Machine learning for charged particle tracking

Jan Stark
Laboratoire des 2 Infinis - Toulouse (L2IT)

Learning to discover, Institut Pascal, Université Paris-Saclay
Workshop on representation learning, April 19th 2022

Acknowledgements

During the preparation of these slides, I have made extensive use of the following talks:

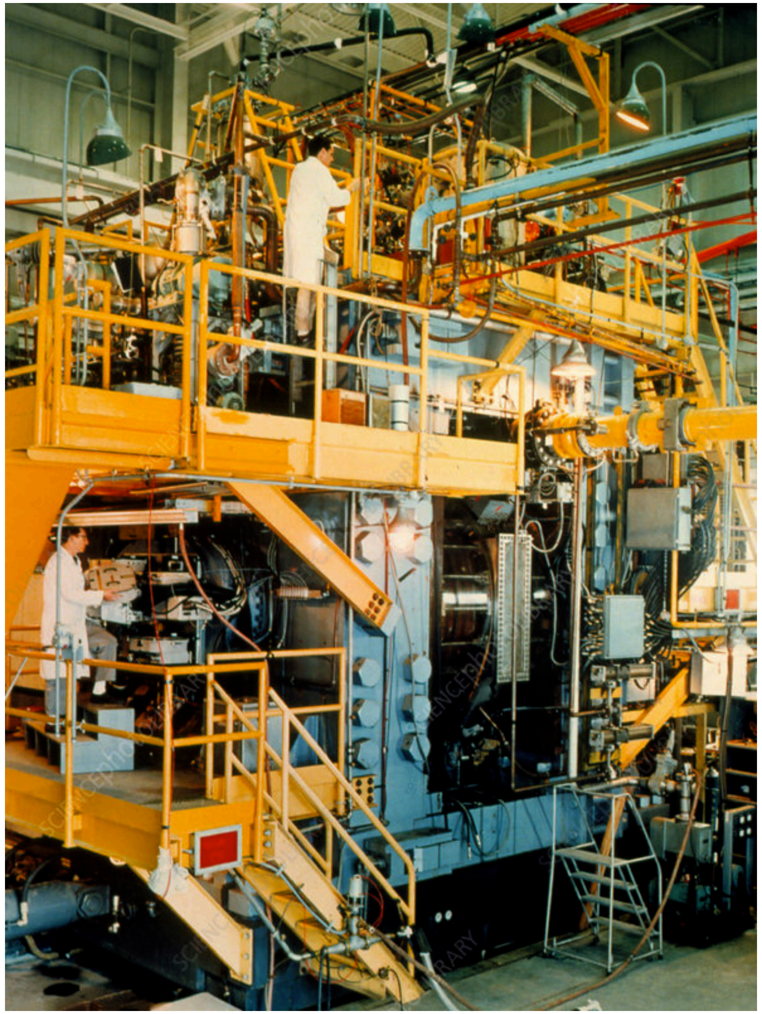
- Charline Rougier, “Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC”, vCHEP 2021, May 17-21 2021
  (link)

- Catherine Biscarat et Sylvain Caillou, « Comment la future phase de haute luminosité du collisionneur LHC du CERN, qui doit permettre l'étude des interactions du boson de Higgs, bouscule notre façon de calculer », Séminaire SFP de la section Midi-Pyrénées, 29 octobre 2021
  (link)

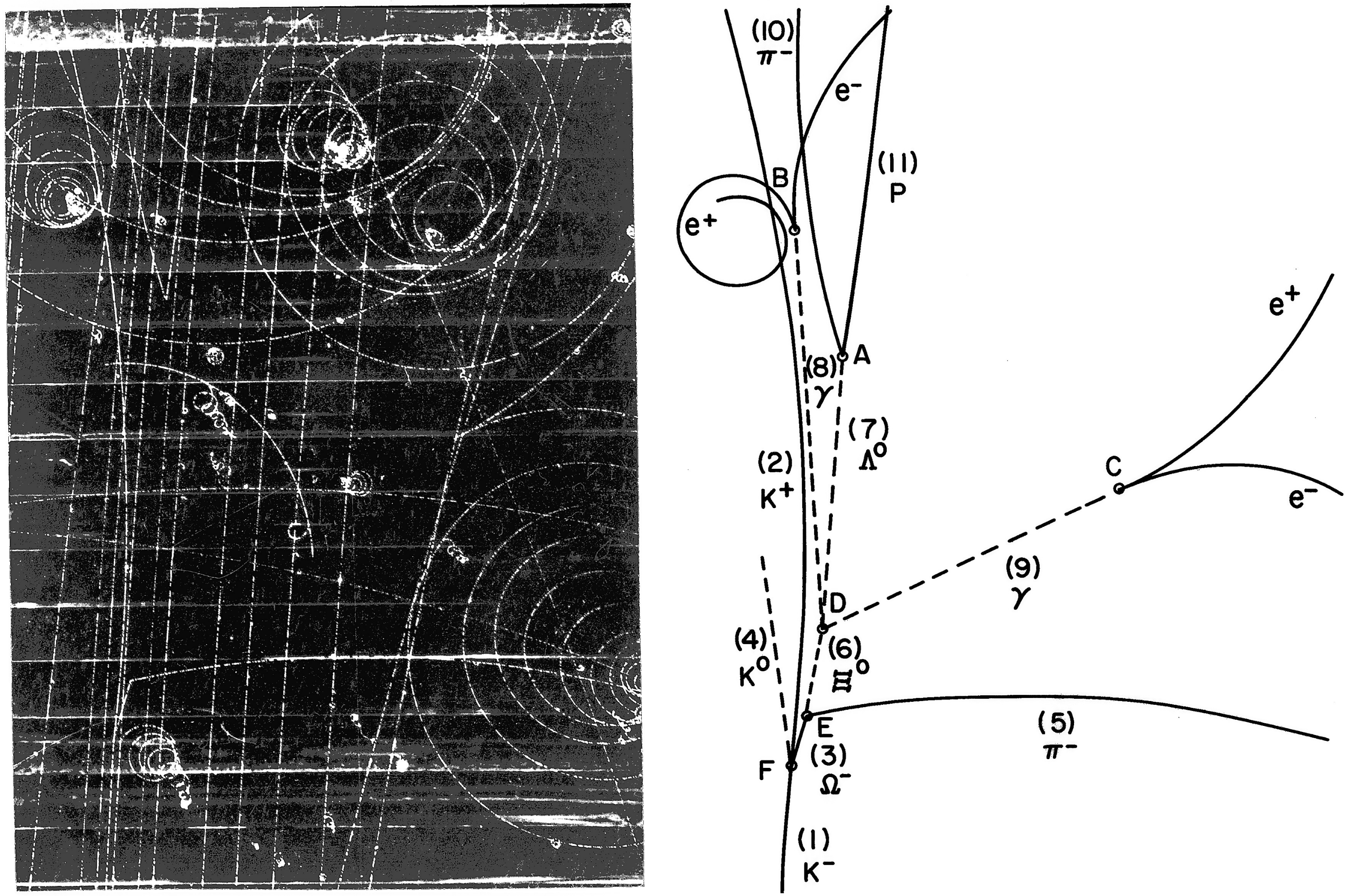
- Markus Elsing, “Pattern recognition in HEP (track reconstruction)”, Learning to Discover: advanced pattern recognition, Institut Pascal, October 14-25 2019
  (link)

- Material from Daniel Murnane (Berkeley LBL) and Charline Rougier (L2IT)

Charged particle tracking – 1960s style



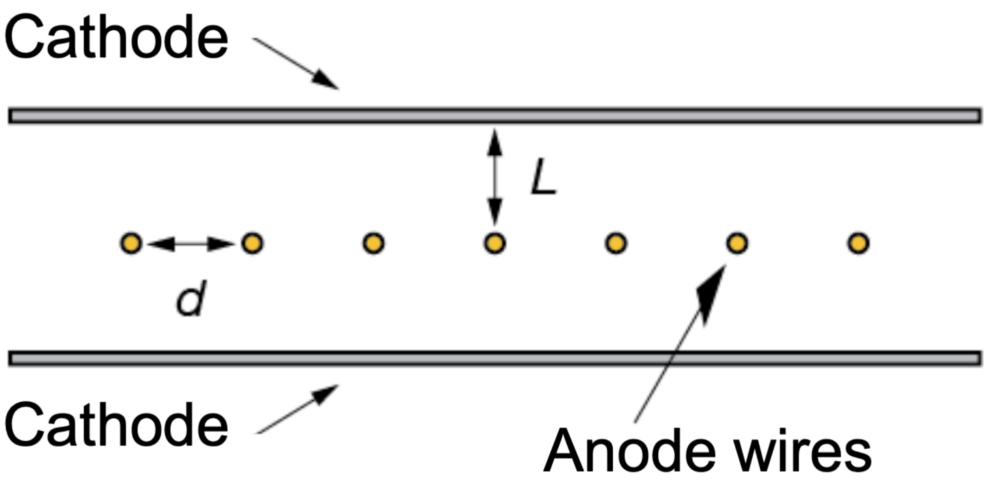
The 80 inch (2.0 m) bubble chamber at BNL



Discovery of the Omega-minus baryon in 1964

Electronic readout !

Multi-wire proportional chamber



Nobel prize
in physics (1992)

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

File : Charpak chambers

#1

EF-1

THE USE OF MULTIWIRE PROPORTIONAL COUNTERS
TO SELECT AND LOCALIZE CHARGED PARTICLES

G. Charpak, R. Bouclier, T. Bressani, J. Favier
and Č. Zupančič

CERN, Geneva, Switzerland.

ABSTRACT

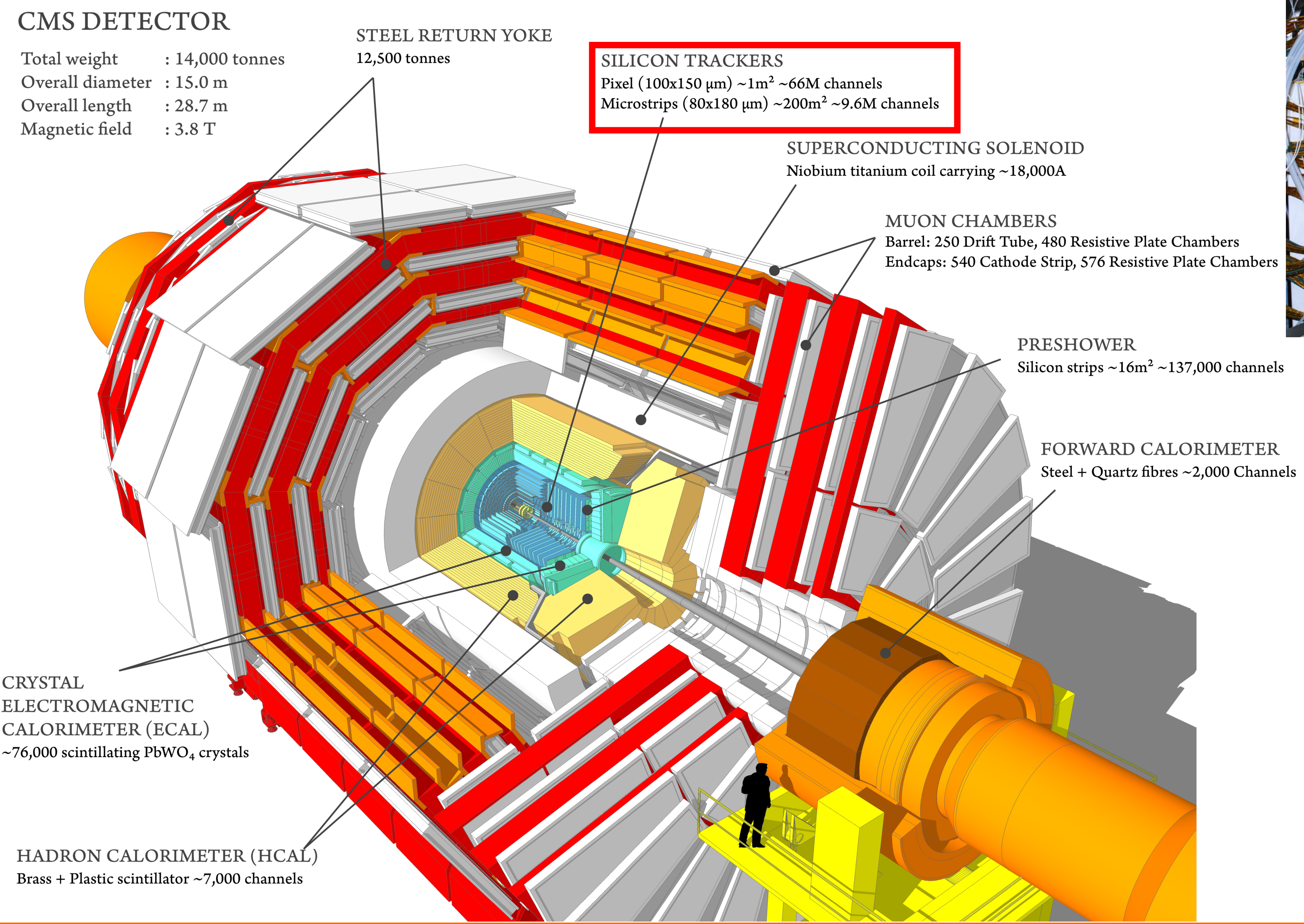
Properties of chambers made of planes of independent wires placed between two plane electrodes have been investigated. A direct voltage is applied to the wires. It has been checked that each wire works as an independent proportional counter down to separation of 0.1 cm between wires.

- Counting rates of 10^5/wire are easily reached.
- Time resolutions of the order of 100 nsec have been obtained in some gases.
- It is possible to measure the position of the tracks between the wires using the time delay of the pulses.
- Energy resolution comparable to the one obtained with the best cylindrical chambers is observed.
- The chambers can be operated in strong magnetic fields.

Geneva - 23 February, 1968
(Submitted to Nucl. Instrum. and Methods)

Messrs. G. Amato and J.P. Papis were of great help in the research into very low-cost amplifiers and were successful in this respect. They showed that less than two dollars of equipment per wire was sufficient to bring the pulses to a level close to 1 volt, where their utilization by logic circuits is easy.

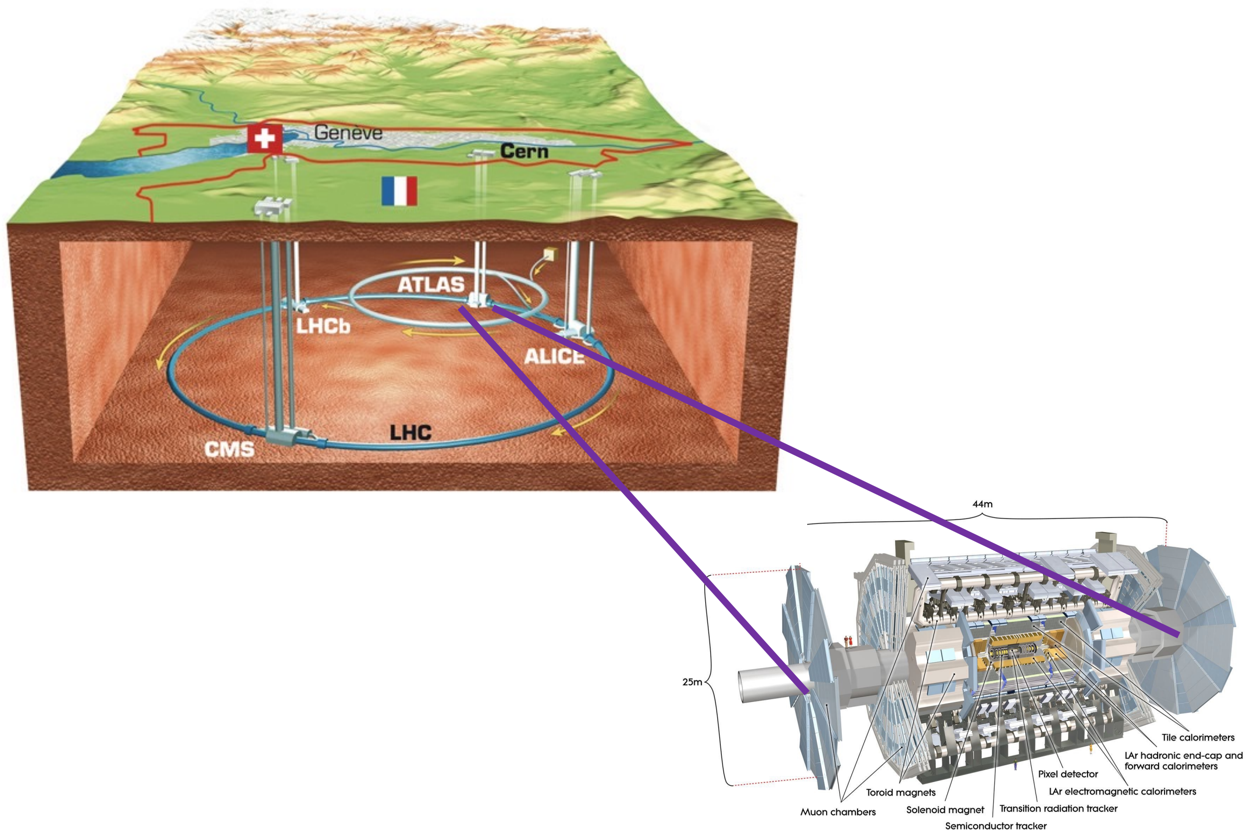
Today: silicon trackers



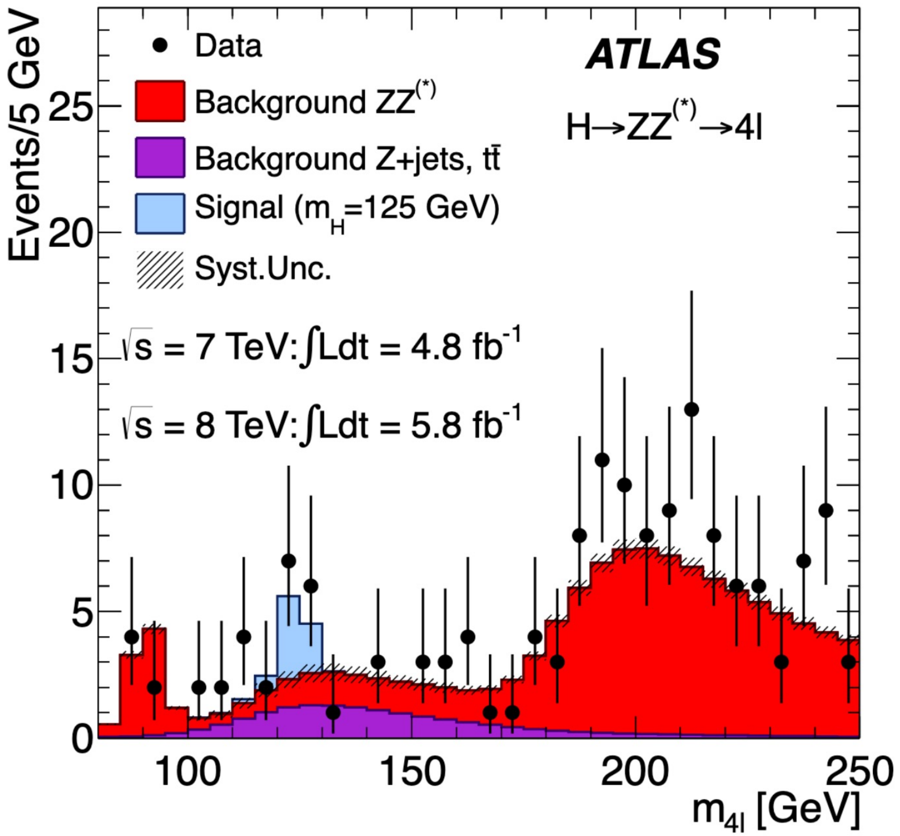
Integration test in 2007

LHC, discovery of the Higgs boson

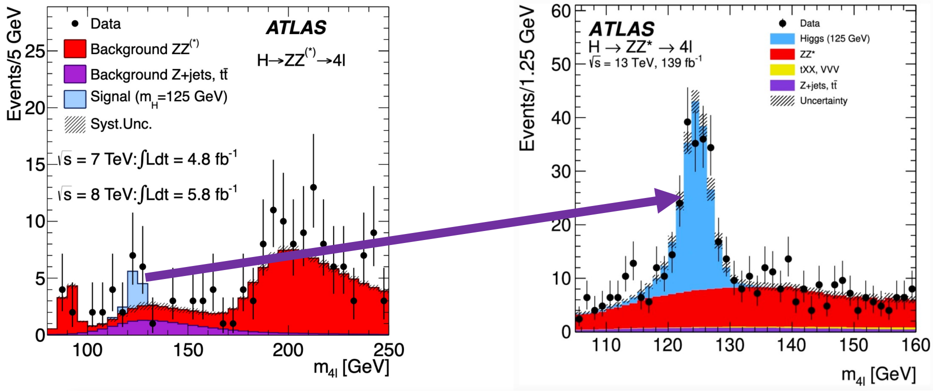
Large Hadron Collider (LHC) at CERN
- Explore small structures -> high energy
- Study rare processes -> high frequency of collisions (40 MHz)
- Two general-purpose detectors: CMS and ATLAS

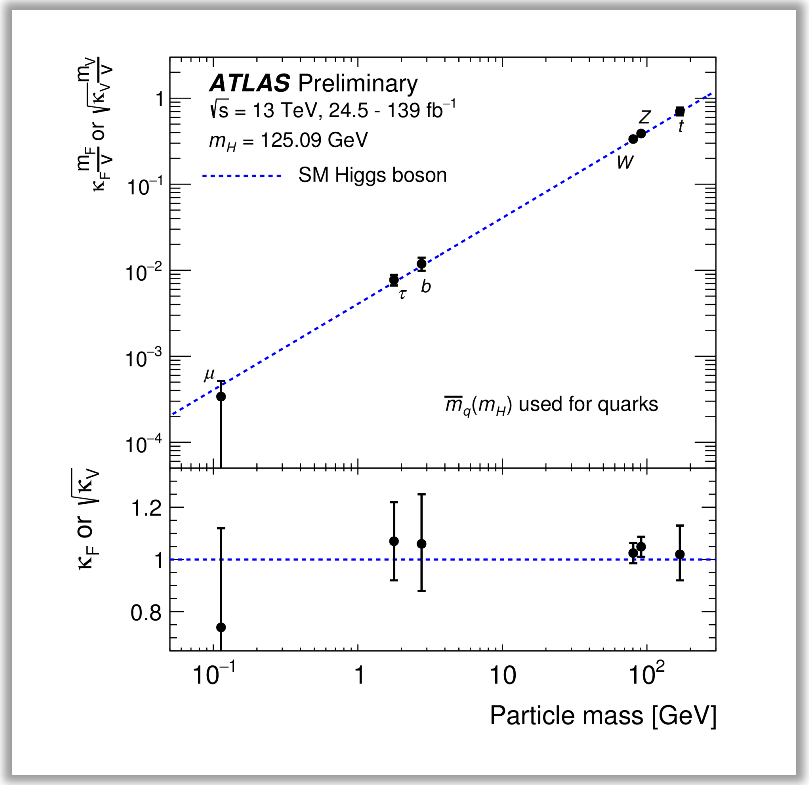


Discovery of the Higgs boson (2012)
- Counting experiment in bins (intervals) of mass
- Comparison to theory prediction



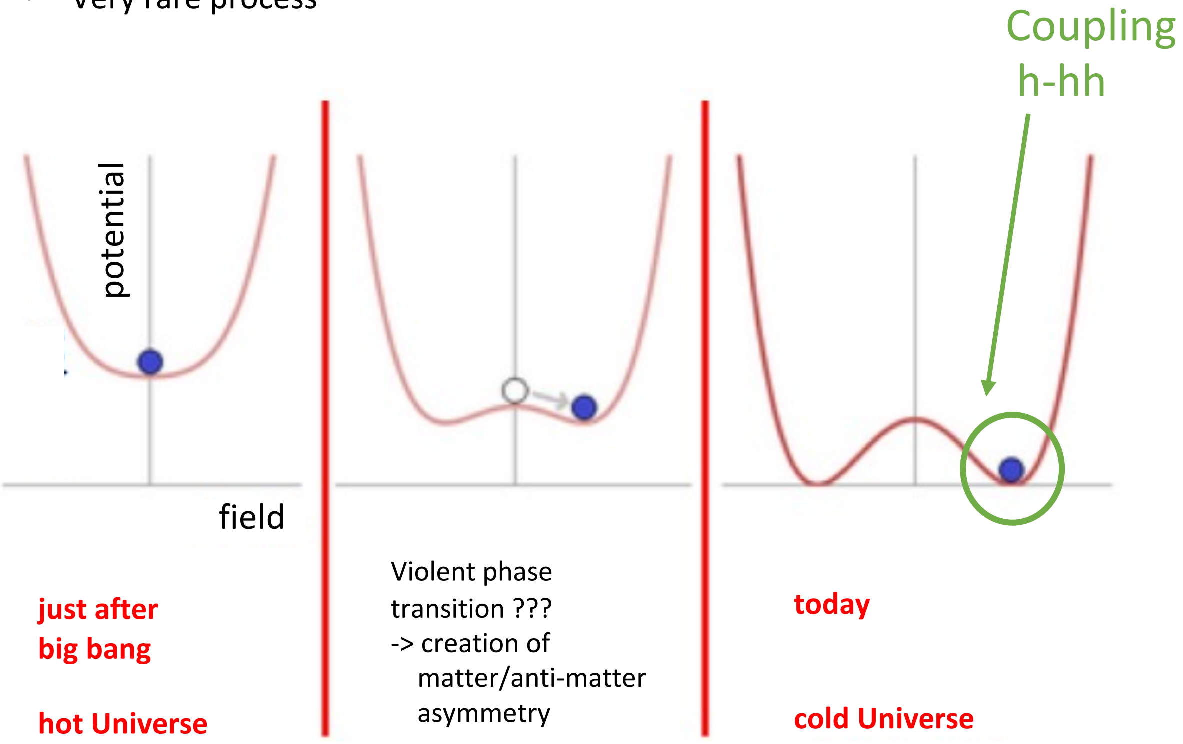
And since then ?





- Since 2012, higher beam energy, accumulation of much more statistics
- Study of the coupling strength of the Higgs boson to many other particles: demonstration of clear correlation between the mass of a given particle and its coupling to the Higgs boson
- Key goal for HL-LHC: first experimental probe of the Higgs potential

  -> expect to gain insight on the first instants of the universe
  - Study Higgs self-coupling: h -> hh
  - Very rare process



Fifteen times more data

HiLumi HL-LHC PROJECT

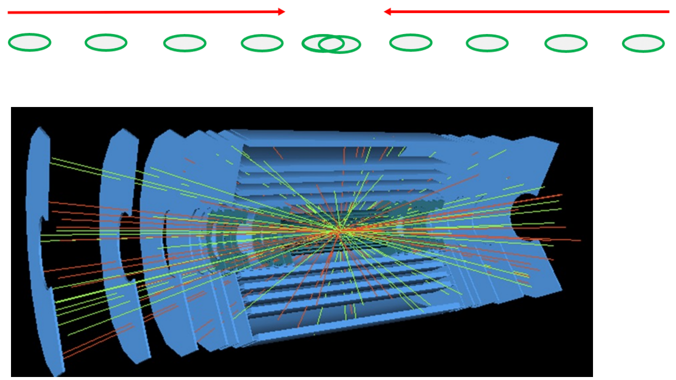
$30\ fb^{-1}$	$190\ fb^{-1}$	$350\ fb^{-1}$	$3000\ fb^{-1}$	Integrated luminosity
2011 -2012	2015 -2018	2022 -2024	2027 -2040	
Run 1	Run 2	Run 3	Run 4-5…	

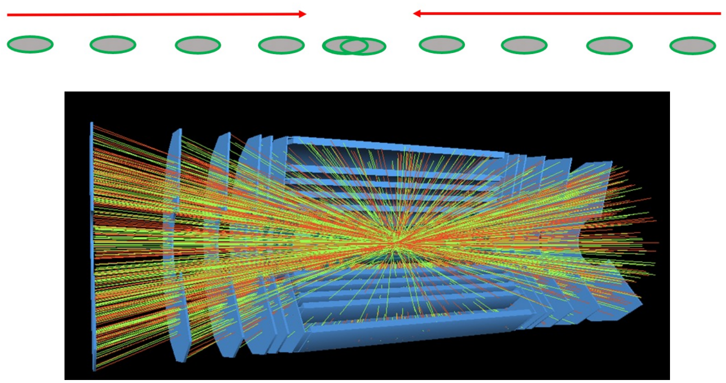
High luminosity phase:

- more data
- 5 times more protons/bunch
- more complex events
- highly granular detectors

40 million crossings of pairs of proton bunches per second !

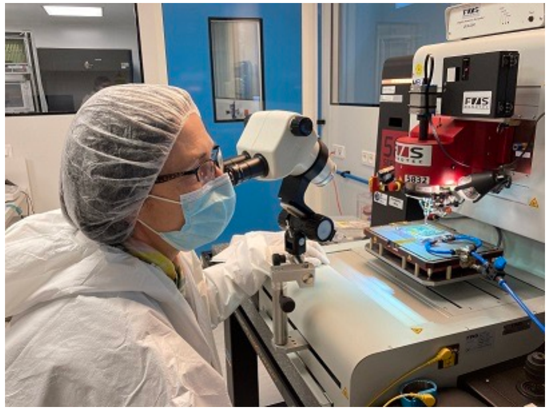
High luminosity: how ? Cannot reduce distance between bunches any further. More protons/bunch !





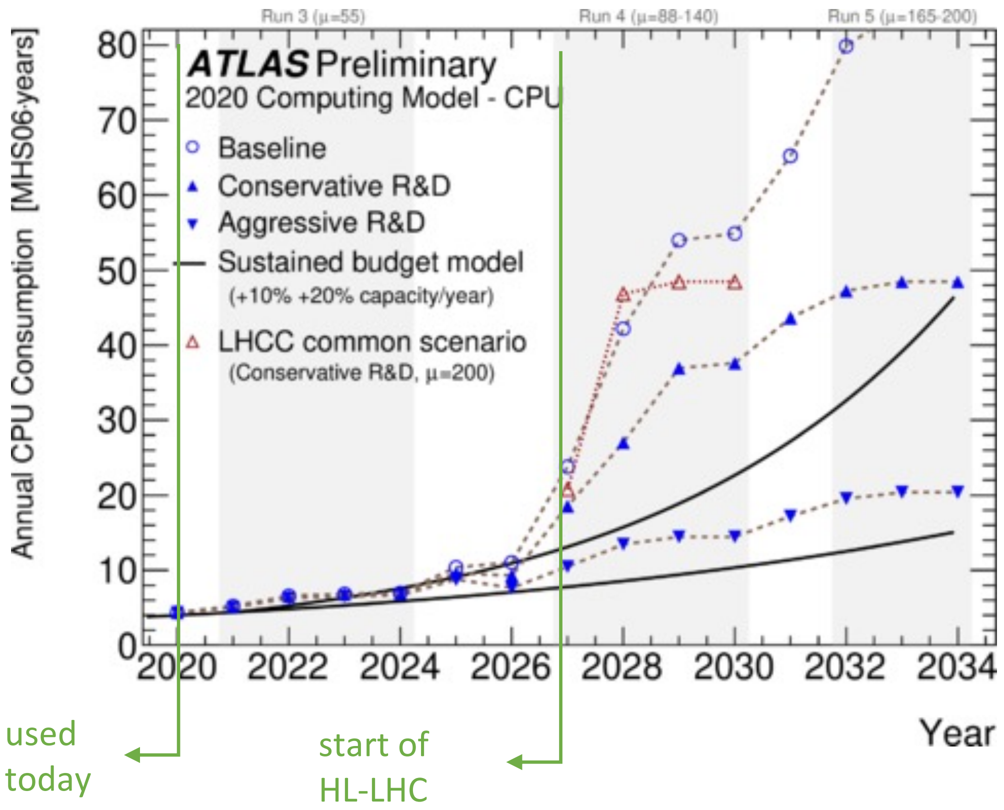
Work towards HL-LHC, today :

- beam injection chain
- construction of new detector components
- design of computing models



Computing resources

Resources used today:
- O(1) million de CPU cores running continuously
- O(1) exabyte of storage



At HL-LHC: with our current computing model, we would face a significant shortage of computing resources

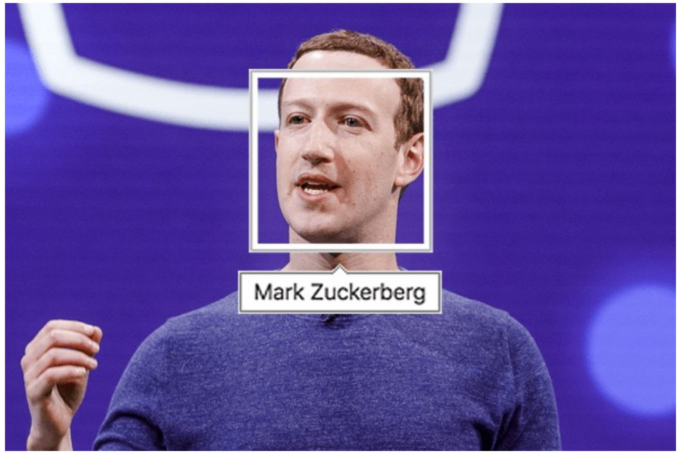
→ **need to make important changes**

→ **… or live with cuts into our physics programme**

ML for track pattern recognition ?

Challenge on Kaggle platform (in 2018): [(link)]

Article in proceedings of CHEP 2018: [(link)]

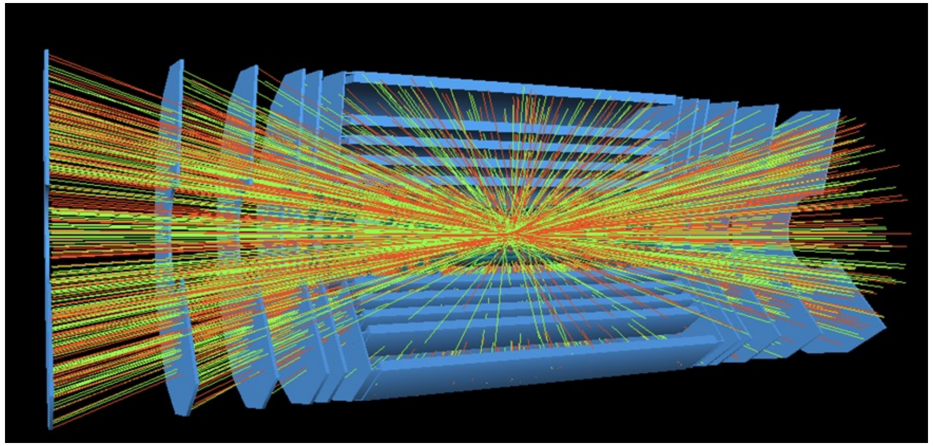


622 * 415 pixels

a large fraction carries information about the person

Can't use the same tools

How to present tracking data to a neural network ?



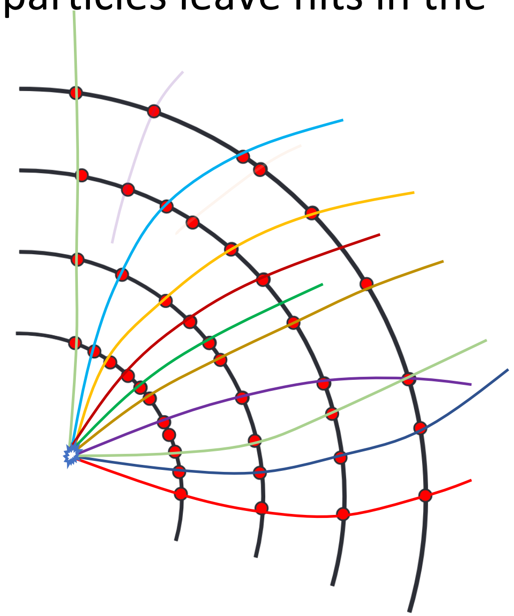
ATLAS tracker for HL-LHC:
$5 * 10^9$ readout channels
$\sim 3 * 10^5$ 3D space-points per event

=> data are *sparse*

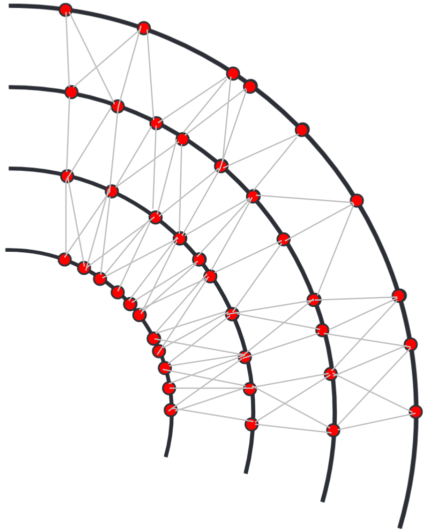
Representing tracking data using graphs

S. Farrell *et al.*, “Novel deep learning methods for track reconstruction”, proceedings of *Connecting the Dots* conference 2018 (link)

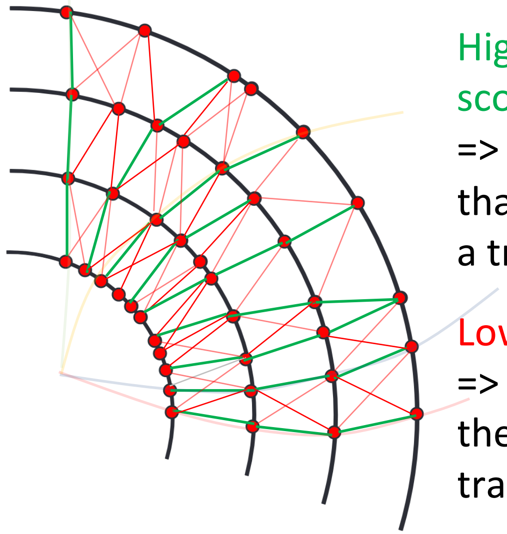
Charged particles leave hits in the detector



Represent the data using a graph



Goal:
classify the edges of the graph



High classification score
=> **high probability** that the edge is part of a track

Low classification score
=> **low probability** that the edge is part of a track

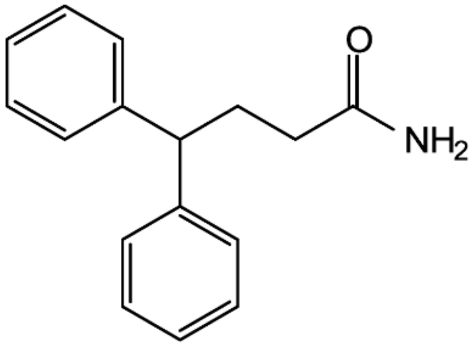
One node of the graph = one hit in the detector

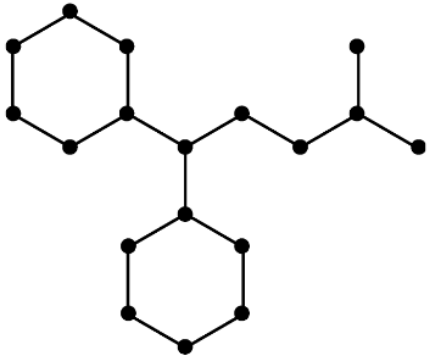
Connect two nodes using an edge
if “it seems possible” that the two hits
are two (consecutive) hits on a track

Graph creation

A classic use case for graph neural networks:

Study molecules and their chemical bonds





"I suppose I'll be the one to mention the elephant in the room."

In our tracking example, using the TrackML dataset, we have O(100k) hits per event.

$\Rightarrow$ A fully connected graph would have O(100k) nodes and O(10^{10}) edges. This is not going to fly.

Keep in mind that we want to run this at high throughput.
Efficient graph creation becomes an area of study on its own.

Graph creation: "module map"

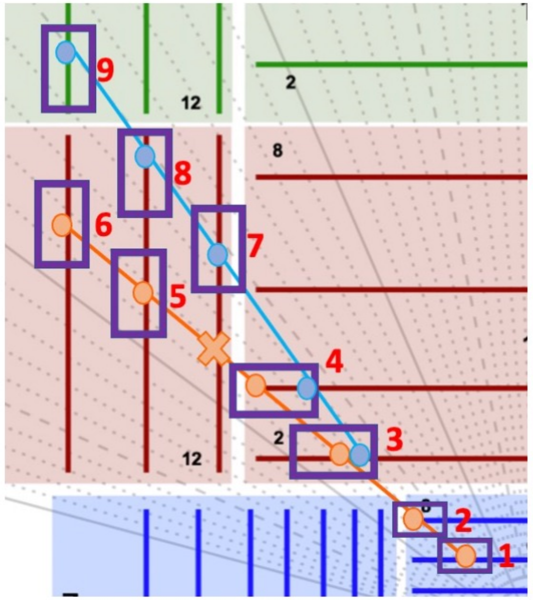
C. Biscarat *et al.*, "Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC", proceedings of the *vCHEP2021* conference (link)

Results for TrackML detector

New **data-driven** graph construction method:

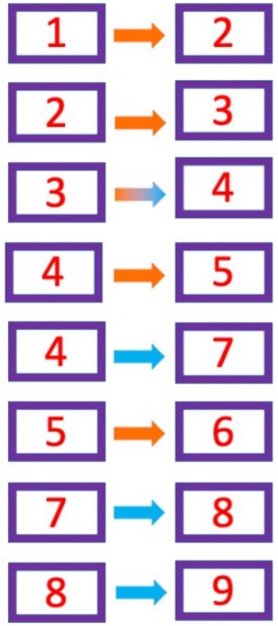
- build graphs starting from a list of possible connections from a *zone* to another *zone*: the *module map*
- done using 1000 $t\bar{t}$ events considering tracks with p_T>0.5 GeV and leaving at least 3 hits

Particles leaving hits



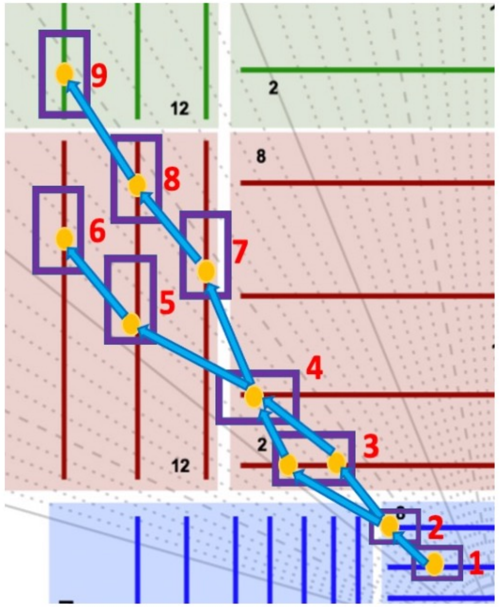
Done once

Module map creation



For event reconstruction

Graph creation



239 699 module connections found

Graph creation: "module map"

C. Biscarat *et al.*, "Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC", proceedings of the *vCHEP2021* conference [(link)]

Results for TrackML detector

- Graphs created with this module map have $O(10^5)$ nodes and $O(10^7)$ edges

➡ Need to reduce further !

Edge selection on geometric parameters:

- $z_0 = z_{h1} - r_{h1} \times \left(\frac{\Delta z}{\Delta r}\right)$
- $\phi_{slope} = \frac{\Delta\phi}{\Delta r}$
- $\Delta\phi = \phi_{h2} - \phi_{h1}$
- $\Delta\eta = \eta_{h2} - \eta_{h1}$

$h_{1,2}$ being the hits connected by a given edge

	Mean number of edges	ε
Initial	18.5×10^6	—-
z_0	3.56×10^6	19%
z_0, ϕ_{slope}	1.32×10^6	7.1%
$z_0, \phi_{slope}, \Delta\phi$	1.10×10^6	5.8%
$z_0, \phi_{slope}, \Delta\phi, \Delta\eta$	1.05×10^6	5.6%

- **Input** graph

$$Node\ features = (r, \phi, z)$$
$$Edge\ features = (\Delta\eta, \Delta\phi, \Delta r, \Delta z)$$

- **Target** graph

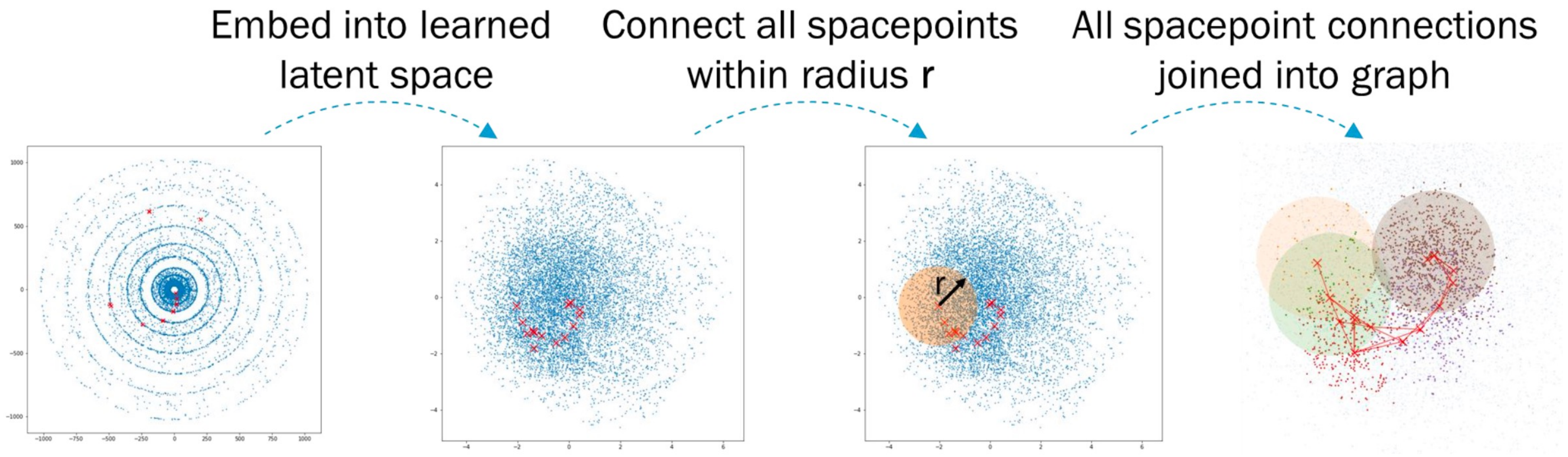
$$Edge\ features = (boolean\ truth\ flag)$$

Graph creation: metric learning

X. Yu *et al.*, “Performance of a geometric deep learning pipeline for HL-LHC particle tracking”, Eur. Phys. J. C 81 (2021) 876 (link)

First Step: metric learning

- ➢ For all hits, embed features (coordinates, cell direction, ...) with multi-layer perceptron (MLP) into N-dimensional space
- ➢ Associate hits on same track as close as N-dimensional distance
- ➢ Score each neighbour hit within embedding neighbourhood against the “source” hit at centre
- ➢ Create edges between the source hit at centre and the neighbouring hits above a given threshold on the score.



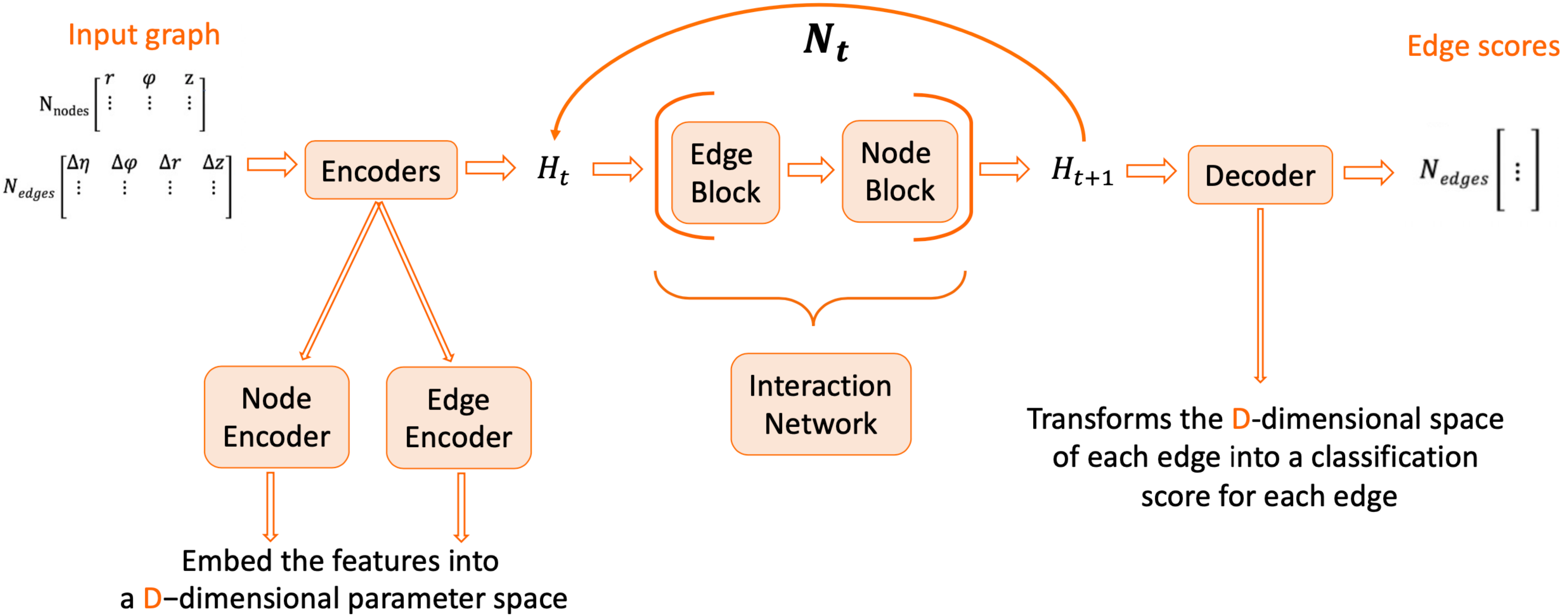
Second step: filtering

Reduce the number of edges using an MLP that looks separately at each edge (the features of the two nodes).

GNN architectures

S. Farrell *et al.*, "Novel deep learning methods for track reconstruction", proceedings of *Connecting the Dots* conference 2018 [(link)]

Also used in Biscarat *et al.* (vCHEP2021).

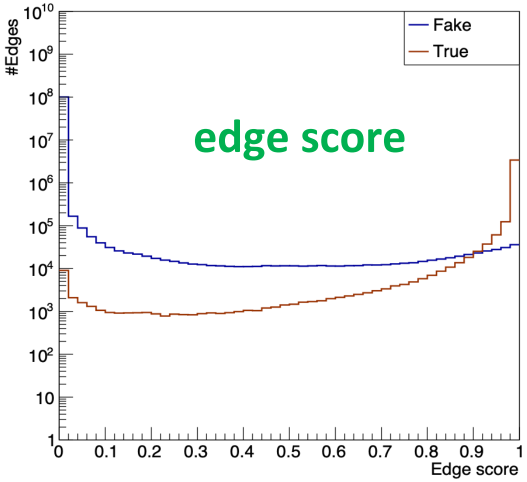


An alternative GNN architecture ("Recurrent Attention Message Passing") is presented in N. Choma *et al.* (CTD 2020) [(link)]

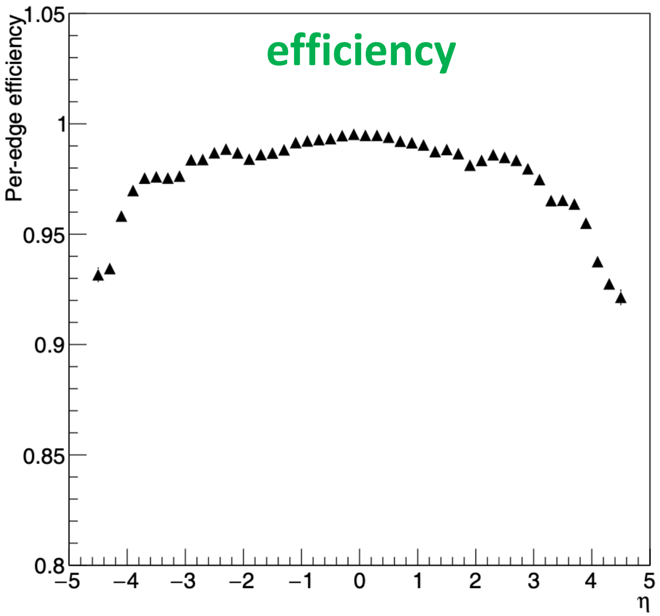
Edge-level performance

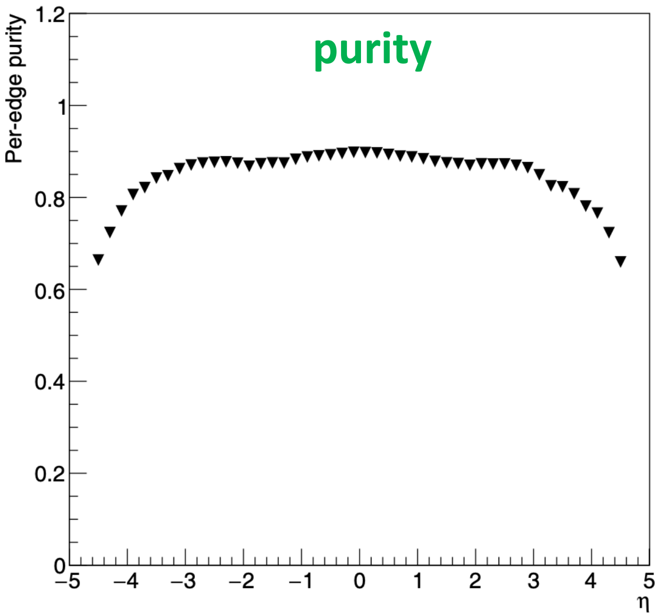
C. Biscarat *et al.*, "Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC", proceedings of the *vCHEP2021* conference [(link)]

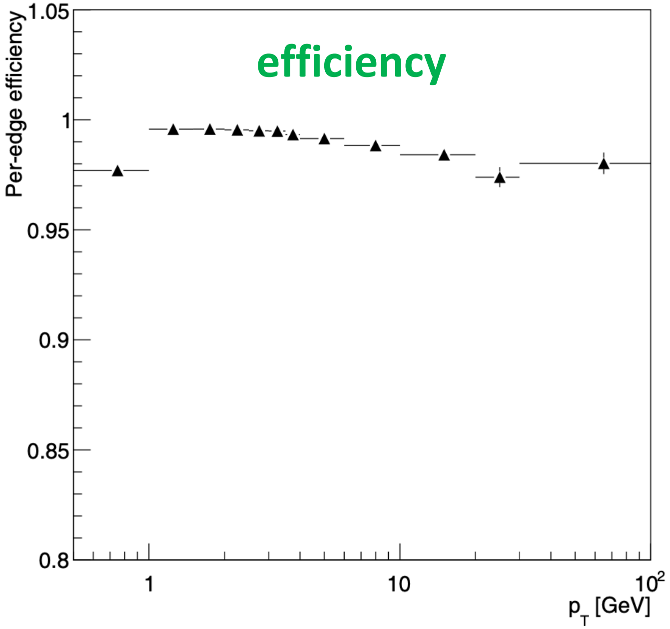
Results for TrackML dataset



Cut	0.5	0.7	0.8
Per-edge efficiency	0.992	0.987	0.982
Per-edge purity	0.916	0.937	0.950

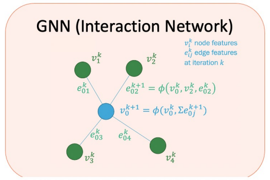


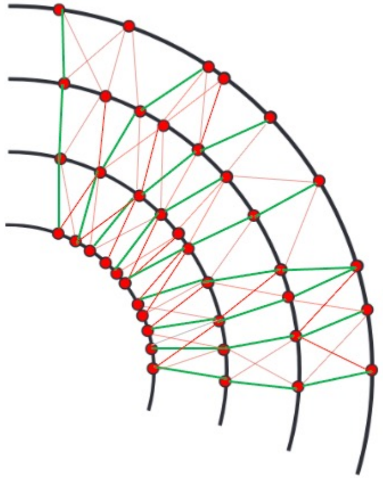


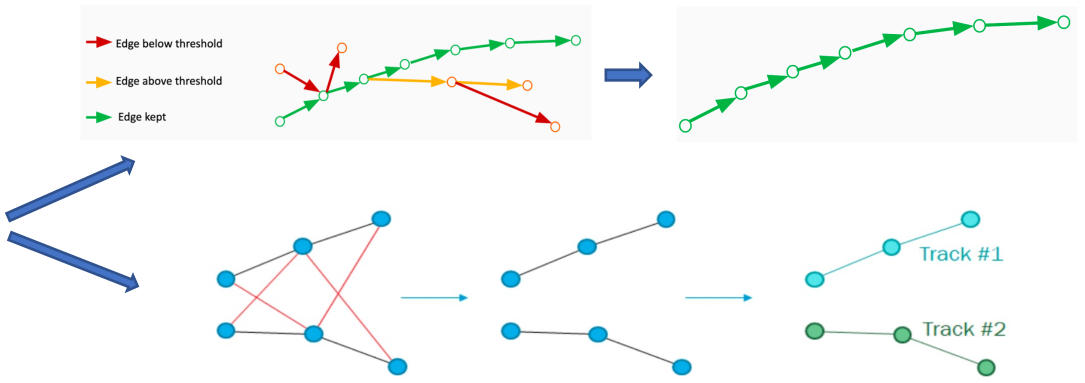


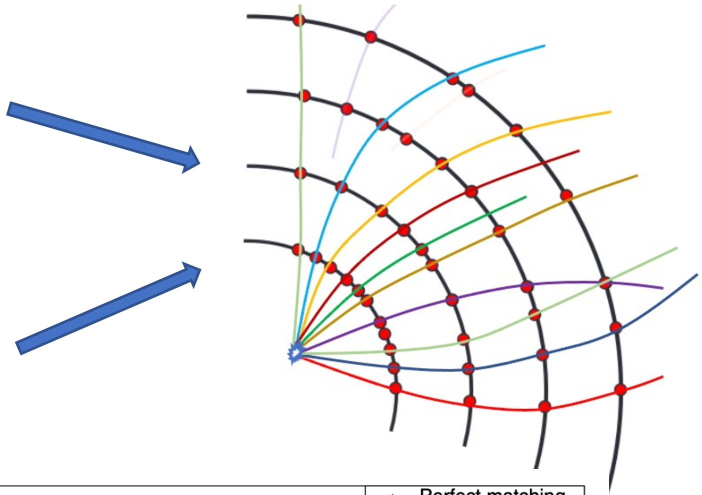
Track building starting from graph with edge scores

GNN









Matching Criteria

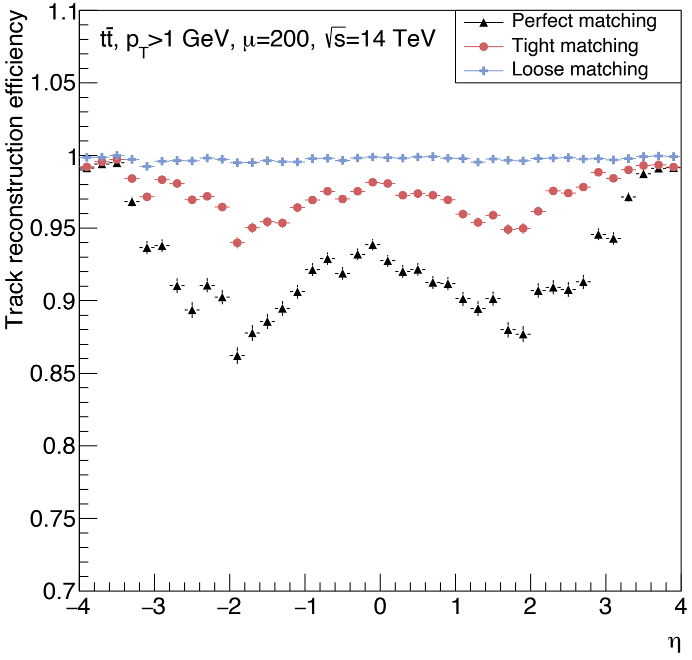
Particle :

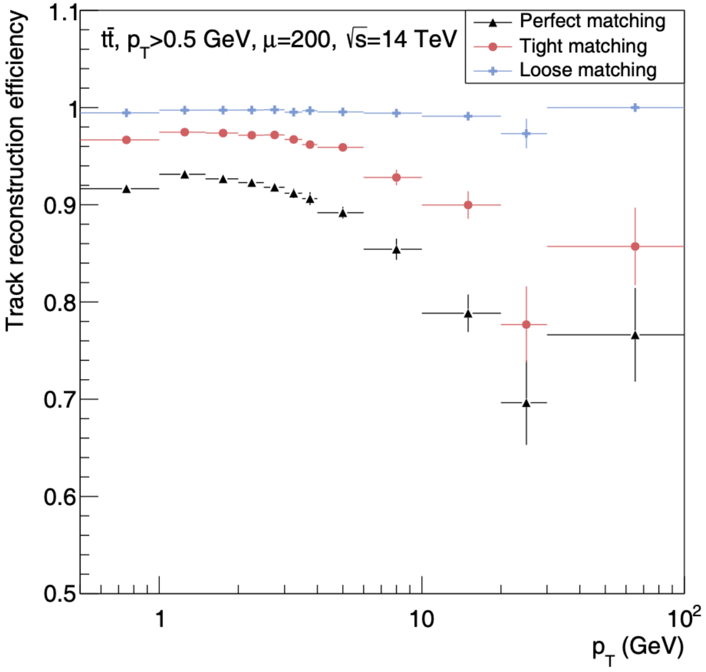
Loose matching :

Tight matching :

Perfect matching :

Results for TrackML dataset





C. Biscarat *et al.*, "Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC", proceedings of the *vCHEP2021* conference [(link)]

Inference speed

Article in proceedings of ACAT2021 conference (link)

In many studies, no attempt is made to optimise execution speed (demonstrate feasibility first).

Constraints imposed by the need to run the final algorithm at high throughput are kept in mind.

Competitive execution speed has been demonstrated for one complete chain of algorithms.

Substantial gains expected from future implementations with custom CUDA kernels.

(Any volunteer for coding the module map in CUDA ?)

Accelerating the Inference of the Exa.TrkX Pipeline

Alina Lazar[1], Xiangyang Ju[2], Daniel Murnane[2], Paolo Calafiura[2], Steven Farrell[2], Yaoyuan Xu[3], Maria Spiropulu[4], Jean-Roch Vlimant[4], Giuseppe Cerati[5], Lindsey Gray[5], Thomas Klijnsma[5], Jim Kowalkowski[5], Markus Atkinson[6], Mark Neubauer[6], Gage DeZoort[7], Savannah Thais[7], Shih-Chieh Hsu[8], Adam Aurisano[9], Jeremy Hewes[9], Alexandra Ballow[1], Nirajan Acharya[1], Chun-yi Wang[10], Emma Liu[11], Alberto Lucas[12]

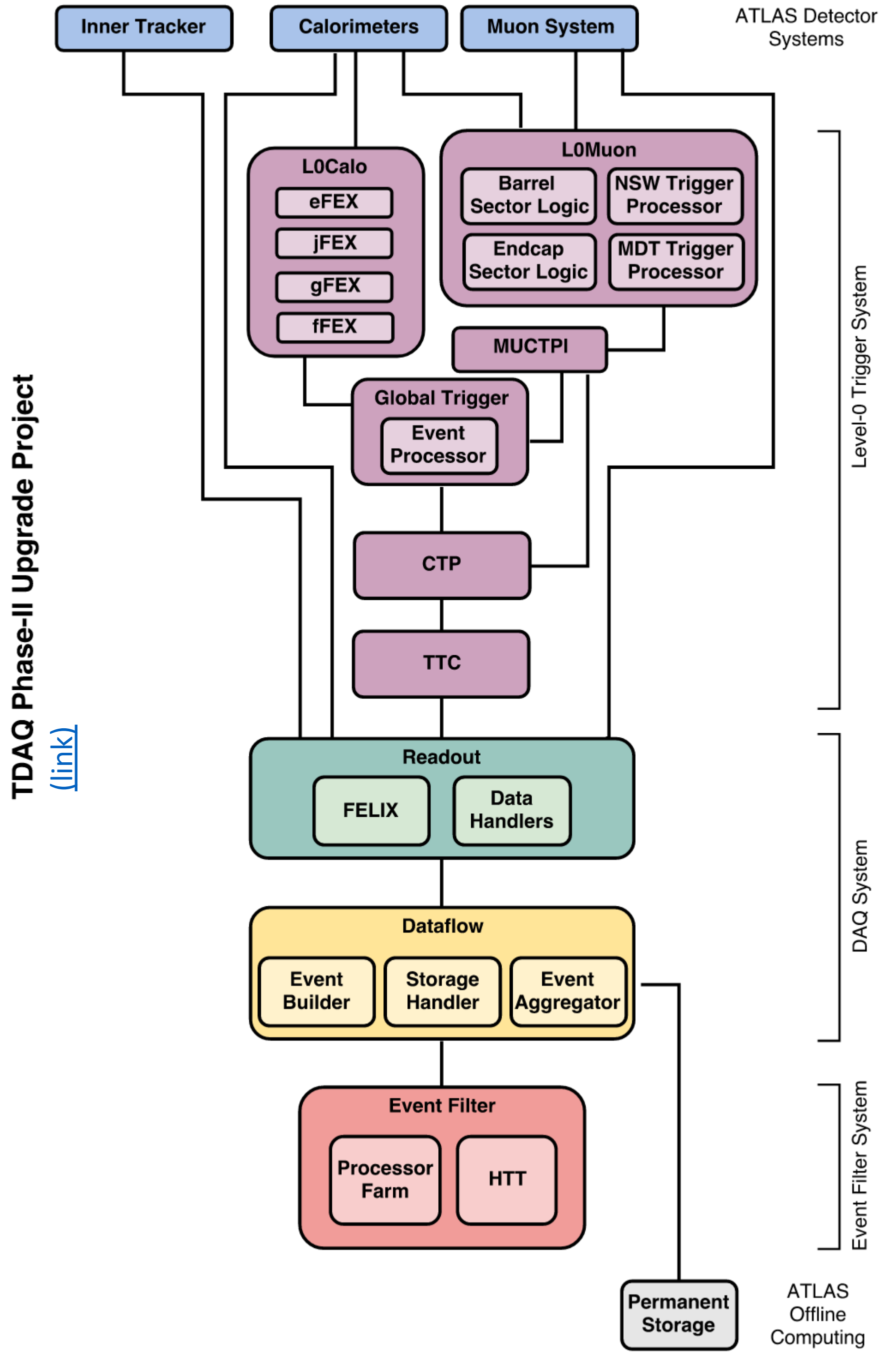
[1]Youngstown State University, [2]Lawrence Berkeley National Lab, [3]University of California-Berkeley, [4]California Institute of Technology, [5]Fermi National Accelerator Laboratory, [6]University of Illinois Urbana-Champaign, [7]Princeton University, [8]University of Washington, [9]University of Cincinnati, [10]National Tsing Hua University [11]University of California, Los Angeles [12]California State University, Monterey Bay

Table 1. Wall time of the Python-based Inference pipeline for the baseline and optimized implementations. The time is calculated with 500 events on an Nvidia Volta 100 GPU with a memory of 16 GB. The reported times are the average time and the standard deviation of the time in the unit of seconds.

	Baseline	Faiss	cuGraph	AMP	FRNN
Data Loading	0.0022 ± 0.0003	0.0021 ± 0.0003	0.0023 ± 0.0003	0.0022 ± 0.0003	0.0022 ± 0.0003
Embedding	0.02 ± 0.003	0.02 ± 0.003	0.02 ± 0.003	0.0067 ± 0.0007	0.0067 ± 0.0007
Build Edges	12 ± 2.64	0.54 ± 0.07	0.53 ± 0.07	0.53 ± 0.07	0.04 ± 0.01
Filtering	0.7 ± 0.15	0.7 ± 0.15	0.7 ± 0.15	0.37 ± 0.08	0.37 ± 0.08
GNN	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03	0.17 ± 0.03
Labeling	2.2 ± 0.3	2.1 ± 0.3	0.11 ± 0.01	0.09 ± 0.008	0.09 ± 0.008
Total time	$15 \pm 3.$	3.6 ± 0.6	1.6 ± 0.3	1.2 ± 0.2	0.7 ± 0.1

Triggering (ATLAS)

“Recording data at the LHC is like drinking from a fire hose”



Event rate: 40 MHz

after hardware-based L0 trigger: 100 kHz

detailed detector readout after L0 accept

after event filter (to tape): 1.5 kHz

Simplify GNN -> inference on FPGA

Computing and Software for Big Science (2021) 5:26
https://doi.org/10.1007/s41781-021-00073-z (link)

ORIGINAL ARTICLE

Charged Particle Tracking via Edge-Classifying Interaction Networks

Gage DeZoort[1] · Savannah Thais[1] · Javier Duarte[2] · Vesal Razavimaleki[2] · Markus Atkinson[3] · Isobel Ojalvo[1] ·
Mark Neubauer[3] · Peter Elmer[1]

Received: 12 July 2021 / Accepted: 13 October 2021 / Published online: 15 November 2021

Abstract
Recent work has demonstrated that geometric deep learning methods such as graph neural networks (GNNs) are well suited to address a variety of reconstruction problems in high-energy particle physics. In particular, particle tracking data are naturally represented as a graph by identifying silicon tracker hits as nodes and particle trajectories as edges, given a set of hypothesized edges, edge-classifying GNNs identify those corresponding to real particle trajectories. In this work, we adapt the physics-motivated interaction network (IN) GNN toward the problem of particle tracking in pileup conditions similar to those expected at the high-luminosity Large Hadron Collider. Assuming idealized hit filtering at various particle momenta thresholds, we demonstrate the IN's excellent edge-classification accuracy and tracking efficiency through a suite of measurements at each stage of GNN-based tracking: graph construction, edge classification, and track building. The proposed IN architecture is substantially smaller than previously studied GNN tracking architectures; this is particularly promising as a reduction in size is critical for enabling GNN-based tracking in constrained computing environments. Furthermore, the IN may be represented as either a set of explicit matrix operations or a message passing GNN. Efforts are underway to accelerate each representation via heterogeneous computing resources towards both high-level and low-latency triggering applications.

Work has also been done to accelerate the inference of deep neural networks with heterogeneous resources beyond GPUs, like field-programmable gate arrays (FPGAs) [49–57]. This work extends to GNN architectures [29, 58]. Specifically, in Ref. [29], a compact version of the IN was implemented for $p_T > 2$ GeV segmented geometric graphs with up to 28 nodes and 37 edges, and shown to have a latency less than 1 μs, an initiation interval of 5 ns, reproduce the floating-point precision model with a fixed-point precision of 16 bits or less, and fit on a Xilinx Kintex UltraScale FPGA.

While this preliminary FPGA acceleration work is promising, there are several limitations of the current FPGA implementation of the IN:

1. This fully pipelined design cannot easily scale to beyond $\mathcal{O}(100)$ nodes and $\mathcal{O}(1000)$ edges. However, if the initiation interval requirements are loosened, it can scale up to $O(10, 000)$ nodes and edges.
2. The neural network itself is small, and while it is effective for $p_T > 2$ GeV graphs, it may not be sufficient for lower-p_T graphs.
3. The FPGA design makes no assumptions about the pos-

Triggering (ATLAS)

ATLAS EXPERIMENT

(link)

CERN-LHCC-2022-004 / ATLAS-TDR-029-ADD-1
01/03/2022

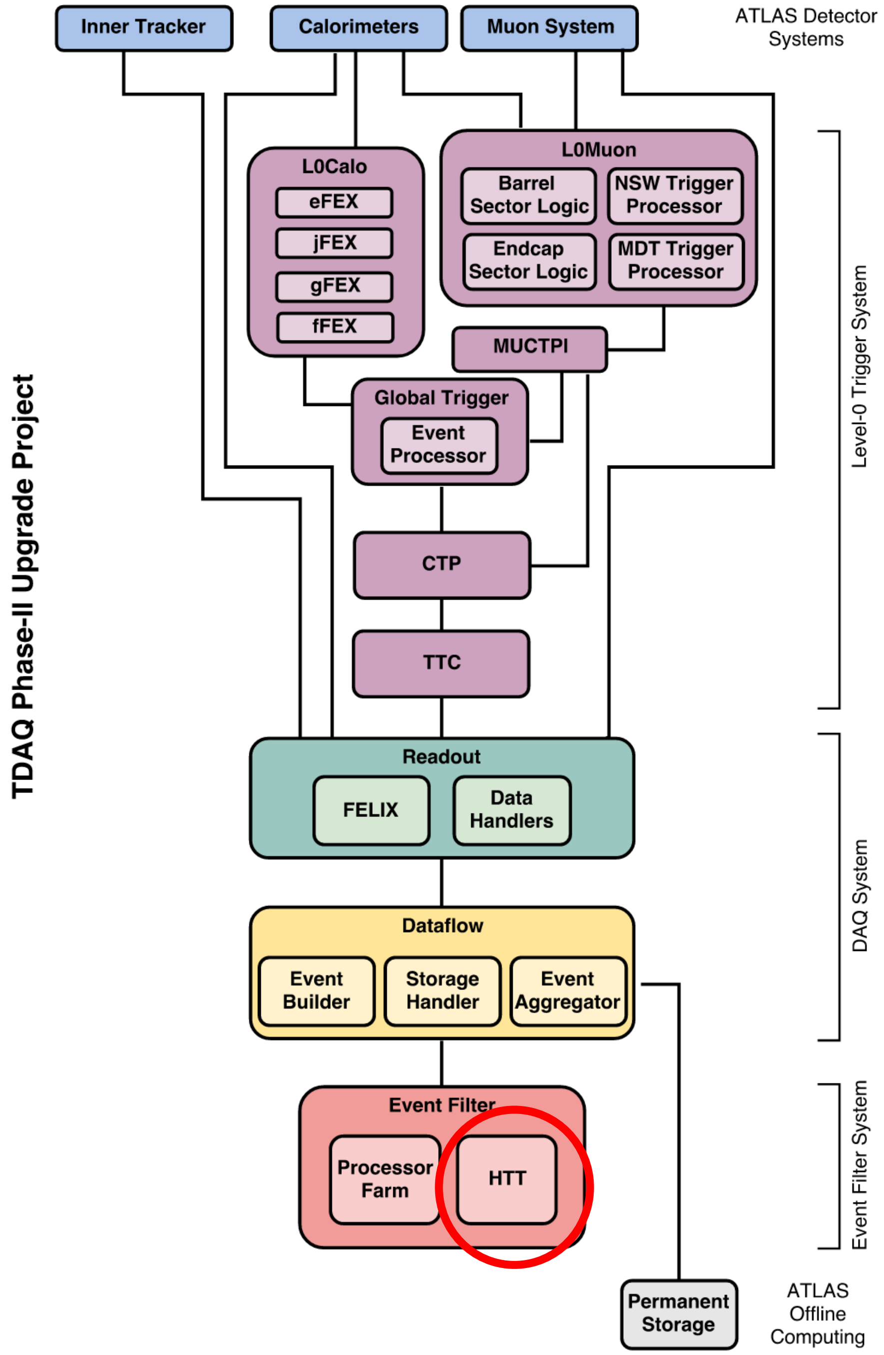
Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System - EF Tracking Amendment

The ATLAS Collaboration

Reference: v1.2

Created: March 1, 2022
Last modified: March 1, 2022
Prepared by: The ATLAS Collaboration

Abstract
This Technical Design Report Amendment describes revised plans for Event Filter Tracking in the upgrade of the ATLAS Trigger and Data Acquisition system for the High Luminosity LHC. The motivation to change the baseline for Event Filter Tracking is explained. Next, a description of the requirements for Event Filter Tracking and the definition of the proposed baseline to meet these requirements are presented. Several demonstrations using various hardware and software are reported in support of this proposal. Finally, the organization and resources needed to deliver the new baseline are set out.



Event rate: 40 MHz

after hardware-based L0 trigger: 100 kHz

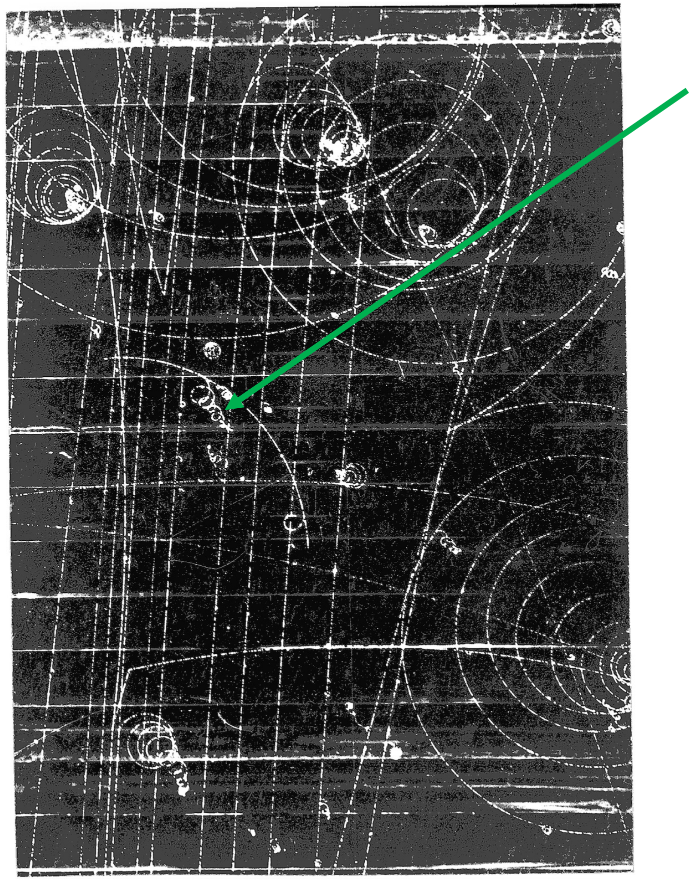
detailed detector readout after L0 accept

after event filter (to tape): 1.5 kHz

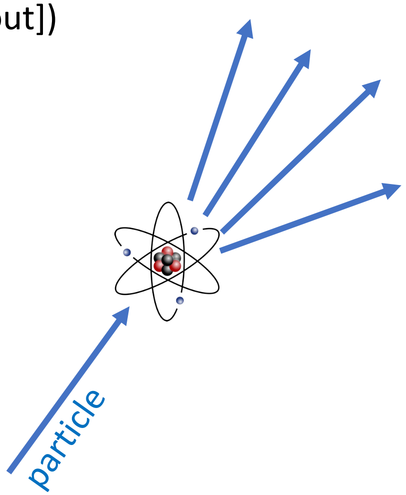
From the TrackML dataset to the real world

At least two major complications have been (deliberately) omitted from the TrackML dataset.

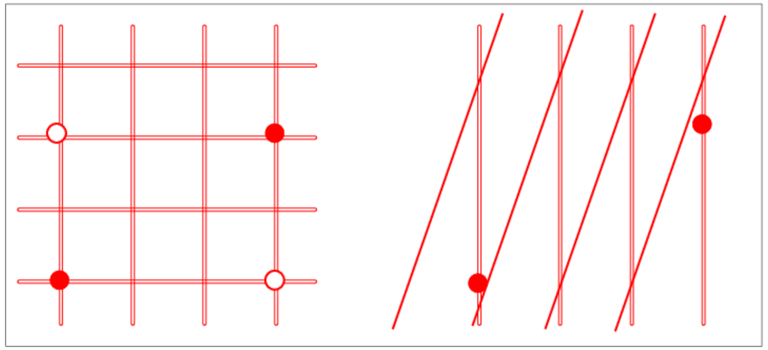
Secondary particles from interactions with material in detector



hadronic interactions:
(lots of relatively high-Z material in modern trackers [support, cooling, readout])



Silicon strip detectors



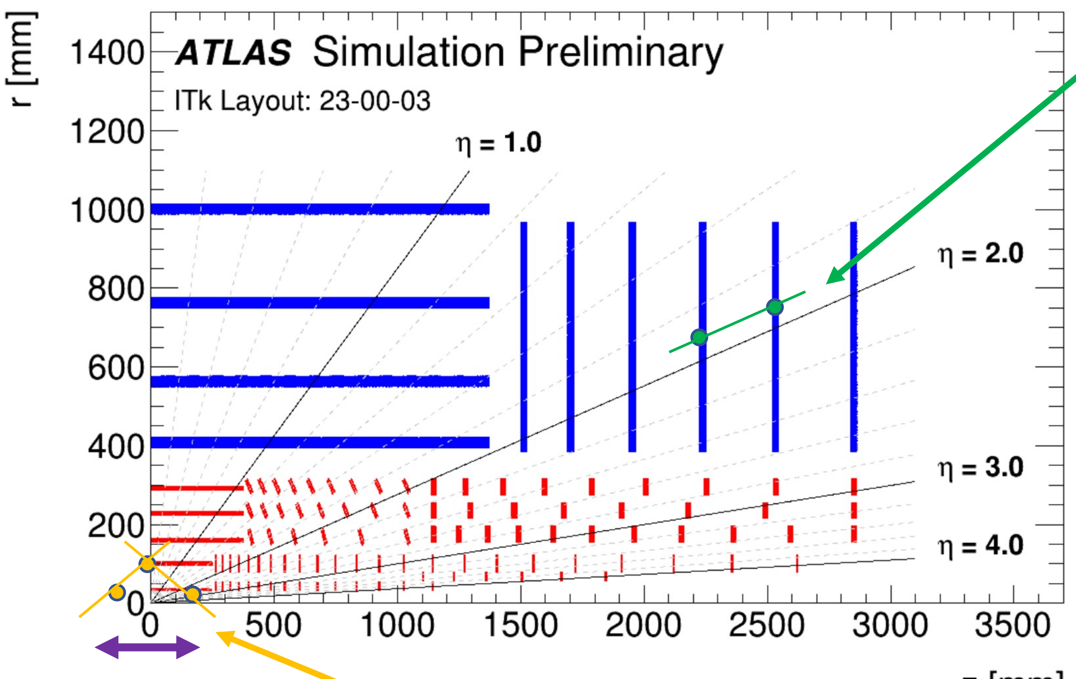
strip (one strip = 1 readout channel)

● hit from a charged particle

○ “ghost” (accidental crossing of strips)

It would be extremely useful to have a new open dataset that includes these effects.
Maybe even a dataset released by CMS and/or ATLAS ?

Comment on GNN design



In this region, it is relatively clear in which direction to look for the next hit.

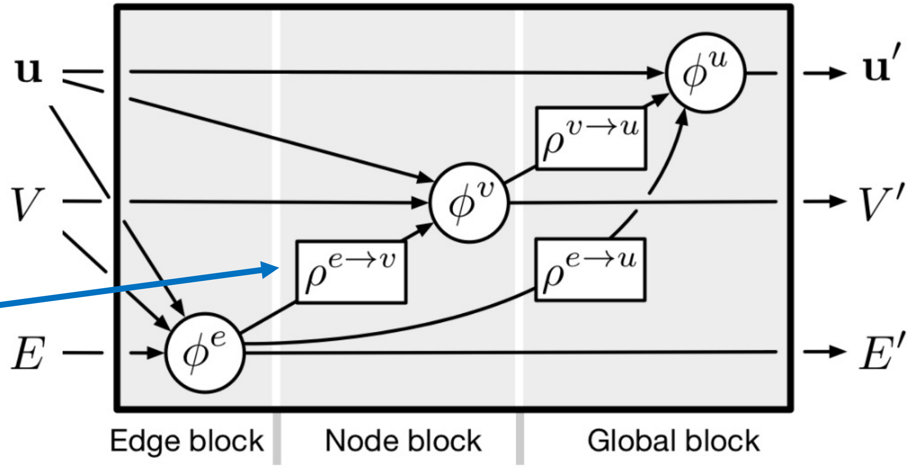
luminous region:
-200 < z < 200
at r = 0

In this region, the direction is less clear.
In addition, this is where the density of hits is largest.

One can easily have >10 edges on a given node.

In this aggregation function, we simply add "messages" over all incoming edges. Is a straight sum over >10 edges a good idea ? Should we pay more attention to the most interesting edges ?

Battaglia *et al.* (2018) (link)



(a) Full GN block

Conclusion

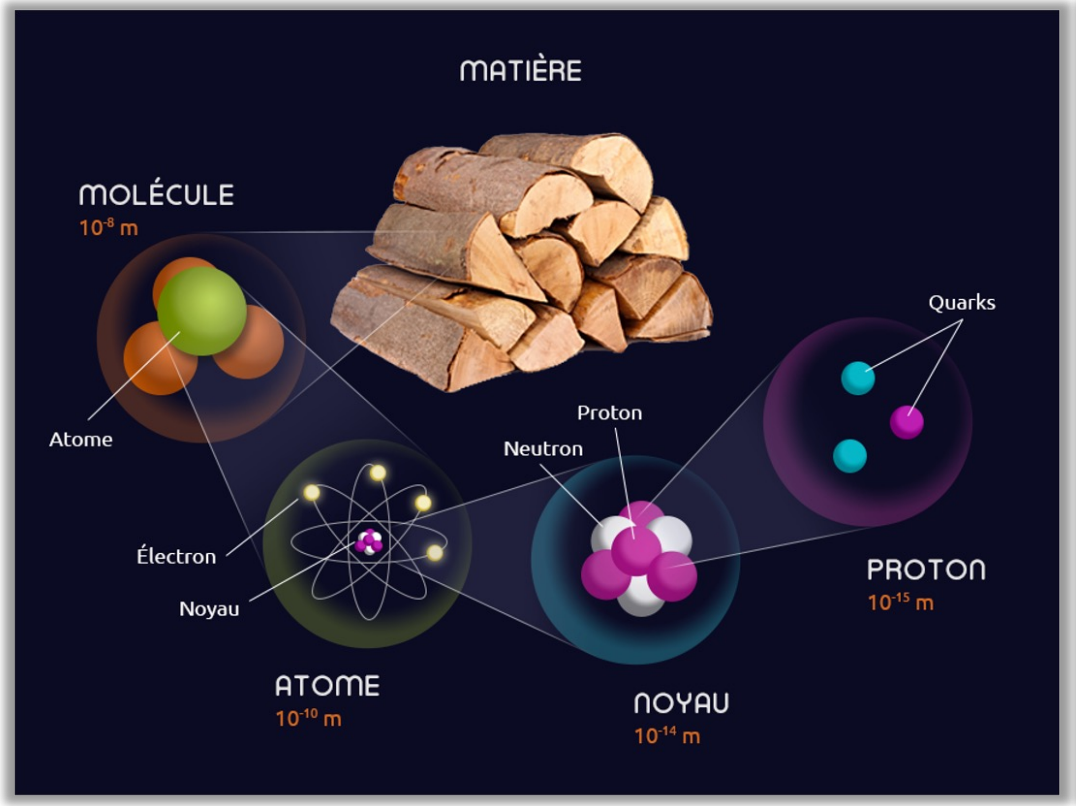
- Lots of exciting physics results since the Higgs boson discovery in 2012; the Higgs boson as new tool for precision tests of the standard model
- Even more exciting outlook for the HL-LHC: among others, probe the first moments of the Universe right after the Big Bang (EW baryogenesis)
- TrackML challenge in 2018 was huge boost to get the “ML for tracking” effort going. The dataset from then is still being used today.
- Feasibility of complete GNN-based tracking solutions has been demonstrated in 2021.
- These conceptual solutions are trying to face the real world right now.
  - A new open dataset that is close to the real world would really help.
- Lot’s of variants:
  - Fast, use in trigger decision
  - Seeding only
  - Large radius tracking
  - …

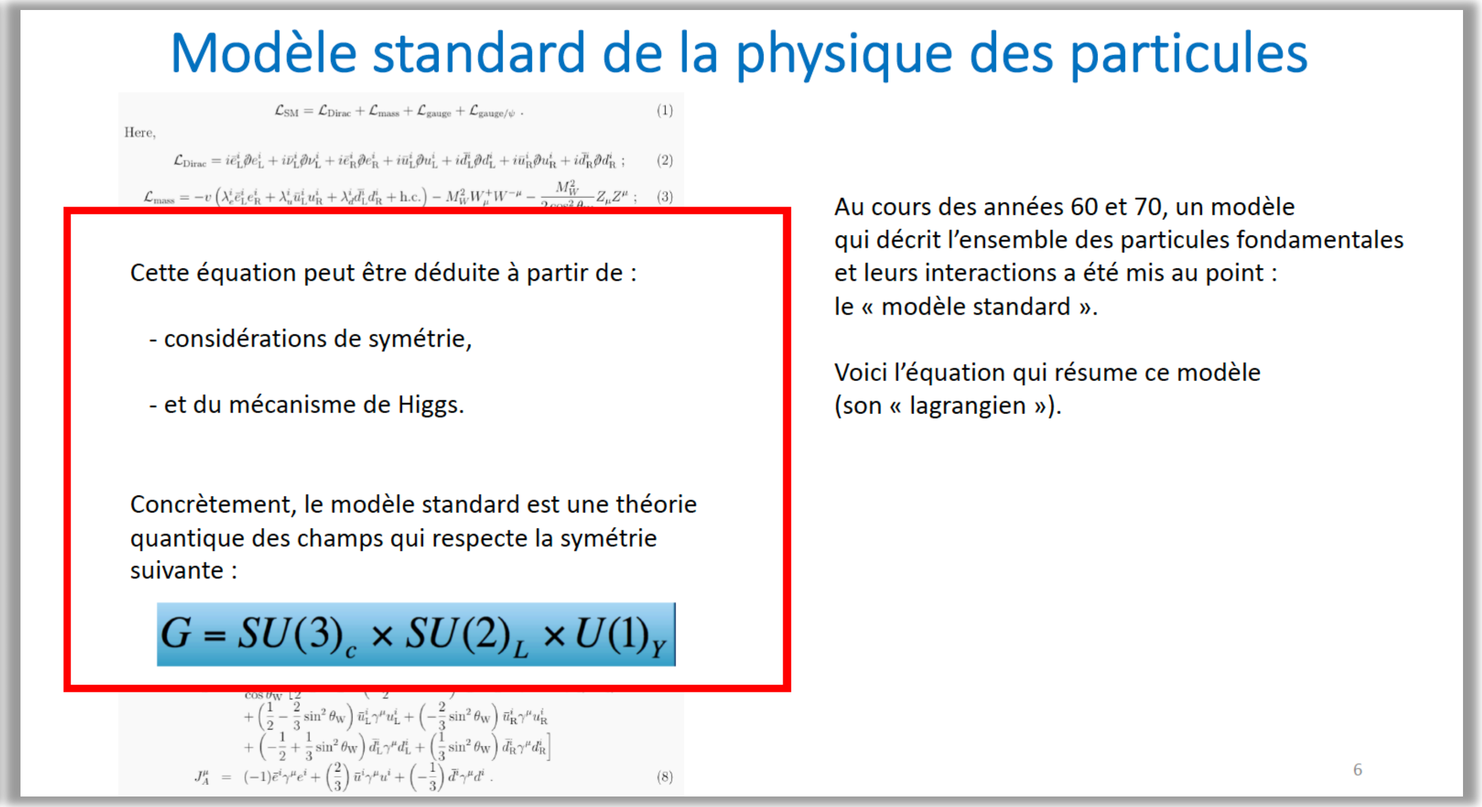
Backup material

La physique des particules et le boson de Higgs

- Comprendre les constituants fondamentaux de la matière et les interactions qui les régissent

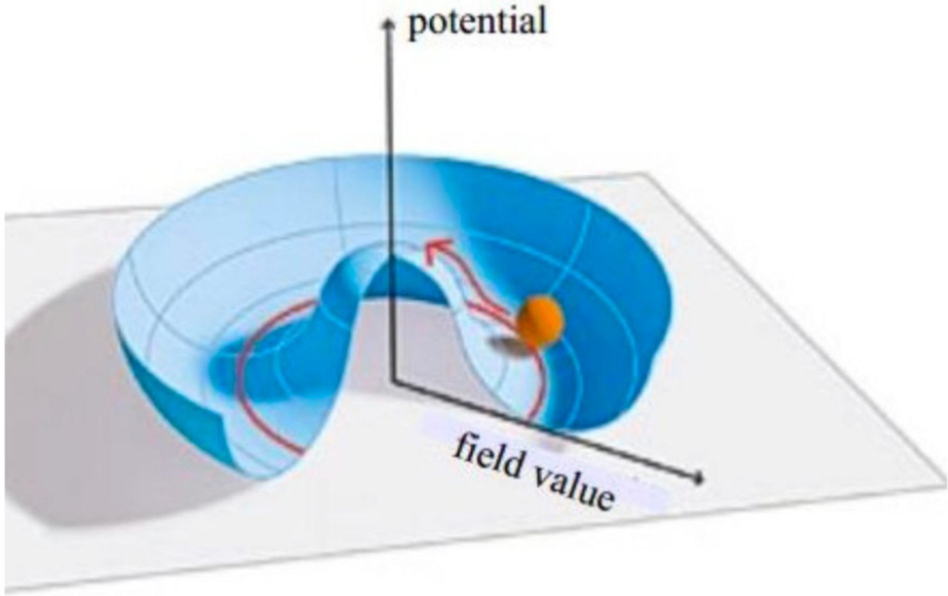
Séminaire SFP MP, Jan Stark (L2IT), déc. 2019





Higgs potential

A measurement of the Higgs self-coupling is the only way to experimentally reconstruct the Higgs potential (reconstruct its shape close to the minimum).

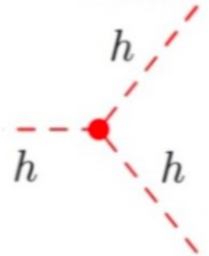


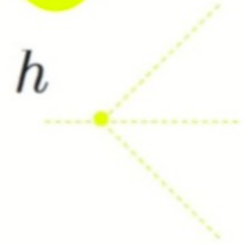
Higgs potential in the standard model:

$$V(\Phi) = \mu^2 \Phi^+ \Phi + \eta (\Phi^+ \Phi)^2$$

expansion around the minimum

$$\frac{1}{2} m_H^2 h^2 + \sqrt{\frac{\eta}{2}}\, m_H h^3 + \frac{\eta}{4} h^4$$

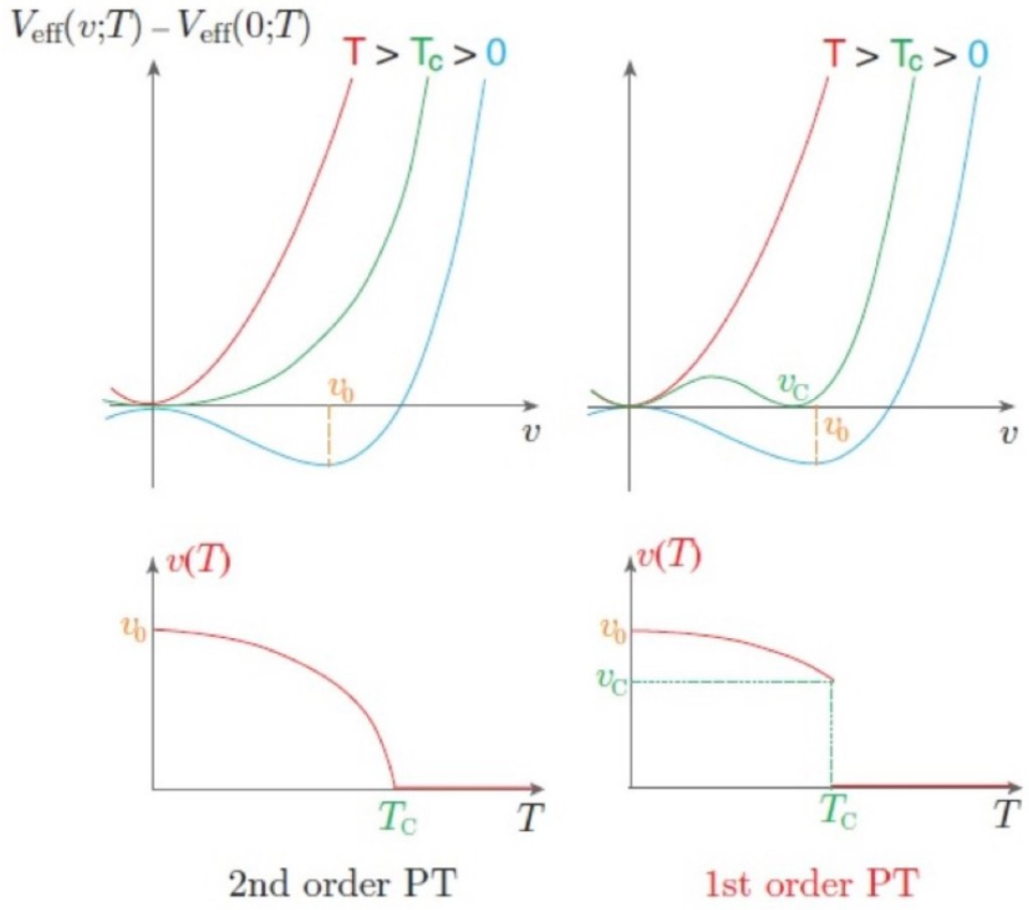




Electroweak baryogenesis ?

To get the observed baryon asymmetry of the universe from an initially baryon-symmetric universe, Sacharow's conditions must be satisfied.

(1) Baryon number (B) violation
(2) C and CP violation
(3) Out of equilibrium



It is not easy to construct a credible mechanism that meets these conditions.

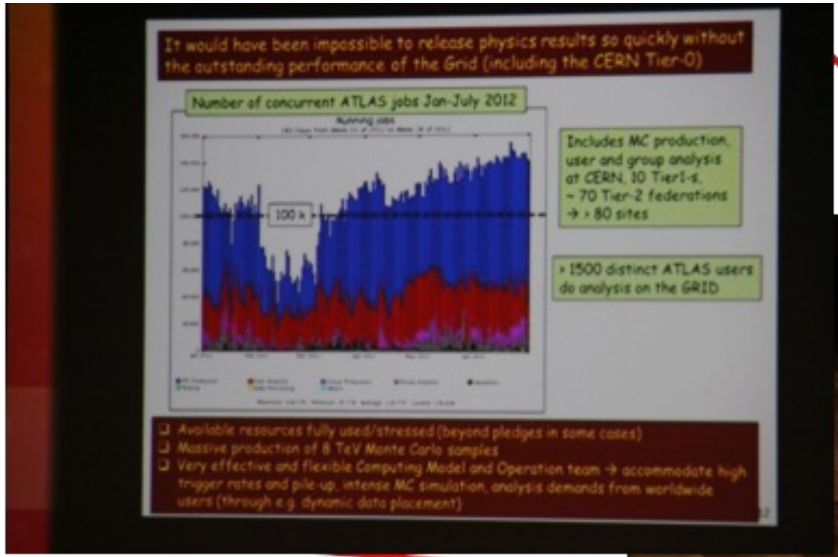
The mechanism that meets these conditions and that is considered to be the most credible one is electroweak baryosynthesis.

An effective potential (free energy density) is used to describe the Higgs potential during the electroweak phase transition.

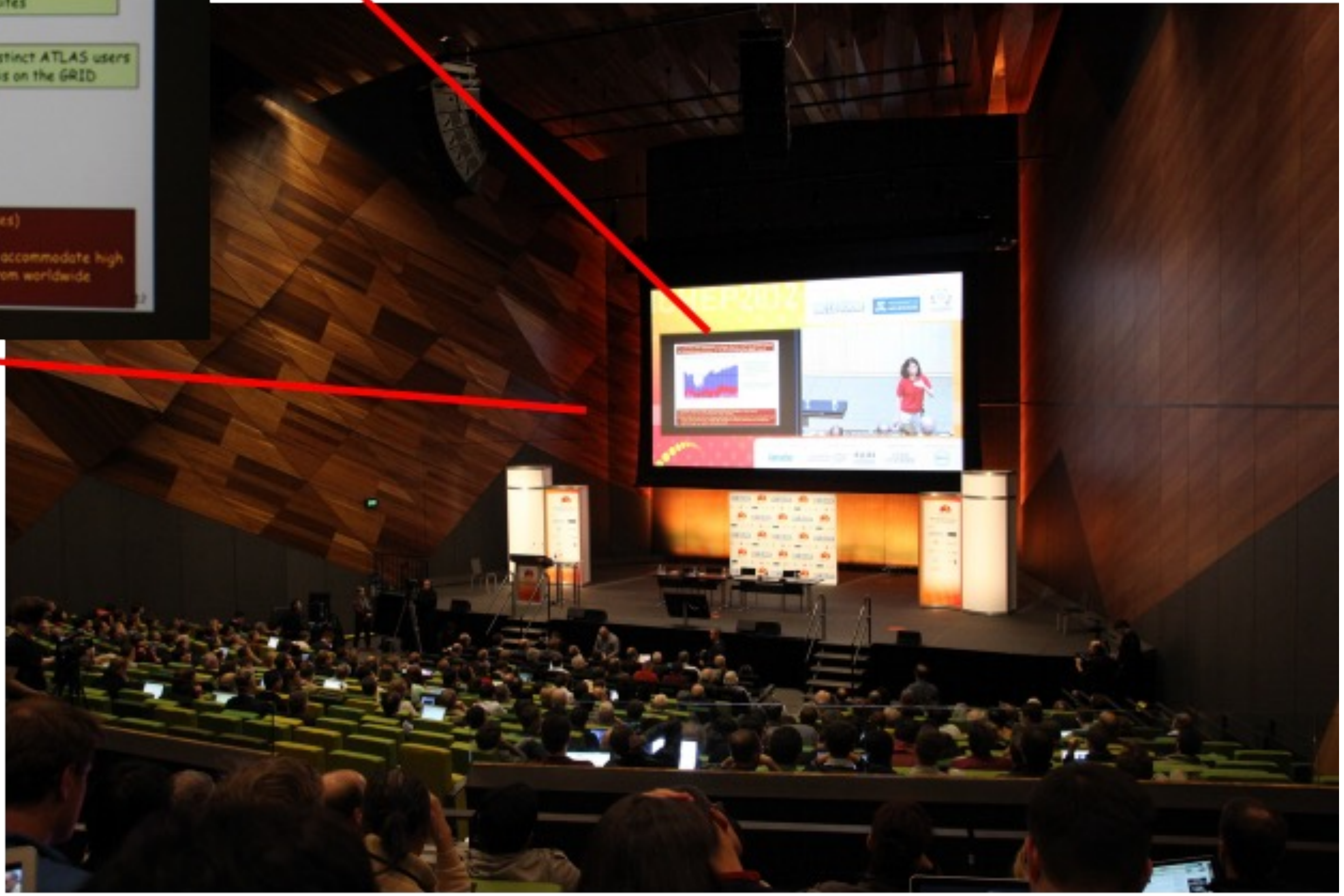
Electroweak baryosynthesis can only work if the electroweak phase transition is a a first order phase transition (PT).

First order PTs imply a system that is out of equilibrium (violent transition, large creation of entropy).

« Computing enables physics »



Photography: C. Biscarat



Annonce de la découverte du maillon manquant de notre Modèle Standard, le boson de Higgs.

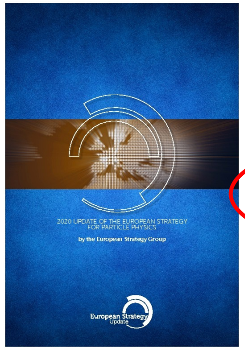
CERN seminar, July 4th 2012, retransmitted at ICHEP (Melbourne)

10 ans pour nous préparer

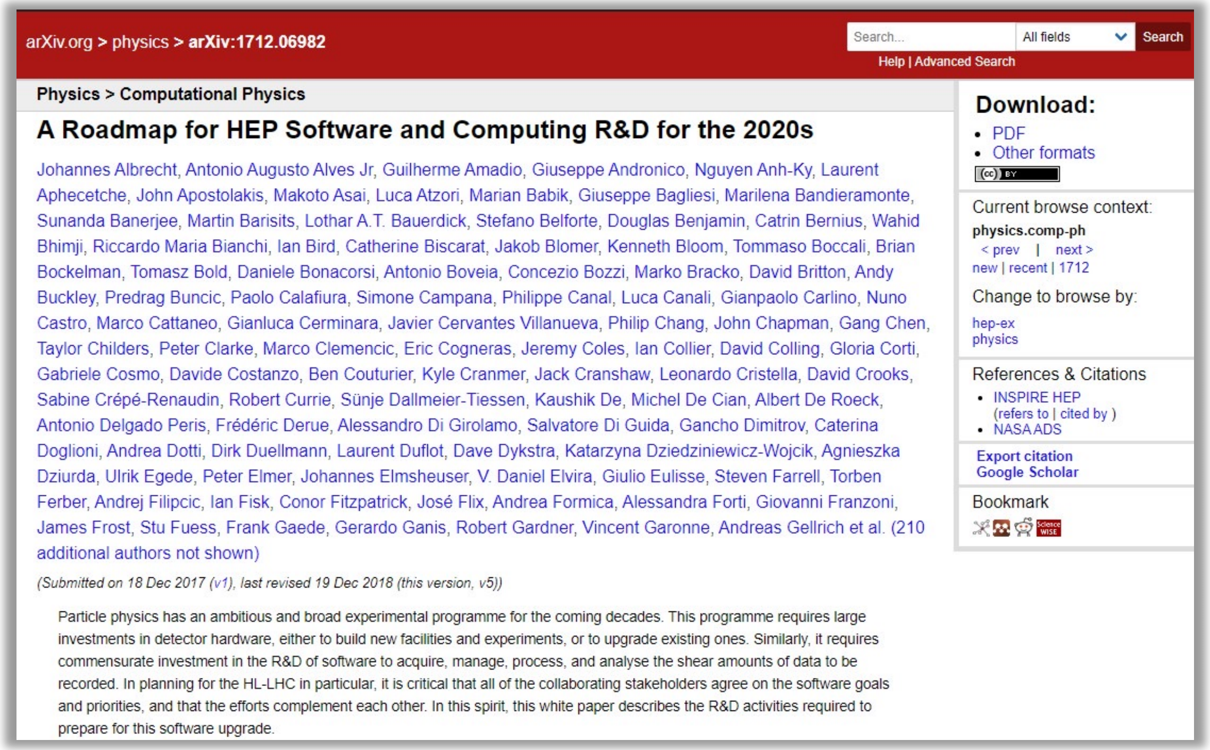
- *Community white paper* (2017)
  - Algorithmes, infrastructures, data access…

- Des actions concrètes :
  - HEP Software Foundation (HSF)
  - Software Institute for Data-Intensive Sciences (SIDIS)
  - Création d'une revue scientifique « Computing and software for big Science » (Springer)
  - IRIS-HEP (Projet NSF U.S.A.)
  - Projet international Data Organization, Management and Access (DOMA)

- The 2020 update of the EU strategy for particle physics



D. Large-scale data-intensive software and computing infrastructures are an essential ingredient to particle physics research programmes. The community faces major challenges in this area, notably with a view to the HL-LHC. As a result, the software and computing models used in particle physics research must evolve to meet the future needs of the field. ***The community must vigorously pursue common, coordinated R&D efforts in collaboration with other fields of science and industry, to develop software and computing infrastructures that exploit recent advances in information technology and data science. Further development of internal policies on open data and data preservation should be encouraged, and an adequate level of resources invested in their implementation.***



arXiv.org > physics > arXiv:1712.06982

Search... All fields Search

Help | Advanced Search

Physics > Computational Physics

A Roadmap for HEP Software and Computing R&D for the 2020s

Johannes Albrecht, Antonio Augusto Alves Jr, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, John Apostolakis, Makoto Asai, Luca Atzori, Marian Babik, Giuseppe Bagliesi, Marilena Bandieramonte, Sunanda Banerjee, Martin Barisits, Lothar A.T. Bauerdick, Stefano Belforte, Douglas Benjamin, Catrin Bernius, Wahid Bhimji, Riccardo Maria Bianchi, Ian Bird, Catherine Biscarat, Jakob Blomer, Kenneth Bloom, Tommaso Boccali, Brian Bockelman, Tomasz Bold, Daniele Bonacorsi, Antonio Boveia, Concezio Bozzi, Marko Bracko, David Britton, Andy Buckley, Predrag Buncic, Paolo Calafiura, Simone Campana, Philippe Canal, Luca Canali, Gianpaolo Carlino, Nuno Castro, Marco Cattaneo, Gianluca Cerminara, Javier Cervantes Villanueva, Philip Chang, John Chapman, Gang Chen, Taylor Childers, Peter Clarke, Marco Clemencic, Eric Cogneras, Jeremy Coles, Ian Collier, David Colling, Gloria Corti, Gabriele Cosmo, Davide Costanzo, Ben Couturier, Kyle Cranmer, Jack Cranshaw, Leonardo Cristella, David Crooks, Sabine Crépé-Renaudin, Robert Currie, Sünje Dallmeier-Tiessen, Kaushik De, Michel De Cian, Albert De Roeck, Antonio Delgado Peris, Frédéric Derue, Alessandro Di Girolamo, Salvatore Di Guida, Gancho Dimitrov, Caterina Doglioni, Andrea Dotti, Dirk Duellmann, Laurent Duflot, Dave Dykstra, Katarzyna Dziedziniewicz-Wojcik, Agnieszka Dziurda, Ulrik Egede, Peter Elmer, Johannes Elmsheuser, V. Daniel Elvira, Giulio Eulisse, Steven Farrell, Torben Ferber, Andrej Filipcic, Ian Fisk, Conor Fitzpatrick, José Flix, Andrea Formica, Alessandra Forti, Giovanni Franzoni, James Frost, Stu Fuess, Frank Gaede, Gerardo Ganis, Robert Gardner, Vincent Garonne, Andreas Gellrich et al. (210 additional authors not shown)

(Submitted on 18 Dec 2017 (v1), last revised 19 Dec 2018 (this version, v5))

Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

Download:
- PDF
- Other formats

Current browse context:
physics.comp-ph
< prev | next >
new | recent | 1712

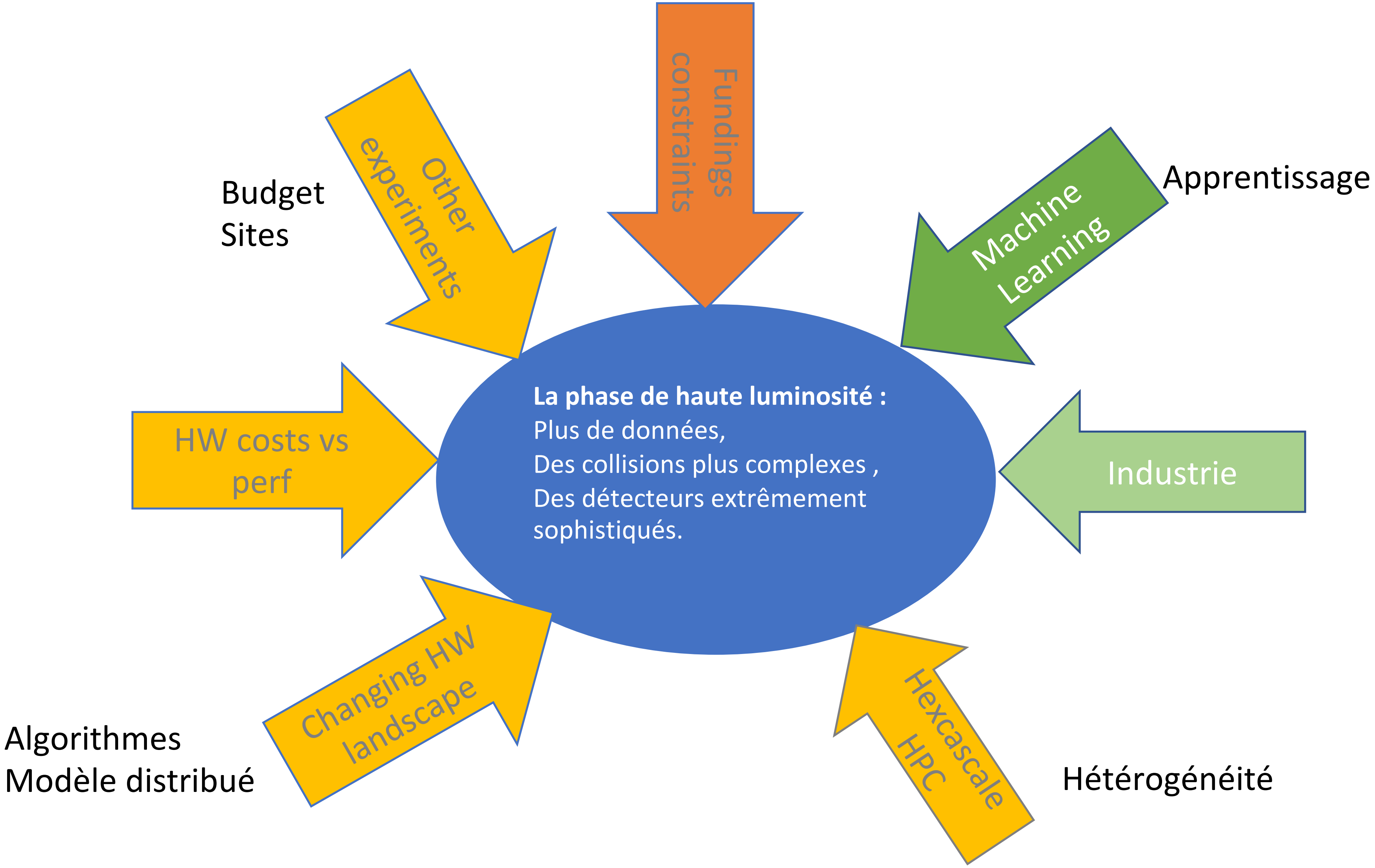
Change to browse by:
hep-ex
physics

References & Citations
- INSPIRE HEP (refers to | cited by)
- NASA ADS

Export citation
Google Scholar

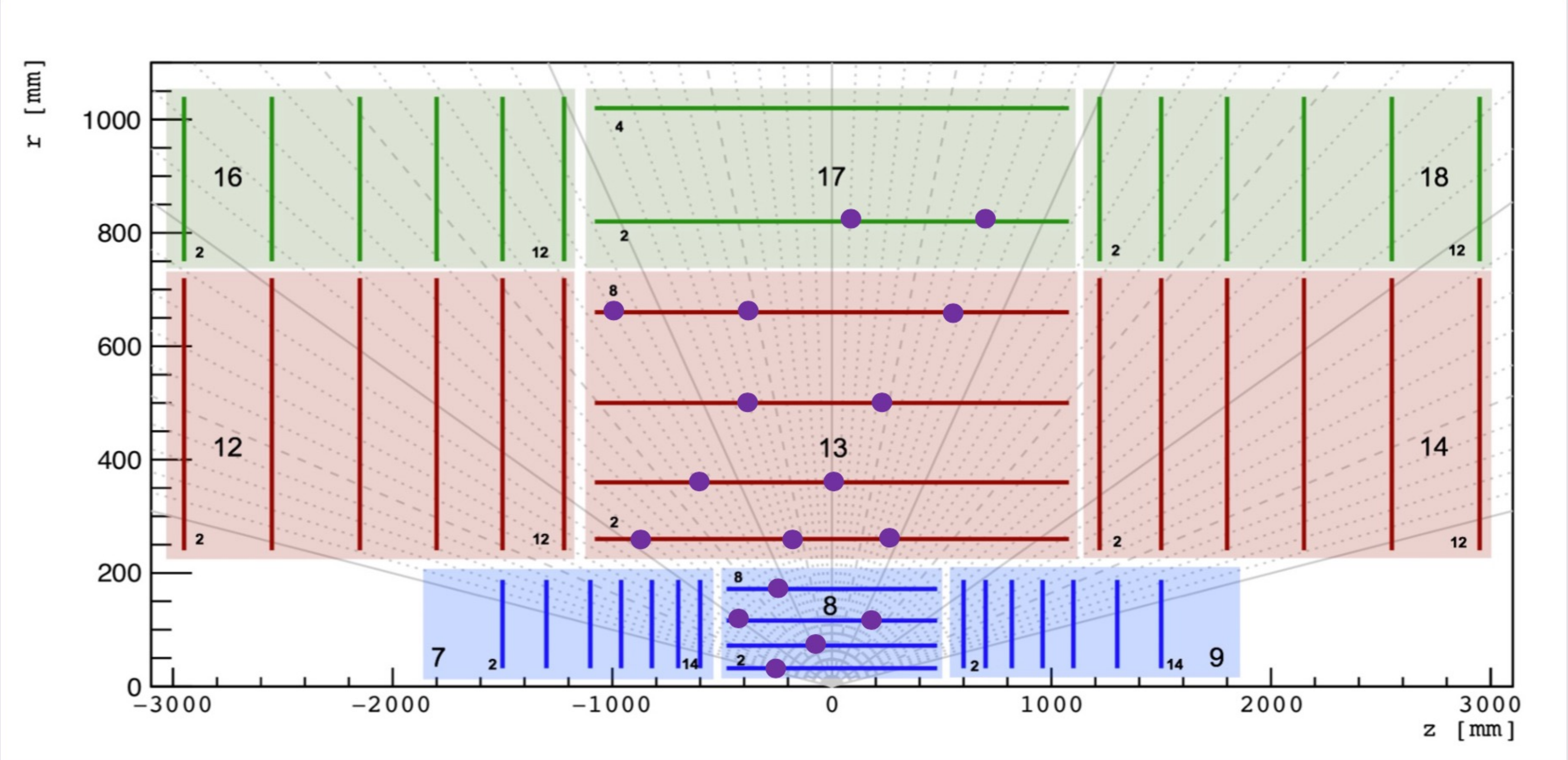
Bookmark

Contraintes et opportunités



The TrackML dataset

- **Generation (Pythia8):** 1000 $t\bar{t}$ events from *pp* collisions
  - $\sqrt{s} = 14\ TeV, \mu = 200$ (HL-LHC conditions), pile-up modeling using A3 tune
- **Simulation:** *Generic detector* simulated with fast simulation of ACTS framework

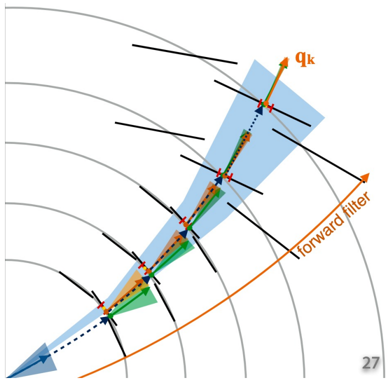


18 728 silicon modules

hit

Le défi de HL-LHC pour la reconstruction de traces

- Reconstruction actuellement résolu par des algorithmes basés sur des filtres de Kalman
- Estimation des paramètres de la trajectoire hélicoïdal
- Très bonne performance et optimisé depuis des années
- La partie la plus coûteuse **en complexité de calcul (donc en ressources CPU)** dans la reconstruction d'un événement
- La combinatoire va exploser avec HL-LHC (pileup ~20 => pileup ~200)
- Va entrainer une augmentation très importante du volume et de la complexité des données
- Les algorithmes actuels ne suffiront pas



1. propagate p_{k-1} and its covariance C_{k-1} :

$$q_{k|k-1} = f_{k|k-1}(q_{k-1|k-1})$$
$$C_{k|k-1} = F_{k|k-1} C_{k-1|k-1} F^{\mathrm{T}}_{k|k-1} + Q_k$$

with Q_k ~ noise term (M.S.)

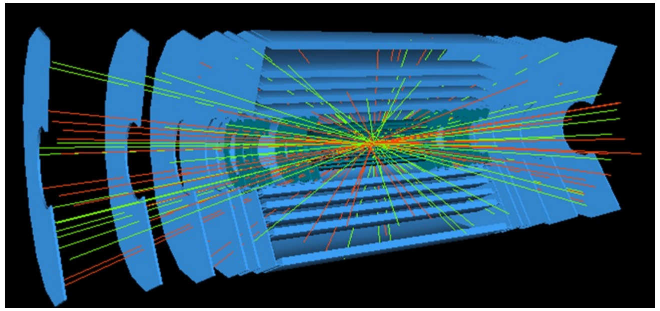
2. update prediction to get $q_{k|k}$ and $C_{k|k}$:

$$q_{k|k} = q_{k|k-1} + K_k[m_k - h_k(q_{k|k-1})]$$
$$C_{k|k} = (I - K_k H_k) C_{k|k-1}$$

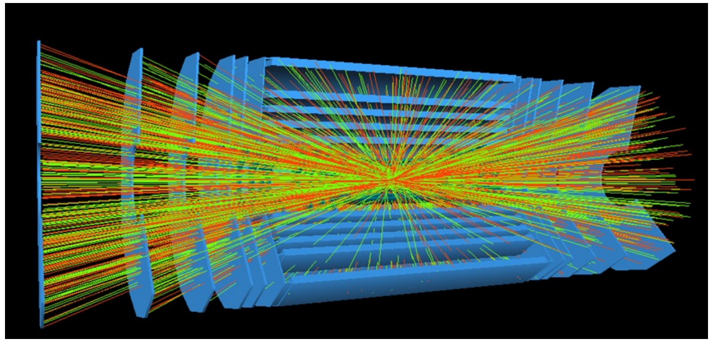
with K_k ~ gain matrix :

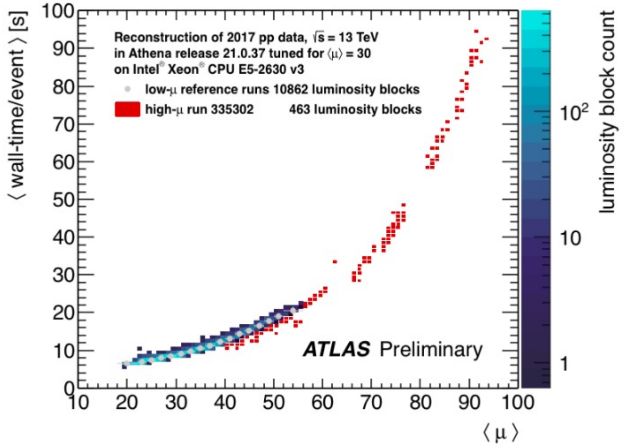
$$K_k = C_{k|k-1} H_k^{\mathrm{T}} (G_k + H_k C_{k|k-1} H_k^{\mathrm{T}})^{-1}$$

LHC (pileup ~20)



HL-LHC (pileup ~200)



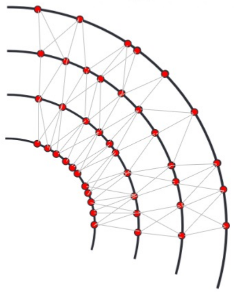


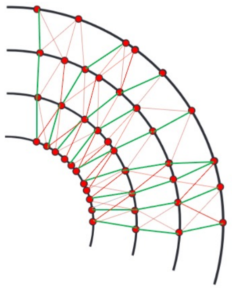
Apprentissage des patterns de traces avec des GNNs

Input graph

Nodes features: (r, φ, z)

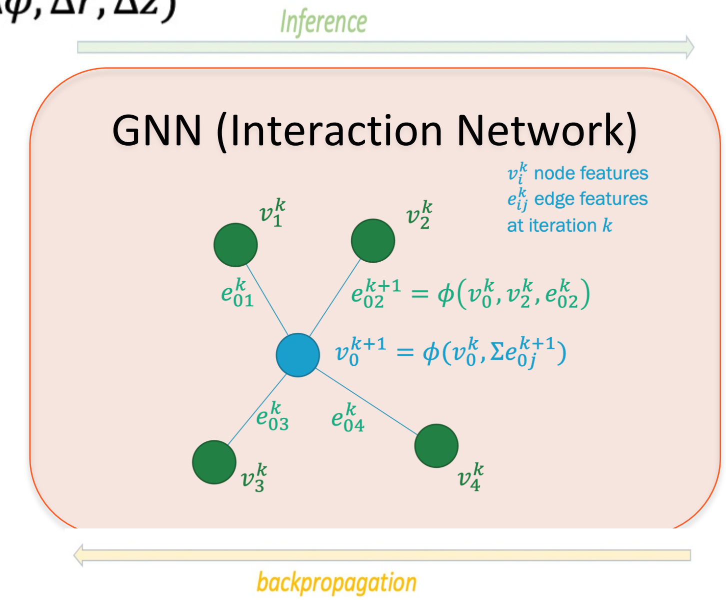
Edges features: $(\Delta\eta, \Delta\varphi, \Delta r, \Delta z)$



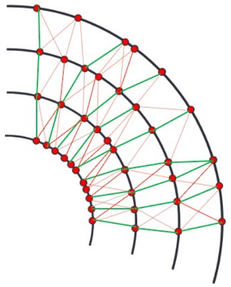


Target graph

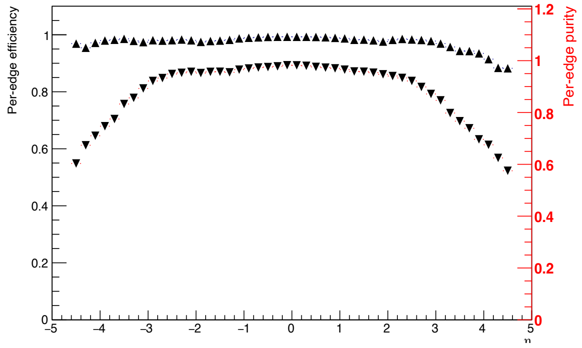
Target: $\begin{cases} \text{label} = 1 \text{ if } \textit{true} \text{ edge} \\ \text{label} = 0 \text{ if } \textit{fake} \text{ edge} \end{cases}$

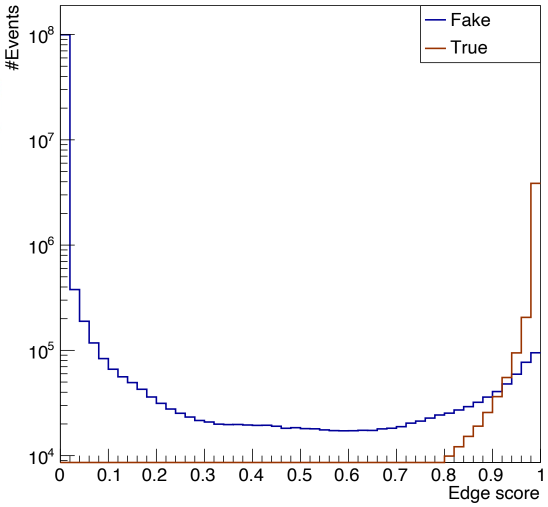


Predict graph



Edges scores





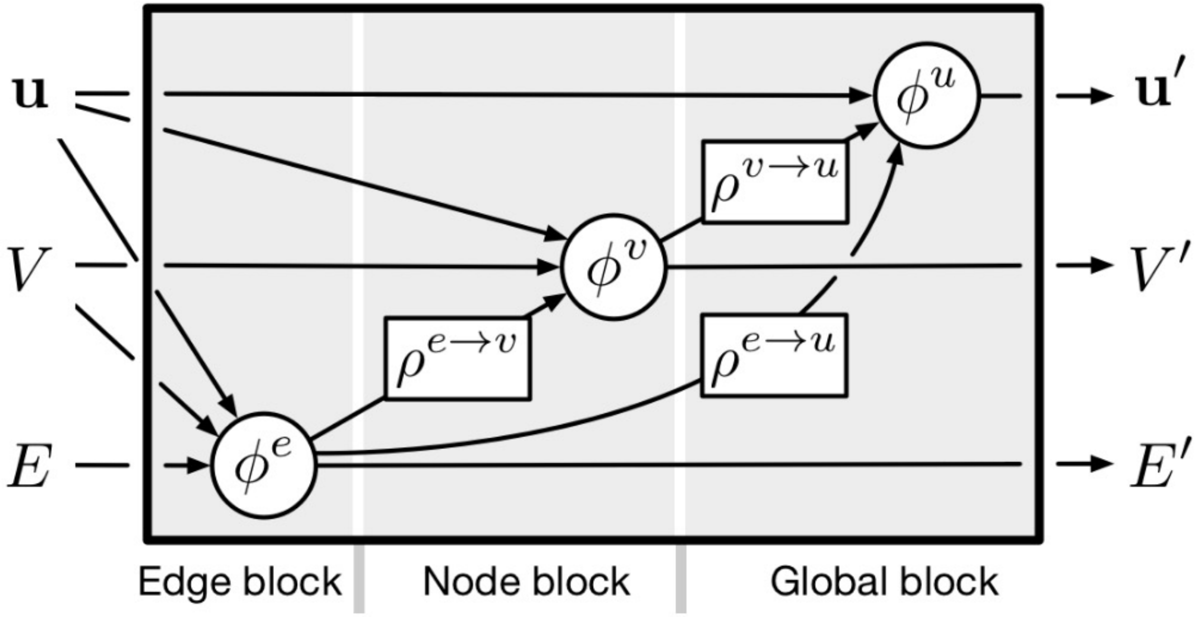
Graph (neural) networks

Oct 2018 (link)

Relational inductive biases, deep learning, and graph networks

Peter W. Battaglia[1*], Jessica B. Hamrick[1], Victor Bapst[1], Alvaro Sanchez-Gonzalez[1], Vinicius Zambaldi[1], Mateusz Malinowski[1], Andrea Tacchetti[1], David Raposo[1], Adam Santoro[1], Ryan Faulkner[1], Caglar Gulcehre[1], Francis Song[1], Andrew Ballard[1], Justin Gilmer[2], George Dahl[2], Ashish Vaswani[2], Kelsey Allen[3], Charles Nash[4], Victoria Langston[1], Chris Dyer[1], Nicolas Heess[1], Daan Wierstra[1], Pushmeet Kohli[1], Matt Botvinick[1], Oriol Vinyals[1], Yujia Li[1], Razvan Pascanu[1]

[1]DeepMind; [2]Google Brain; [3]MIT; [4]University of Edinburgh



(a) Full GN block

A GN block contains three "update" functions, ϕ, and three "aggregation" functions, ρ,

$$
\begin{aligned}
\mathbf{e}'_k &= \phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right) & \bar{\mathbf{e}}'_i &= \rho^{e\to v}\left(E'_i\right) \\
\mathbf{v}'_i &= \phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right) & \bar{\mathbf{e}}' &= \rho^{e\to u}\left(E'\right) \\
\mathbf{u}' &= \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}\right) & \bar{\mathbf{v}}' &= \rho^{v\to u}\left(V'\right)
\end{aligned}
$$

where $E'_i = \{(\mathbf{e}'_k, r_k, s_k)\}_{r_k=i,\ k=1:N^e}$, $V' = \{\mathbf{v}'_i\}_{i=1:N^v}$, and $E' = \bigcup_i E'_i = \{(\mathbf{e}'_k, r_k, s_k)\}_{k=1:N^e}$.

Graph (neural) networks

Oct 2018 (link)

Relational inductive biases, deep learning, and graph networks

Peter W. Battaglia[1]*, Jessica B. Hamrick[1], Victor Bapst[1], Alvaro Sanchez-Gonzalez[1], Vinicius Zambaldi[1], Mateusz Malinowski[1], Andrea Tacchetti[1], David Raposo[1], Adam Santoro[1], Ryan Faulkner[1], Caglar Gulcehre[1], Francis Song[1], Andrew Ballard[1], Justin Gilmer[2], George Dahl[2], Ashish Vaswani[2], Kelsey Allen[3], Charles Nash[4], Victoria Langston[1], Chris Dyer[1], Nicolas Heess[1], Daan Wierstra[1], Pushmeet Kohli[1], Matt Botvinick[1], Oriol Vinyals[1], Yujia Li[1], Razvan Pascanu[1]

[1]DeepMind; [2]Google Brain; [3]MIT; [4]University of Edinburgh

Interaction networks

Interaction Networks (Battaglia et al., 2016; Watters et al., 2017) and the Neural Physics Engine Chang et al. (2017) use a full GN but for the absence of the global to update the edge properties:

$$
\begin{aligned}
\phi^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}\right) &:= f^e\left(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}\right) = \mathrm{NN}_e\left([\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}]\right) \\
\phi^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right) &:= f^v\left(\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}\right) = \mathrm{NN}_v\left([\bar{\mathbf{e}}'_i, \mathbf{v}_i, \mathbf{u}]\right) \\
\rho^{e\to v}\left(E'_i\right) &:= \qquad = \sum_{\{k:\, r_k = i\}} \mathbf{e}'_k
\end{aligned}
$$