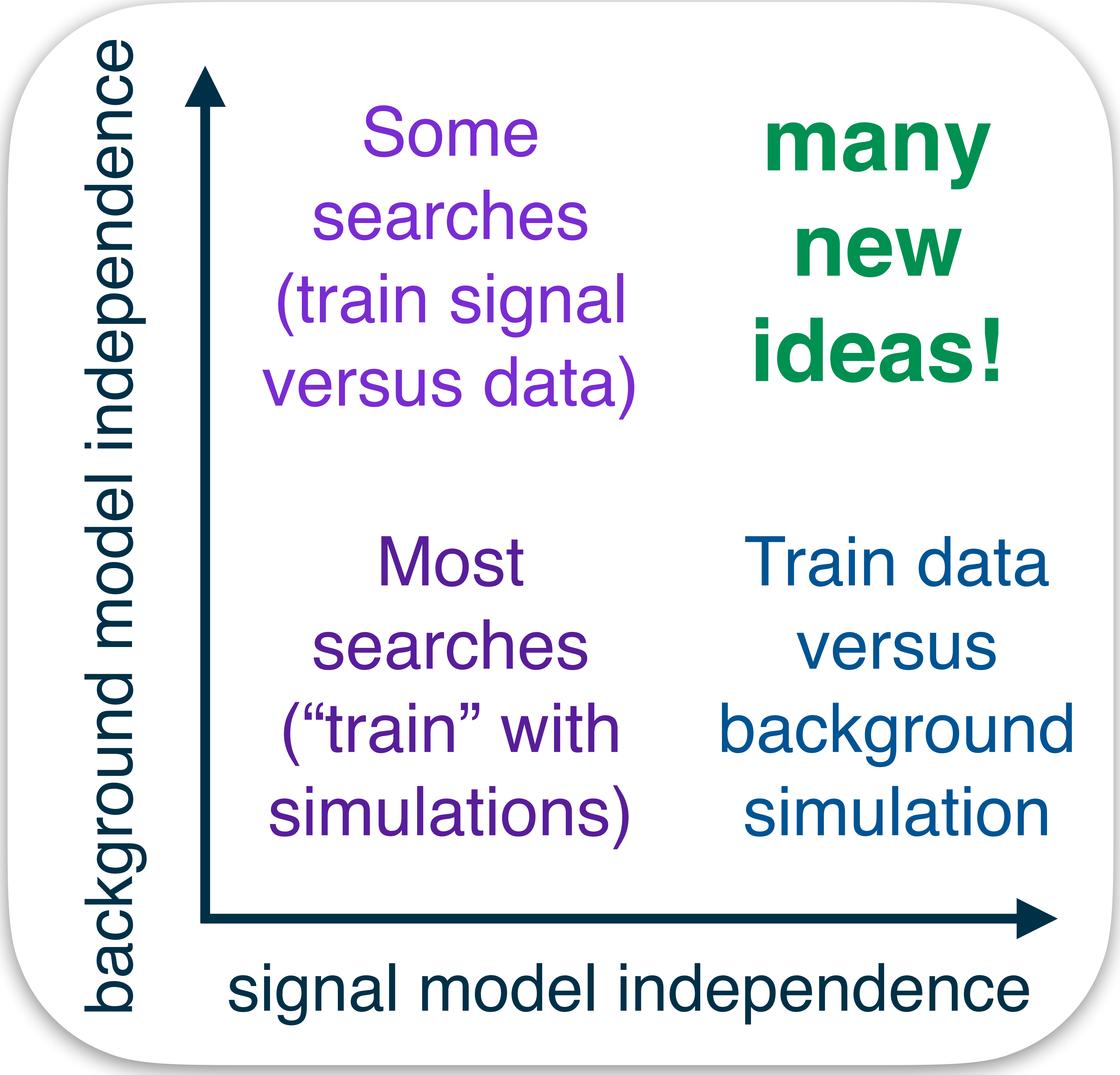
Credit: B. Nachman
https://indico.cern.ch/event/1188153/

- Supervised = full label information



LANDSCAPE OF BSM SEARCHES

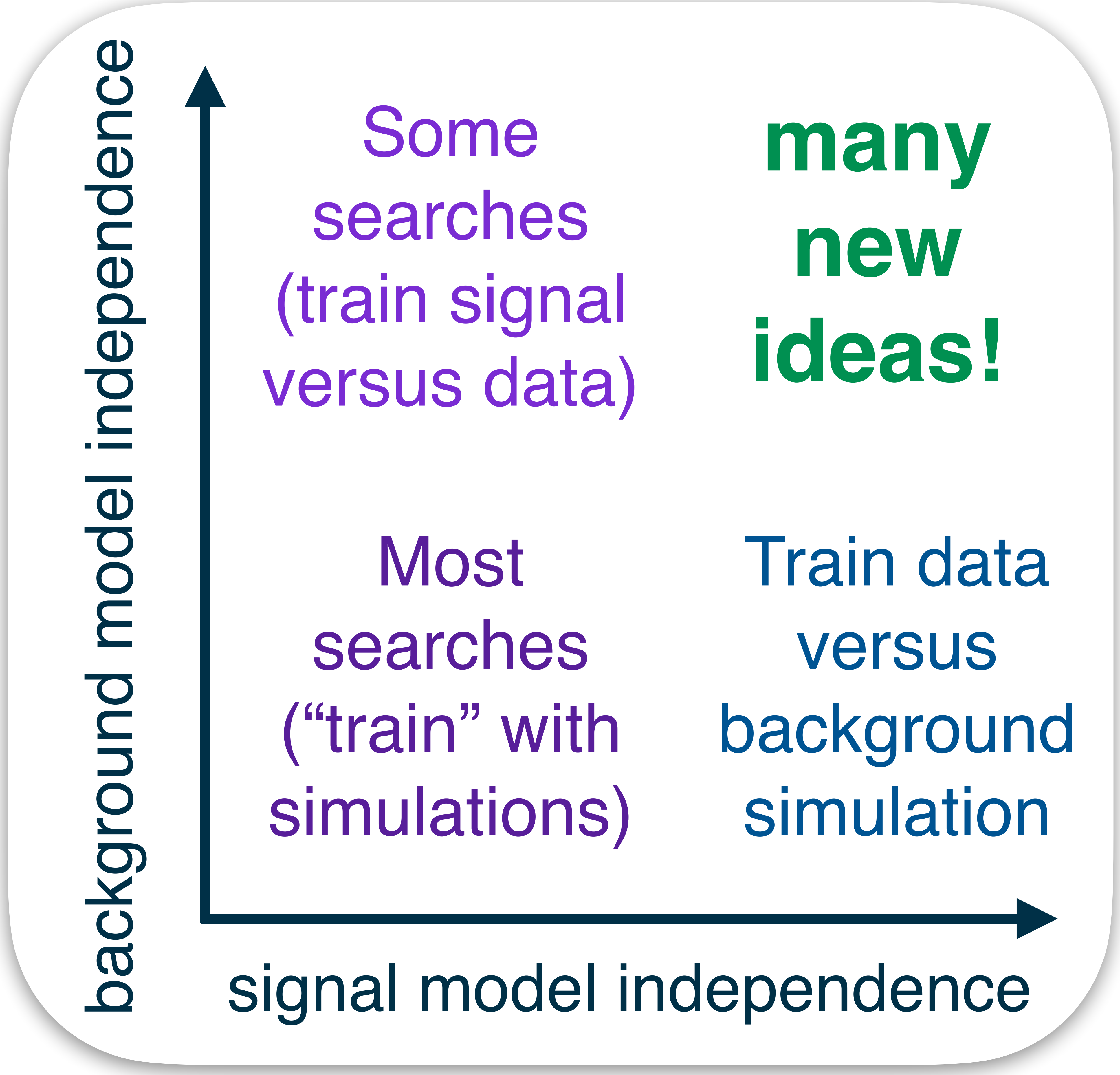
Credit: B. Nachman
https://indico.cern.ch/event/1188153/



- Supervised = full label information
- Semi-supervised = partial labels

LANDSCAPE OF BSM SEARCHES

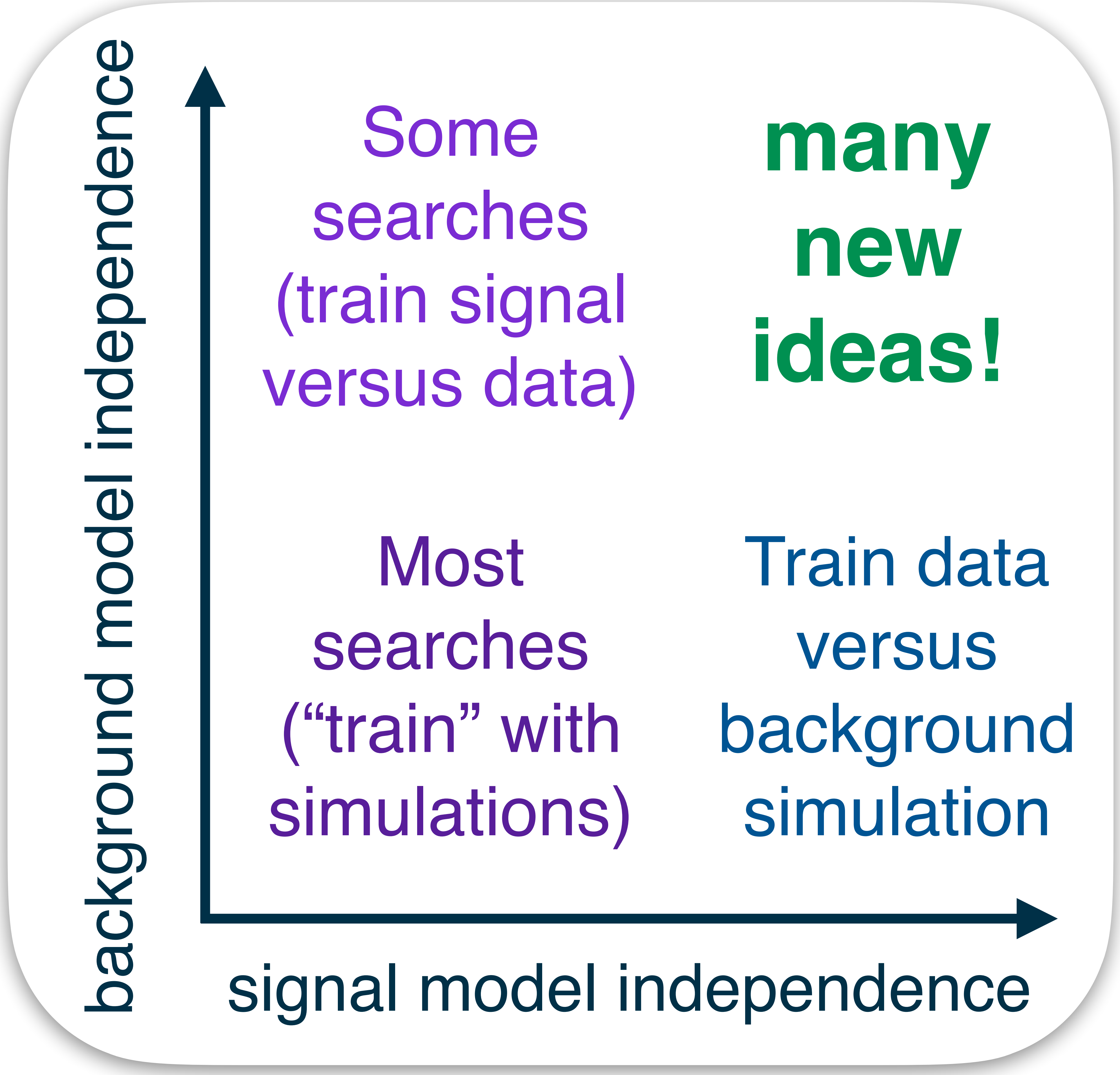
Credit: B. Nachman
https://indico.cern.ch/event/1188153/



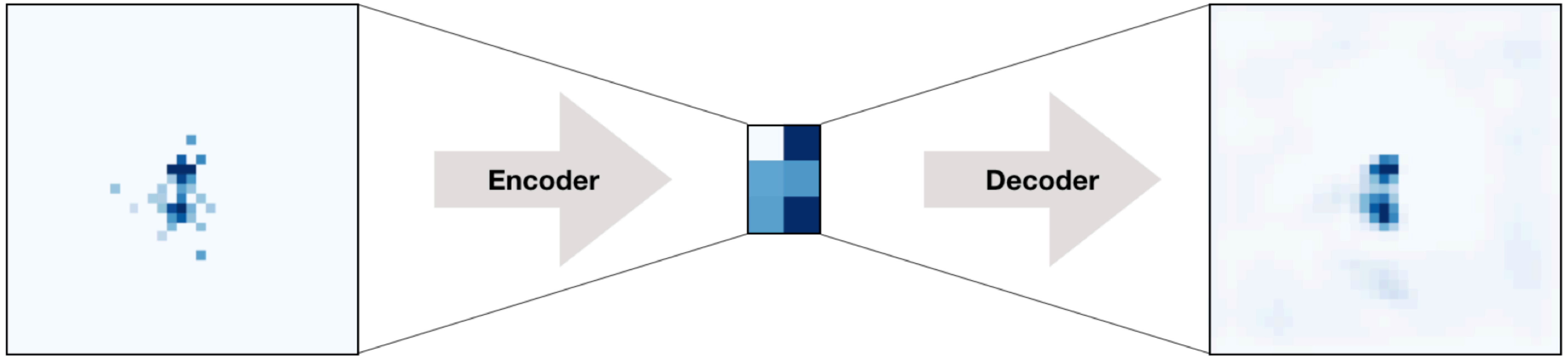
- Supervised = full label information
- Semi-supervised = partial labels
- Weakly-supervised = noisy labels

LANDSCAPE OF BSM SEARCHES

Credit: B. Nachman
https://indico.cern.ch/event/1188153/



- Supervised = full label information
- Semi-supervised = partial labels
- Weakly-supervised = noisy labels
- Unsupervised = no labels
  - Example: autoencoders compress data and then uncompress it
  - Assumption: if x is far from $\mathrm{Decoder}(\mathrm{Encoder}(x))$, then x has low $p_{\mathrm{bkgd}}(x)$



LHC OLYMPICS 2020

- Challenge with "black box" signals run in 2020–2021
- Plethora of new techniques

ML4SIM STRATEGIES

ML4SIM STRATEGIES

- Several different strategies:
  - Replace (part of) FullSim: increase speed, preserve accuracy

ML4SIM STRATEGIES

- Several different strategies:
  - Replace (part of) FullSim: increase speed, preserve accuracy
  - Replace (part of) FastSim: maintain speed, increase accuracy

ML4SIM STRATEGIES

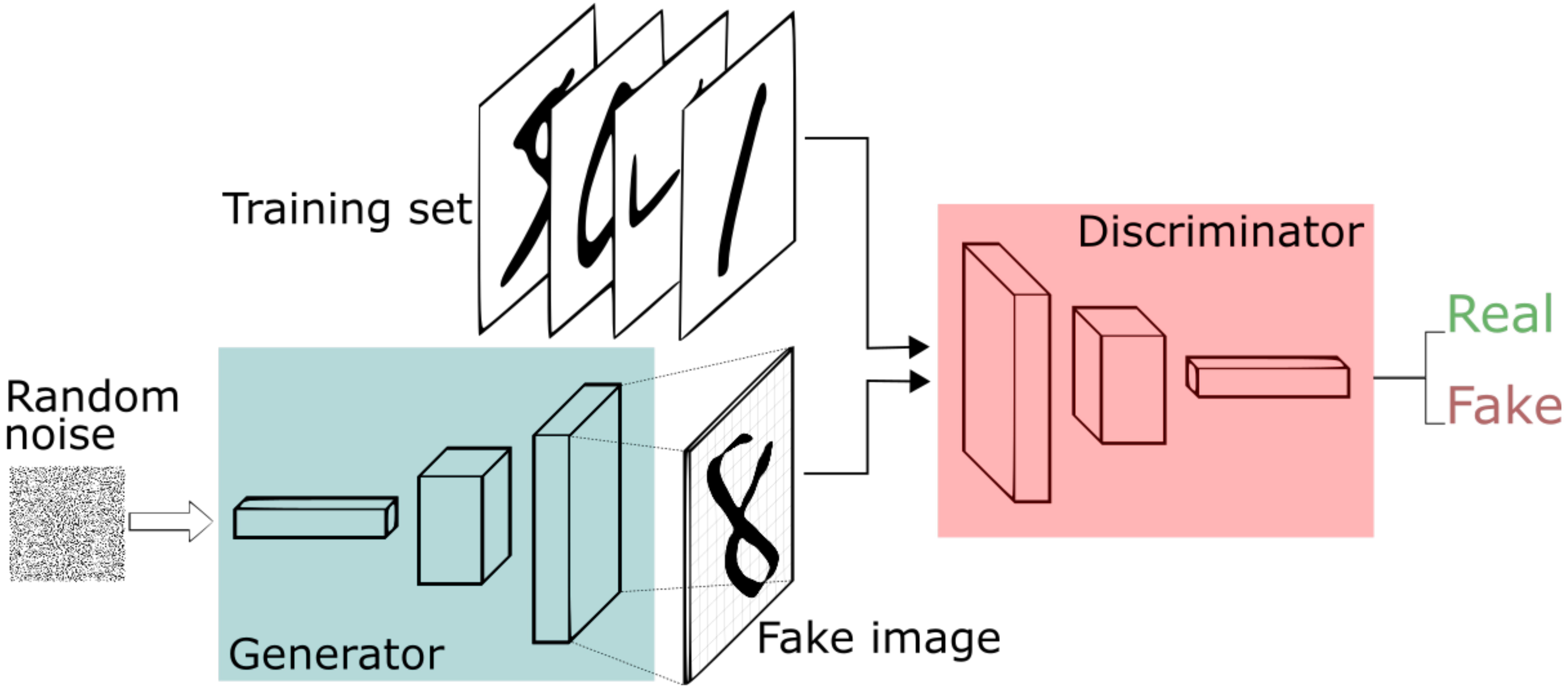
- Several different strategies:
  - Replace (part of) FullSim: increase speed, preserve accuracy
  - Replace (part of) FastSim: maintain speed, increase accuracy
  - Conditional: map generated → reconstructed events

ML4SIM STRATEGIES

- Several different strategies:
  - Replace (part of) FullSim: increase speed, preserve accuracy
  - Replace (part of) FastSim: maintain speed, increase accuracy
  - Conditional: map generated → reconstructed events
  - **End-to-end: map random noise → reconstructed events directly**

GENERATIVE ADVERSARIAL NETWORKS

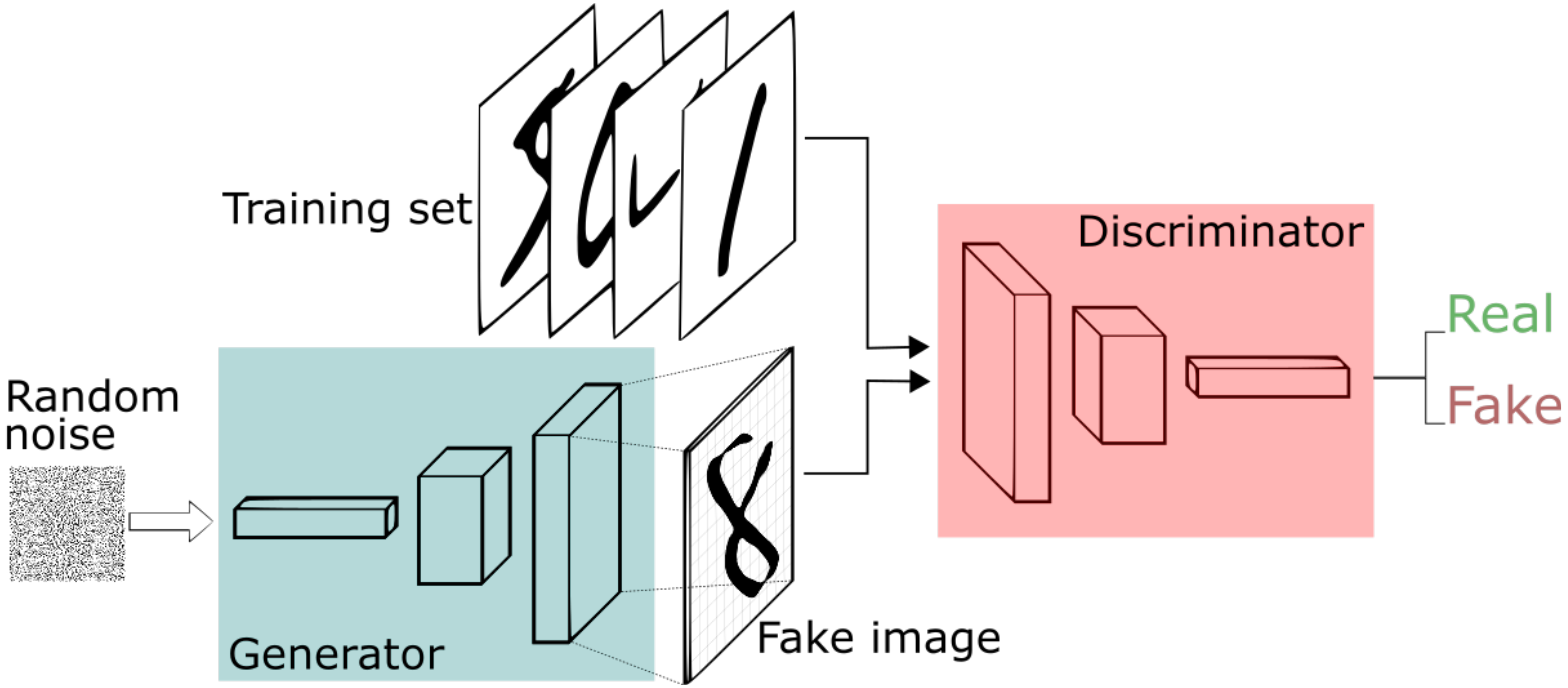
arXiv:1406.2661
arXiv:1912.04958



- Train two neural networks in tandem:
  - one to generate realistic "fake" data
  - the other to discriminate "real" from "fake" data

GENERATIVE ADVERSARIAL NETWORKS

arXiv:1406.2661
arXiv:1912.04958

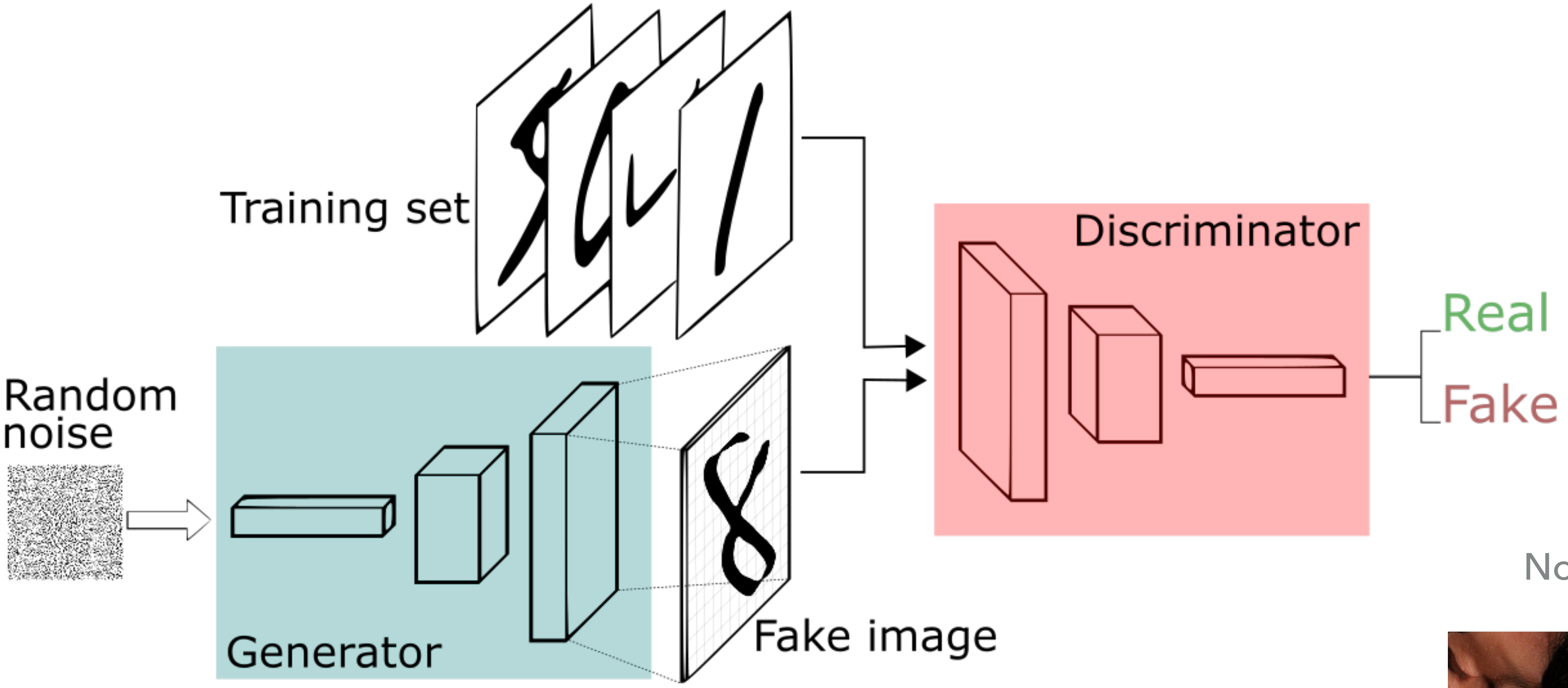


thispersondoesnotexist.com

- Train two neural networks in tandem:
  - one to generate realistic “fake” data
  - the other to discriminate “real” from “fake” data

GENERATIVE ADVERSARIAL NETWORKS

arXiv:1406.2661
arXiv:1912.04958



Note: failure modes!

- Train two neural networks in tandem:
  - one to generate realistic "fake" data
  - the other to discriminate "real" from "fake" data

GENERATIVE AI EVALUATION METRICS

- Evaluation of generative models is in general difficult
- We want to evaluate quantitatively:
  - the **quality** of the data
  - the **diversity** of the data
  - ultimately, **physics performance**
- To do so, we proposed with four physics- and computer-vision-inspired metrics

GENERATIVE AI EVALUATION METRICS

- Evaluation of generative models is in general difficult
- We want to evaluate quantitatively:
  - the **quality** of the data
  - the **diversity** of the data
  - ultimately, **physics performance**
- To do so, we proposed with four physics- and computer-vision-inspired metrics

	Minimum Matching Distance	Coverage	Fréchet ParticleNet Distance	1-Wassersstein Distance (W_1)
Quality	✅		✅	✅
Diversity		✅	✅	✅
Physics Perf.				✅

GENERATIVE AI EVALUATION METRICS

New print on arXiv this week!

arXiv:2012.00173
arXiv:2106.11535

- Evaluati
- We want
  - the qu
  - the di
  - ultima
- To do so … etrics

	…sserstein …ance (W_1)
Qualit…	✅
Diversi…	✅
Physics F…	✅

On the Evaluation of Generative Models in High Energy Physics

Raghav Kansal,* Anni Li, and Javier Duarte
University of California, San Diego

Nadezda Chernyavskaya, Maurizio Pierini
European Center for Nuclear Research (CERN)

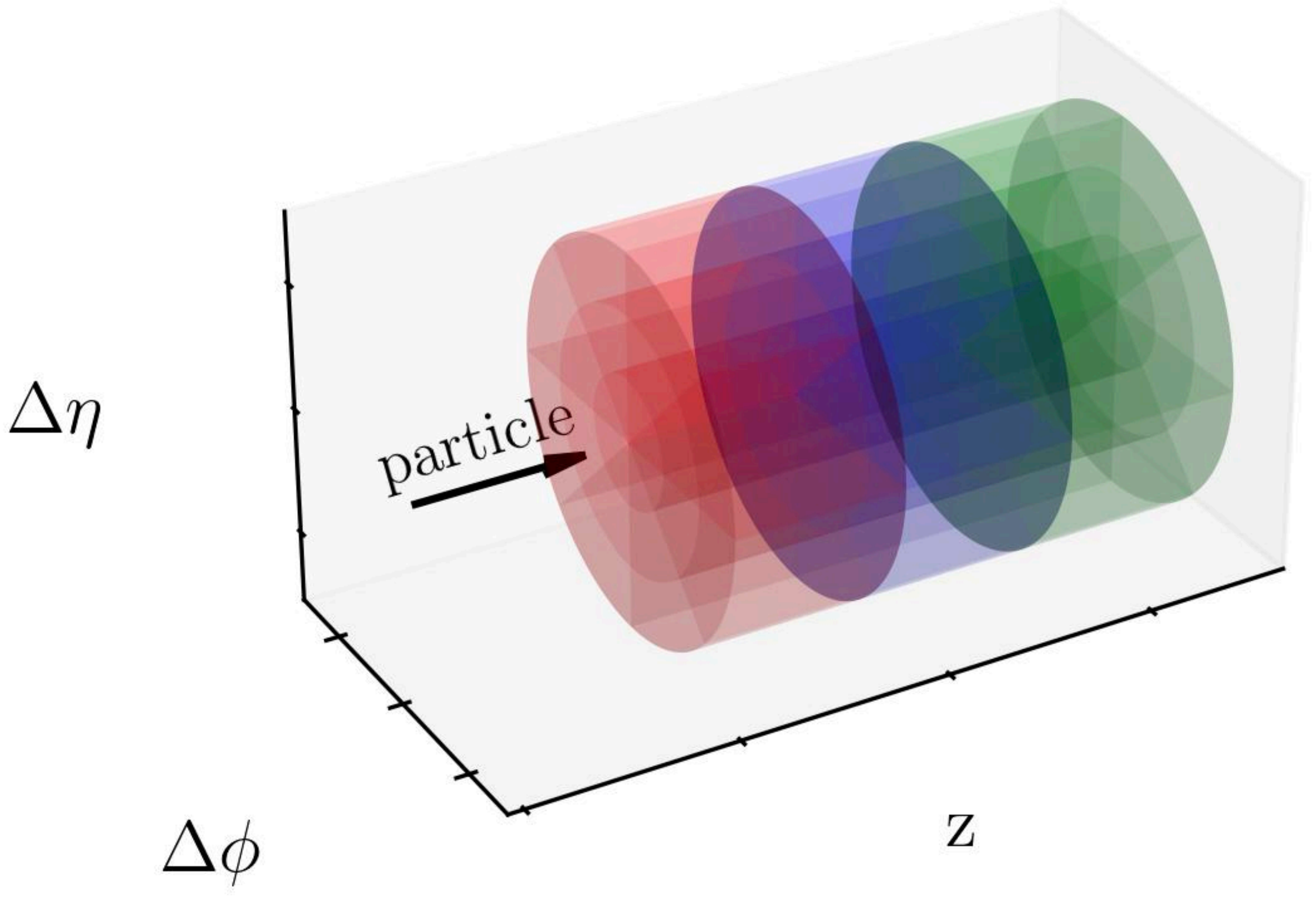
Breno Orzari, Thiago Tomei
Universidade Estadual Paulista, São Paulo/SP

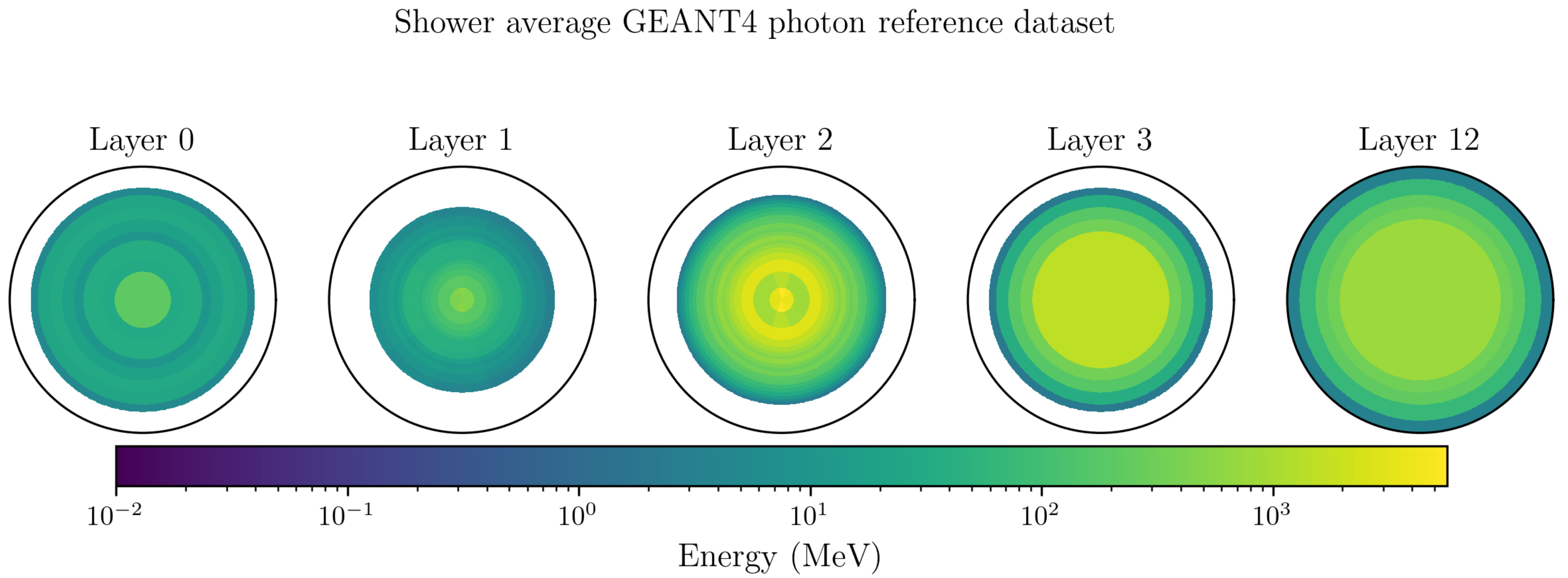
(Dated: November 16, 2022)

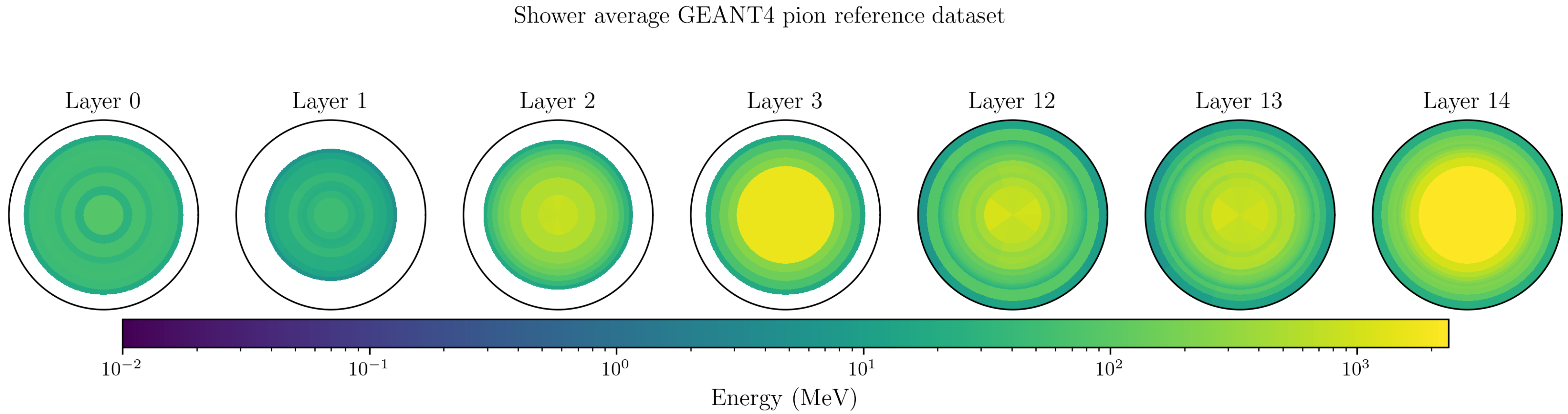
There has been a recent explosion in research into machine-learning- (ML-) based generative modeling to tackle computational challenges for simulations in high energy physics (HEP). In order to use such alternative simulators in practice, we need a well defined metrics to compare different generative models and evaluate their discrepancy from the true distributions. We present the first systematic review and investigation into evaluation metrics and their sensitivity to failure models of generative models, using the framework of two-sample goodness-of-fit testing, and their relevance and viability for HEP. Inspired by previous work in both physics and computer vision, we propose two new metrics, the Fréchet and Kernel Physics Distances (FPD and KPD), and perform a variety of experiments measuring their performance on simple Gaussian-distributed, and simulated high energy jet datasets. We find FPD, in particular, to be the most sensitive metric to all alternative jet distributions tested and recommend its adoption, along with KPD and Wasserstein distances between individual feature distributions, for evaluating generative models in HEP. We finally demonstrate the efficacy of these proposed metrics in evaluating and comparing a novel attention-based generative model, GAPT, to the state-of-the-art MPGAN jet simulation model.

CALO CHALLENGE

- Ongoing challenge for generative modeling of calorimeter showers in HEP!
- Many new approaches presented at ML4Jets 2022: https://indico.cern.ch/event/1159913/



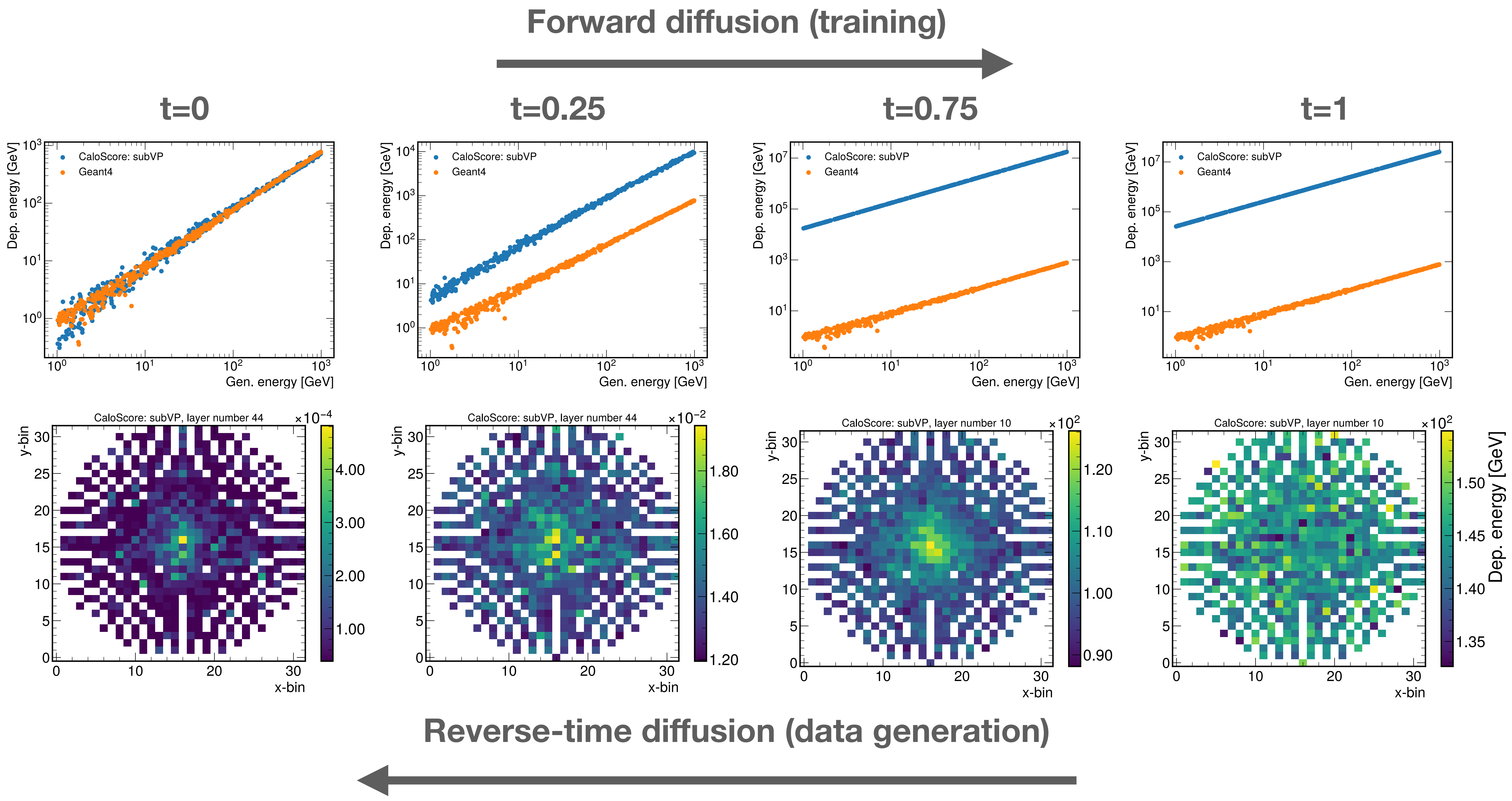




DIFFUSION MODELS IN HEP

[arXiv:2011.13456](https://arxiv.org/abs/2011.13456)
[arXiv:2206.11898](https://arxiv.org/abs/2206.11898)

- Diffusion models have very recently dethroned GANs for natural images
- Generative model is trained using a diffusion process that slowly perturbs the data by adding noise – model learns to **denoise**
- Generation of new samples by reversing the diffusion process



- Distribution of deposited energies for generated particle energies (top) and the energy deposition in a single layer of a calorimeter (bottom) vs time step

I. DATA REPRESENTATIONS
& SYMMETRIES

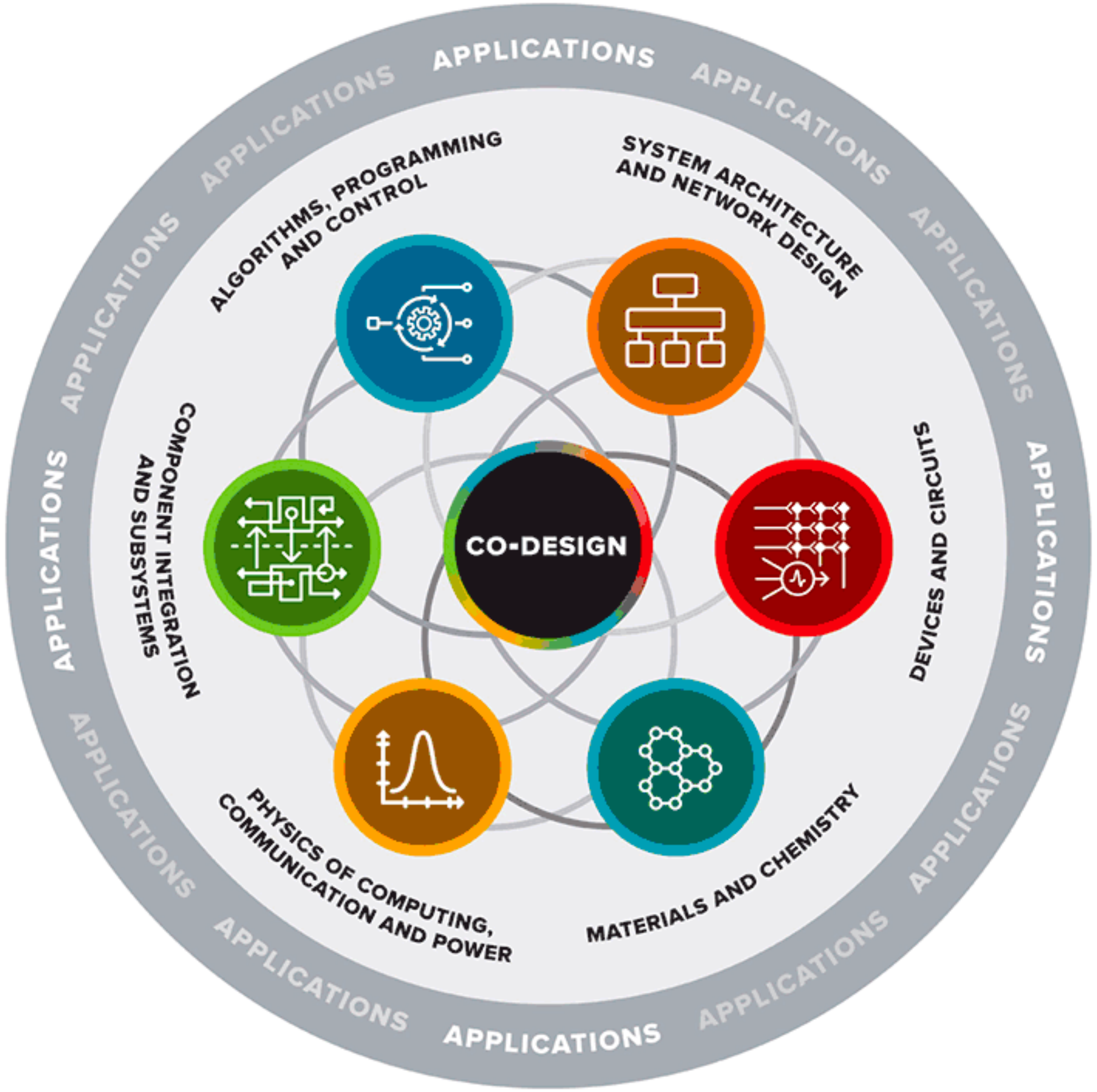
II. ANOMALY DETECTION

III. GENERATIVE MODELING

III. FAST INFERENCE

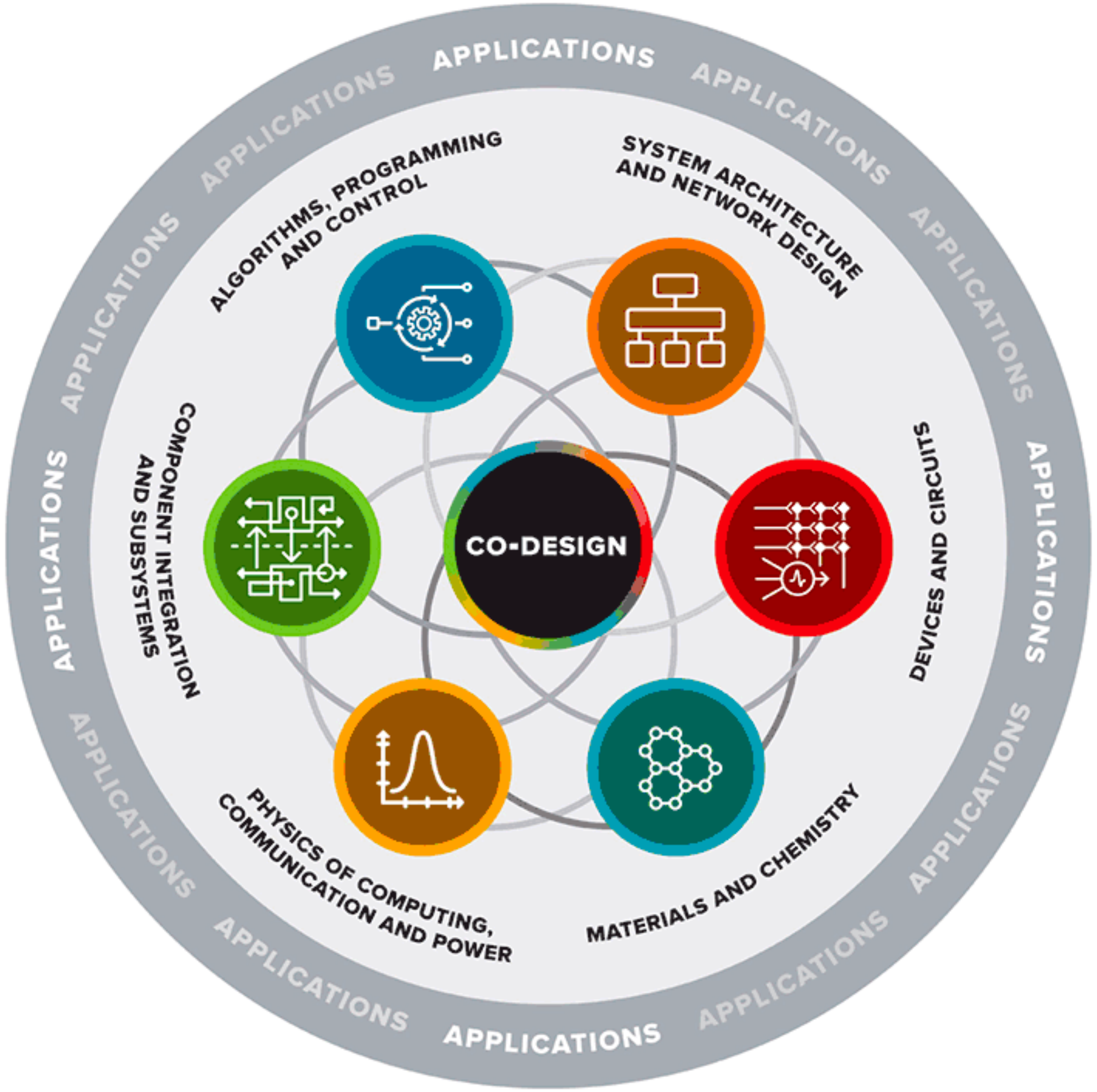
VI. SUMMARY & OUTLOOK

CODESIGN


APPLICATIONS
ALGORITHMS, PROGRAMMING AND CONTROL
SYSTEM ARCHITECTURE AND NETWORK DESIGN
COMPONENT INTEGRATION AND SUBSYSTEMS
CO-DESIGN
DEVICES AND CIRCUITS
PHYSICS OF COMPUTING, COMMUNICATION AND POWER
MATERIALS AND CHEMISTRY

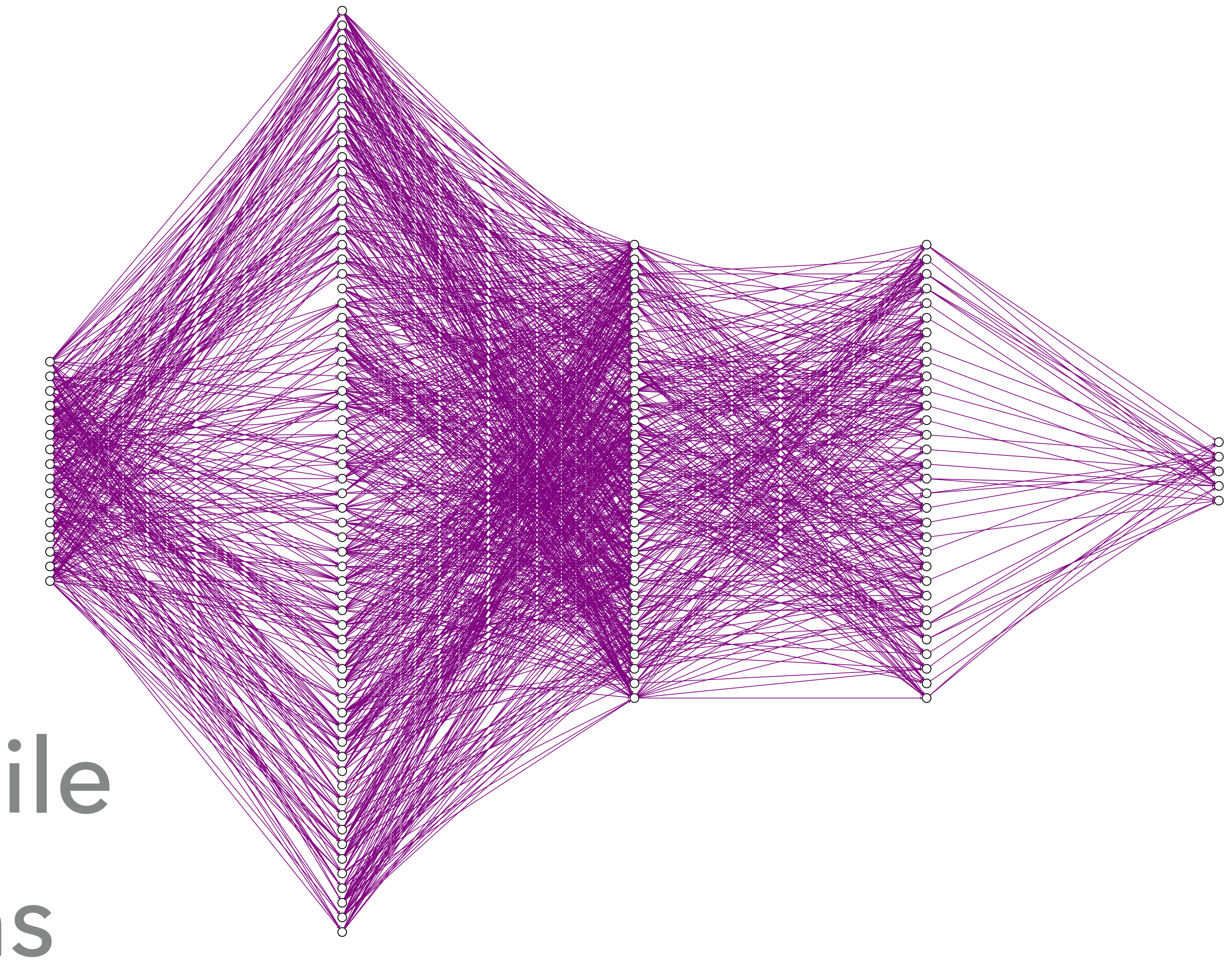
CODESIGN

- **Codesign**: intrinsic development loop between algorithm design, training, and implementation



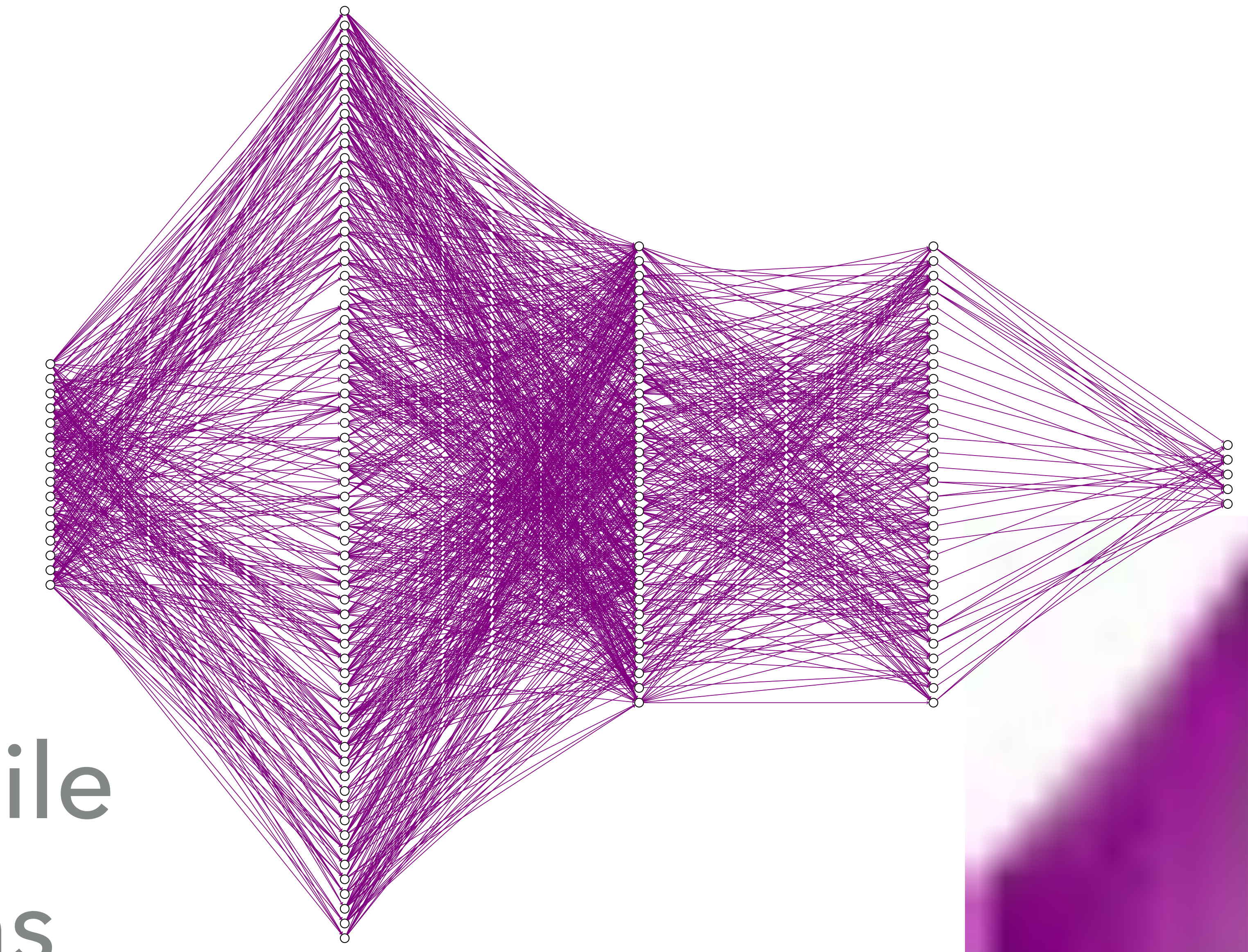
CODESIGN

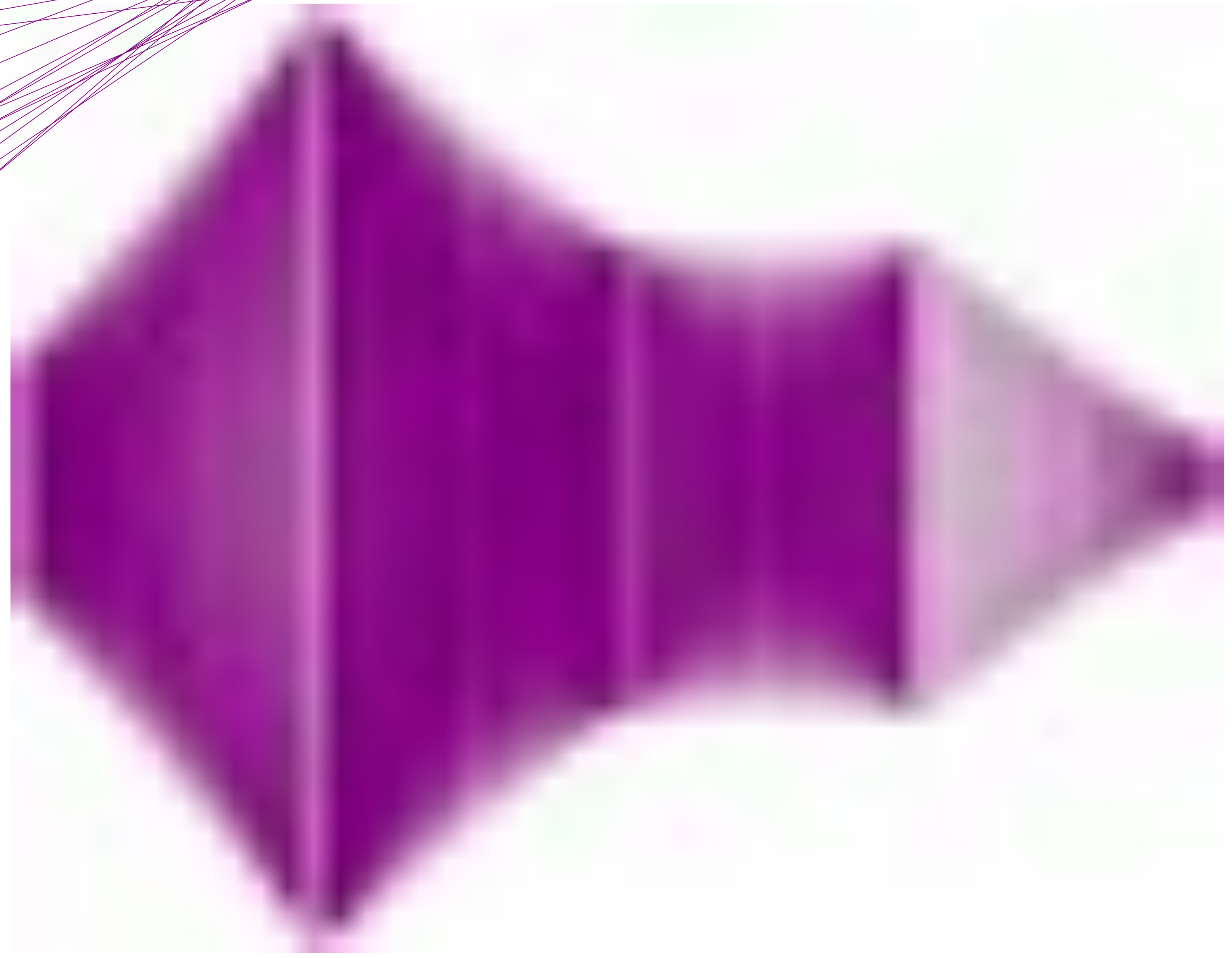
- **Codesign**: intrinsic development loop between algorithm design, training, and implementation
- Compression
  - Maintain high performance while removing redundant operations



CODESIGN

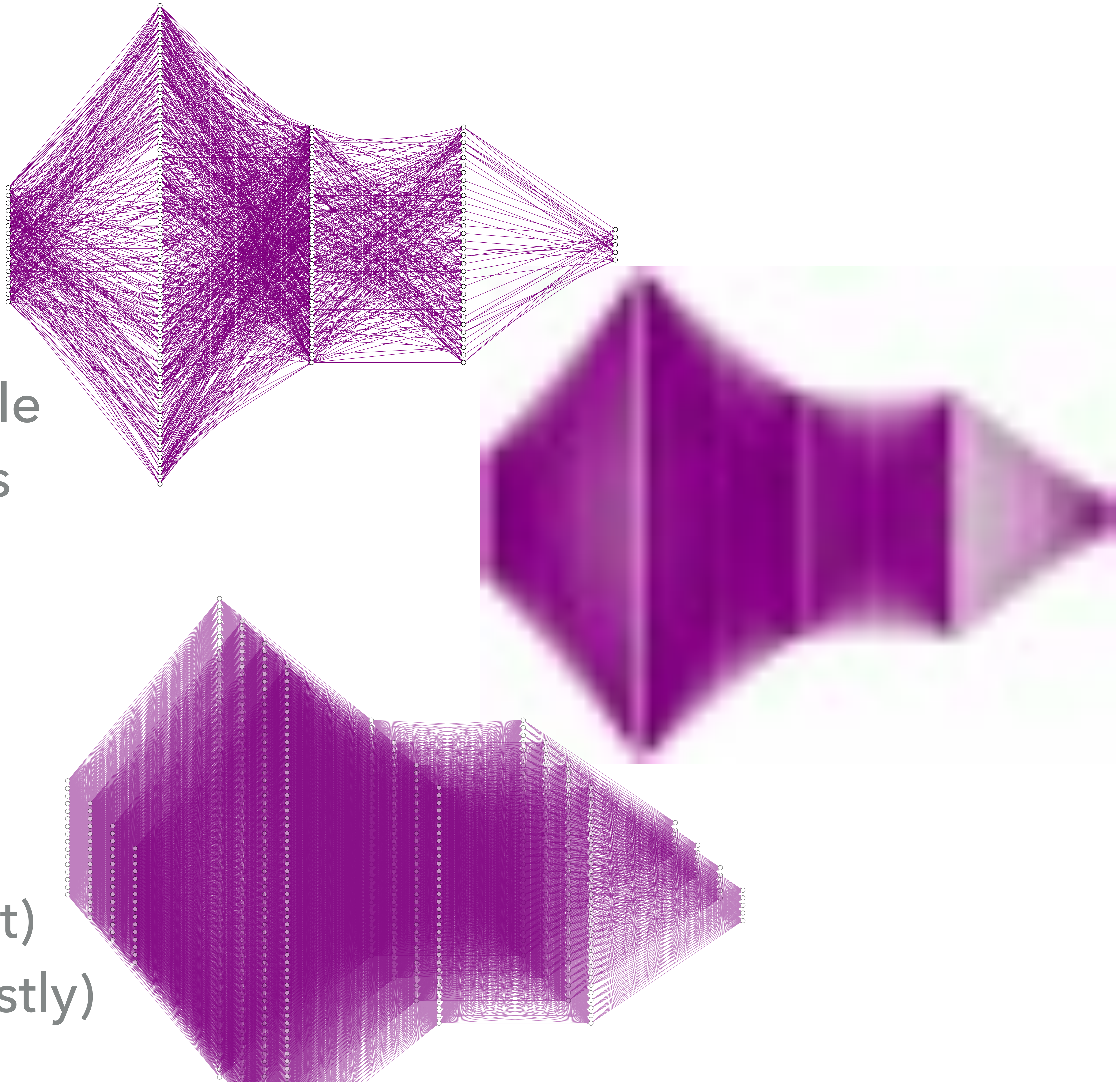
- **Codesign**: intrinsic development loop between algorithm design, training, and implementation
- Compression
  - Maintain high performance while removing redundant operations
- Quantization
  - Reduce precision from 32-bit floating point to 16-bit, 8-bit, …





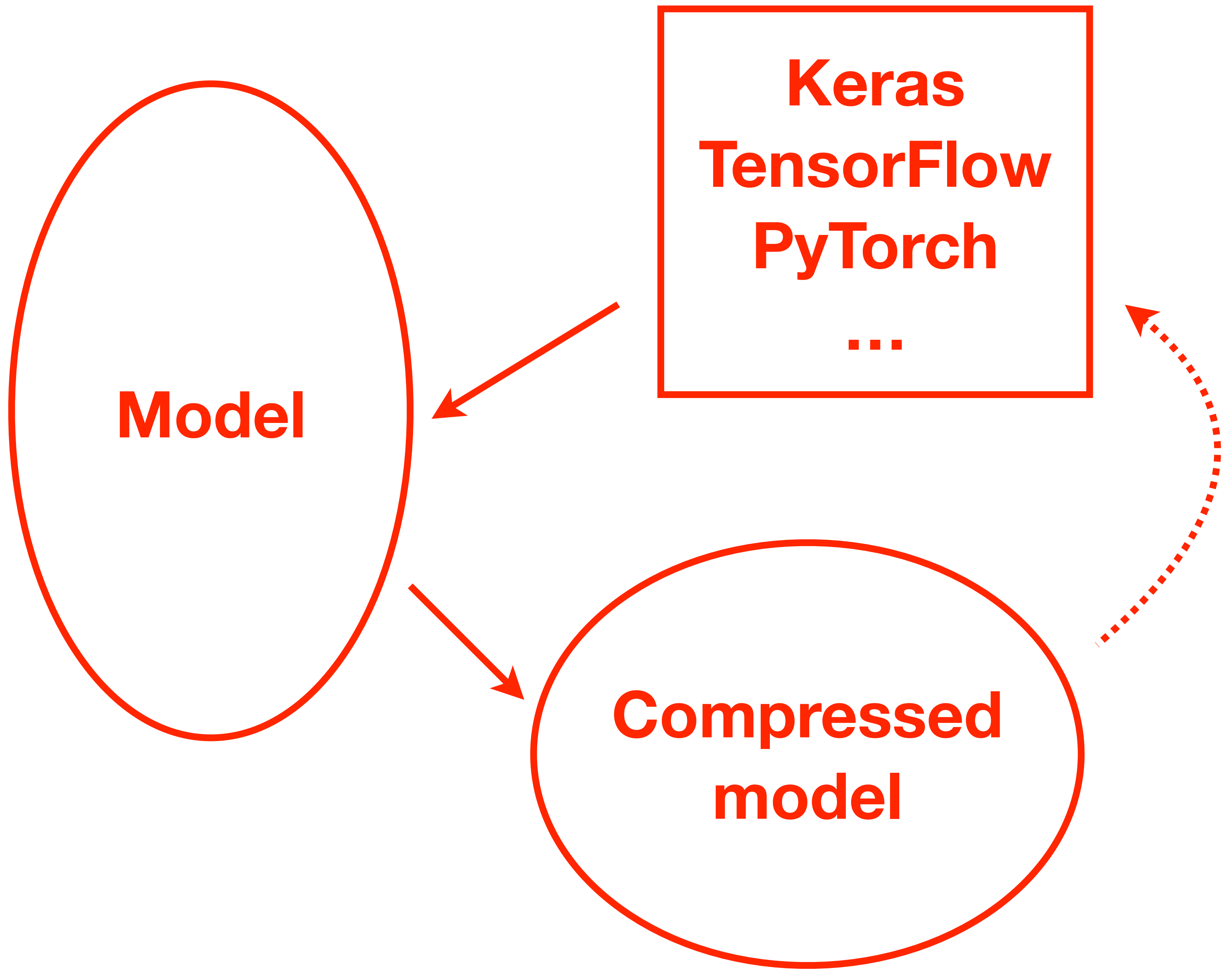
CODESIGN

- **Codesign**: intrinsic development loop between algorithm design, training, and implementation
- Compression
  - Maintain high performance while removing redundant operations
- Quantization
  - Reduce precision from 32-bit floating point to 16-bit, 8-bit, …
- Parallelization
  - Balance parallelization (how fast) with resources needed (how costly)



DESIGN EXPLORATION WITH HLS4ML

- hls4ml for scientists or ML experts to translate ML algorithms into RTL firmware

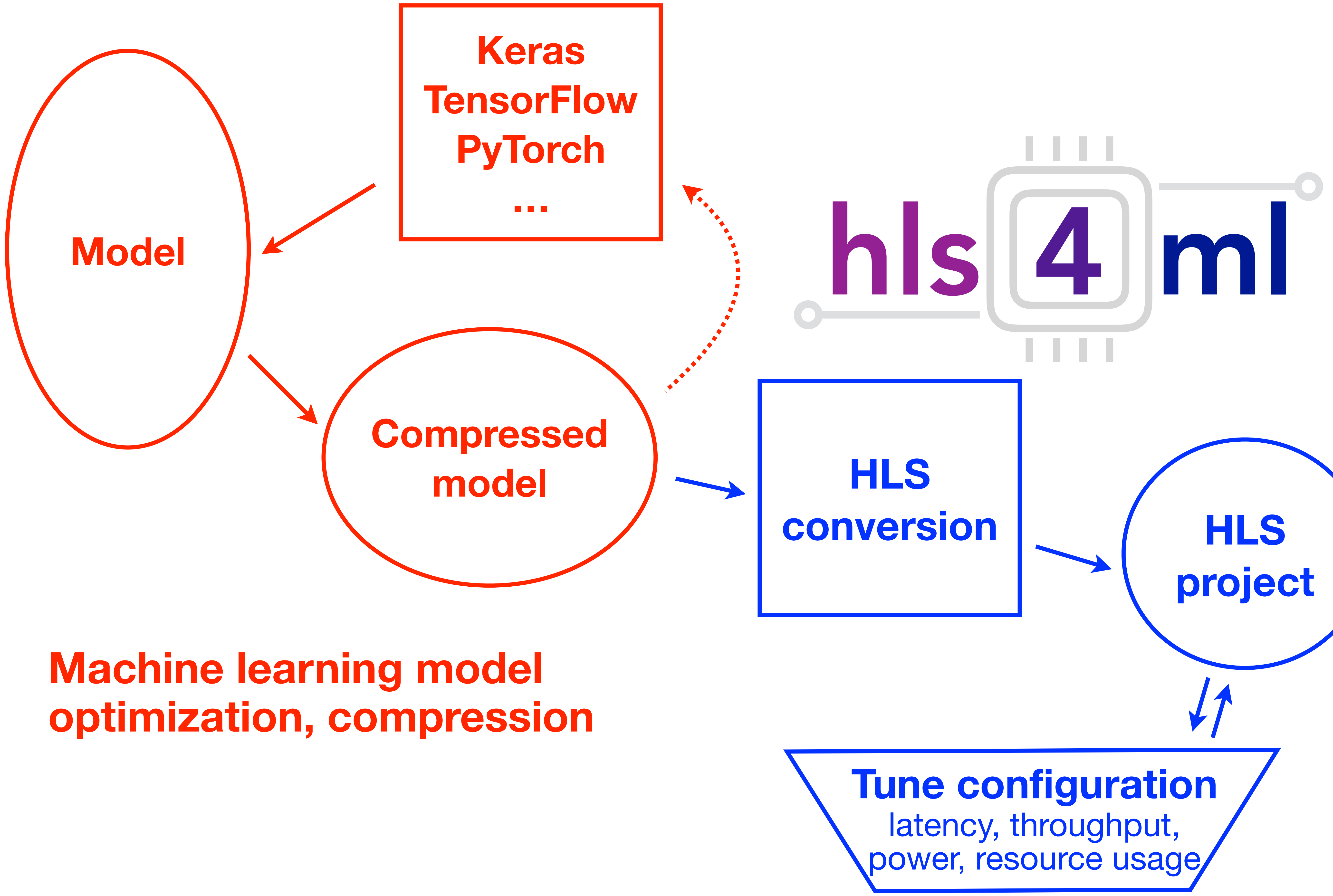


Machine learning model optimization, compression

DESIGN EXPLORATION WITH HLS4ML

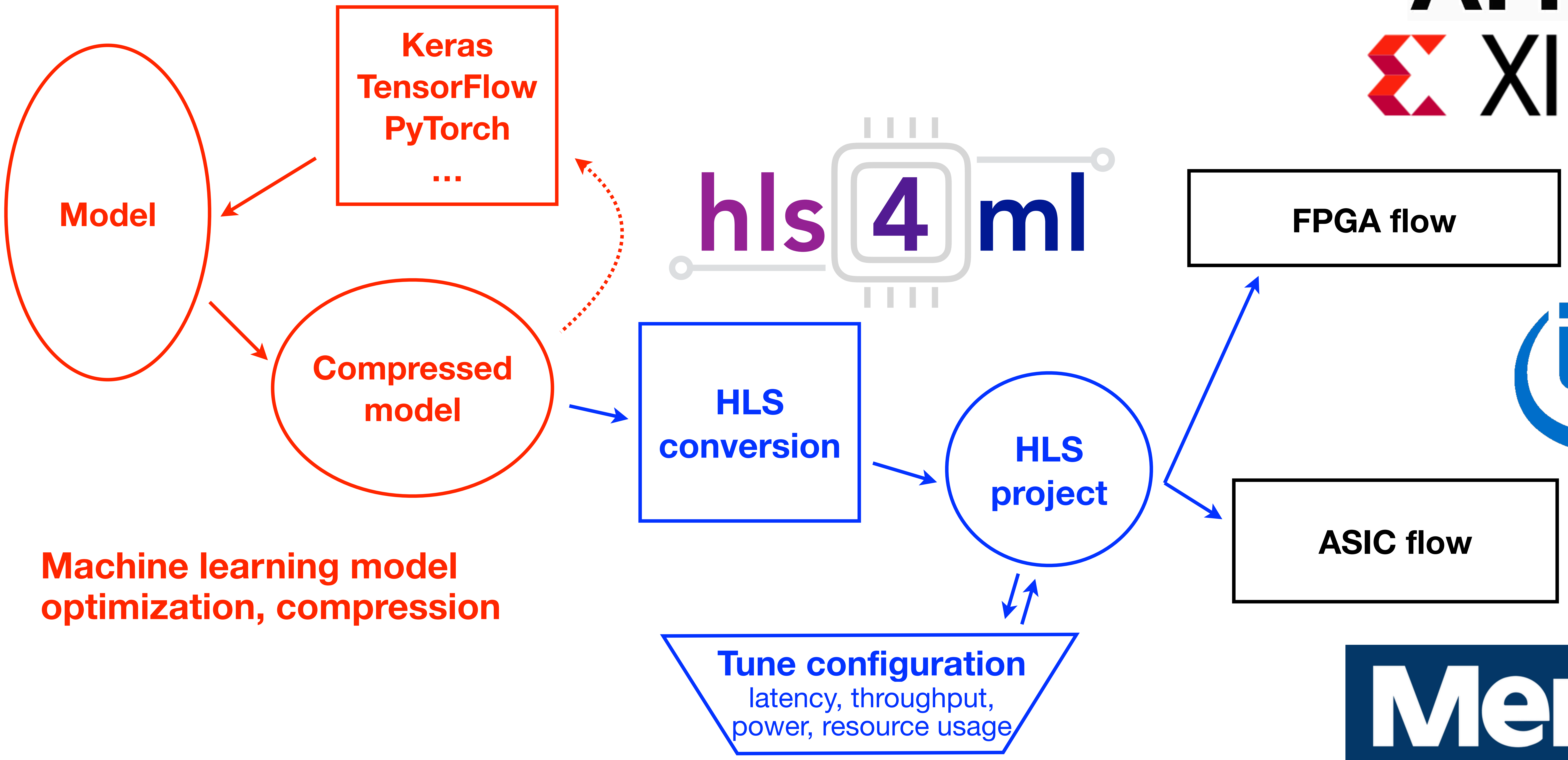
J. Instrum. 13, P07027 (2018)

- hls4ml for scientists or ML experts to translate ML algorithms into RTL firmware



DESIGN EXPLORATION WITH HLS4ML

- hls4ml for scientists or ML experts to translate ML algorithms into RTL firmware



APPLICATION: ANOMALY DETECTION AT 40 MHZ

arXiv:2108.03986

Data challenge: mpp-hep.github.io/ADC2021

APPLICATION: ANOMALY DETECTION AT 40 MHZ

arXiv:2108.03986

Data challenge: mpp-hep.github.io/ADC2021

- Challenge: if new physics has an unexpected signature that doesn't align with existing triggers, precious BSM events may be discarded at trigger level

APPLICATION: ANOMALY DETECTION AT 40 MHZ

arXiv:2108.03986
Data challenge: mpp-hep.github.io/ADC2021

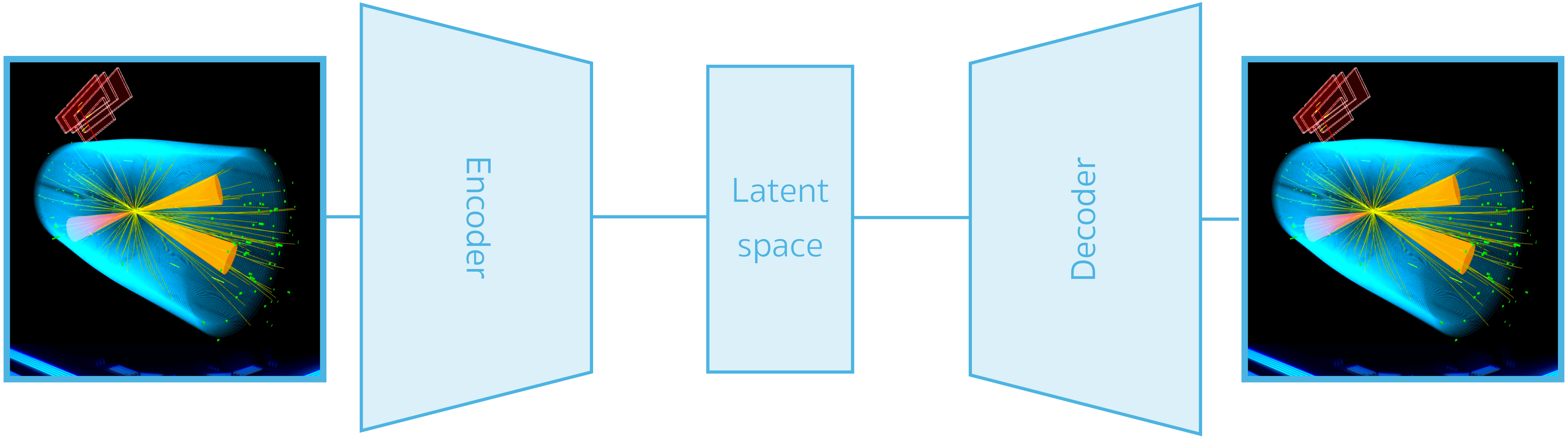
- Challenge: if new physics has an unexpected signature that doesn't align with existing triggers, precious BSM events may be discarded at trigger level
- Can we use unsupervised algorithms to detect non-SM-like anomalies?

APPLICATION: ANOMALY DETECTION AT 40 MHZ

arXiv:2108.03986

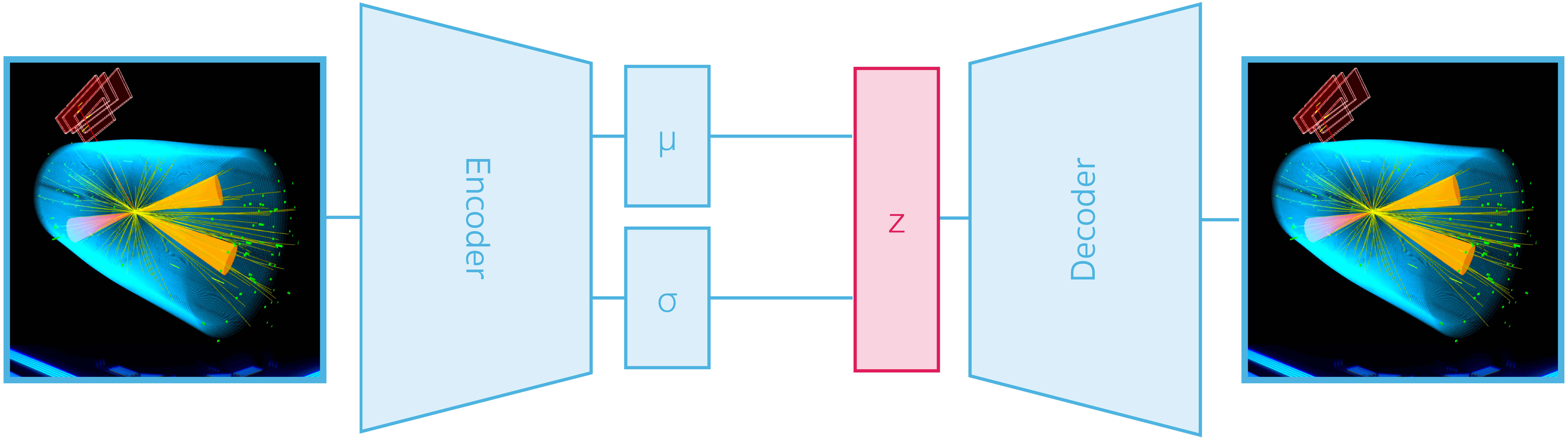
Data challenge: mpp-hep.github.io/ADC2021

- Challenge: if new physics has an unexpected signature that doesn't align with existing triggers, precious BSM events may be discarded at trigger level
- Can we use unsupervised algorithms to detect non-SM-like anomalies?
  - Autoencoders (AEs): compress input to a smaller dimensional latent space then decompress and calculate difference



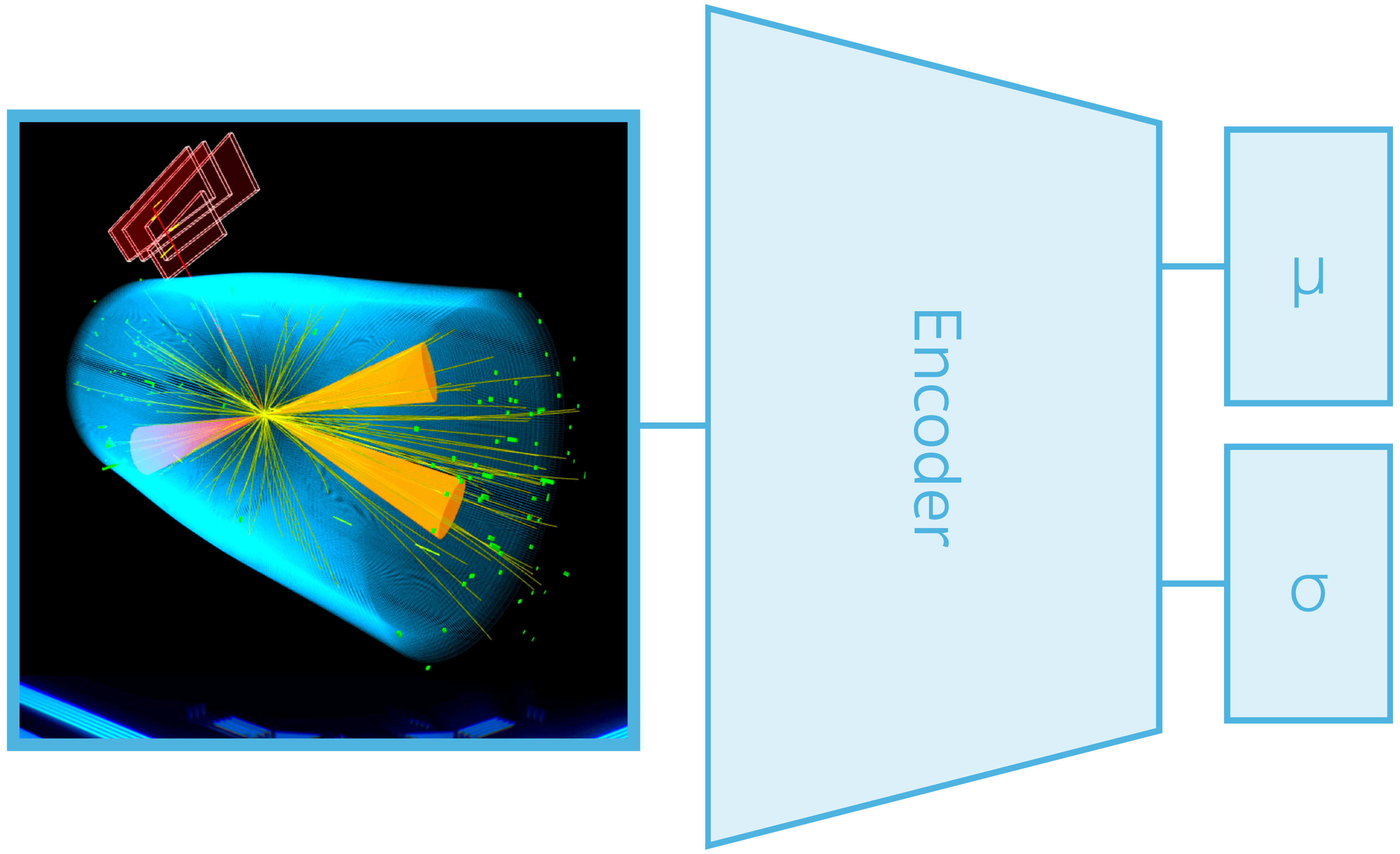
APPLICATION: ANOMALY DETECTION AT 40 MHZ

- Challenge: if new physics has an unexpected signature that doesn't align with existing triggers, precious BSM events may be discarded at trigger level
- Can we use unsupervised algorithms to detect non-SM-like anomalies?
  - Autoencoders (AEs): compress input to a smaller dimensional latent space then decompress and calculate difference
  - Variational autoencoders (VAEs): model the latent space as a probability distribution; possible to detect anomalies purely with latent space variables



APPLICATION: ANOMALY DETECTION AT 40 MHZ

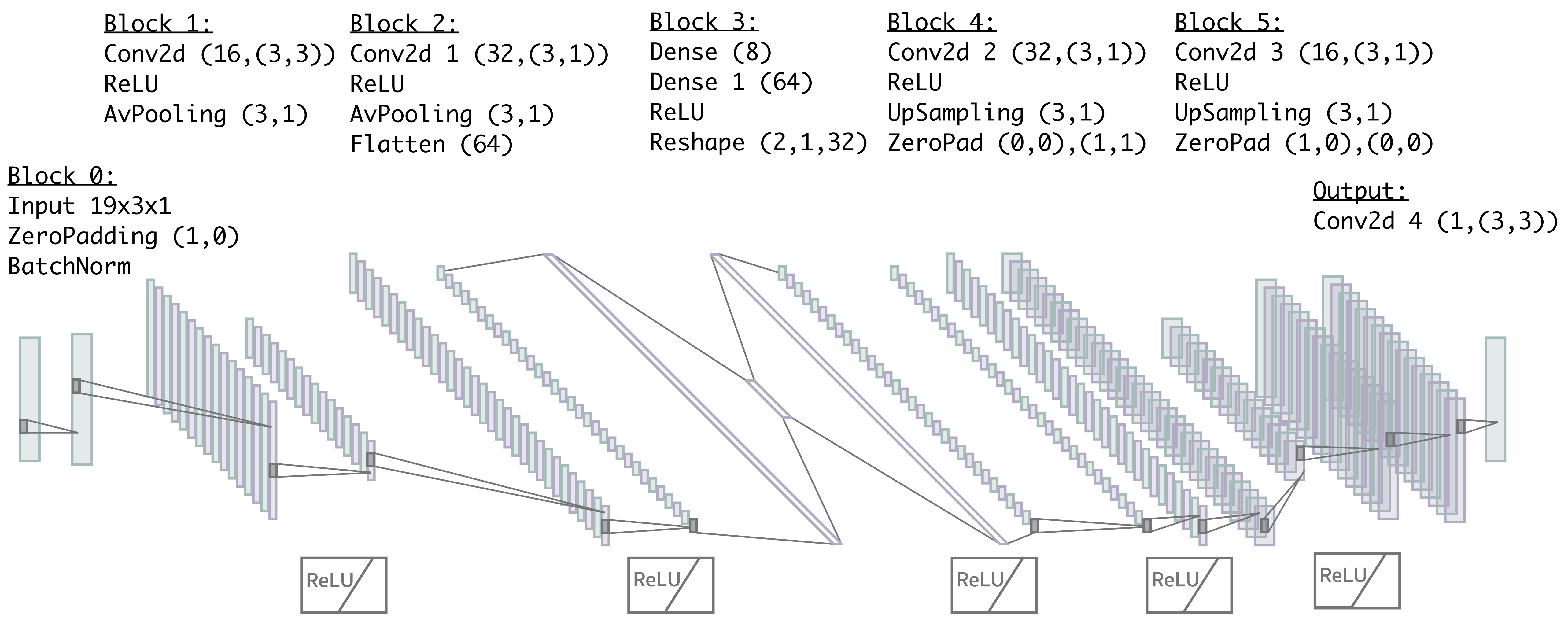
- Challenge: if new physics has an unexpected signature that doesn't align with existing triggers, precious BSM events may be discarded at trigger level
- Can we use unsupervised algorithms to detect non-SM-like anomalies?
  - Autoencoders (AEs): compress input to a smaller dimensional latent space then decompress and calculate difference
  - Variational autoencoders (VAEs): model the latent space as a probability distribution; possible to detect anomalies purely with latent space variables



Key observation: Can build an anomaly score from the latent space of VAE directly! No need to run decoder!

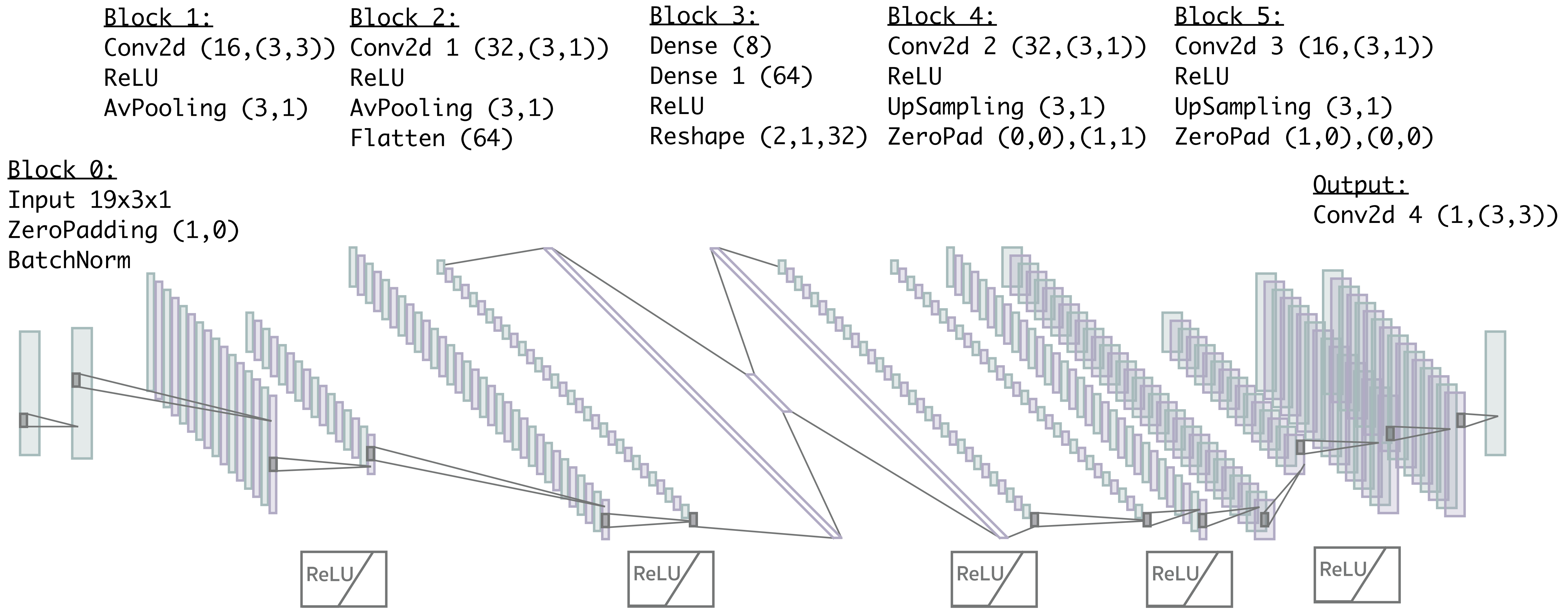
$$R_z = \sum_i \frac{\mu_i^2}{\sigma_i^2}$$

FPGA IMPLEMENTATION



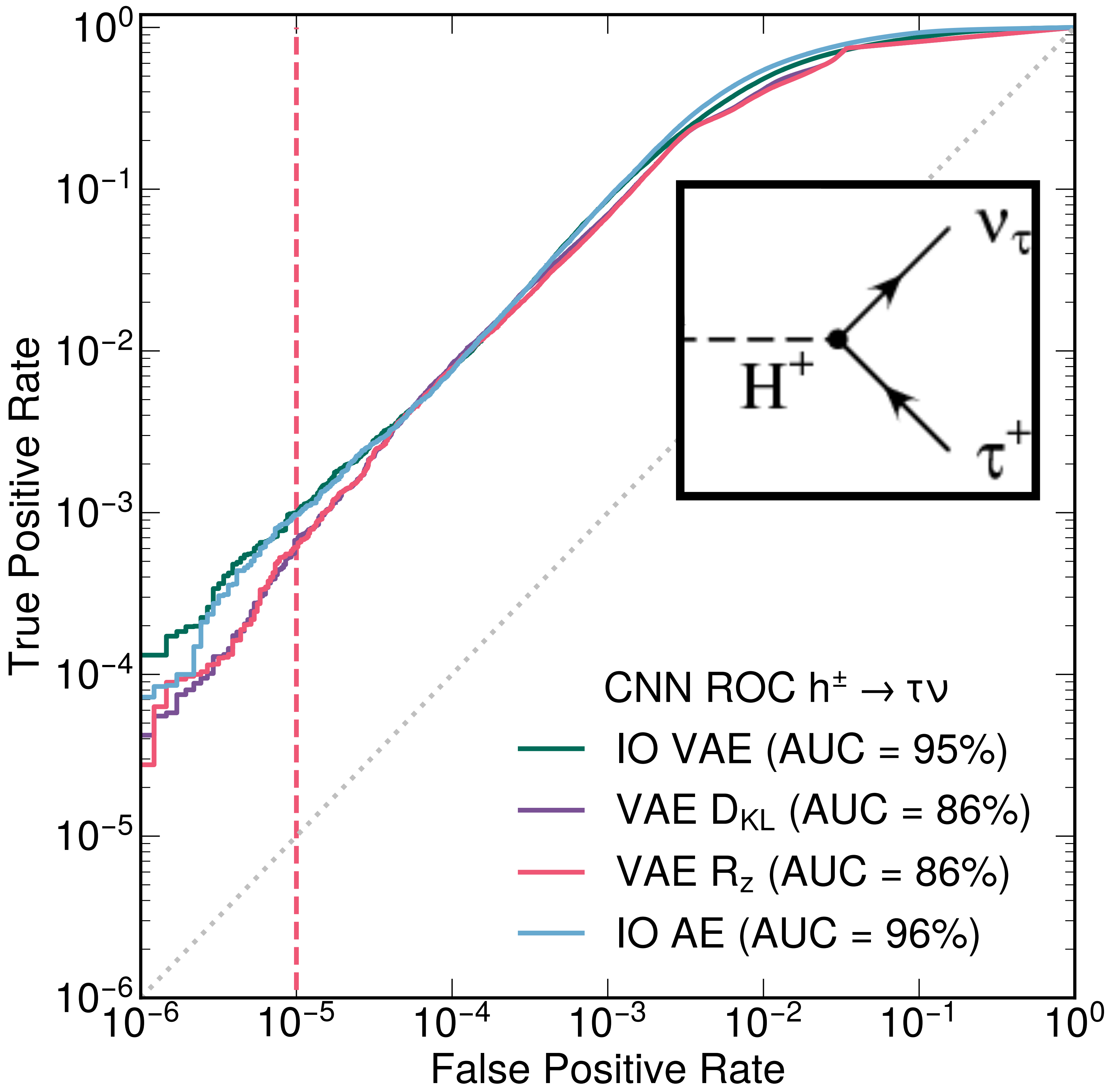
FPGA IMPLEMENTATION

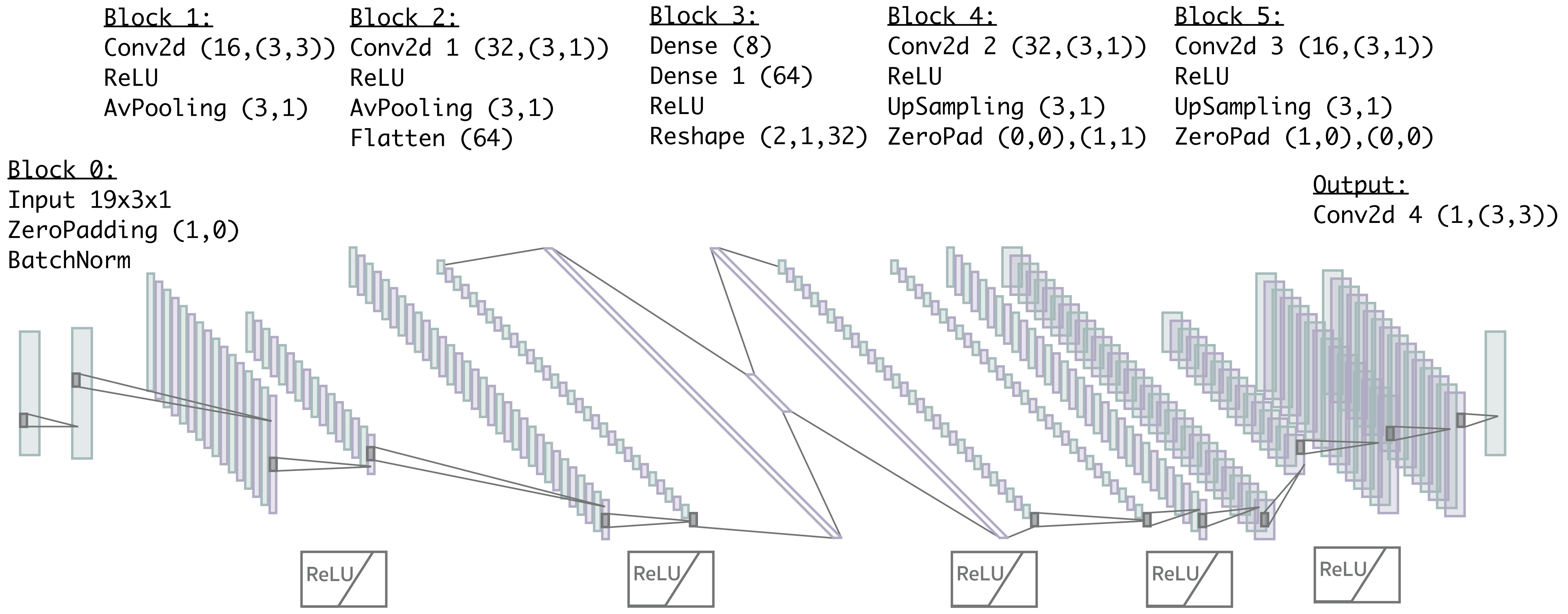
- CNNs as the basis for (V)AEs for anomaly detection



FPGA IMPLEMENTATION

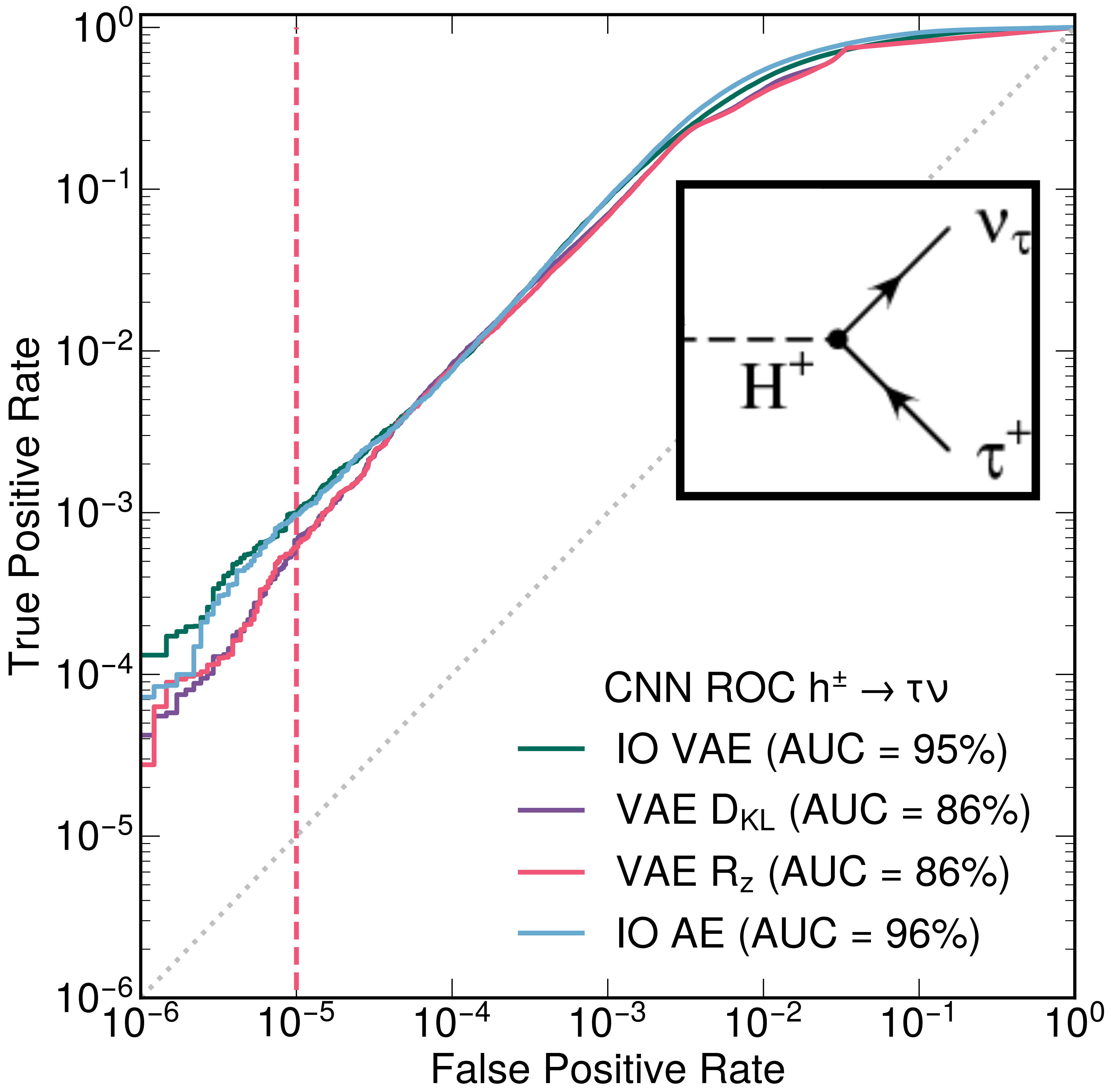
- CNNs as the basis for (V)AEs for anomaly detection
- Good anomaly detection performance for unseen signals ($LQ \rightarrow b\tau$, $A \rightarrow 4l$, $\mathbf{h^{\pm} \rightarrow \tau\nu}$, $h^0 \rightarrow \tau\tau$)

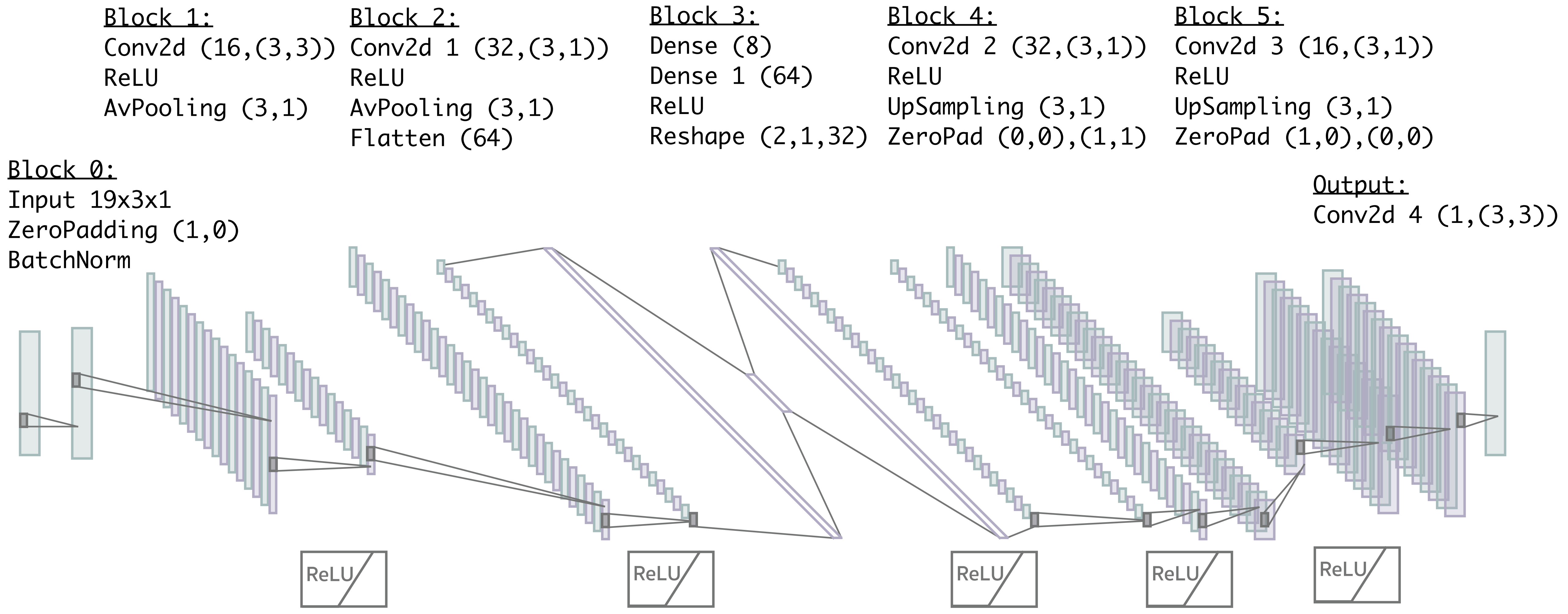




FPGA IMPLEMENTATION

- CNNs as the basis for (V)AEs for anomaly detection
- Good anomaly detection performance for unseen signals ($LQ \to b\tau$, $A \to 4l$, $\mathbf{h^\pm \to \tau\nu}$, $h^0 \to \tau\tau$)
- **VAE** fits in latency and resource requirements for HL-LHC!





Model	DSP [%]	LUT [%]	FF [%]	BRAM [%]	Latency [ns]	II [ns]	AUC [%]	TPR @ FPR=10^{-5}
CNN VAE R_z	10	**12**	**4**	**2**	**365**	**115**	86	0.06%
CNN AE	**7**	47	5	6	1480	895	**96**	**0.10%**

I. DATA REPRESENTATIONS
& SYMMETRIES

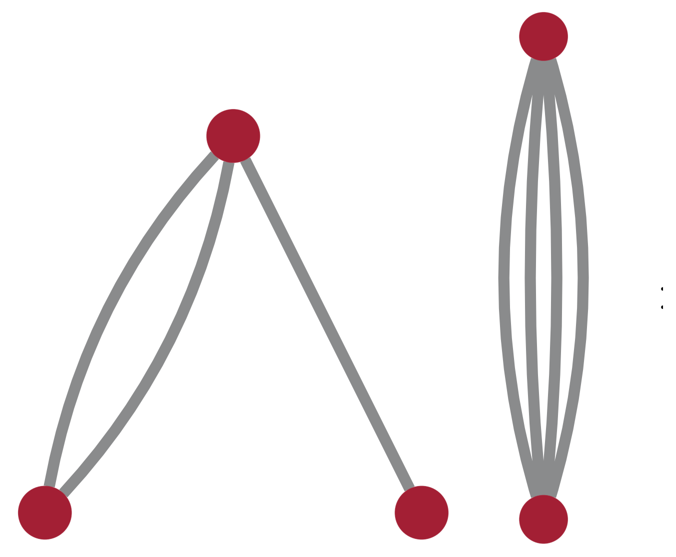
II. ANOMALY DETECTION

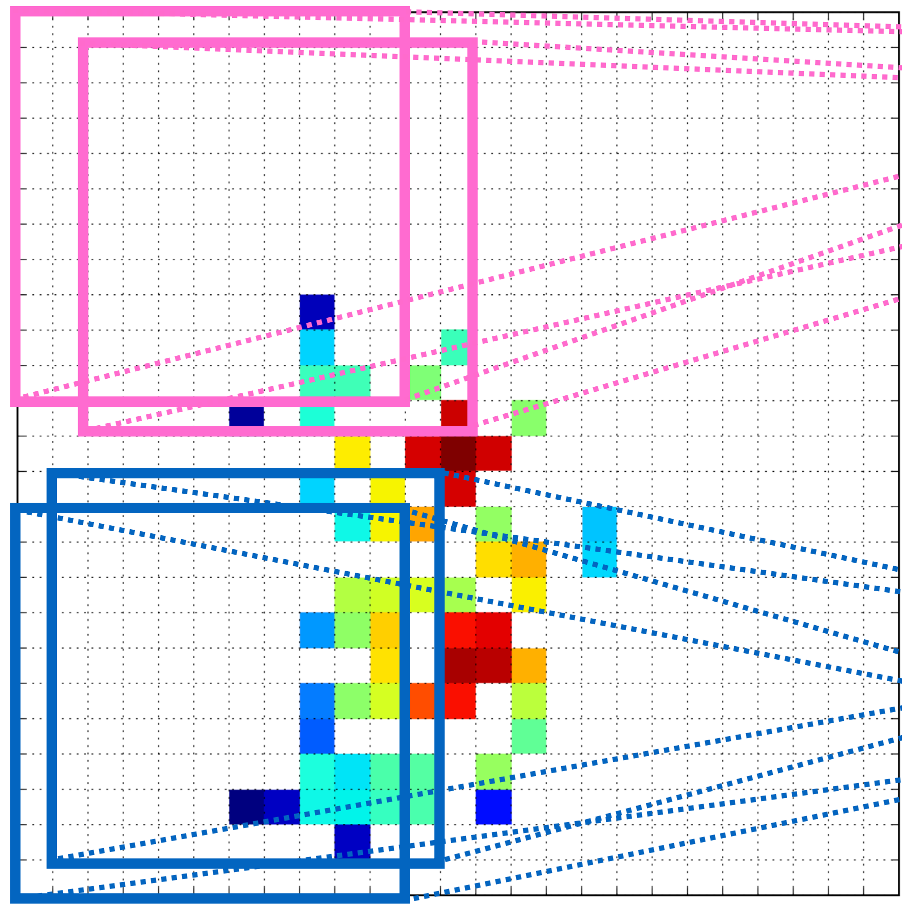
III. GENERATIVE MODELING

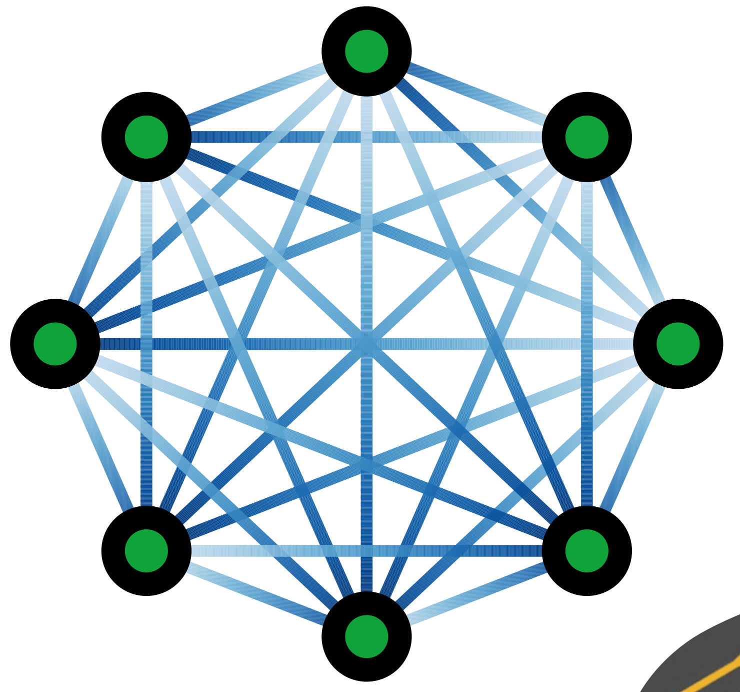
III. FAST INFERENCE

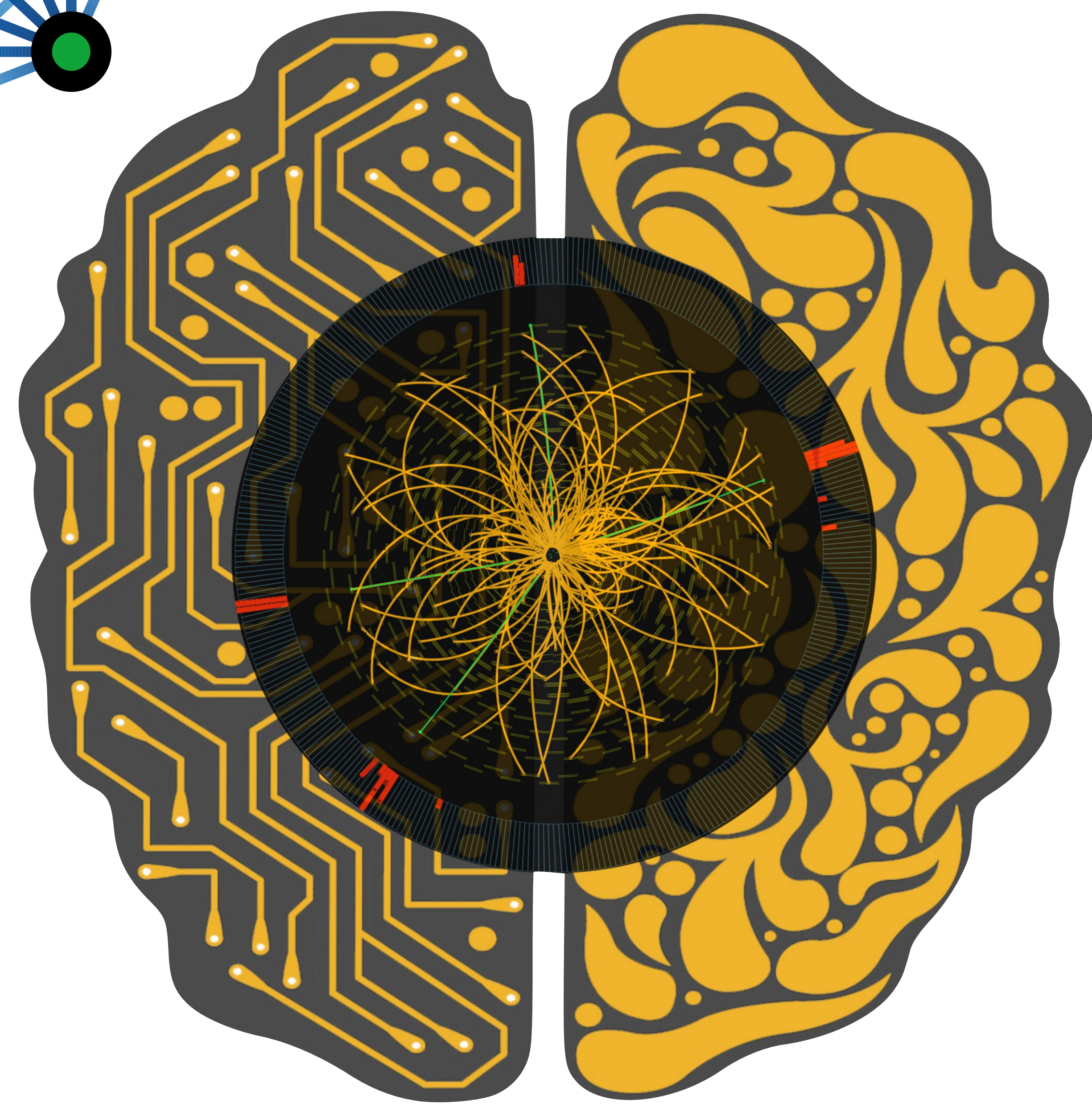
VI. SUMMARY & OUTLOOK

SUMMARY AND OUTLOOK



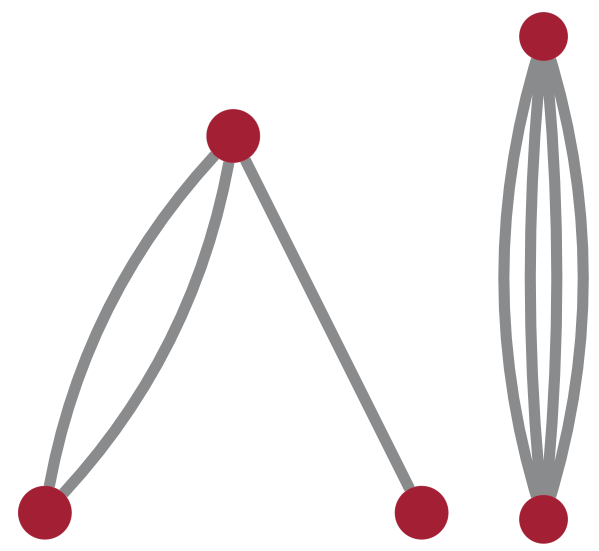


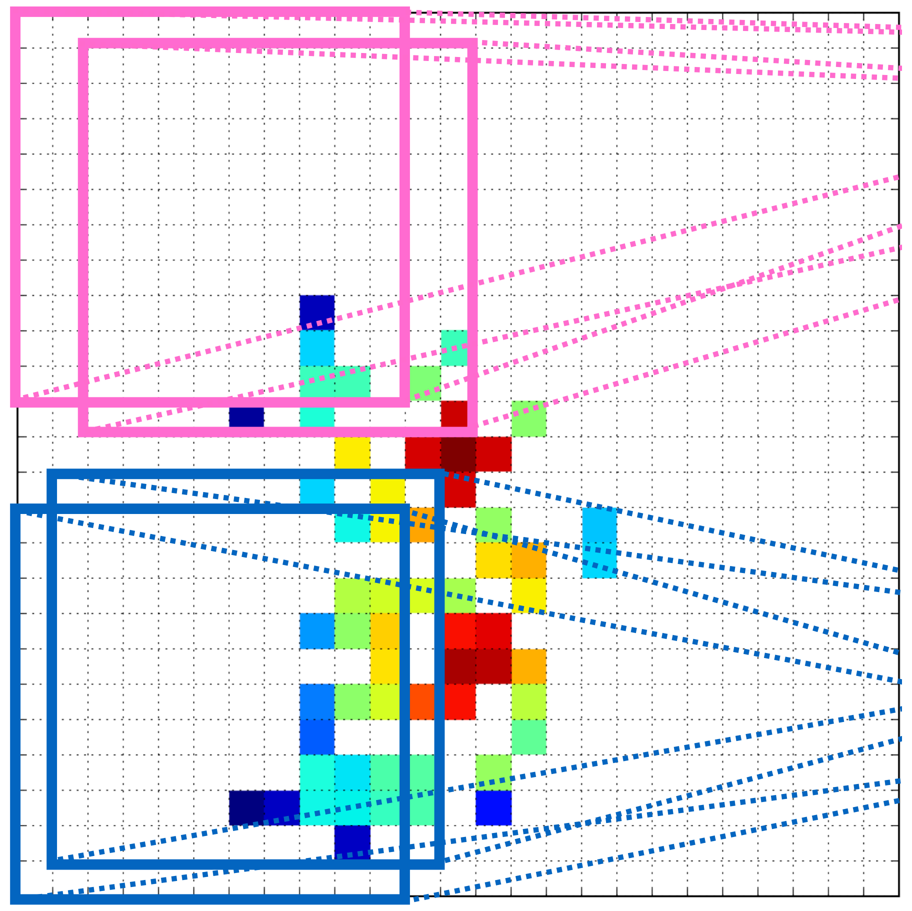


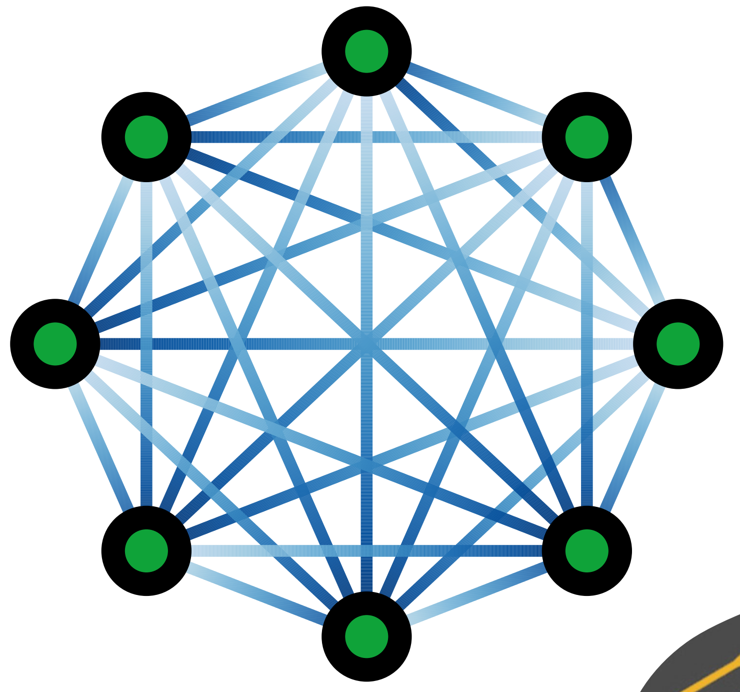


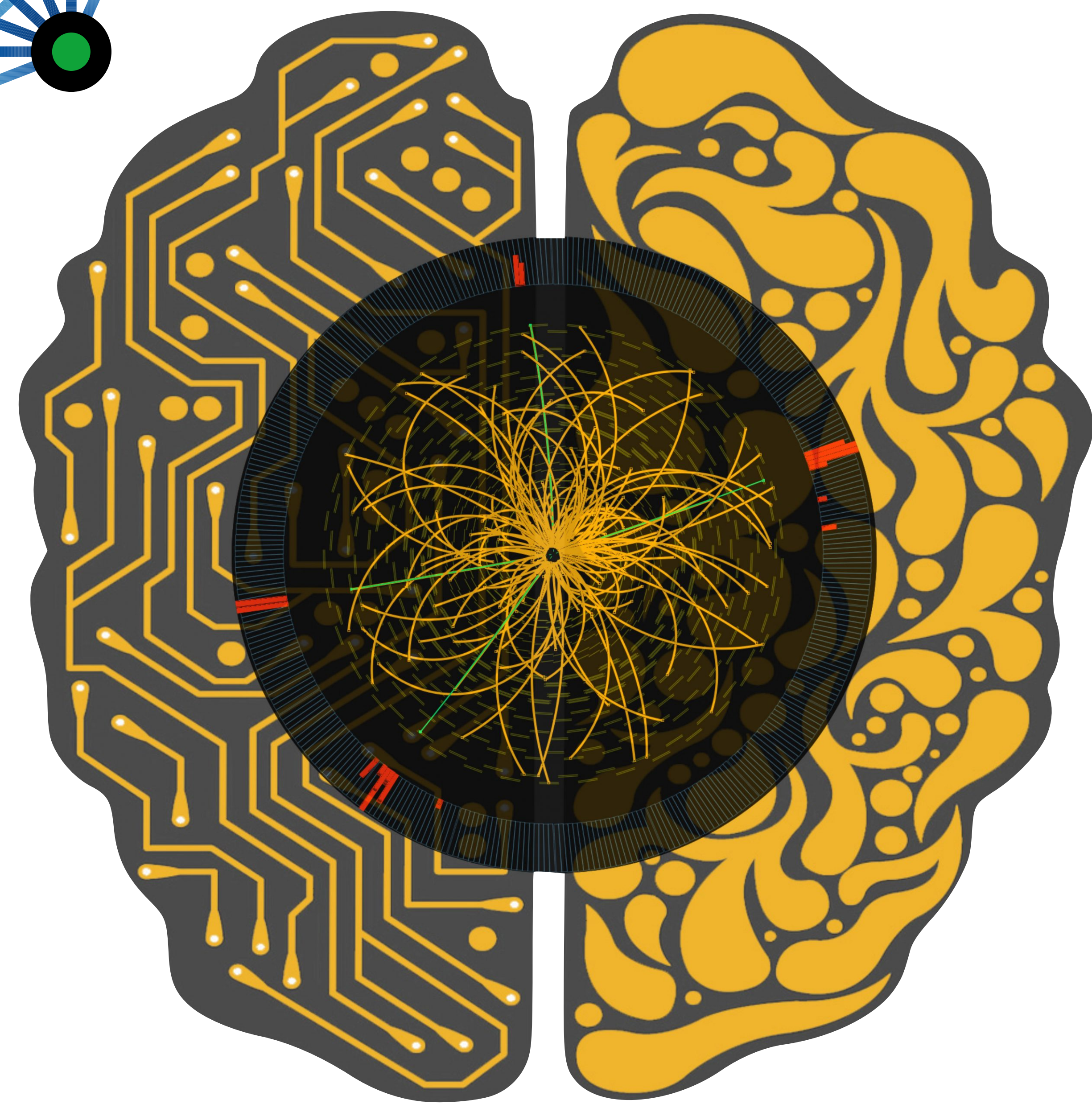
SUMMARY AND OUTLOOK

- Different representations of HEP data, from tabular data, image data, set data, graph data, paired with corresponding algorithms can achieve excellent performance



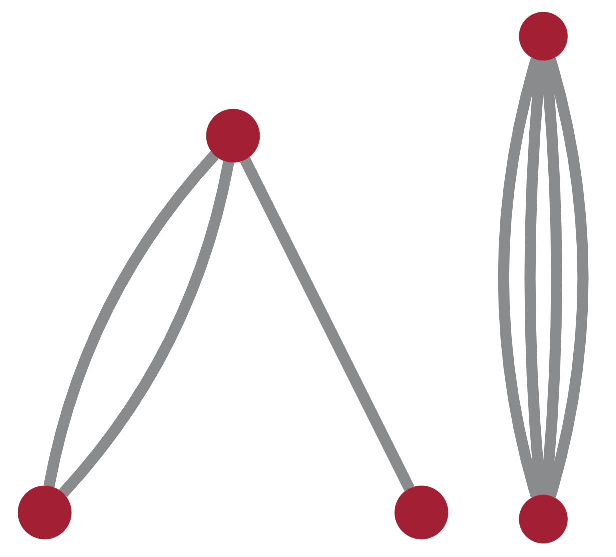


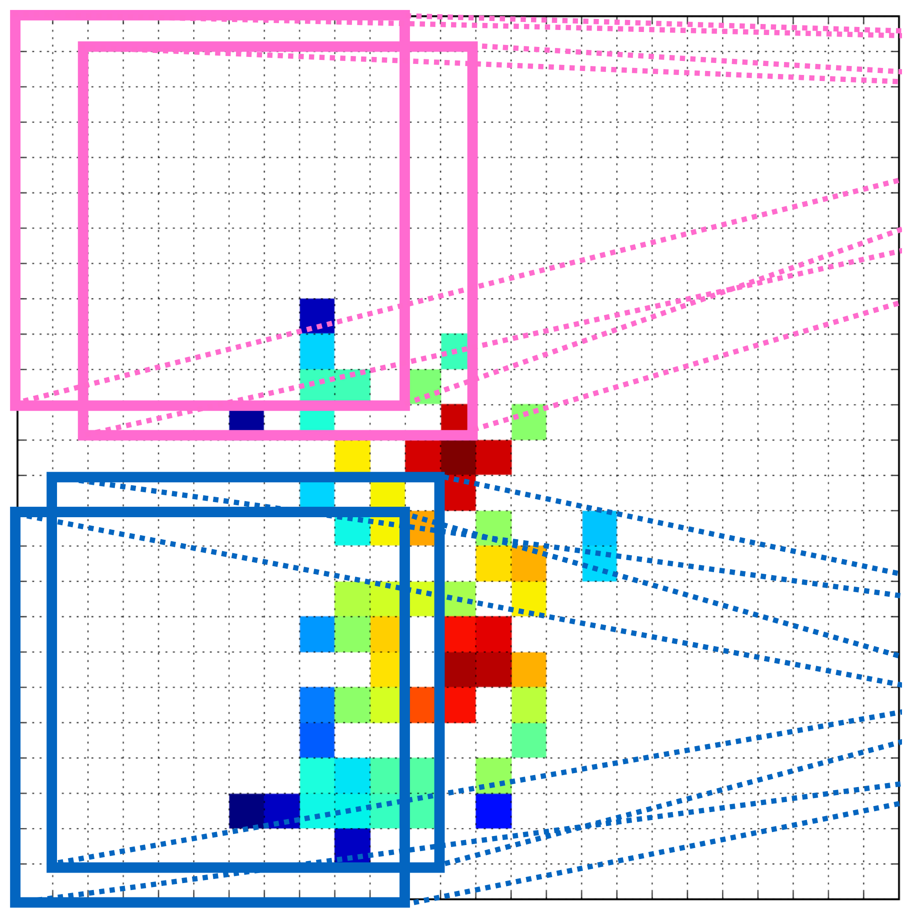


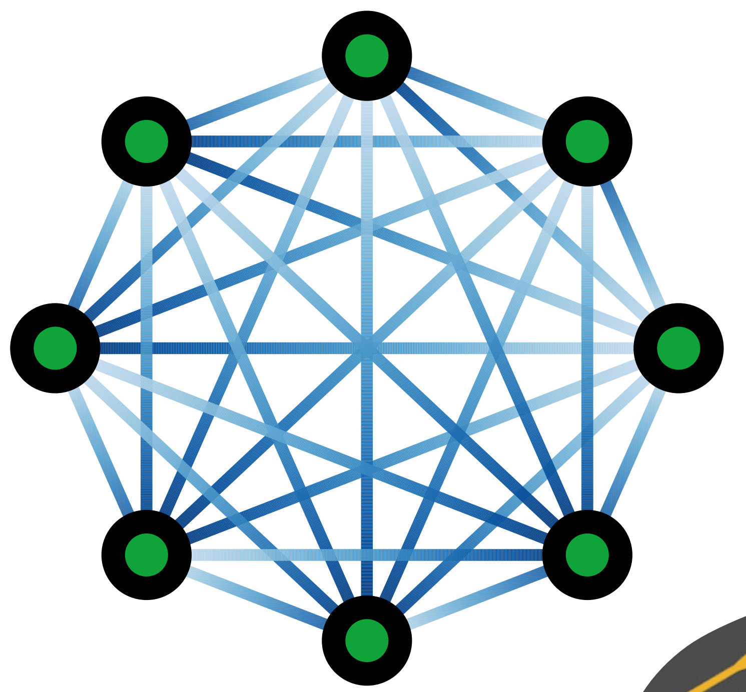


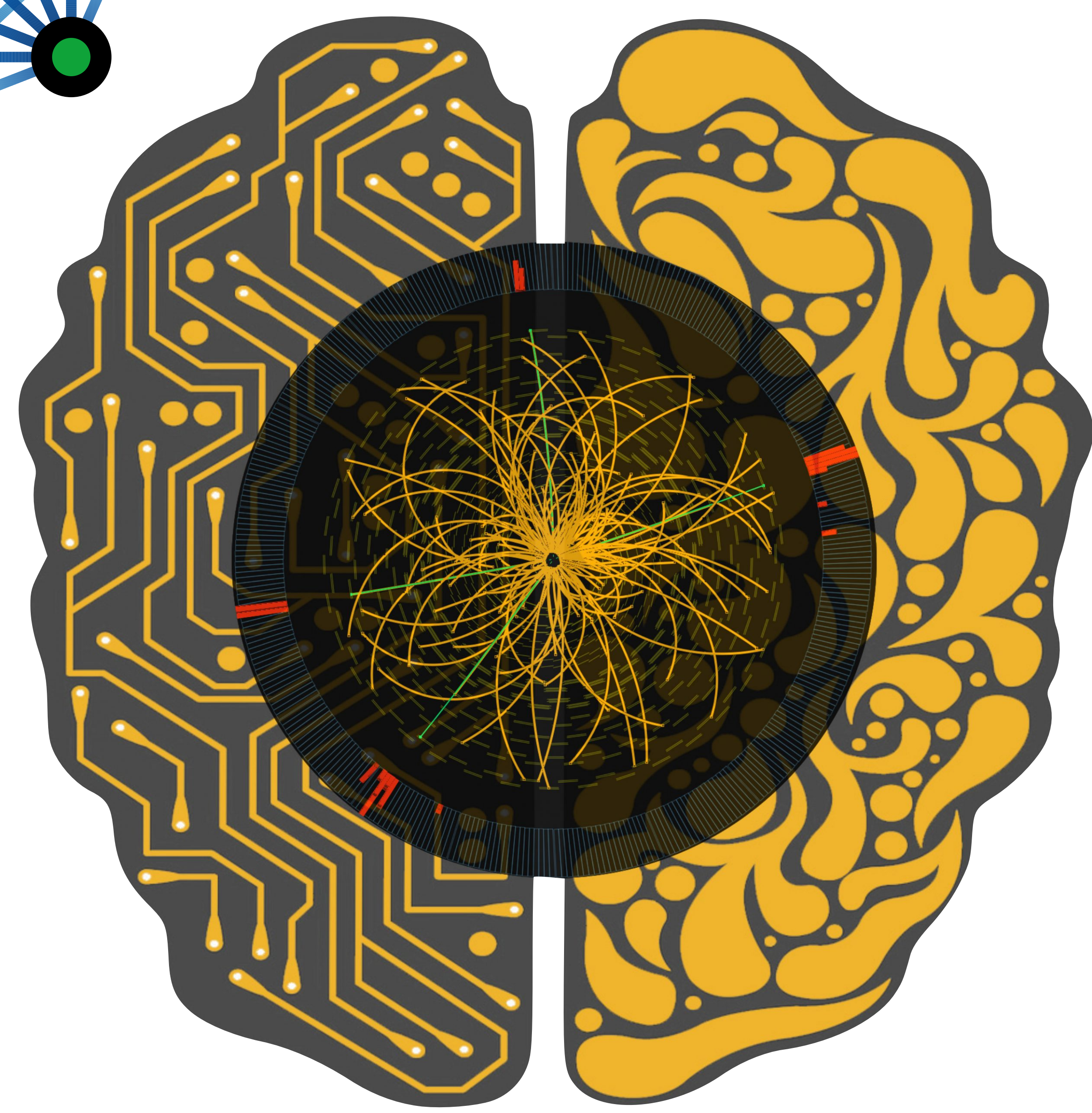
SUMMARY AND OUTLOOK

- Different representations of HEP data, from tabular data, image data, set data, graph data, paired with corresponding algorithms can achieve excellent performance
- Plethora of ML techniques in HEP from anomaly detection to generative modeling have exploded in recent years
  - Availability of public datasets and challenges have advanced the state-of-the-art



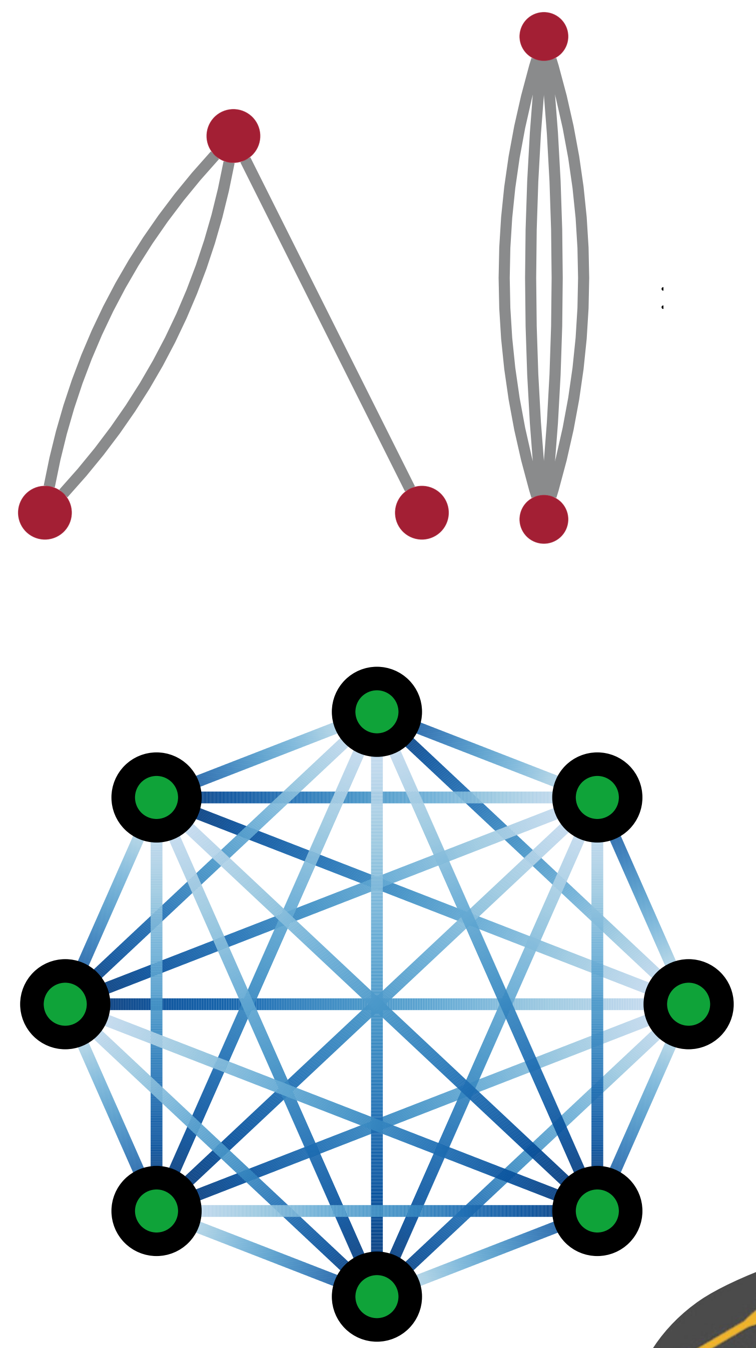


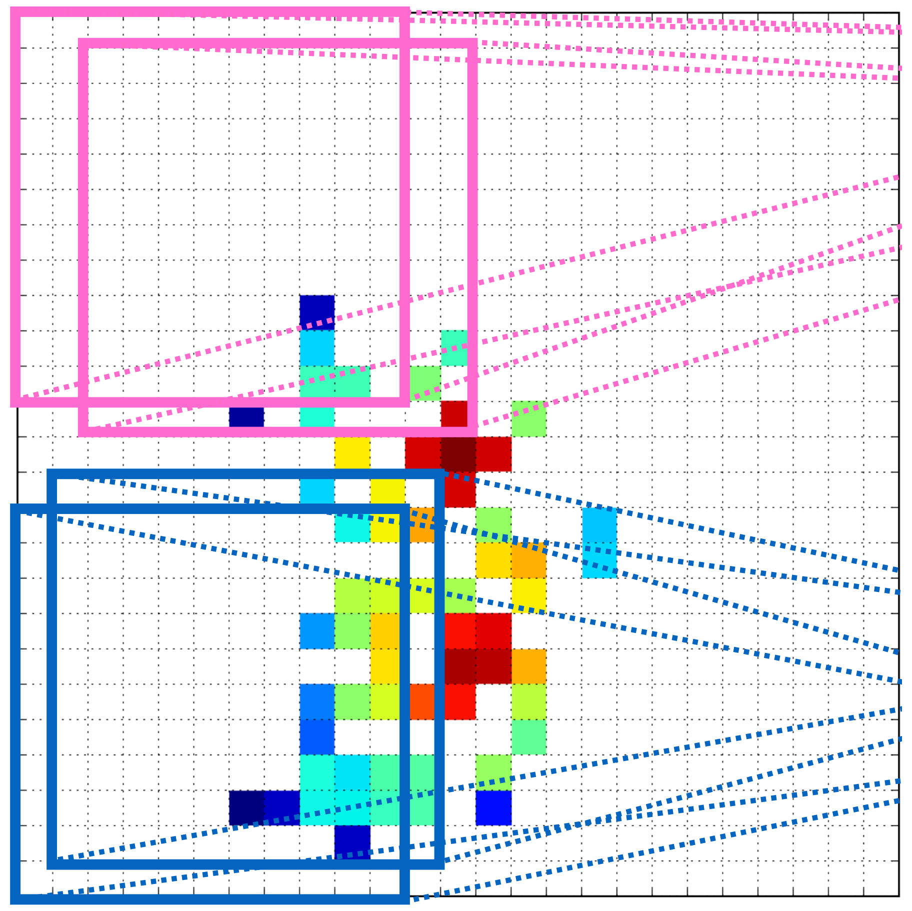


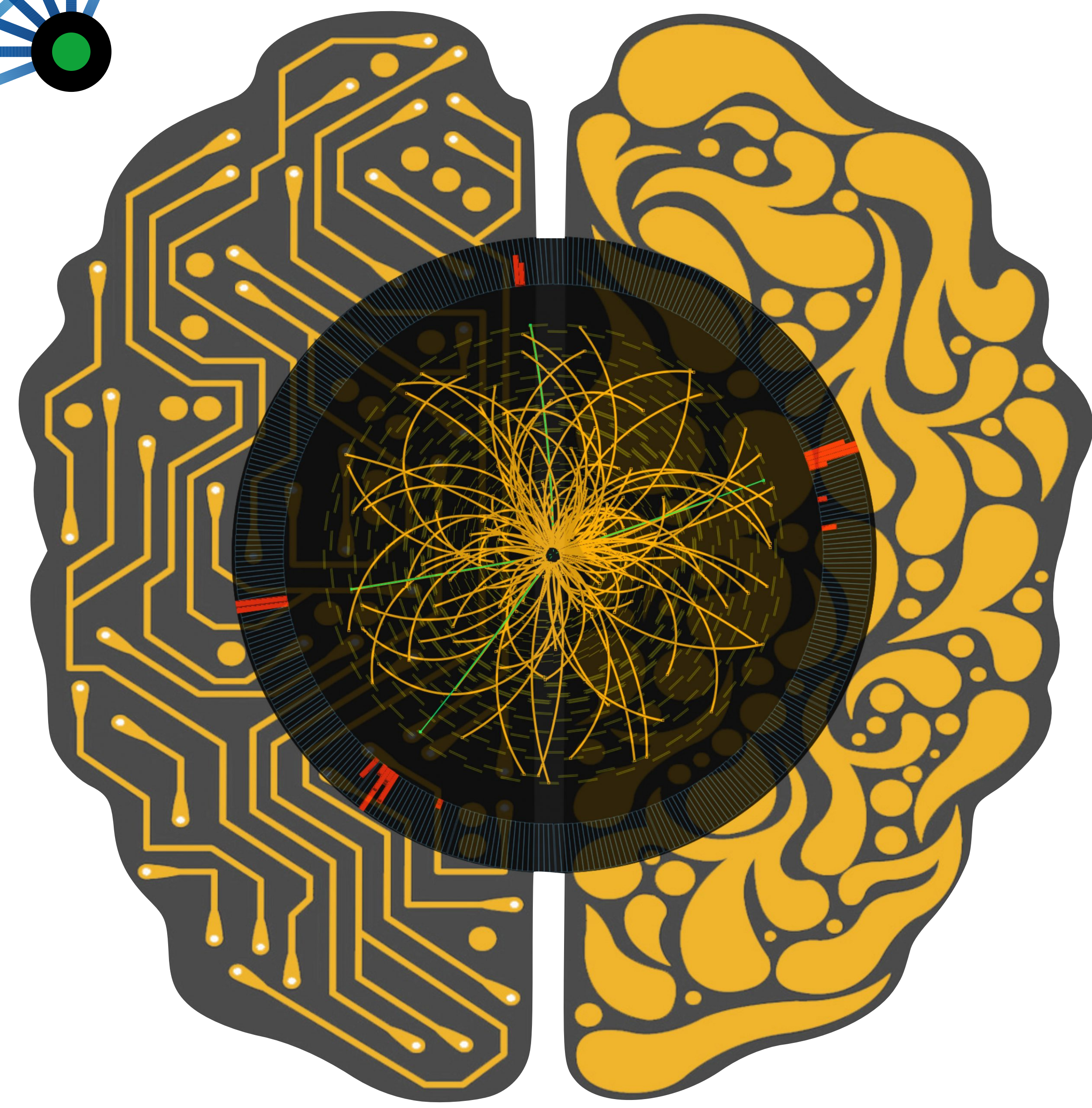


SUMMARY AND OUTLOOK

- Different representations of HEP data, from tabular data, image data, set data, graph data, paired with corresponding algorithms can achieve excellent performance
- Plethora of ML techniques in HEP from anomaly detection to generative modeling have exploded in recent years
  - Availability of public datasets and challenges have advanced the state-of-the-art
- Fast ML can accelerate science allowing us to test hypotheses faster, enhance performance of detectors/accelerators, and save potentially overlooked data







BACKUP

JAVIER DUARTE

DARK INTERACTIONS WORKSHOP

NOVEMBER 16, 2022