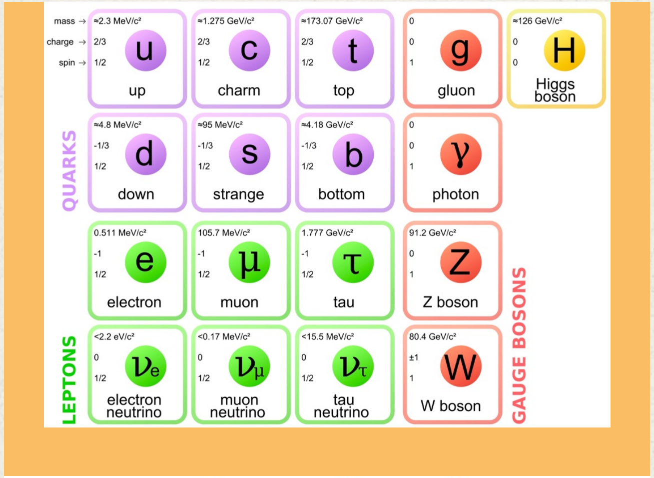
From Neural Networks to Foundation Models: The Evolving Role of AI in Particle Physics

Savannah Thais, Columbia University

A Quick Refresher on Particle Physics

01.

And its computational challenges



$$
\begin{aligned}
L = & -\tfrac{1}{4} W_{\mu\nu} W^{\mu\nu} - \tfrac{1}{4} B_{\mu\nu} B^{\mu\nu} - \tfrac{1}{4} G_{\mu\nu} G^{\mu\nu} \\
& + \psi_j \gamma^{\mu} (i\,\delta_{\mu} - g^{\tau} W_{\mu} - g^{i} Y_j B_{\mu} - g_j T_j G_{\mu}) \psi_j \\
& + | D_{\mu} \phi |^2 + \mu^2 | \phi |^2 - \lambda | \phi |^4 \\
& - (y_i \psi_{iL} \phi \psi_{jR} + y_i \psi_{jL} \phi_c \phi_{jR} + \text{conjugate})
\end{aligned}
$$

The Standard Model

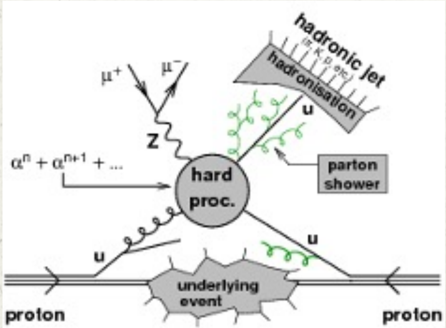
- Makes up all visible matter in the universe
- Creates (almost) all known forces and mass
  - Electricity, magnetism, strong, and weak
- Helps us understand the origins and evolution of the universe
- All interactions governed by mathematical rules (Lagrangian)

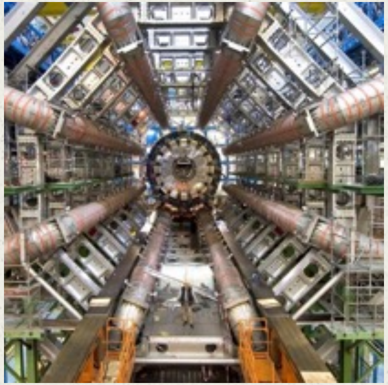
But....

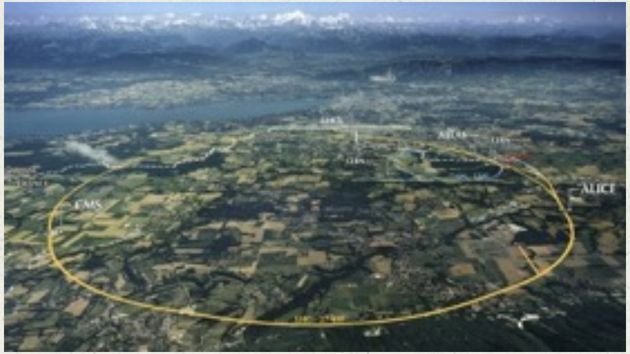
- What is dark matter?
- Where is the graviton?
- Is super symmetry real?
- How do neutrinos change flavor?
- Are there more generations of matter?
- What is dark energy?
- Why is the top so heavy? The Higgs so light?
- Are there other Higgs bosons?

The Large Hadron Collider

- 17 mile proton-proton collider under the the French-Swiss border
- Produces 1000 million collisions per second
- Allows us to study the fundamental constituents of matter



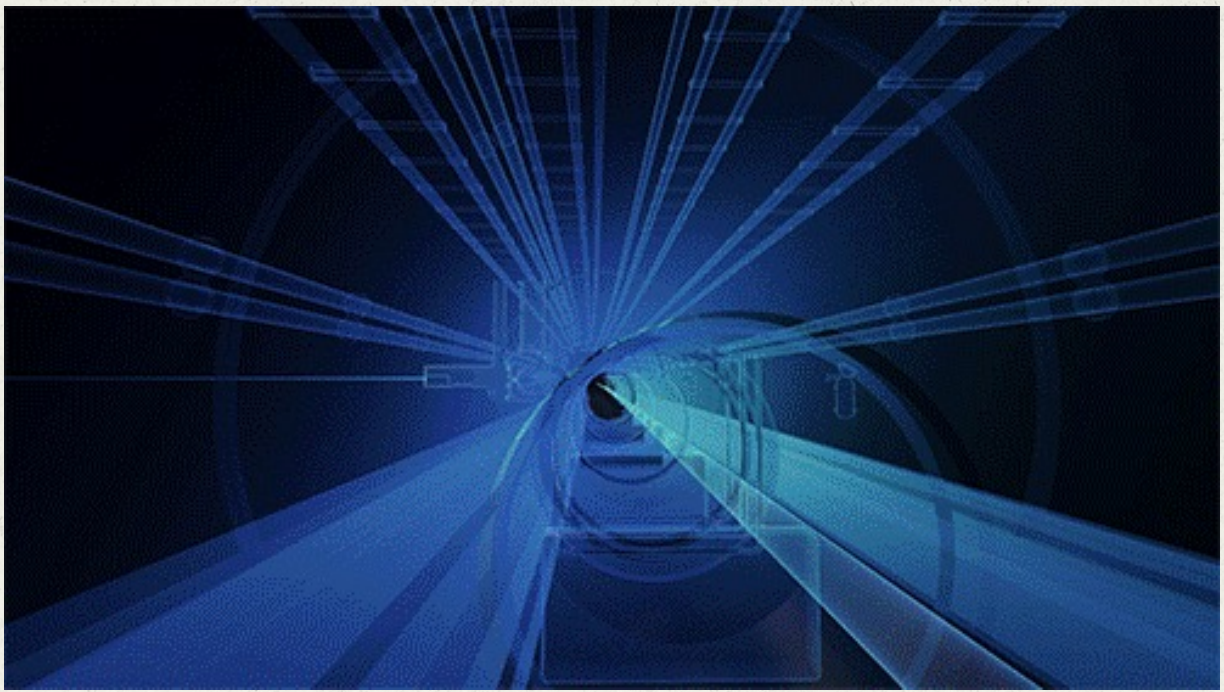




Particle Collisions:

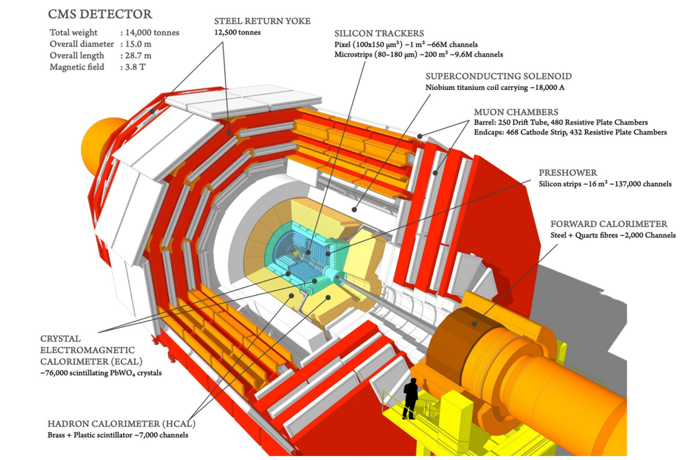
- Accelerate protons to .99x the speed of light
- Collide the accelerated particles
- E=mc2, so the high energy collisions create rare, exciting particles
- Measure the decay products with specialized detectors.

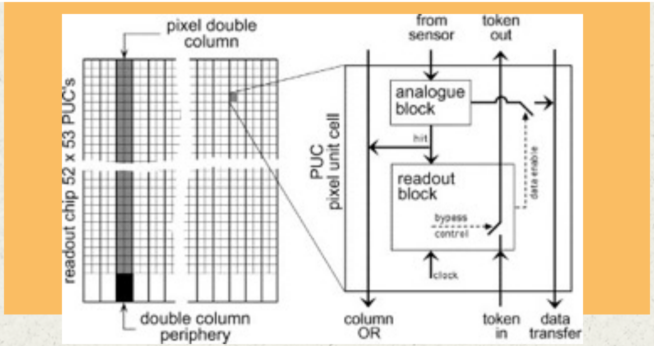
Particle Collisions



Physics at the LHC

- Theorists model new particles that fit with the SM to explain these phenomena
- If they exist (and are accessible at LHC energy scales) we can create them and record their SM decay products
- Increasing collision energy means a wider variety of particles and increasing luminosity means more data!
- Can also make more precise checks of the SM





LHC Data

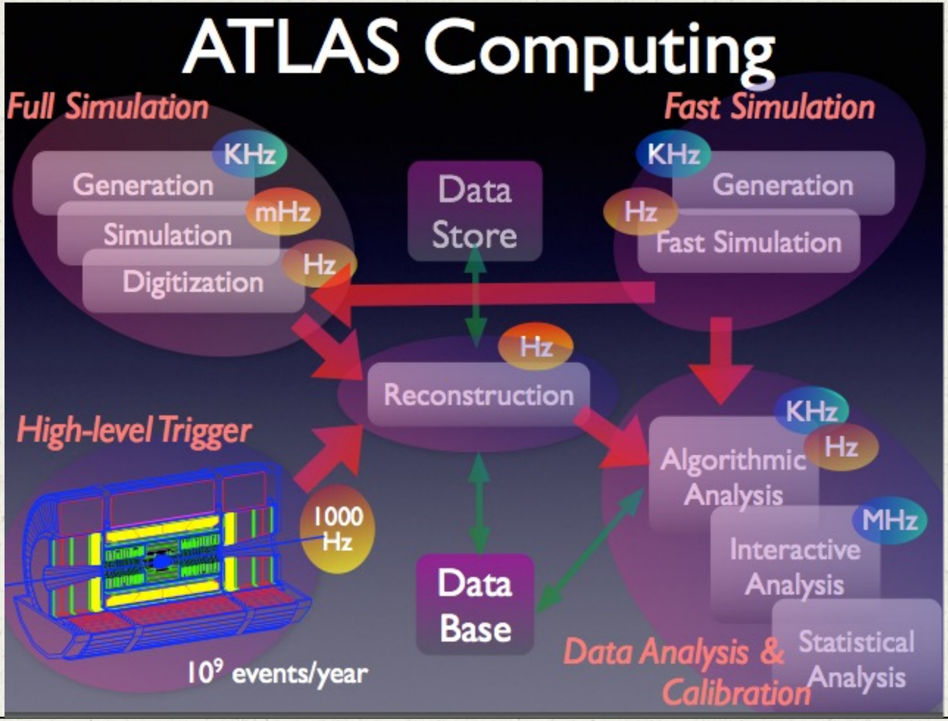
- Collision data measured by dedicated subsystems
  - Quantifies interactions with highly granular detectors
  - Readouts must be reconstructed into particle components (tracks, clusters) then full particle candidates and event information
- We can only measure SM particles, so we must (accurately) extrapolate what happened during the initial collision
- Poses many computing challenges
  - Non-fixed size, heterogenous data
  - Varying density/sparsity
  - Very tight computing time and resource constraints

AI and Particle Physics

02.

Symbiotic evolution

The LHC Involves Extensive Software



AI Can Be Utilized In Nearly Every Aspect

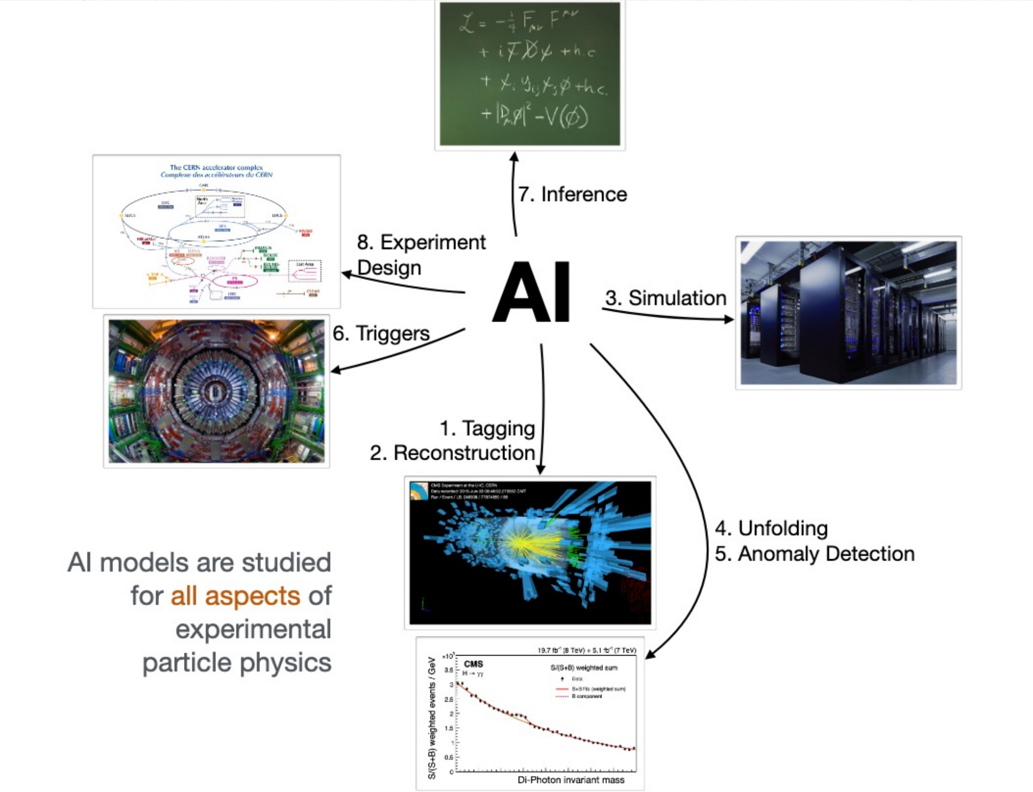
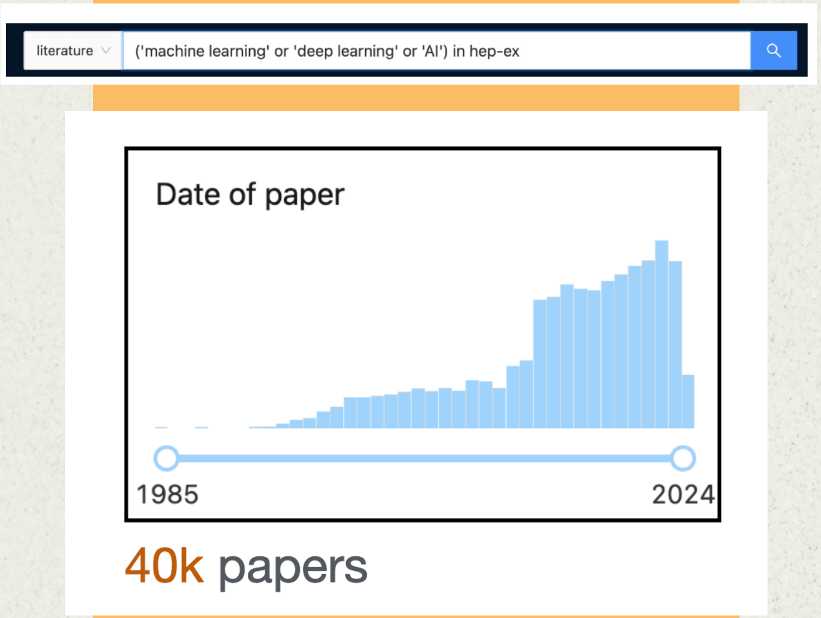


Diagram from Gregor Kasieczka



An Evolving Field

- AI has been used in particle physics, in some form, for nearly 40 years
- Particle physics has often been at the forefront of adopting and innovating novel AI methods
  - See Kyle Cranmer's keynote at NeurIPS 2016!
  - And our ongoing ML and the Physical Sciences Workshop (large particle physics component)
- The methods have substantially evolved over the years, from decision trees to neural networks to transformers
  - The data representations have similarly evolved (see next slides!)
- However, much remains the same
  - Overall focus has been on leveraging the full, low level data
  - Many unanswered questions still remain

Current R&D in AI and Particle Physics

03.

A wide variety of
approaches for shared
goals

Common Themes

DATA FORMATS

Preserve information and enable effective learning

GEOMETRIC DL

Leverage relationships and structure between data points

ANOMALIES

Identify potential beyond the standard model events

GENERATIVE AI

Improve or accelerate simulations for training and analysis

UNFOLDING

Accurately experimental observations to nature

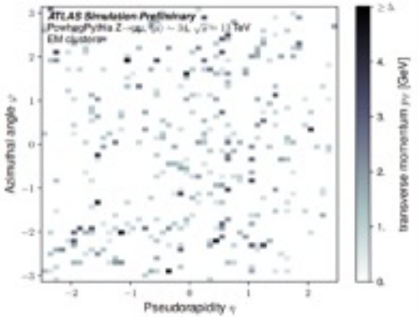
FOUNDATION MODELS

Perform multiple tasks with the same model

Data Representations

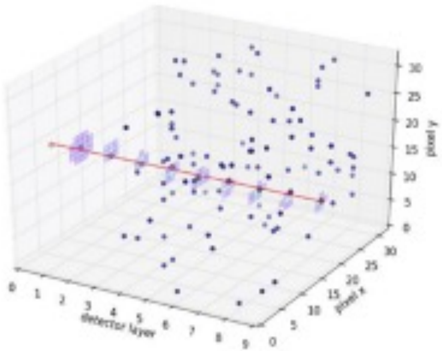
How Should We Represent Particle Collisions?

Image?



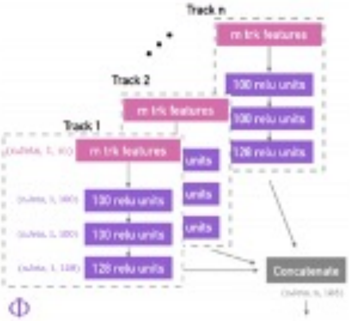
Convolutional Neural Networks with Event Images..., ATLAS Collab.

Sequence?



Particle Track Reconstruction with Deep Learning, Farrell et al

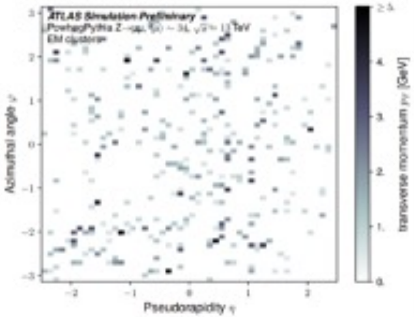
Set/Point Cloud?



Deep Sets based Neural Networks for Impact Parameter..., ATLAS Collab

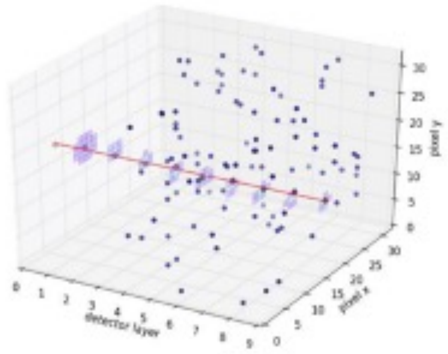
How Should We Represent Particle Collisions?

Image?



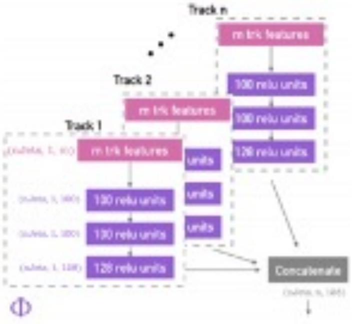
Convolutional Neural Networks with Event Images..., ATLAS Collab.

Sequence?



Particle Track Reconstruction with Deep Learning, Farrell et al

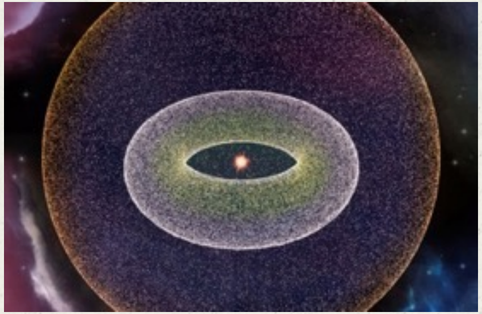
Set/Point Cloud?

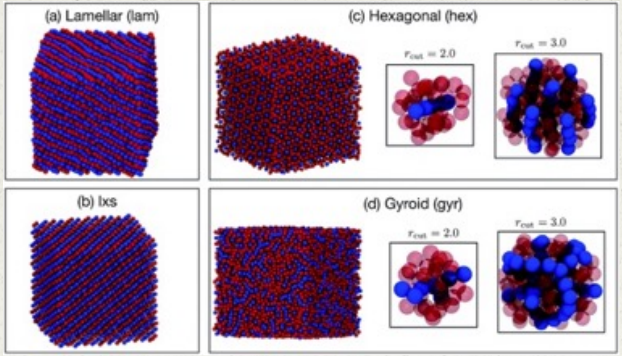


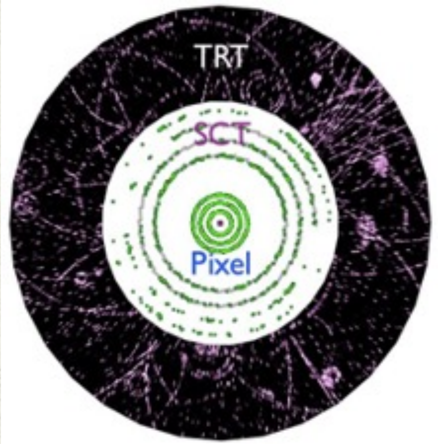
Deep Sets based Neural Networks for Impact Parameter..., ATLAS Collab

Language? Maybe! We'll return to this later….

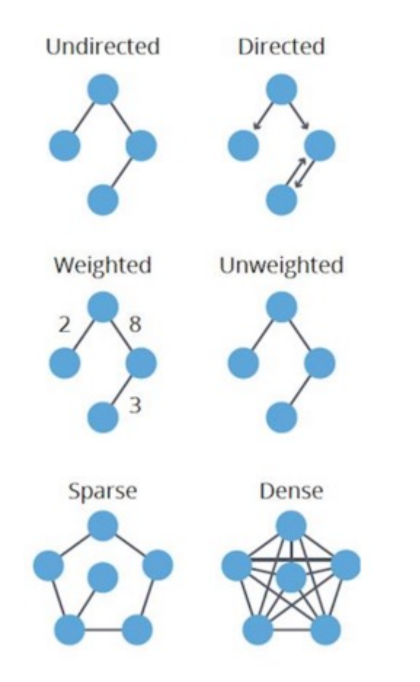
Point Clouds







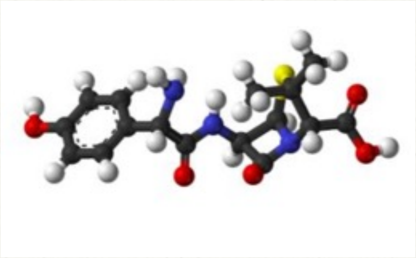
Can naturally represent many types of data

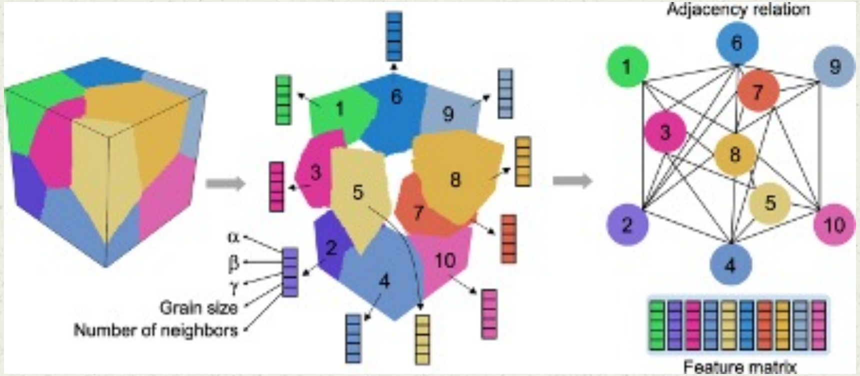


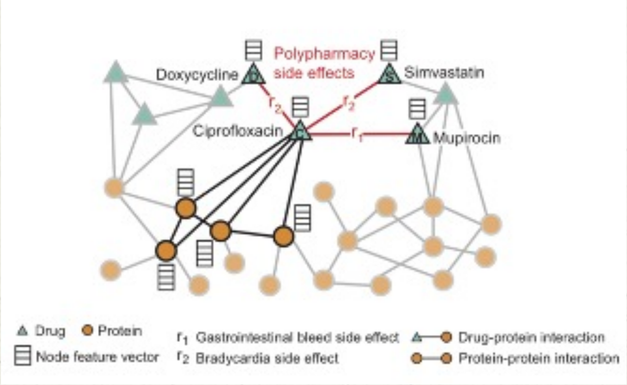
Graphs

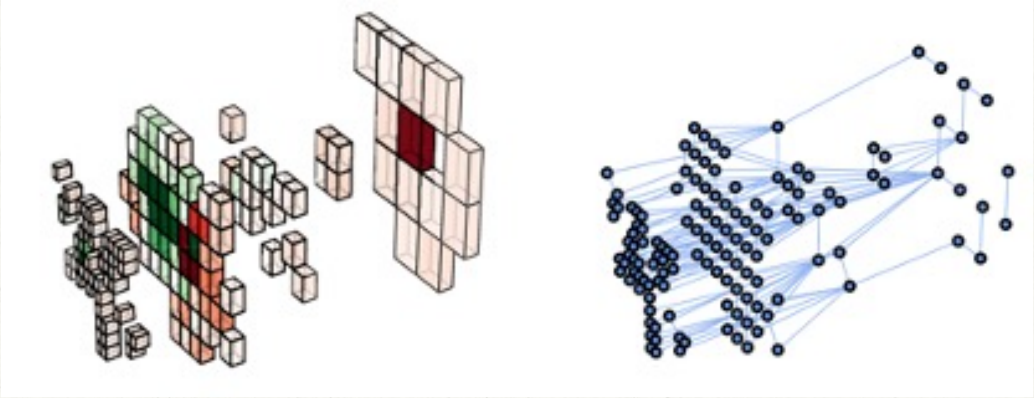
- If we add relational information to a point cloud we get a graph
  - Nodes: vertices $u \in V$ with associated information $x_u \in \mathbb{R}^{dv}$
  - Spatial coordinates, features, etc
  - Edges: connections between nodes $(u,v) \in E$
  - Can be directed or undirected, can have associated information $e_{u,v} \in \mathbb{R}_{dv}$
- Graphs can represent many types of relational/geometric data
- Inherent geometric inductive bias
  - By including edges we encode information about data structure and can localize computation

Graphs





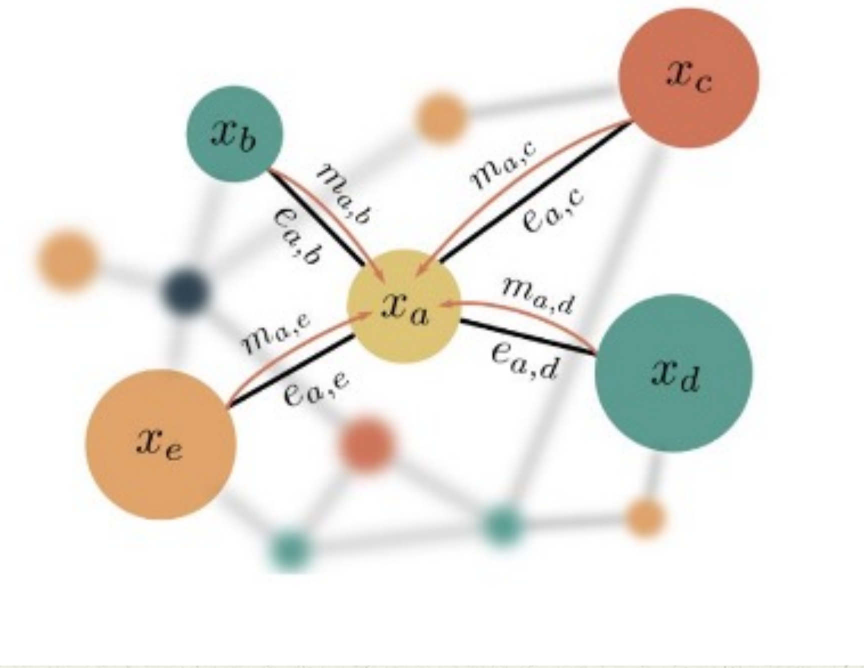




An intuitive representation for all kinds of geometric, structured, variable length physics data

Geometric Deep Learning

Graph Neural Networks



Message Passing (MPNN) Layers:
Framework for many equivariant graph updates

At each layer *k*, compute messages in each node's neighborhood:

$$\boldsymbol{m}_{uv}^{(k)} = \psi^{(k)}\left(\boldsymbol{h}_u^{(k-1)}, \boldsymbol{h}_v^{(k-1)}, \boldsymbol{e}_{uv}^{(k-1)}\right)$$

Aggregate messages in a permutation-invariant way:

Messages passed only from u's direct neighbors

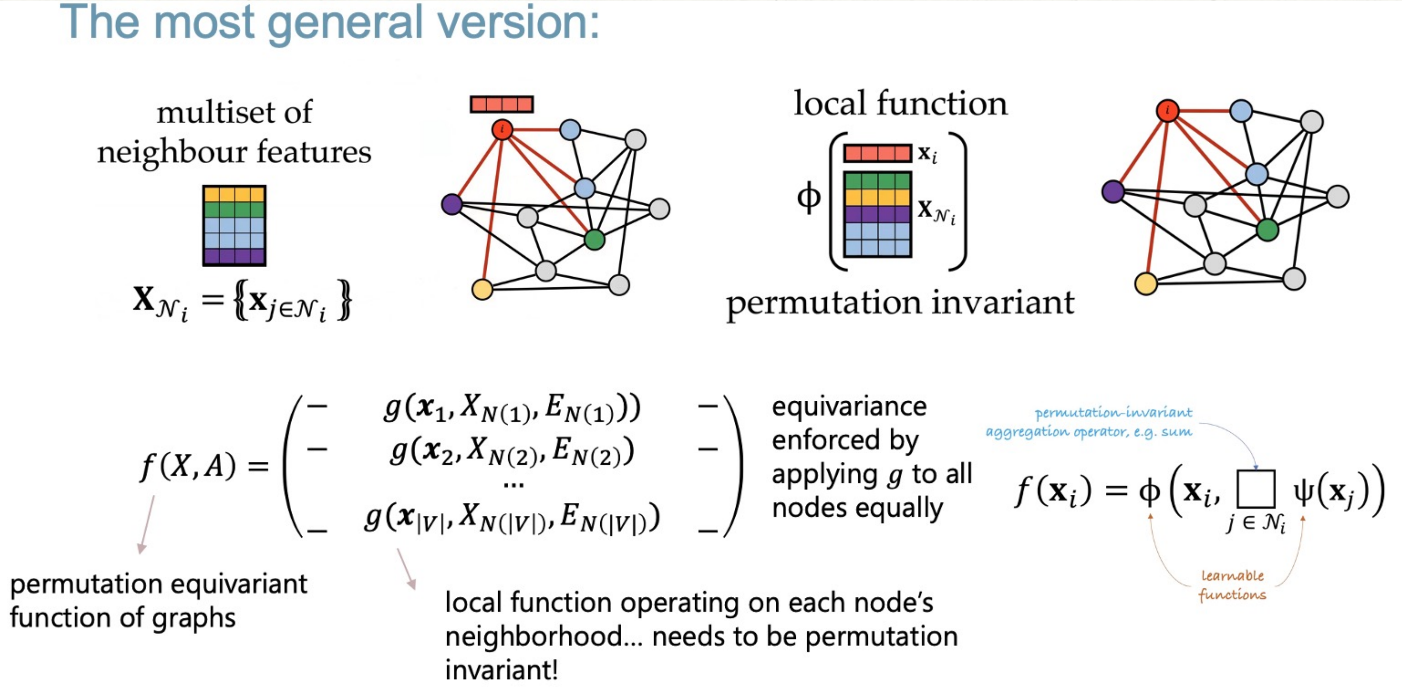
$$\boldsymbol{a}_u^{(k)} = \bigoplus_{v \in N(u)} \boldsymbol{m}_{uv}^{(k)}$$

Any permutation invariant operation (e.g. sum, mean, max)

Update the node's state based on the messages it received:

$$\boldsymbol{h}_u^{(k)} = \phi^{(k)}(\boldsymbol{h}_u^{(k-1)}, \boldsymbol{a}_u^{(k)})$$

Graph Neural Networks

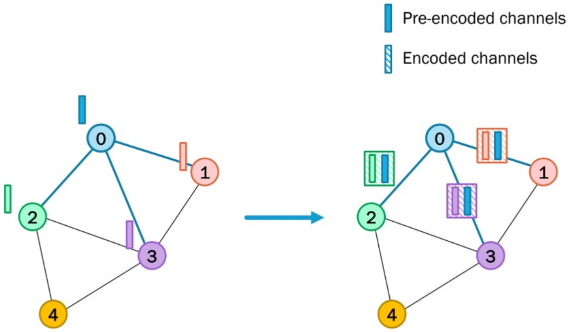


The goal of a (or at least some) GNN(s) is to learn a smart re-embedding of the graph data that preserves the relational structure but makes it easier to solve some downstream task

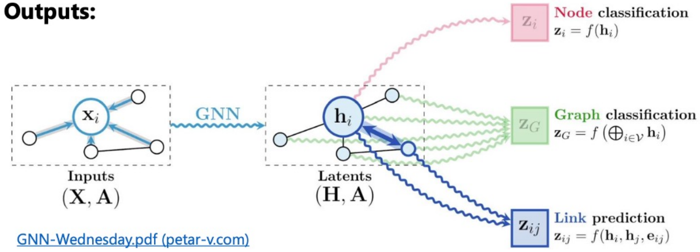
Graph Neural Networks

We can also update the graph edges

- Isotropic message passing can't differentiate importance of neighbors
- *An*isotropic message passing: encode a combination of node and neighbor along each edge



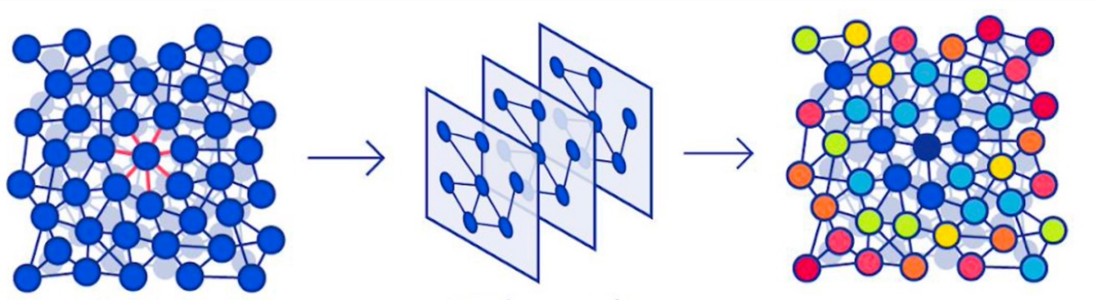
Outputs:



And we can use the updated graph in many ways

GNN-Wednesday.pdf (petar-v.com)

Graphs and GNNs for Physics

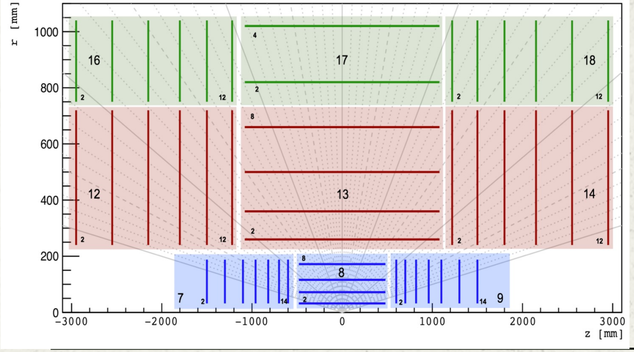


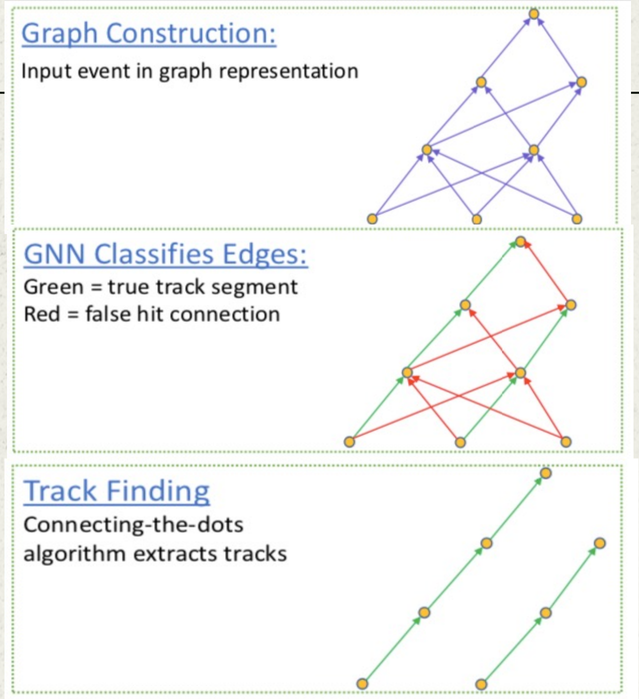
- Many physics datasets have inherent relational structure and/or no inherent ordering
    - We get permutation equivariance by construction
- Grids, sequences, etc. can't naturally represent irregular geometries
    - Graphs can handle sparsity, different data size, different measuring devices
- Many experimental data sets are heterogeneous
    - Data recorded from multiple subdetectors or even experiments
    - Different types of objects
- Graph representations help address the curse of dimensionality and include a geometric prior

GNNs for Tracking

Basic procedure

1. Form initial graph from spacepoints/hits (pre-processing)
2. Process with GNN to get probabilities of all edges
3. Apply post-processing algorithm to link edges together into tracks and get parameters

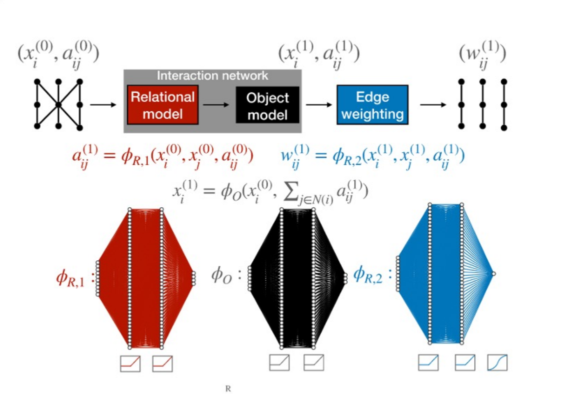


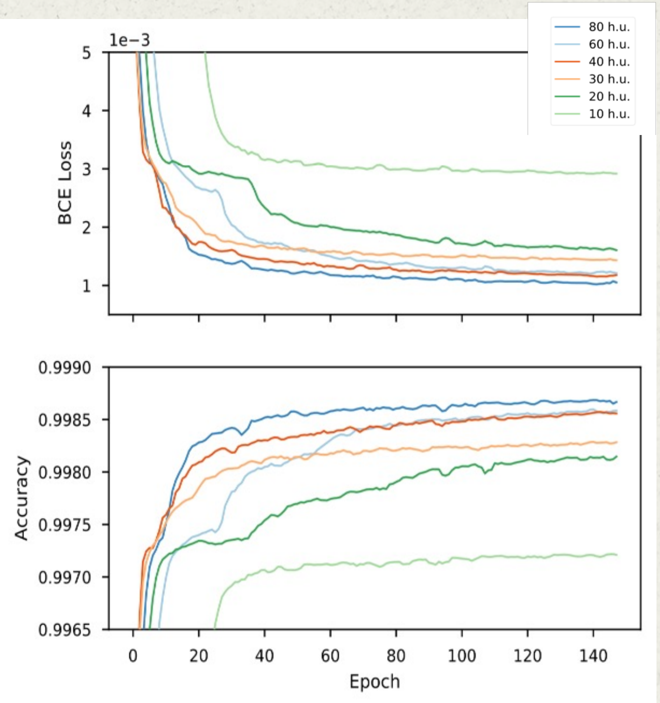


- Many places to improve/innovate
  - Graph construction, architectures, data augmentation…
- Work shown here uses TrackML dataset
  - Open, experiment agnostic
  - 200 PU, silicon semiconductor detector

Interaction Networks

- Originally developed for next time step predictions of physical systems
- Our implementation adds an additional relational model to predict edge weights
- Includes geometric edge features
- Total of ~6,000 learnable parameters
  - Smaller than many other architectures





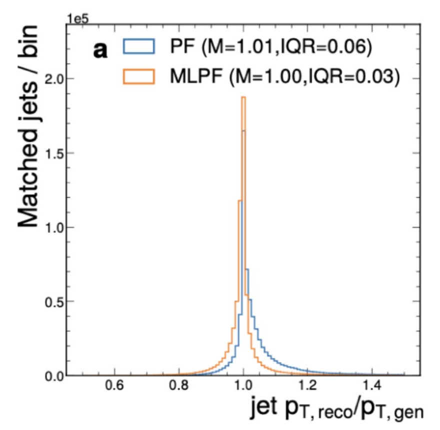
Trained with standard BCE loss

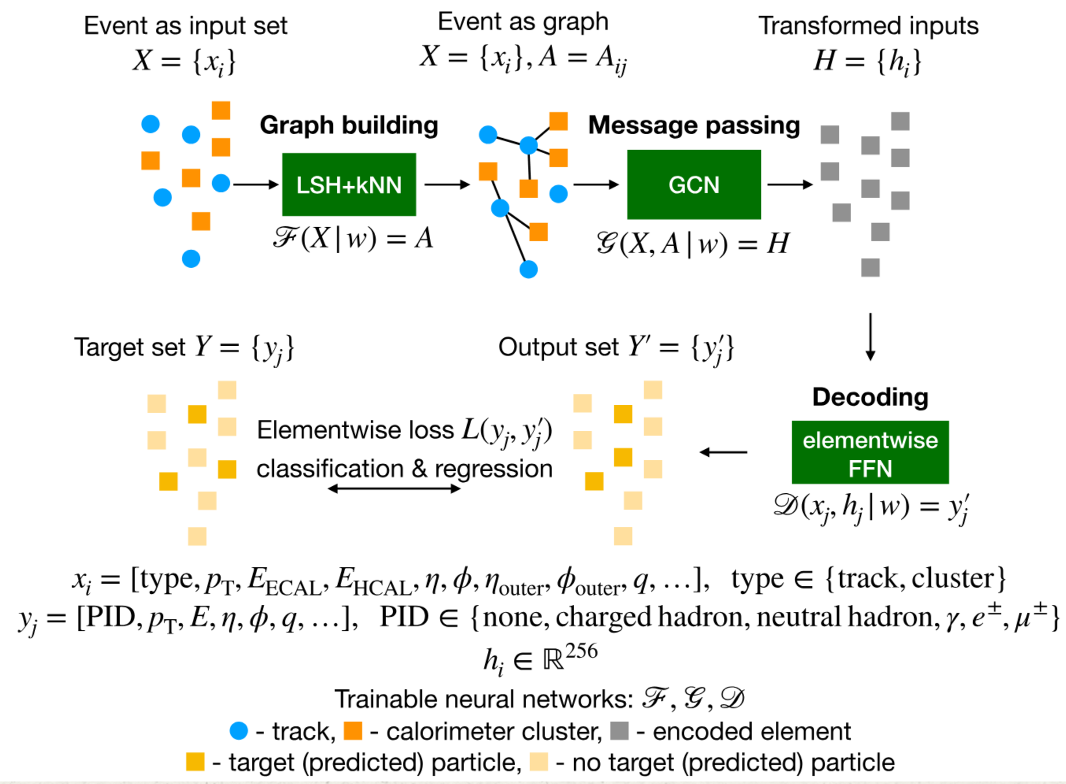
$$\mathcal{L}_w(y_j, w_j) = -\sum_{j=1}^{|\mathcal{E}|} \left(y_j \log w_j + (1 - y_j)\log(1 - w_j)\right)$$

Our Paper, Original Paper

Particle Flow

- GNN based framework that constructs particle candidates
- Improves on previous rule based methods





Pata et al, 2309.06782

Many Many More Examples!

Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges

Savannah Thais*[1], Paolo Calafiura[2], Grigorios Chachamis[3], Gage DeZoort[1], Javier Duarte[4], Sanmay Ganguly[5], Michael Kagan[6], Daniel Murnane[2], Mark S. Neubauer[7], and Kazuhiro Terao[6]

Technical Review | Published: 17 April 2023

Graph neural networks at the Large Hadron Collider

Gage DeZoort, Peter W. Battaglia, Catherine Biscarat & Jean-Roch Vlimant

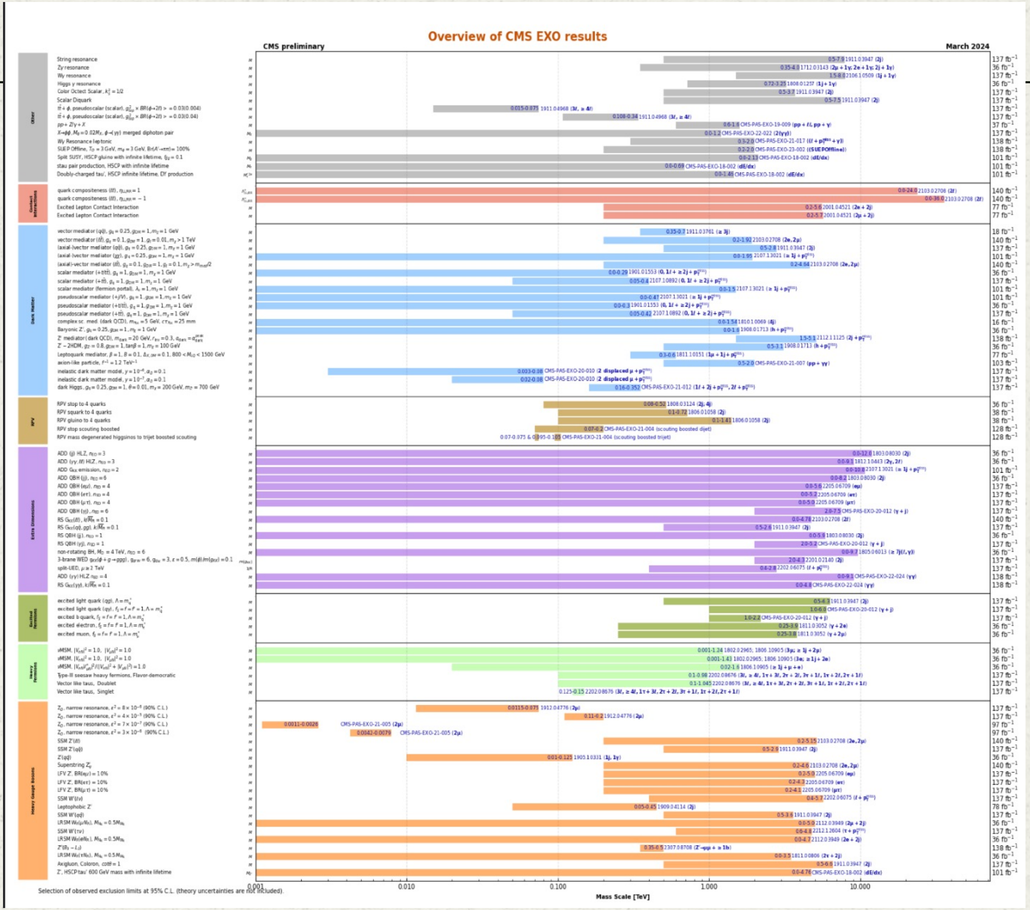
Nature Reviews Physics **5**, 281–303 (2023) | Cite this article

Thais et al 2203.12852 and DeZoort et al

Anomaly Detection

Anomalies

- We know there are unanswered particle physics questions, but so far no evidence of BSM physics
- Anomaly detection aims to identify ANY events that do not fit with the SM
  - As opposed to identifying a single BSM model and conducting a dedicated search



Three Main Approaches

Overdensity Estimation

Learn approximation of the likelihood ratio between background (SM) and signal (BSM). Typically uses classifier between signal enriched region data and background model

$$p_{\mathrm{data}}(x) = (1 - \epsilon)p_{\mathrm{bg}}(x) + \epsilon p_{\mathrm{sig}}(x),$$

Outlier Detection

Looks for out of distribution samples in any area of kinematic phase space. Often uses VAEs or GMMs.

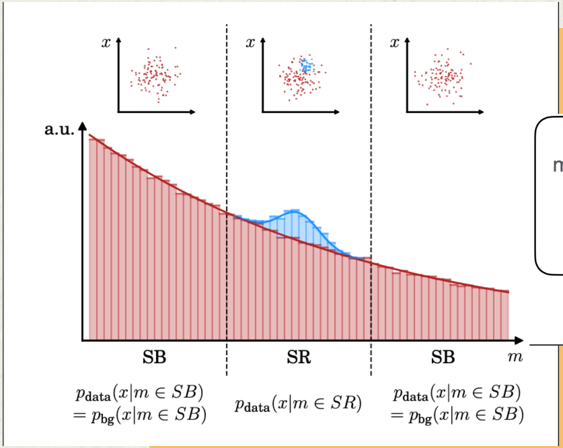
Parameterizing

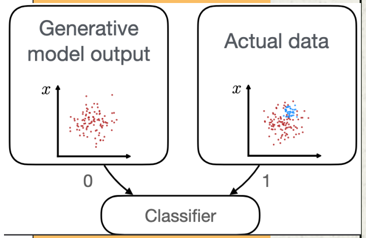
Trains a model on data and reference sample (with anomalies) and learns reference sample as small perturbations away from reference. Returns ratio between best fit of data and reference distribution.

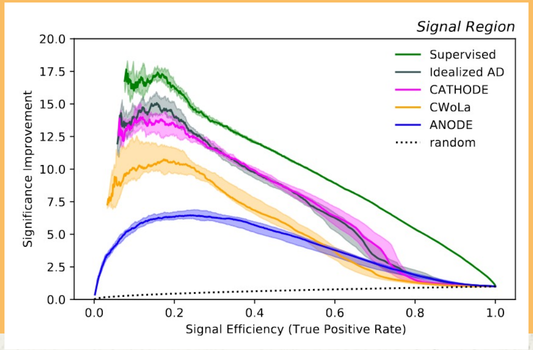
Golling et al 2307.11157

Finke et al

Gross et al 2305.14137





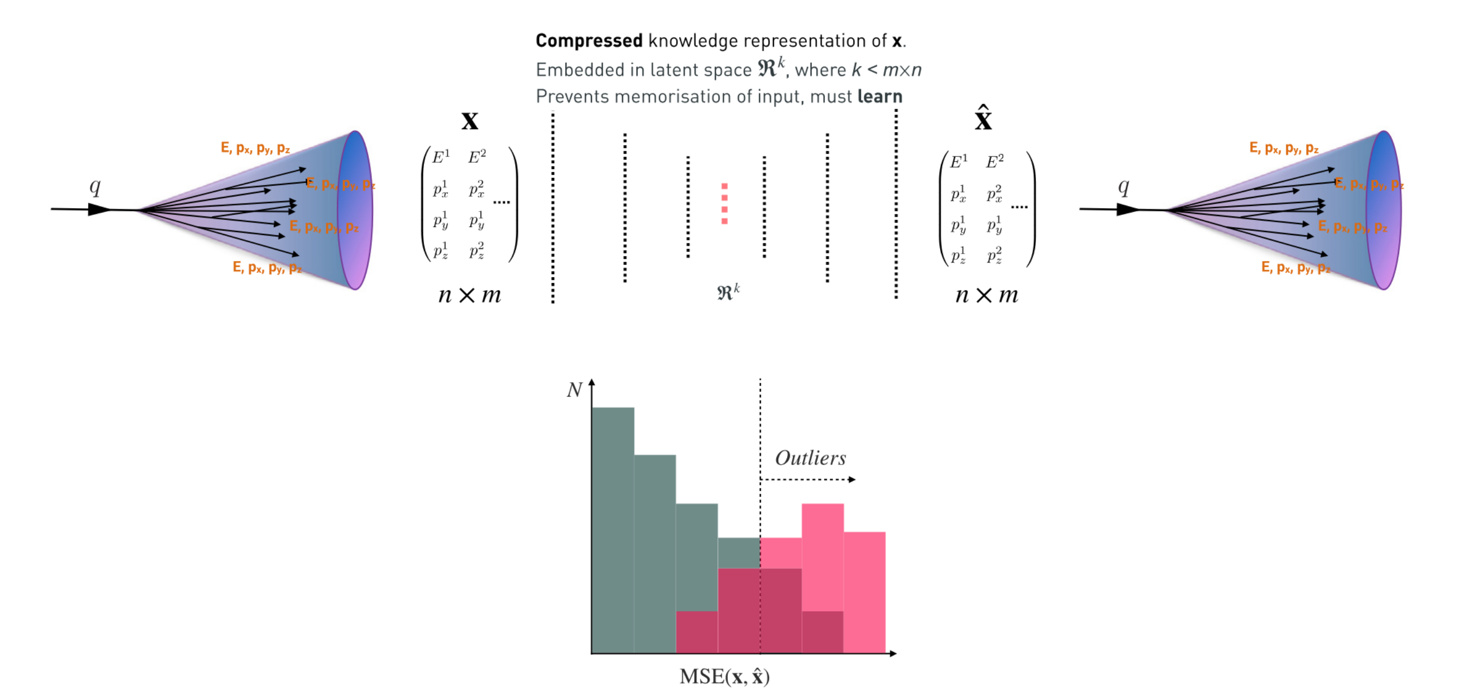


CATHODE

1. Don’t assume mass or type of resonant (BSM) particle but assume decay products
2. Train a generative model conditional on resonant feature (here m)
3. Interpolate and sample in SR
4. Train classifier on prediction vs data

Hallin et al 2109.00546

Variational Autoencoders

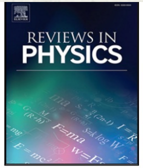


Many More Examples!

Contents lists available at ScienceDirect

Reviews in Physics

journal homepage: www.elsevier.com/locate/revip



Machine learning for anomaly detection in particle physics

Vasilis Belis, Patrick Odagiu, Thea Klaeboe Aarrestad *

Institute for Particle Physics and Astrophysics, ETH Zurich, 8093 Zurich, Switzerland

iHARP

NSF HDR ML Challenge

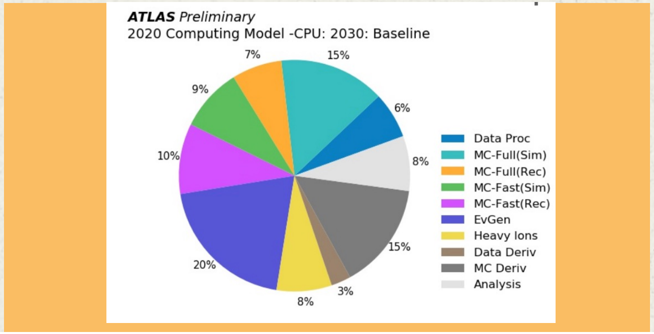
Scientific discovery often involves finding an **inconsistent pattern** within our data. Data that behaves differently from what is expected can indicate that the underlying science is different. Different behavior can result from a number of effects, but ultimately this could imply that we have observed something **new** 🌟!

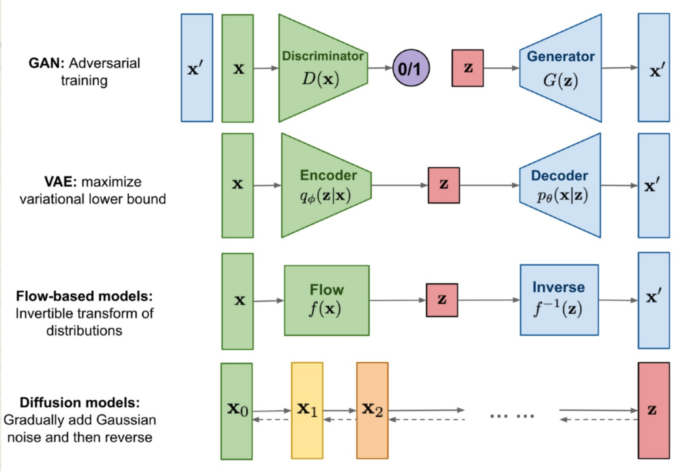
Depending on the scientific domain, a **new, unpredictable object/event** could have a profound impact. This could be a new type of material, the discovery of a **new astrophysical object** 🌌, the observation of **unusual climate behavior** 🌦️, or the discovery of a **new species** 🦋. The observation of something different, incongruous with the data, is what we call **anomaly detection** 🔍. Looking for anomalies is often quite different than other tasks since we do not know what exactly to look for, we just need to look for **something different**.

The challenge of scientific anomaly detection is one of the main focuses. **Using machine learning** to identify these anomalies 🤖

Belis et al and HDR Challenge

Generative AI

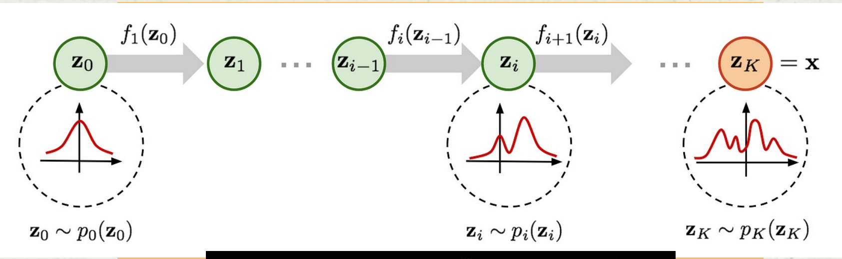




GenAI for Simulation

- Simulation is essential for training models and for connecting theory predictions with experimental data
- But simulation is very computationally expensive
- Aim to use GenAI trained on physics-driven simulation or data to augment traditional simulation
- All simulators attempt to (implicitly or explicitly) learn an approximation of p(x)

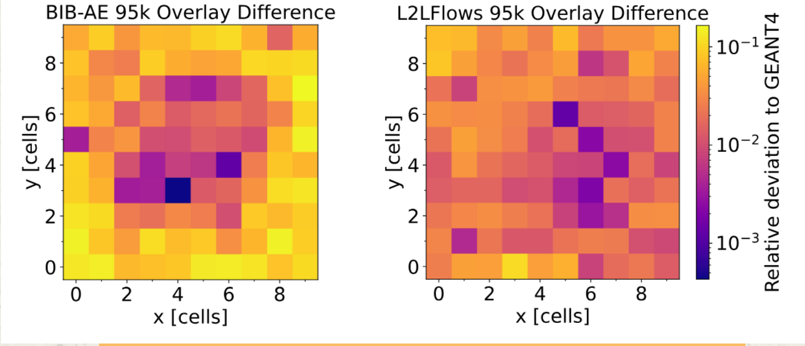
Hallin et al 2109.00546



$$\log p_\theta(x) = \log p_\theta(z) + \sum_{i=1}^{K} \log \left| \det\left(\frac{\partial f_i^{-1}}{\partial z_i} \right) \right|$$

exact log-likelihood evaluation

exact posterior inference (via $z = f^{-1}(x)$)



Normalizing Flows

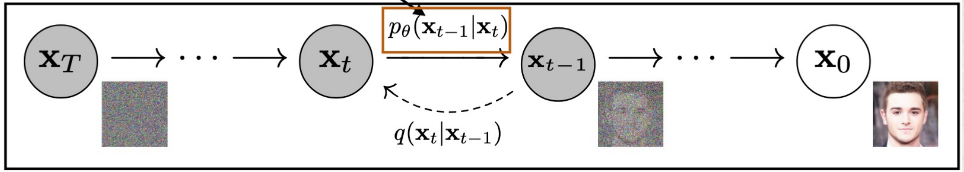
- Rather than learning to decode the encoder as in VAEs, Normalizing Flows attempt to exactly learn the likelihood
- Progressively add bijective and invertible functions to a simple distribution
- Use Jacobian of the transformations to evaluate probability density
- Should be higher fidelity than GANs or VAEs because it is learning exact likelihood
  - But requires some tricks (mainly data splitting) to train on high dimensional data

Diefenbacher et al 2302.11594

Diffusion Models

To improve the generative fidelity, move to a point cloud diffusion model

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



Forward
(Data → Noise)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Individual step

Noise schedule
(hyper-parameter)

$$\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}) = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boxed{\boldsymbol{\epsilon}} \text{ for } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Rewrite: State at any time

Will try to predict

$$\alpha_t := 1 - \beta_t \qquad \bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$$

Backward
(Noise → Data)

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boxed{\boldsymbol{\epsilon}_\theta}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$

Noisy image

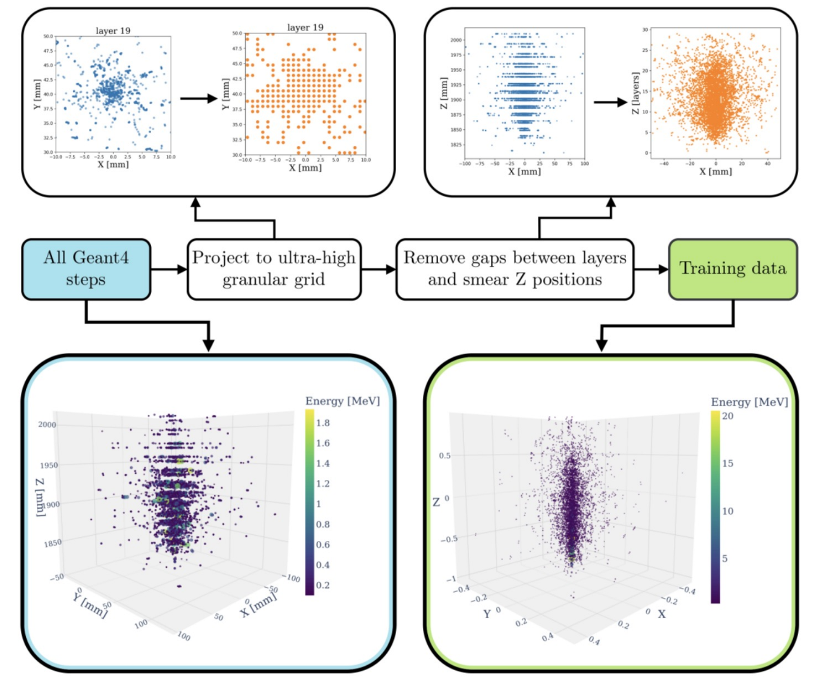
Reminder: Forward diffusion to time t

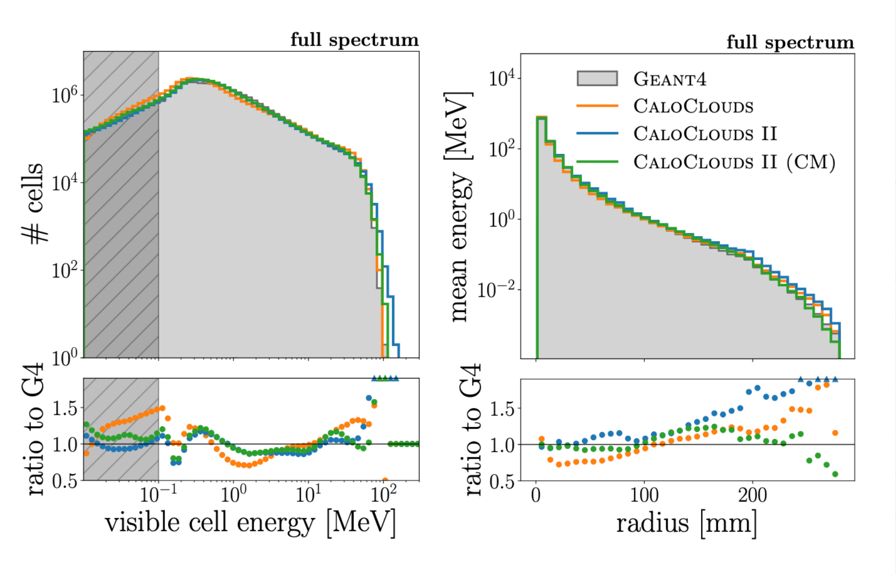
$$\mathbf{x}_t(\mathbf{x}_0, \boldsymbol{\epsilon}) = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$$

Timestep

Diagrams from Gregor Kasieczka

Calo Clouds



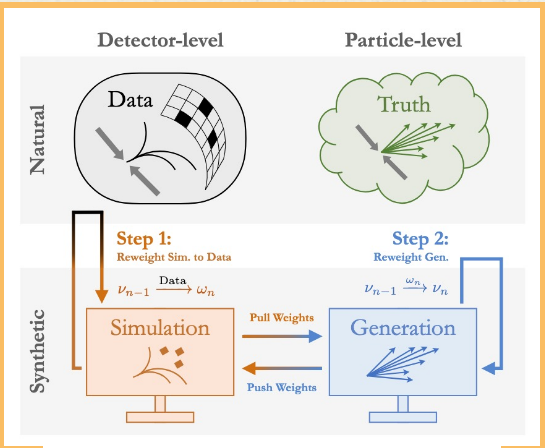


Buhmann et al 2309.05704

A Very Active Area of Research!

Recent Talk from Gregor Kasieczka

Unfolding



1. $\omega_n(m) = \nu_{n-1}^{\text{push}}(m)\, L[(1, \text{Data}), (\nu_{n-1}^{\text{push}}, \text{Sim.})](m),$

2. $\nu_n(t) = \nu_{n-1}(t)\, L[(\omega_n^{\text{pull}}, \text{Gen.}), (\nu_{n-1}, \text{Gen.})](t).$

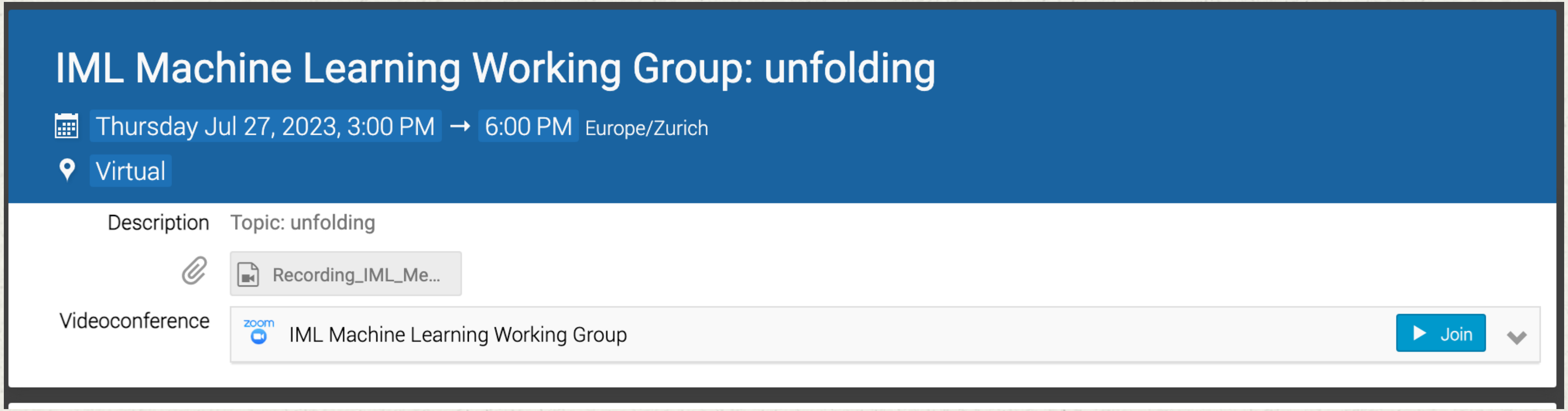
	Observable					
Method	m	M	w	$\ln \rho$	τ_{21}	z_g
OmniFold	**2.77**	**0.33**	0.10	**0.35**	0.53	0.68
MultiFold	3.80	0.89	**0.09**	0.37	**0.26**	**0.15**
UniFold	8.82	1.46	0.15	0.59	1.11	0.59
IBU	9.31	1.51	0.11	0.71	1.10	0.37
Data	24.6	130	15.7	14.2	11.1	3.76
Generation	3.62	15	22.4	19	20.8	3.84

OmniFold

- Even our best simulations differ from nature. These effects must be accounted for in order to trust our physics results.
  - Traditional approach uses a weight function developed separately for each variable
- Unfolding tries to learn generalized corrections
  - Either using trained networks or diffusion models for reweighting
- Omnifold pushes particle weights to detector weights, learns $p_{data}(m)/p_{sim}(m)$, pulls back to particle weights and calculates new weighting function
  - Push and pull ‘functions’ are trained NNs

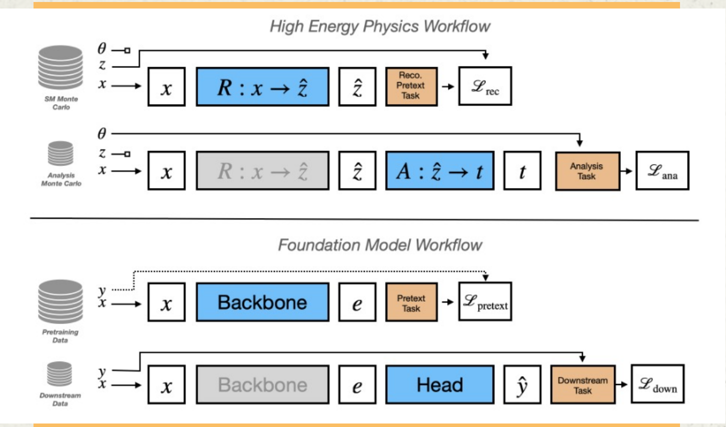
Andreassen et al 11911.09107

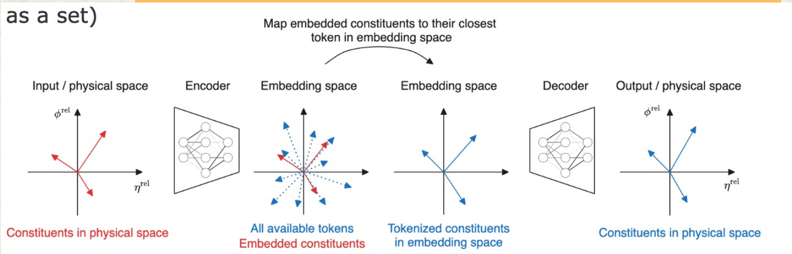
Active Area of Research



Recent IML Workshop

Foundation Models

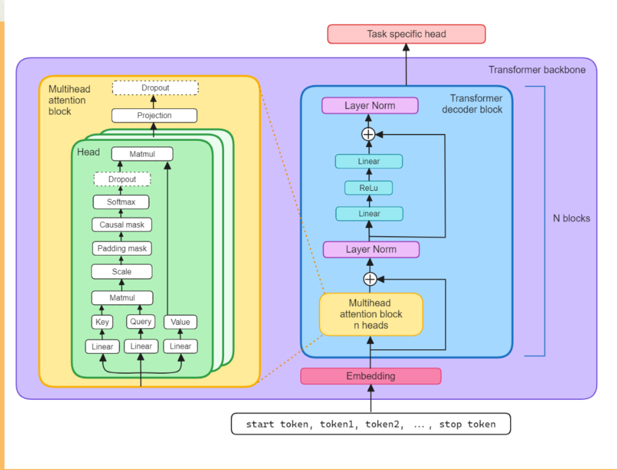


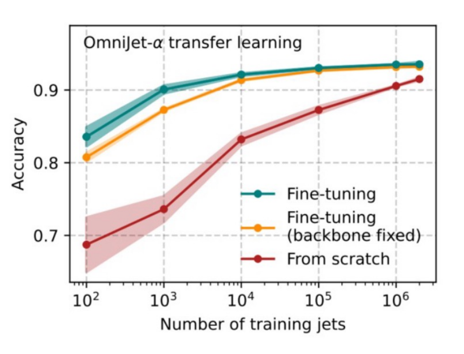


Foundation Models

- Goal is (like in LLMs) to pretrain on a large dataset for a certain task then finetune for a different dataset or task
- Enables sharing of models and data
  - Potentially even across experiments
- Could enable discovery of new physics
- Need to tokenize physics data
  - Binning
  - Vector quantization with VAE

OmniJet





- Uses generative pretraining (while learning to generate, model also learns physics)
  - Based on GPT1 Transformer model
- Transformer backbone takes tokens as input, sends output to task specific head
- Causal mask to prevent attention to future tokens
- Transferred to task of classifying q/g vs t->bqq' jets

Birk et al

Open Questions

04.

How do we continue to
improve science with AI

Common Themes

UNCERTAINTY

How do we characterize and propagate uncertainty?

INDUCTIVE BIAS

How do we incorporate physics knowledge into AI models? Does it help?

EXPLAINABILITY

Can we reliably describe what the model is learning?

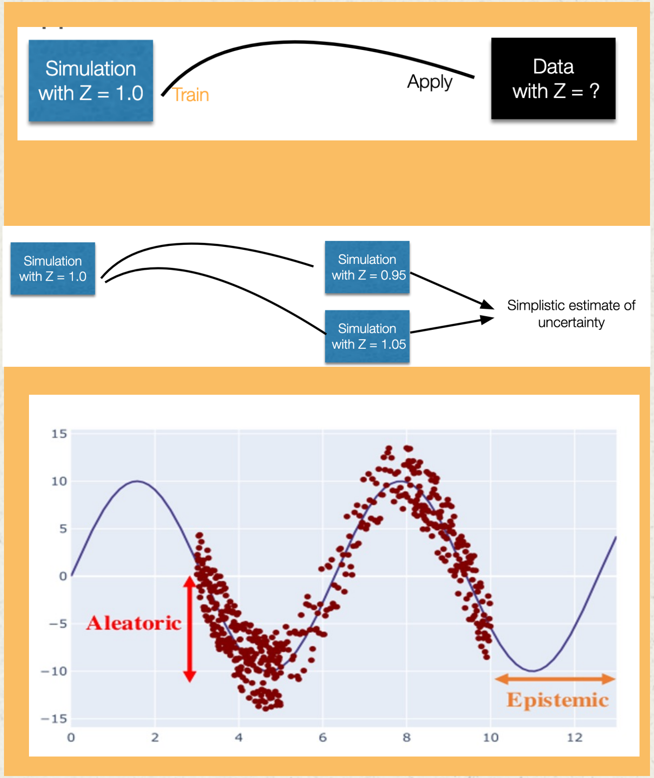
PHYSICS FOR AI

Can physics help us better understand AI?

NATURE OF SCIENCE

What does it mean to do physics with AI?

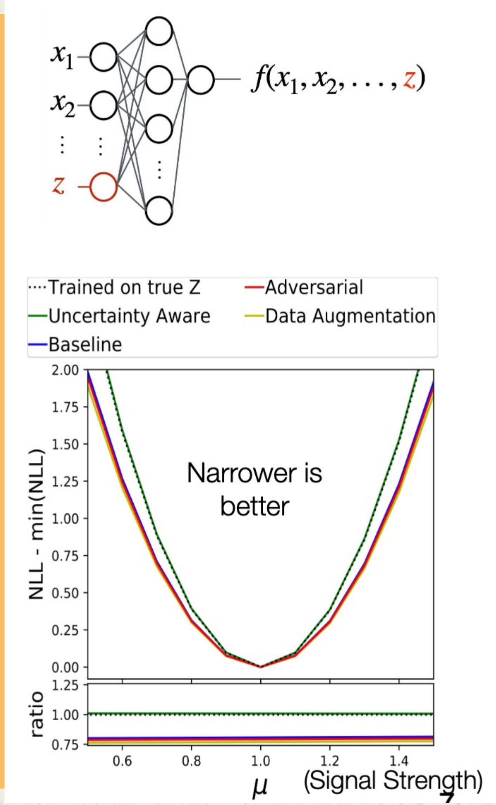
Uncertainty



Uncertainty

- Simulation has inherent uncertainty (systematics) that needs to be propagated through trained model
  - But uncertainties of actual detector data is unknown
- Current common approach is train model on normal simulation (Z=1) then estimate uncertainties with alternate simulations (shift Z) and look at impact on model outcomes
- In the language of ML, this is aleatoric uncertainty (from the data). There is also epistemic uncertainty due to model
  - Handling this is still an open question in ML. See this talk.

A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Abdar et al



New Approaches

- Adversarial decorrelation
  - Train a model to predict nuisance parameter using output of classifier
- Uncertainty aware learning
  - Parameterize the classifier based on Z
- Inference aware neural optimization
  - Include uncertainty on parameters of interest in loss function

Inductive Bias

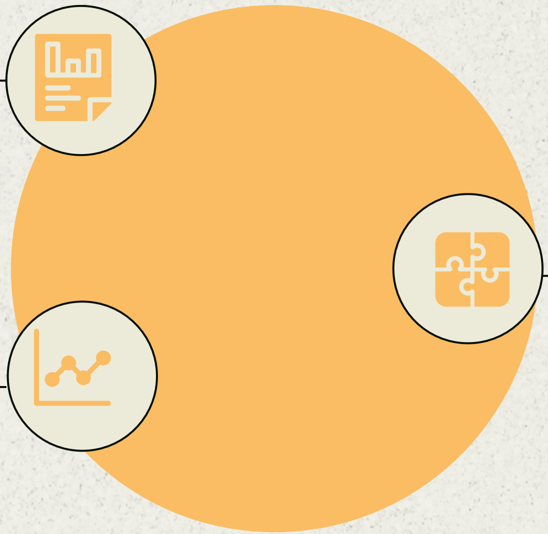
Physical Inductive Bias

DATA STRUCTURE

Relational structure, ordering, feature selection, pre-processing, etc

MODEL CONSTRAINTS

Restricting model weights, learned function, propagated information, etc



TASK FORMULATION

Physics informed neural networks, incorporating conservation laws or equations through loss function design, etc

Task Formulation

A simple inductive bias: Inertial dynamics

$x^{t+1} = \mathbf{NN}(x^t, v^t)$

Has to learn to predict static motion

Position: $x(t)$

Velocity: $v(t)$

$$\sum \mathbf{F} = m\mathbf{a} = m\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2}$$

Static prior

$x^{t+1} = x^t + \mathbf{NN}(x^t, v^t)$

Trivial to predict static motion

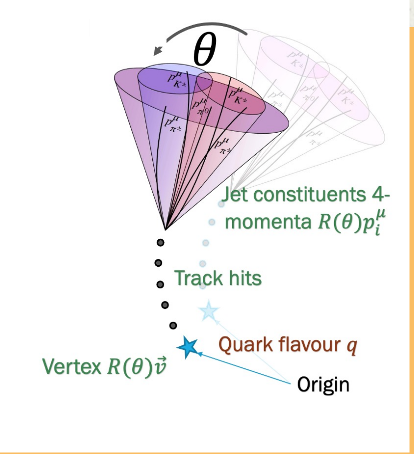
Has to learn to predict inertial motion

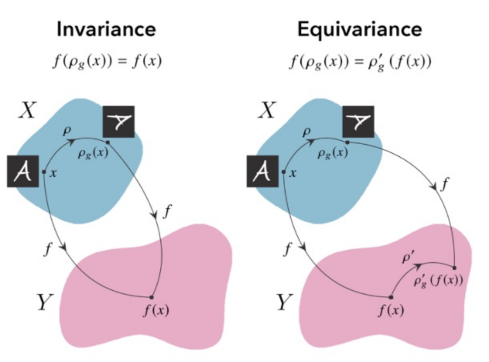
Inertial prior

$x^{t+1} = x^t + \Delta t \cdot v^t + \mathbf{NN}(x^t, v^t)$

Trivial to predict inertial motion!

Slide from Alvaro Sanchez



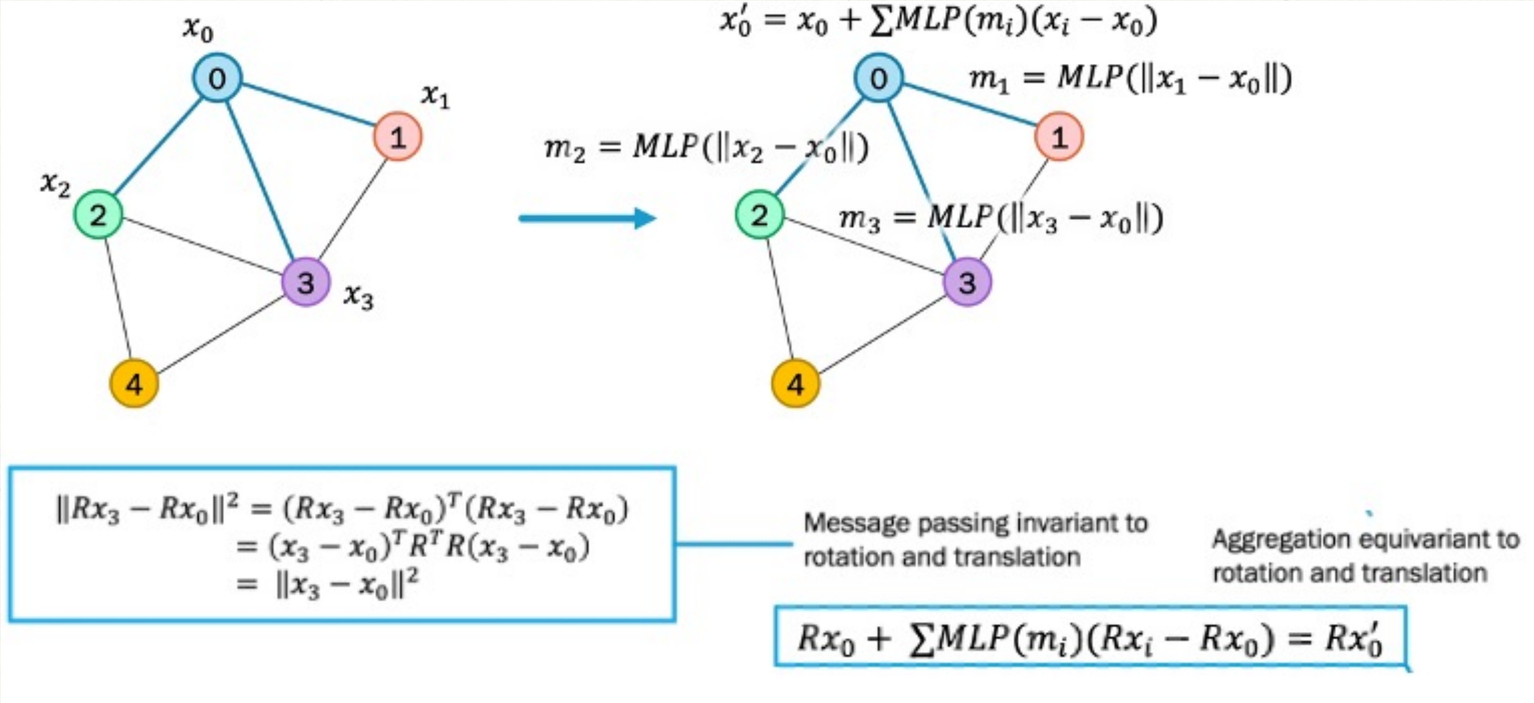


Including Symmetries

- Physics has many inherent symmetries, thus a popular inductive bias approach is enforcing symmetry conservation
- Consider rotating a jet by angle ϕ, using rotation matrix $R(\theta)$
  - Some predictions like the production vertex will rotate with the transformation: “equivariant”
  - Some predictions like the jet flavor should not be affected: “invariant”

Bogatskiy et al 2203.06153

Equivariance



Satorras et al 2102.09844

Potential Benefits of Equivariance

ACCURACY

- Most published models achieve SotA accuracy and attribute it to design choices
- In practice, equivariant models performance varies across formulations

GENERALIZABILITY

- Models should learn complete symmetry orbit from one example
- Demonstrated in practice, but other models can generalize well too

MODEL EFFICENCY

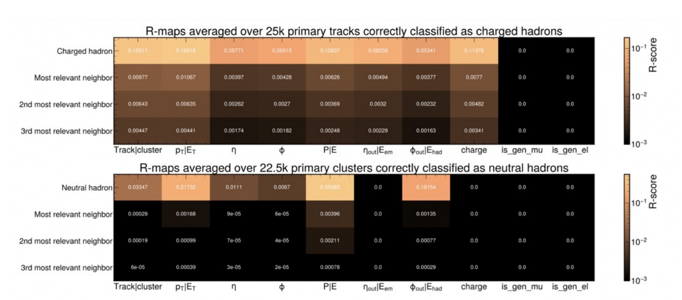
- Models may have an 'easier' time learning an optimal function
- Using ant factor, we find that equivariant models are not the most efficient

DATA EFFICENCY

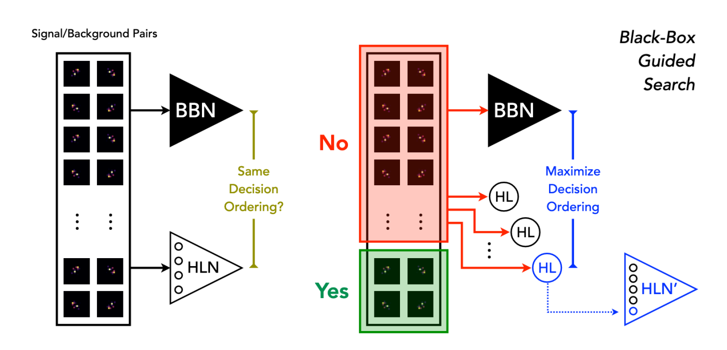
- Models don't need to rely on data augmentation to learn symmetries
- Most replicable benefit of equivariance

Thais and Murnane 2311.03094

Explainability



$$\mathbf{R}_j^{(l)} = \sum_k \frac{x_j A_{jk}}{\sum_m x_m A_{mk}} \mathbf{R}_k^{(l+1)}$$



Physics Studies

- We often want to understand what a model is learning
  - To ensure model is obeying known physics
  - To uncover new physics
- Layerwise relevance propagation helps characterize what information the network is leveraging
- Learn surrogate models trained on interpretable features
- Apply symbolic regression to identify the analytic function approximated by the AI model

But There Are Many Limitations…

- No clear way to map relevances to mathematical information
    - Can't understand what a model is learning outside of known features
- No way to know if explanation is correct or due to statistical artifacts
    - No way to know if model is ultimately correct either…
- We don't always have a nice space of features to use for surrogate models
- Symbolic regression doesn't provide guarantees on accuracy of equation
- Explainability methods do not account for uncertainty
- Overall, a very exciting and open area of research in AI as a whole

Physics for AI

AI Has a Reliability Problem

AI and the Everything in the Whole Wide World Benchmark

Inioluwa Deborah Raji
Mozilla Foundation, UC Berkeley
rajiinio@berkeley.edu

Emily M. Bender
Department of Linguistics
University of Washington

Amandalynne Paullada
Department of Linguistics
University of Washington

Emily Denton
Google Research

Alex Hanna
Google Research

Focus on **constructed tasks** and **benchmark data sets** that may be **distant from real world** distributions or goals

The Fallacy of AI Functionality

INIOLUWA DEBORAH RAJI*, University of California, Berkeley, USA
I. ELIZABETH KUMAR*, Brown University, USA
AARON HOROWITZ, American Civil Liberties Union, USA
ANDREW D. SELBST, University of California, Los Angeles, USA

Application to **impossible tasks**, **robustness issues**, **misrepresented** capabilities, **engineering mistakes** or failures

Leakage and the Reproducibility Crisis in ML-based Science

Sayash Kapoor [1] **Arvind Narayanan** [1]

Data **leakage**, incorrect or neglected **testing**, poor **experimental design** practices

Enchanted Determinism:
Power without Responsibility in Artificial Intelligence

ALEXANDER CAMPOLO[1]
UNIVERSITY OF CHICAGO

KATE CRAWFORD[2]
NEW YORK UNIVERSITY, MICROSOFT RESEARCH

Acceptance of **inherent unknowability** of AI systems, willingness to use **imprecise** or **unscientific language**

Physics As a Sandbox

Learning to Pivot with Adversarial Networks

Gilles Louppe New York University g.louppe@nyu.edu
Michael Kagan SLAC National Accelerator Laboratory makagan@slac.stanford.edu
Kyle Cranmer New York University kyle.cranmer@nyu.edu

We know many of the **dependencies** in our data and how our experiments/pre-processing **shape the data** → evaluate **de-biasing methods**

Energy flow polynomials: A complete linear basis for jet substructure

Patrick T. Komiske, Eric M. Metodiev, Jesse Thaler
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
E-mail: pkomiske@mit.edu, metodiev@mit.edu, jthaler@mit.edu

We know some patterns a model should learn and can build **interpretable bases** for some problems → contribute to **mechanistic interpretability**

ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset

The ATLAS Collaboration

We know the **phase space** of our data and **axes** along which it varies → can study **generalizability** of models

Constraint-based Graph Network Simulator

Yulia Rubanova [*1] **Alvaro Sanchez-Gonzalez** [*1] **Tobias Pfaff** [1] **Peter Battaglia** [1]

We can **compare model learned knowledge** to **true generating functions** → evaluate **robustness of new architectures**

Physics and Trustworthy AI

Physics and the empirical gap of trustworthy AI

Savannah Thais

Nature Reviews Physics (2024) | Cite this article

Understanding what cutting-edge AI models are doing 'under the hood' requires not just theoretical research but also well-controlled computational experiments. Savannah Thais explains why physics datasets may be the testing ground that AI developers need and how physicists can play a critical role in developing trustworthy AI.

Paper

The Nature of Science

What Do We Want From Science?

- Using AI for science may be different than other application/co-design areas. What are our priorities?
  - Efficient models? Interpretable models? Accurate models? (what does accuracy mean?) Physics inspired models? Interpretable models?
- Our ultimate goal in science is to extract reliable and robust knowledge about the universe, is our approach to AI helping us get there?
  - Is ML Good or Bad for the Natural Sciences?
  - Artificial Intelligence and Illusions of Understanding in Scientific Research

How Does AI Help Us Get There?

- Is it worth chasing every innovation in AI? Are we focused more on innovation that reliable science?

- What paradigms are most useful to explore? We have a strong mathematical foundation that still doesn’t explain everything. Should we focus more on anomaly detection or latent space explanation?

- Simulation and data are expensive, should we follow the scale approaches in broader AI?

- Are industry language models useful for doing science? Are they reliable? How do we study this?

- How do we continue to build meaningful community around this intersection?

THANKS!

st3565@columbia.edu

@basicsciencegirl