Deep Learning in High Energy Physics

Amir Farbin

UNIVERSITY OF TEXAS ARLINGTON

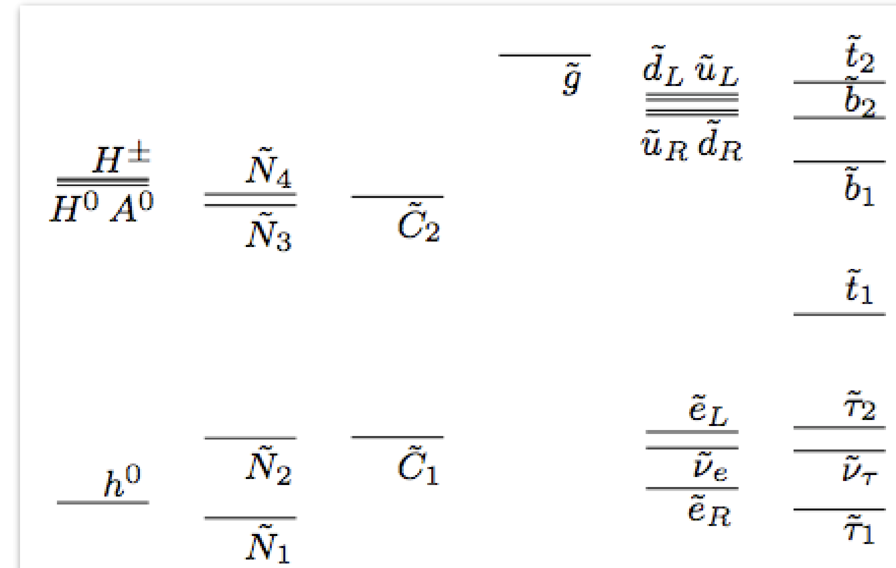
ATLAS EXPERIMENT

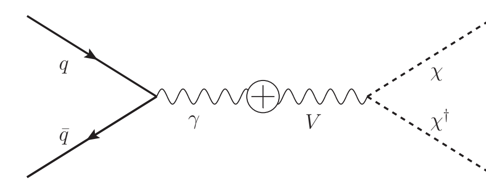
Plan

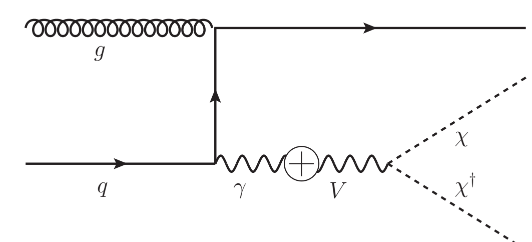
- One slide on Dark Matter in HEP
- One slide introducing Deep Learning
- Deep Learning in HEP: The Big Picture, the Problems, and Examples
- Many interesting Deep Learning topics in HEP I will not mention… e.g. Adversarial Techniques.
- Focus on Images:
  - Three techniques: Feature Learning, Semi-supervised Learning, Generative Model
  - Calorimetry with Deep Learning
  - Jet Physics with Deep Learning

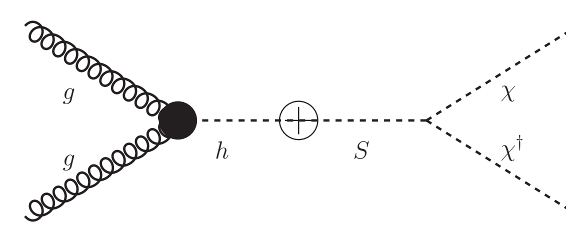
Dark Matter in HEP

- ***Solutions to the Hierarchy Problem***
  - Often evoke some conserved quantity
    - e.g. R-parity in SUSY
  - Leads to a stable particle ~ Dark Matter candidate
- More ***empirical models***: add relevant operators to the Standard Model Lagrangian
- ***Collider-based Experiments***
  - Produce new heavy particles
  - Decay to the DM particle
  - Leave Missing energy signature
- ***Beam Dump Experiments***
  - Produce fast moving DM in the target, look for interaction in detector





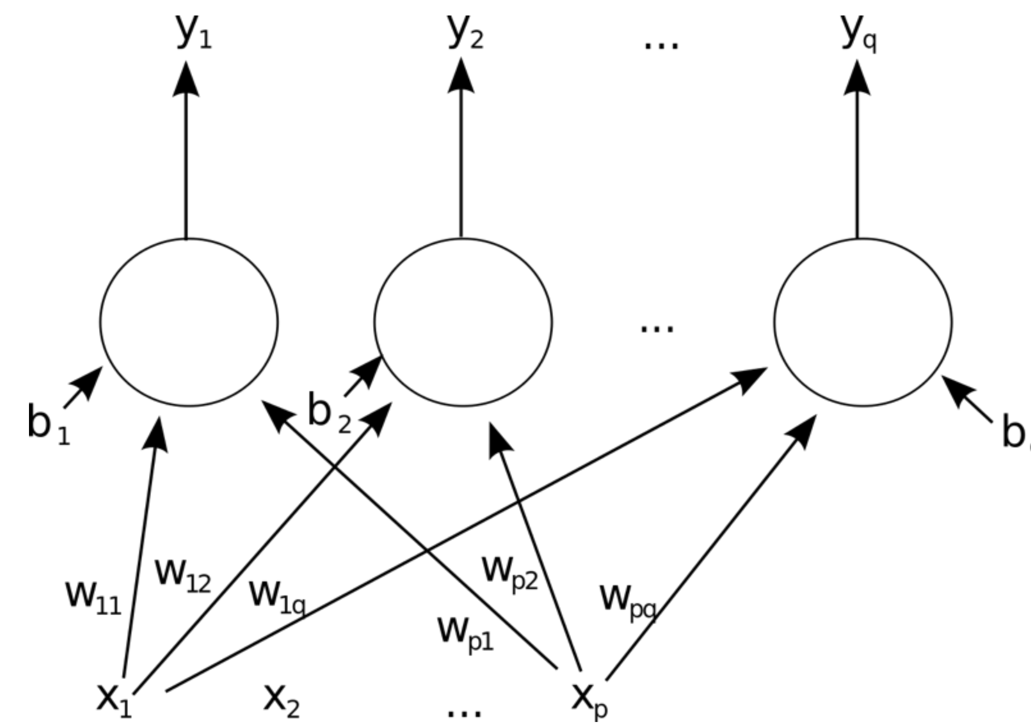


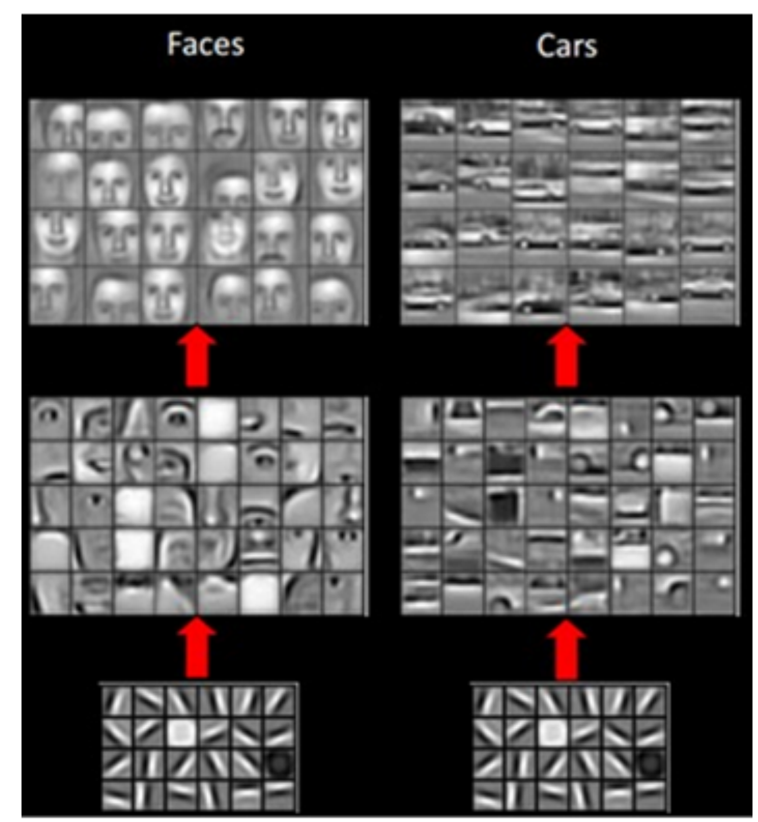


$$pp(n) \to X^* \to \bar{\chi}\chi \ (\text{or } \chi^\dagger\chi)$$

Artificial Neural Networks

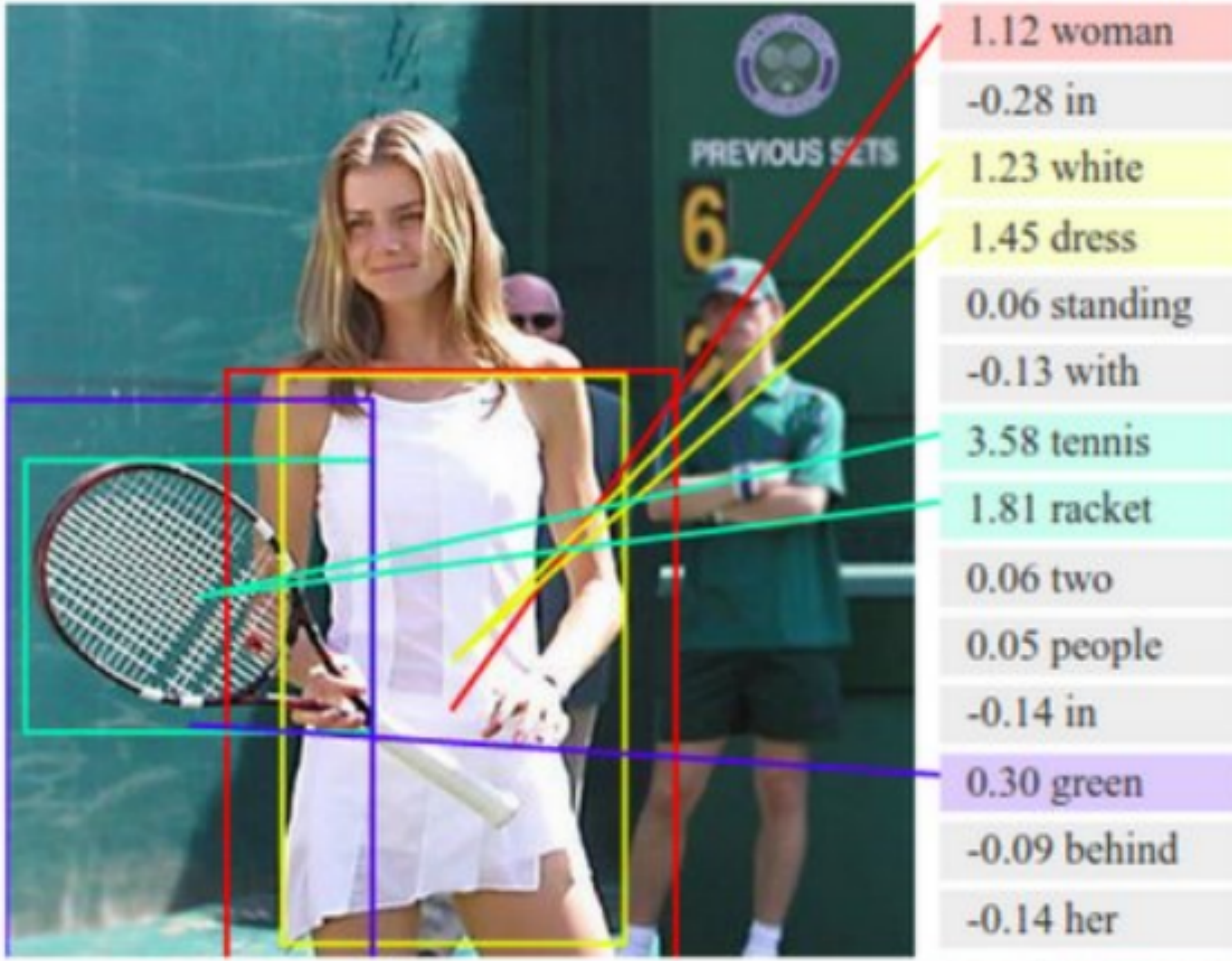
- ***Biologically inspired computation***, (first attempts in 1943)
  - ***Probabilistic Inference***: e.g. signal vs background
  - ***Universal Computation Theorem*** (1989)
- Multi-layer (***Deep***) Neutral Networks:
  - Not a new idea (1965), just impractical to train. ***Vanishing Gradient problem*** (1991)
  - Solutions:
    - New techniques: e.g. better activation or layer-wise training
    - ***More training***: big training datasets and lots of computation … ***big data and GPUs***
  - ***Deep Learning Renaissance***. First DNN in HEP (2014).
  - ***Amazing Feats***: Audio/Image/Video recognition, captioning, and generation. Text (sentiment) analysis. Language Translation. Game playing agents.
  - ***Rich field***: Variety of architectures, techniques, and applications.





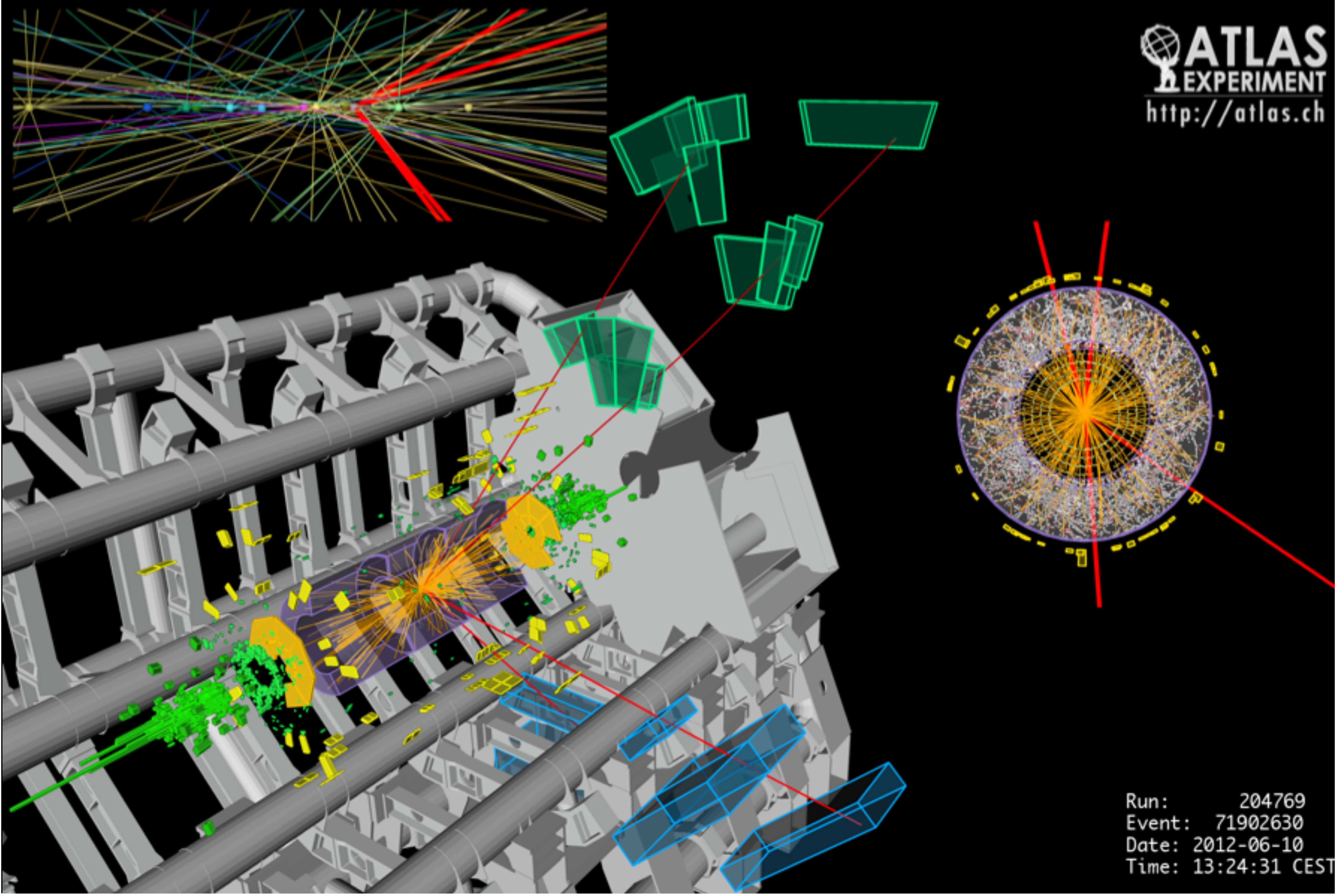
Images from Wikipedia

DL in HEP?



PREVIOUS SETS
6
1.12 woman
-0.28 in
1.23 white
1.45 dress
0.06 standing
-0.13 with
3.58 tennis
1.81 racket
0.06 two
0.05 people
-0.14 in
0.30 green
-0.09 behind
-0.14 her

$$H \to ZZ \to 4l$$

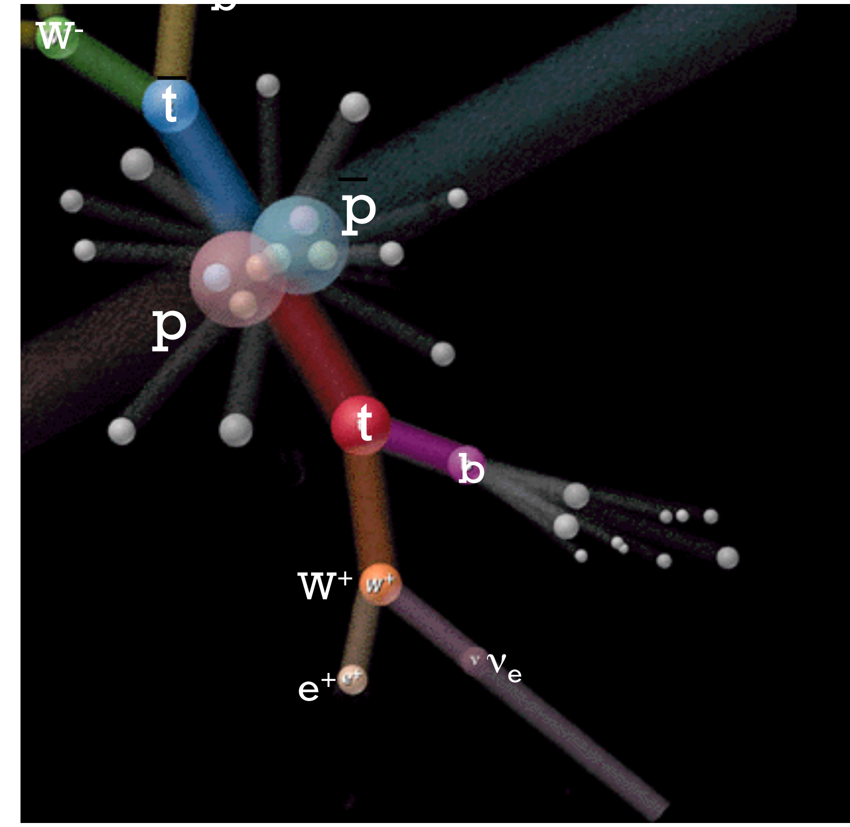


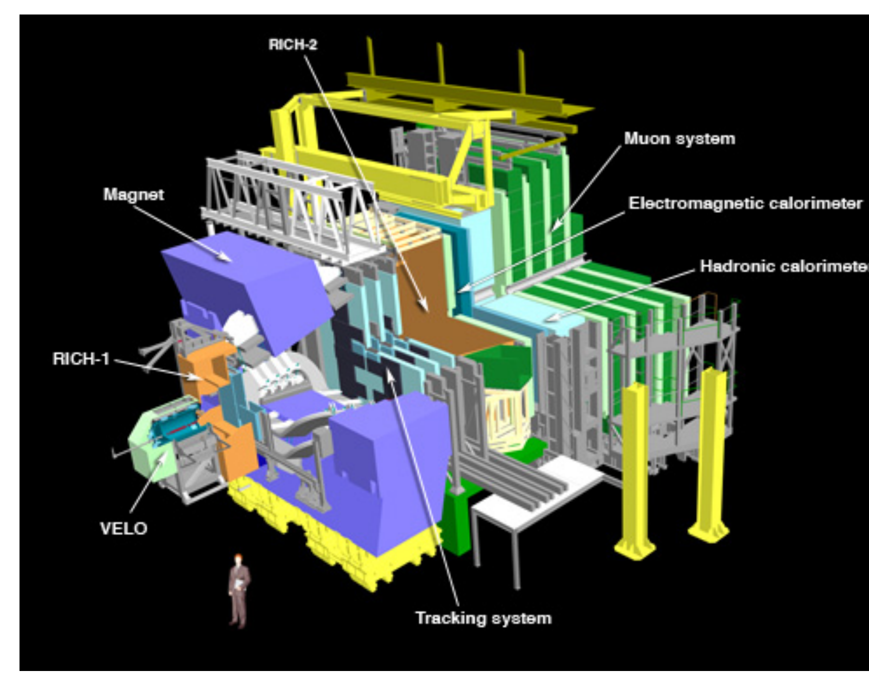
ATLAS
EXPERIMENT
http://atlas.ch
Run: 204769
Event: 71902630
Date: 2012-06-10
Time: 13:24:31 CEST

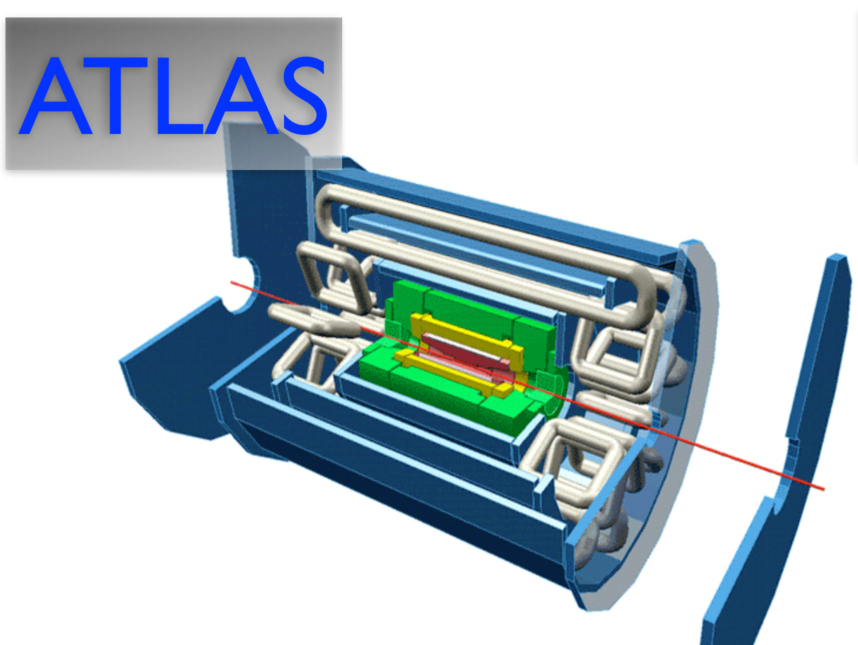
DL in HEP
The Big Picture…

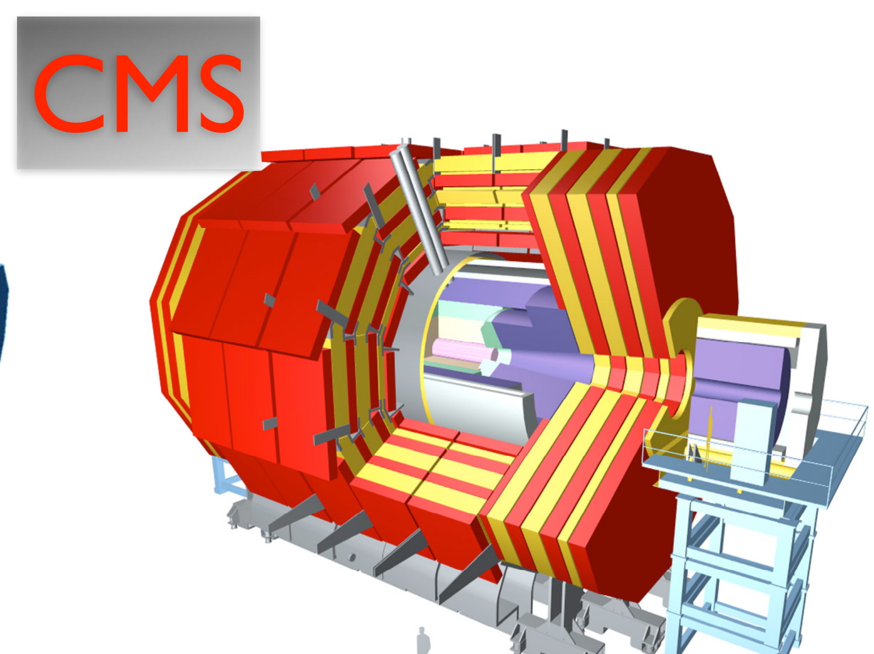
HEP Experiments

- 5 technical components to HEP experiment:
  - ***Accelerator***: e.g. LHC collisions creating quickly decaying heavy particles. Extremely high rate: 40 * O(50) Million collisions/sec.
  - ***Detector***: a big camera. ~ e.g. LHC 1.5 MB/event (60 TB/s)
    - Pictures of long-lived decay products of short lived heavy/ interesting particles.
    - Sub-detectors parts: Tracking, Calorimeters, Muon system, Particle ID (e.g. Cherenkov, Time of Flight)
  - ***DAQ/Trigger:*** Hardware/software
  - ***Software:*** Reconstruction (Raw data -> particle “features”) / Analysis (particles -> “physics”)
  - ***Computing:*** GRID Monarch Model “Cloud” Computing/Data Management (software/hardware)

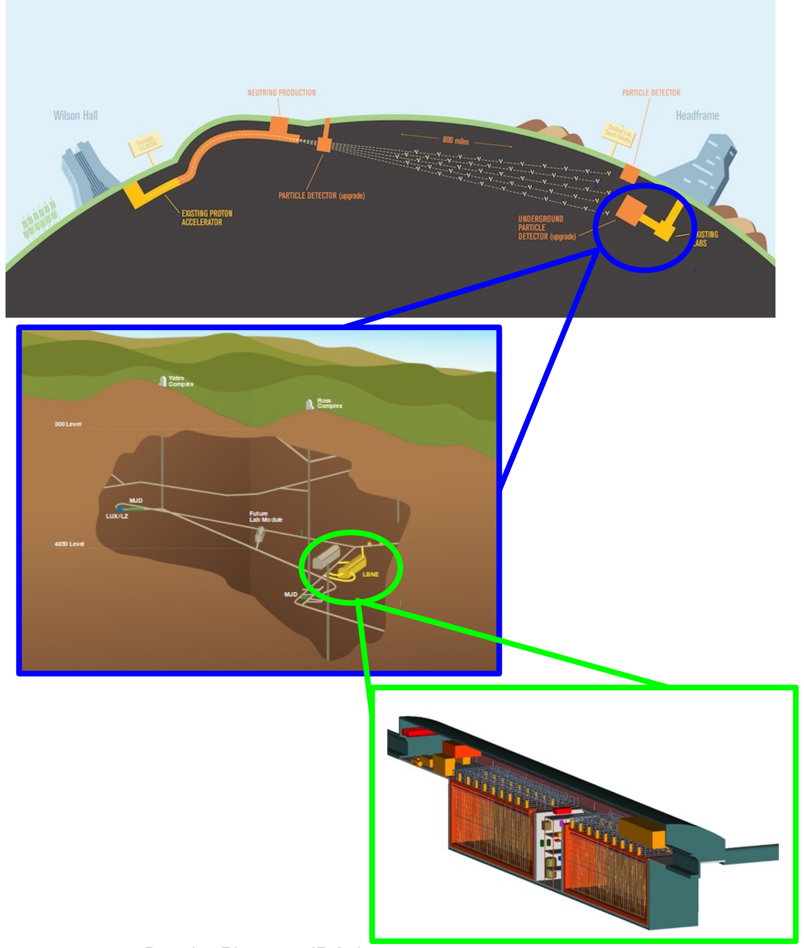




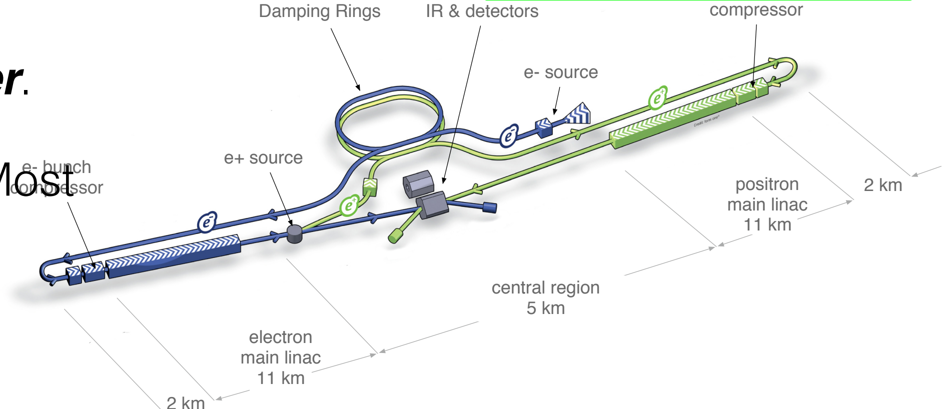




Frontiers



- ***Energy Frontier***: ***Large Hadron Collider (LHC)*** at 13 TeV now, ***High Luminosity (HL)-LHC*** by 2025, perhaps 33 TeV LHC or 100 TeV Chinese machine in a couple of decades.
  - Having found Higgs, moving to studying the SM ***Higgs*** find new Higgses
  - Test ***naturalness*** (Was the Universe and accident?) by searching for New Physics like Supersymmetry that keeps Higgs light without 1 part in 10 fine-tuning of parameters.
  - Find ***Dark Matter*** (reasons to think related to naturalness)
- ***Intensity Frontier***:
  - ***B Factories***: upcoming SuperKEKB/SuperBelle
  - ***Neutrino Beam Experiments***:
    - Series of current and upcoming experiments: Nova, MicroBooNE, SBND, ICURUS
    - ***US's flagship experiment*** in next decade: ***Long Baseline Neutrino Facility (LBNF)/Deep Underground Neutrino Experiment (DUNE) at Intensity Frontier***
  - Measure properties of ***b-quarks*** and ***neutrinos*** (newly discovered mass)… search for ***matter/anti-matter asymmetry***.
  - Auxiliary Physics: Study ***Supernova***. Search for ***Proton Decay*** and ***Dark Matter***.
- ***Precision Frontier***: ***International Linear Collider (ILC)***, hopefully in next decade. Most energetic e e machine.
  - ***Precision studies*** of ***Higgs*** and hopefully ***new particles*** found at LHC.



Why go Deep?

- ***Better Algorithms***
  - DNN-based classification/regression generally **out perform** hand crafted algorithms.
  - In some cases, it may provide a ***solution*** where ***algorithm approach doesn't exist or fails***.
  - ***Unsupervised learning***: make sense of complicated data that we don't understand or expect.
- ***Easier Algorithm Development***: ***Feature Learning*** instead of *Feature Engineering*
  - Reduce time physicists spend writing developing algorithms, ***saving time and cost***. (e.g. ATLAS > $250M spent software)
  - Quickly perform performance ***optimization*** or ***systematic studies***.
- ***Faster Algorithms***
  - After training, DNN inference is often *faster* than sophisticated algorithmic approach.
  - DNN can ***encapsulate expensive computations***, e.g. Matrix Element Method.
  - ***Generative Models*** enable fast simulations.
  - ***Already parallelized*** and optimized for GPUs/HPCs.
  - ***Neuromorphic*** processors.

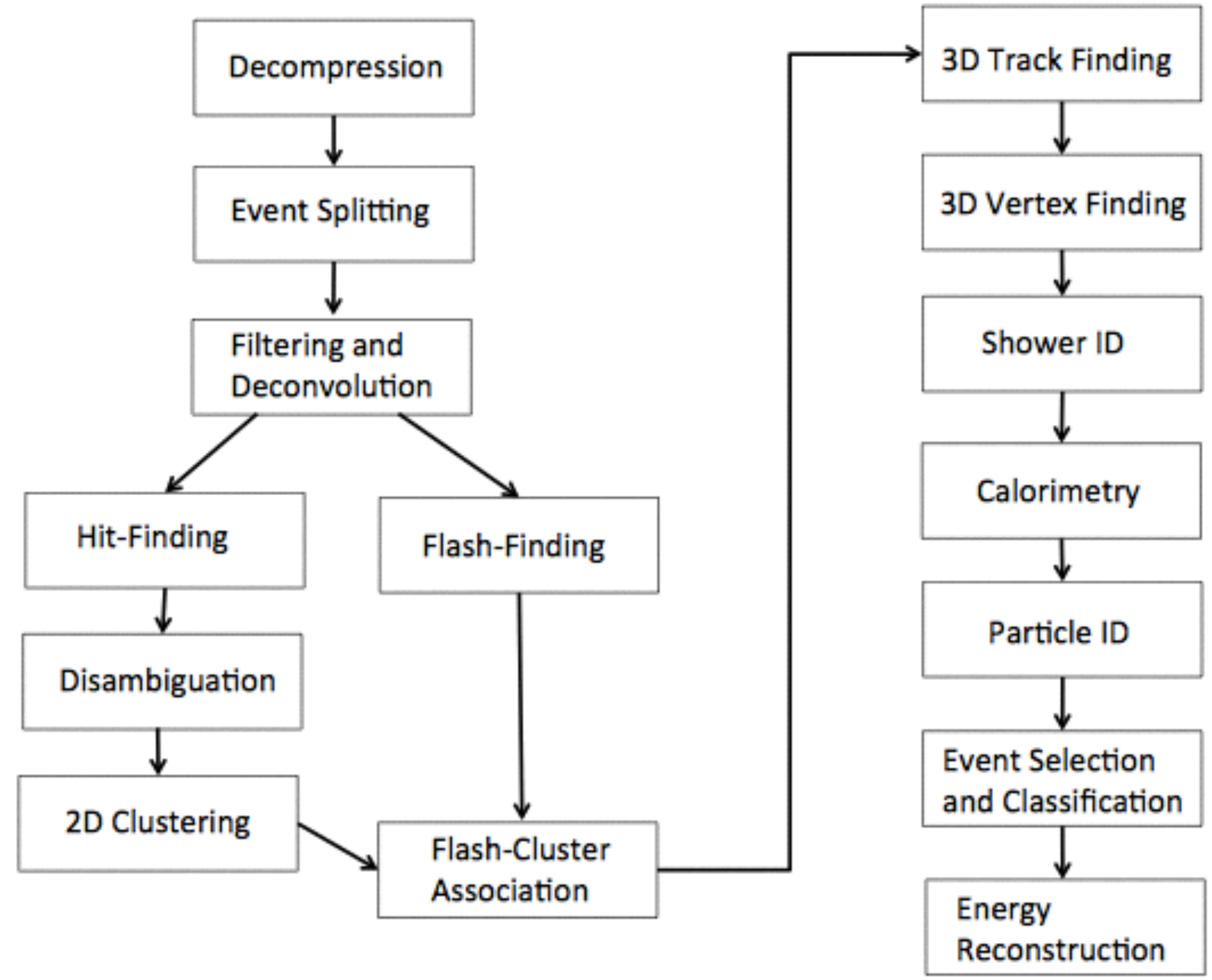
HEP Problems

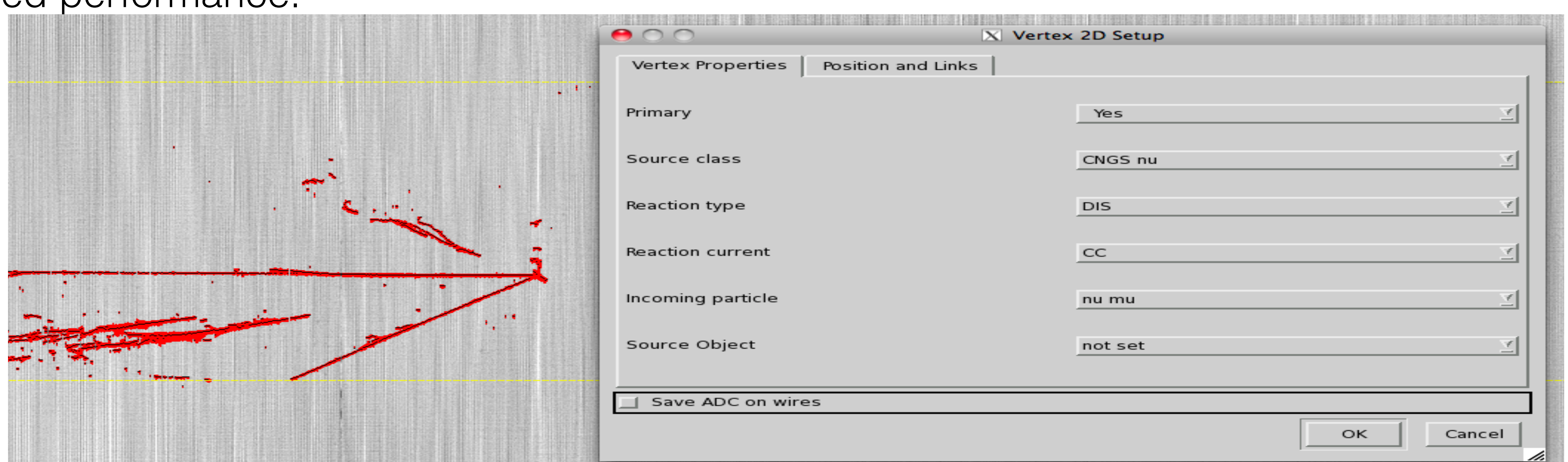
Where is ML needed?

- Traditionally ML Techniques in HEP
    - Applied to Particle/Object Identification
    - Signal/Background separation
    - Here, ML maximizes reach of existing data/detector… equivalent to additional integral luminosity.
    - There is lots of interesting work here… and potential for big impact.
- Now we hope ML can help address looming computing problems of the next decade:
    - ***Reconstruction***
        1. Intensity Frontier- ***LArTPC*** Automatic Algorithmic ***Reconstruction*** still struggling
        2. Energy Frontier- ***HL-LHC Tracking***- Pattern Recognition blows up due to combinatorics
    - ***Simulation***
        3. LHC Calorimetry- Large Fraction of ATLAS CPU goes into ***shower simulation***.

LArTPC Reconstruction

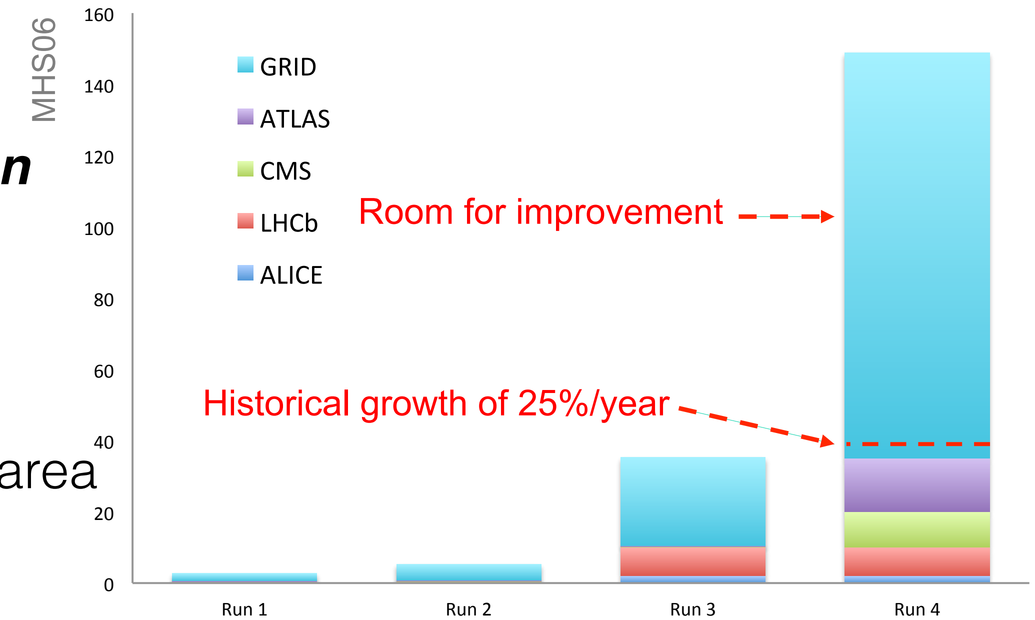
- Neutrino Physics has a long history of *hand scans*.
  - QScan: ICARUS user assisted reconstruction.
- Full automatic reconstruction has yet to be demonstrated.
  - LArSoft project:
    - art framework + LArTPC reconstruction algorithm
    - started in ArgoNeuT and contributed to/used by many experiments.
  - Full neutrino reconstruction is still far from expected performance.



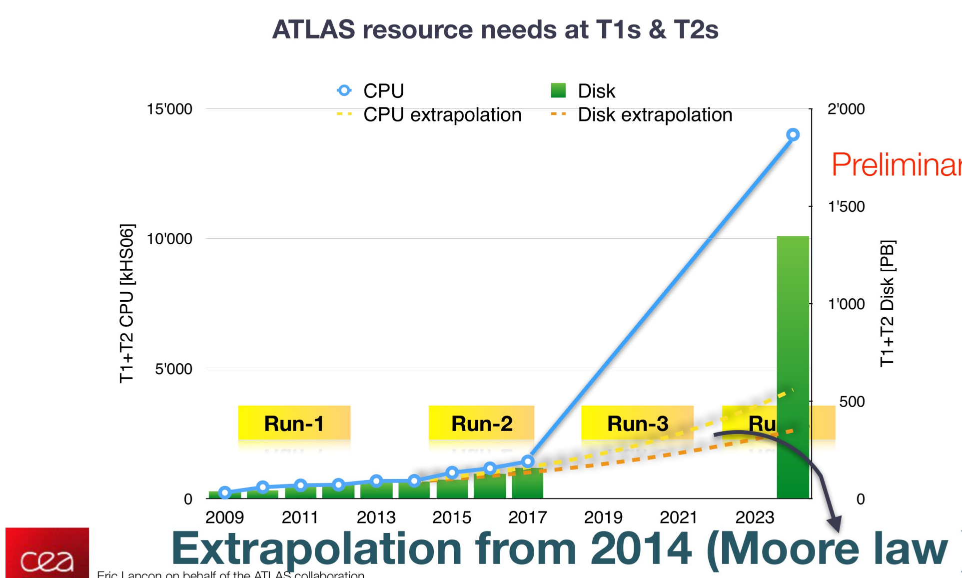


Computing

- ***Computing*** is perhaps the biggest challenge for the HL-LHC
    - ***Higher Granularity*** = larger events.
    - ***O(200) proton collision / crossing: tracking pattern recognition combinatorics becomes untenable***.
    - O(100) times data = multi ***exabyte datasets***.
    - ***Moore's law has stalled***: Cost of adding more transistors/silicon area no longer decreasing.
    - Preliminary estimates of ***HL-LHC computing budget many times larger than LHC***.



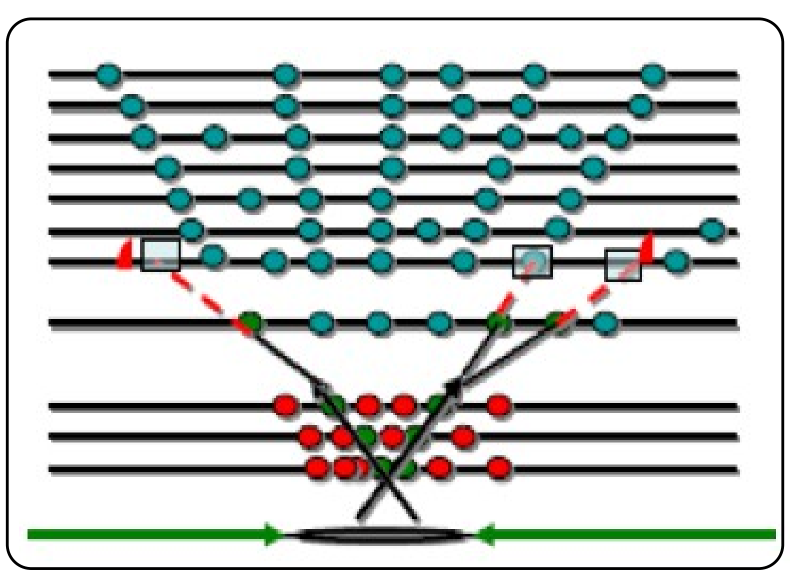
- ***Solutions***:
    - ***Leverage opportunistic resources and HPC*** (most computation power in highly parallel processors).
    - ***Highly parallel processors*** (e.g. GPUs) are already > 10x CPUs for certain computations.
        - Trend is away from x86 towards ***specialized hardware*** (e.g. GPUs, Mics, FPGAs, Custom DL Chips)
        - Unfortunately parallelization (i.e. Multi-core/GPU) has been extremely difficult for HEP.

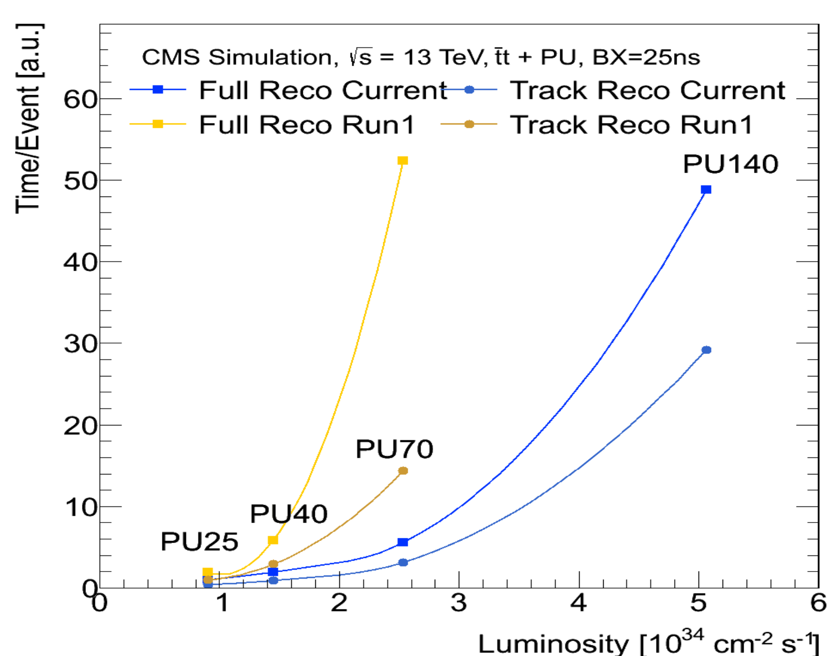


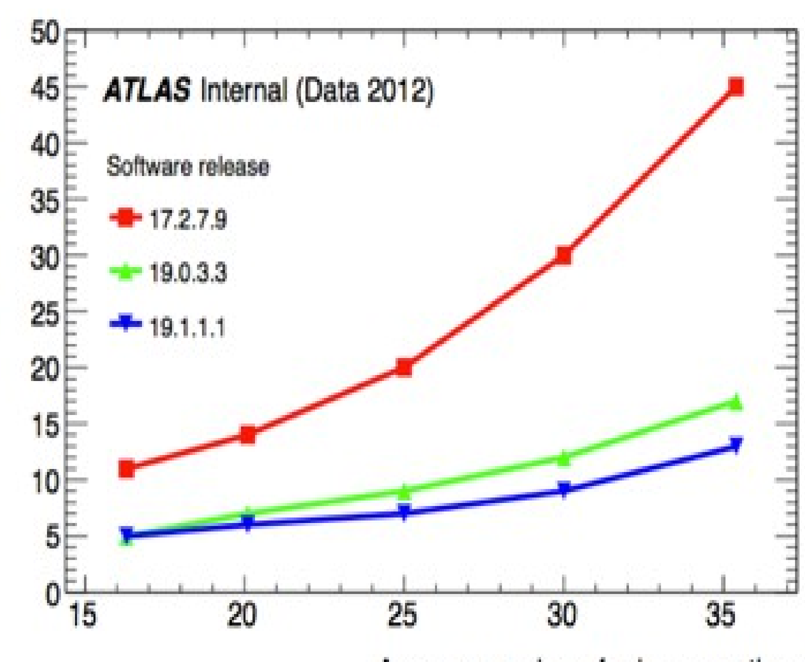
Plots from here.

HL-LHC Tracking

- ***Tracking steps:*** hit prep, seeding, pattern recognition, track fitting, track cleaning
  - Highly optimized already for offline reconstruction for Run 2
    - ~30-50 proton collisions per beam crossing
    - 1 kHz data stream, processed offline.
- ***HL-LHC***: ~ 200 proton collisions per beam crossing
  - combinatorics cause pattern recognition time to grow exponentially
  - Busy environment requires tracking at 40 MHz for trigger
- Need Pattern Recognition that scales better with number of hits. Deep Learning?
- Again an obstacle to applying deep learning techniques is accessibility to the data.
- ***Tracking ML*** (David Rousseau, Andreas Salzberger, …, AF): Hoping to have ML community develop solutions, mirroring the HiggsML Challenge.
  - ACTS: Standalone version of ATLAS Tracking Simulation/Reconstruction developed for this challenge.







Data Analysis

- Objectives:
  - ***Searches*** (hypothesis testing): Likelihood Ratio Test (Neyman-Pearson lemma)
  - ***Measurements***: Maximum Likelihood Estimate

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

  - ***Limits*** (confidence intervals): Also based on Likelihood
- ***Likelihood***

$$p(\{x\}|\theta) = \mathrm{Pois}(n|\nu(\theta)) \prod_{e=1}^{n} p(x_e|\theta)$$

  - n Independent Events (e) with Identically Distributed Observables ($\{x\}$)
  - Significant part of Data Analysis is ***approximating the likelihood*** as best as we can.

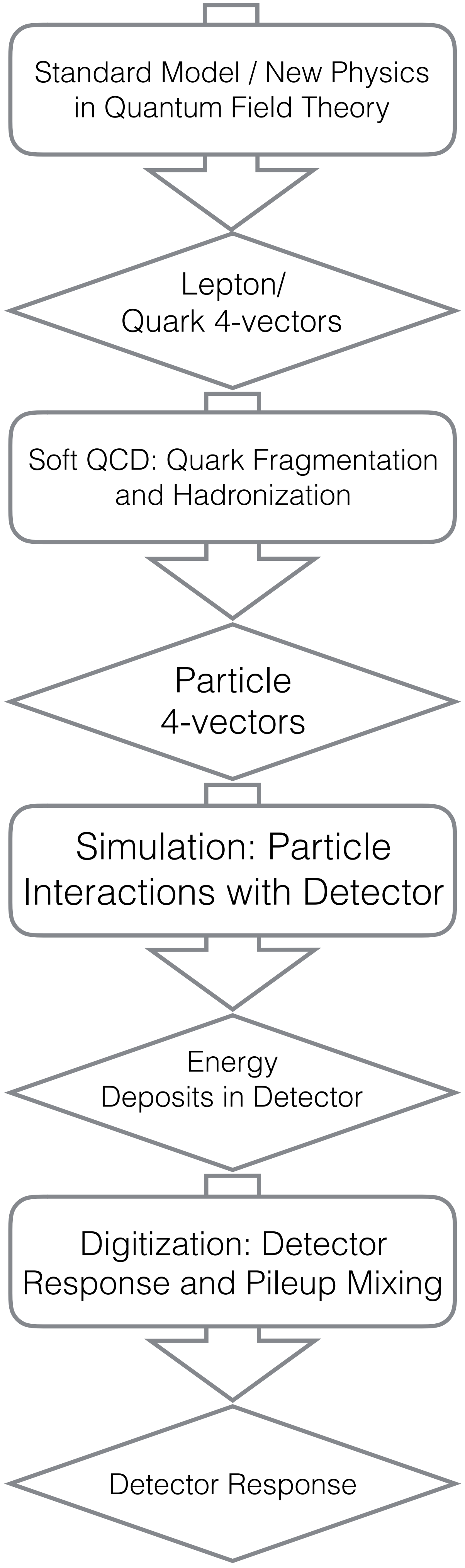
Approximating the Likelihood

- Physics is all about establishing a very precise "model" of the underlying phenomena… so ***we can model our data very well.***

- Enables ***multi-step ab-initio simulations***:

  1. ***Generation***: Standard Model and New Physics are expressed in language of Quantum Field Theory.
     - ➡ Feynman Diagrams simplify perturbative prediction of HEP interactions among the most fundamental particles (leptons, quarks)
  2. ***Hadronization***: Quarks turn to jets of particles via Quantum Chromodynamics (QCD) at energies where theory is too strong to compute perturbatively.
     - ➡ Use semi-empirical models tuned to Data.
  3. ***Simulation***: Particles interact with the Detector via stochastic processes
     - ➡ Use detailed Monte Carlo integration over the "micro-physics"
  4. ***Digitization***: Ultimately the energy deposits lead to electronic signals in the O(100 Million) channels of the detector.
     - ➡ Model using test beam data and calibrations.

- Output is fed through ***same reconstruction as real data***.



Likelihood Approximations

- Need $P(\{x_e\}|\theta)$ of an observed event (e). The better we do, the more sensitive our measurements.
- Steps 2 (Hadronization) and 3 (Simulation) can only be done in the ***forward mode***…
  - ➡ ***cannot evaluate the likelihood.***
- So we simulate a lot of events and use a ***Probability Density Estimator (PDE)***, e.g. a histogram.
  - $\{x_e\}$ = {100M Detector Channels} or even { particle 4-vectors } are too high dimension.
  - Instead we derive $\{x_e\}$ = { small set of physics motivated observables } → ***Lose information.***
    - ***Isolate signal*** dominating regions of $\{x_e\}$ → ***Lose Efficiency.***
    - Sometimes use ***classifiers*** to further reduce dimensionality and improve significance
    - ***Profile the likelihood*** in 1 or 2 (ideally uncorrelated) observables.
- Alternative, try to brute force calculate via ***Matrix Element Method***:

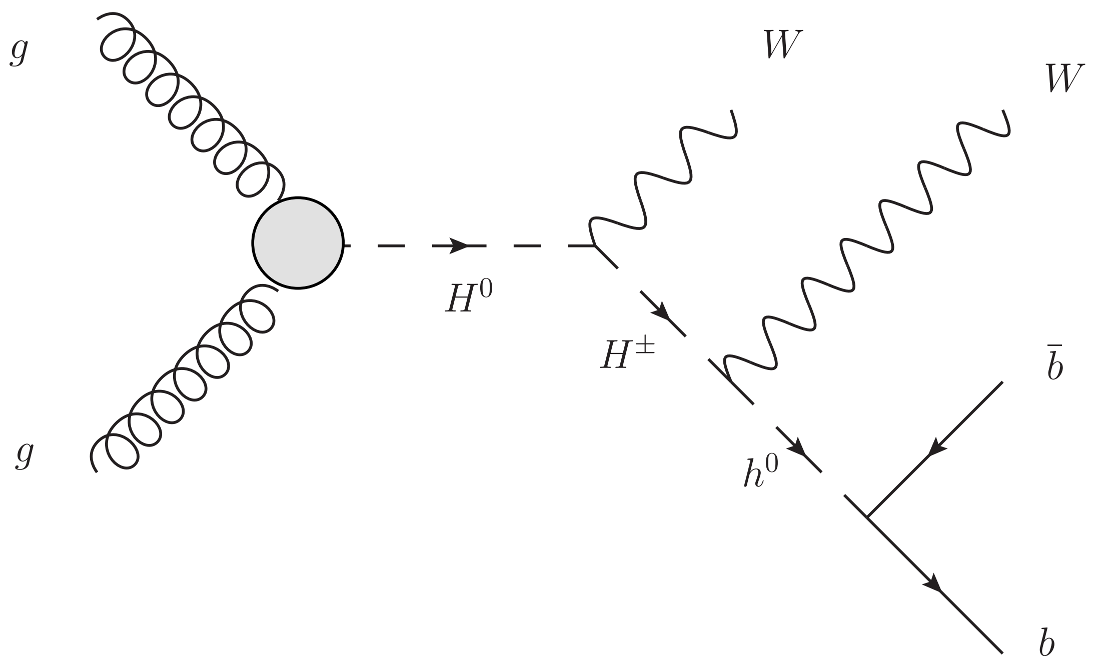
$$\mathcal{P}(\boldsymbol{p}^{vis}|\alpha) = \frac{1}{\sigma_\alpha}\int d\Phi dx_1 dx_2 |M_\alpha(\boldsymbol{p})|^2 W(\boldsymbol{p}, \boldsymbol{p}^{vis})$$

- But it's technically difficult, computationally expensive, mistreats hadronization, and avoids simulation by highly simplifying the detector response.

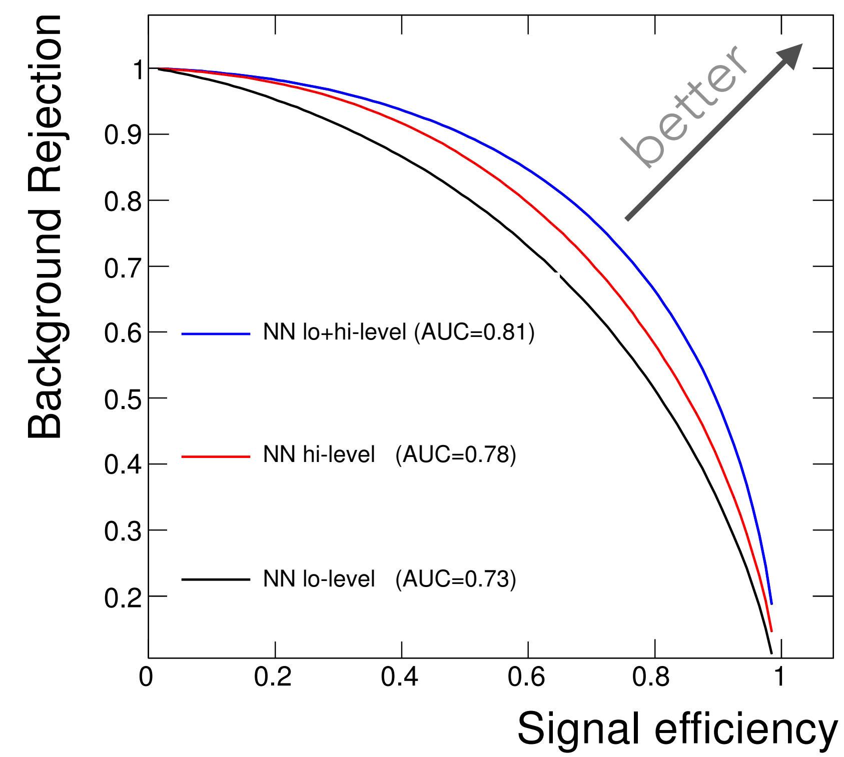
Deep Learning
in
HEP

DEEP LEARNING IN HEP

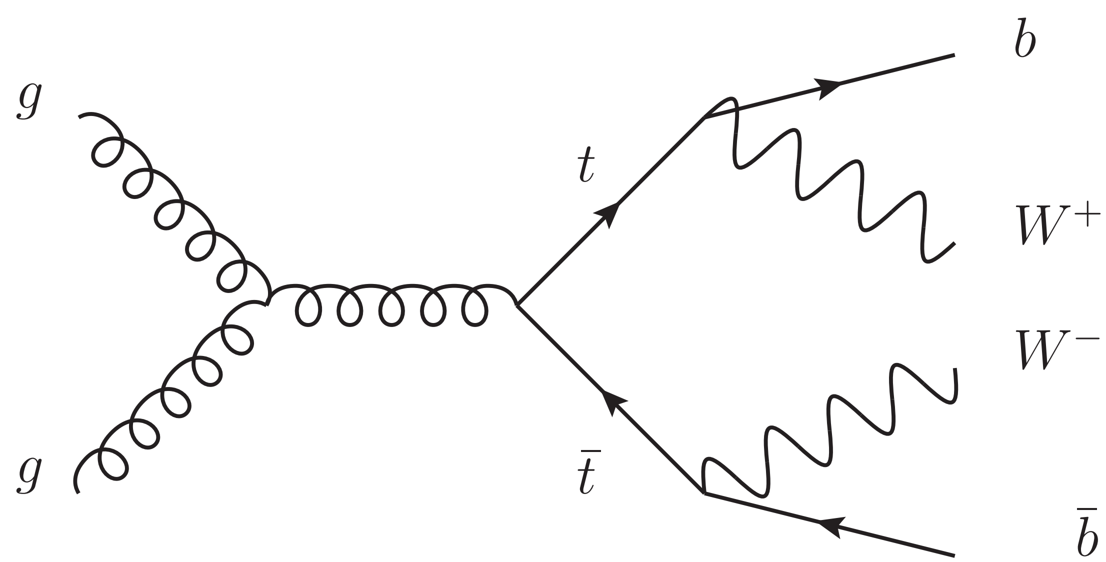
Baldi, Sadowski, Whiteson
arxiv:1402.4735



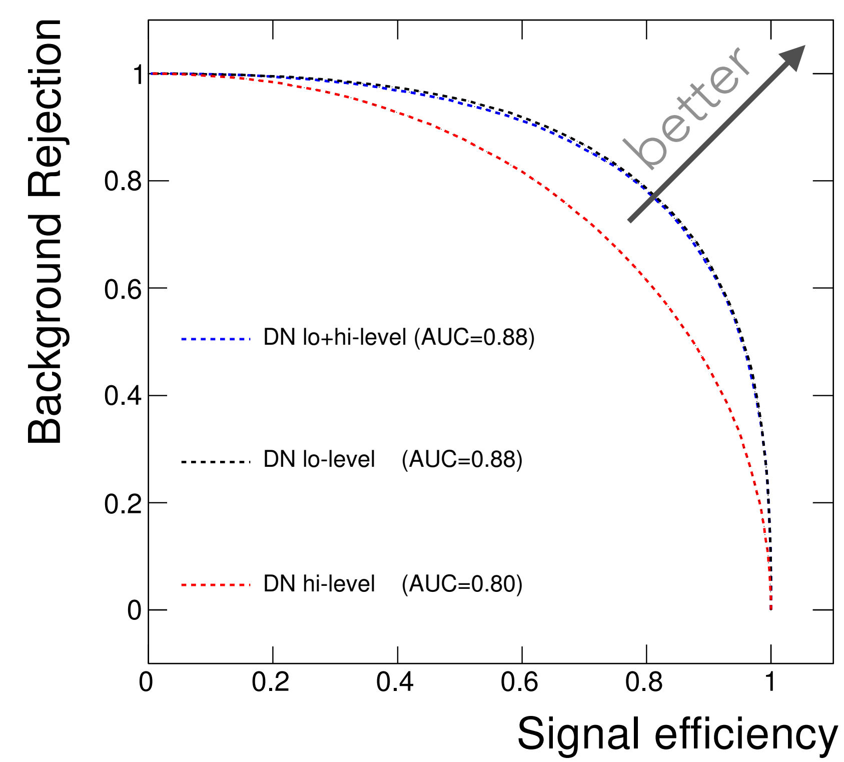
(a)



(a)



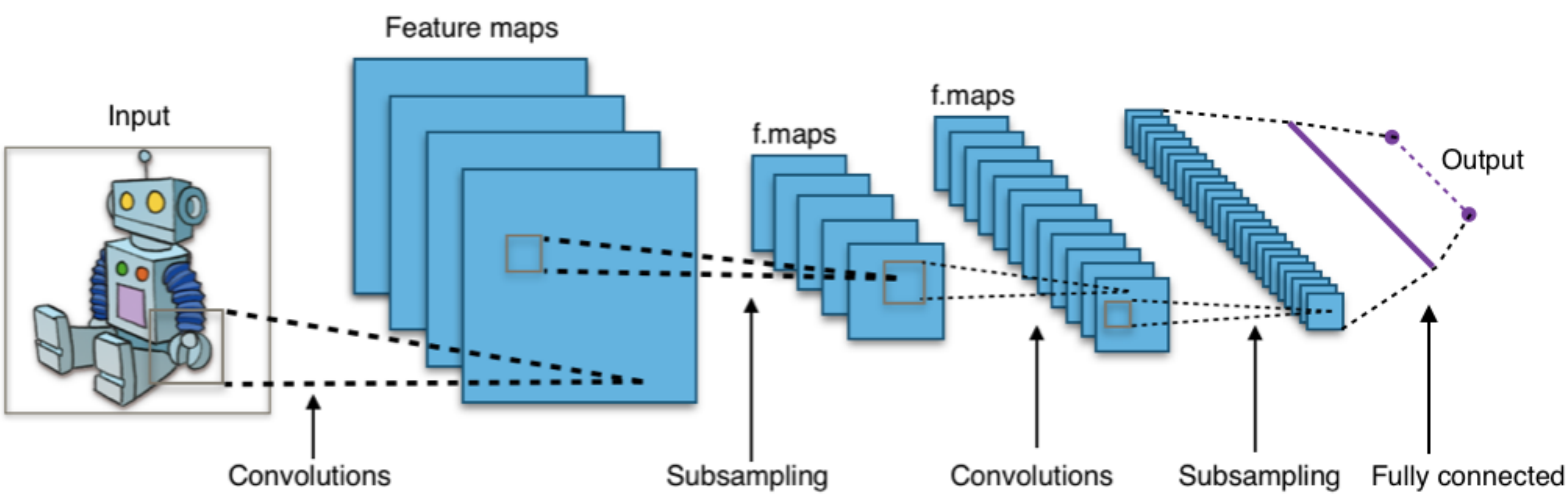
(b)

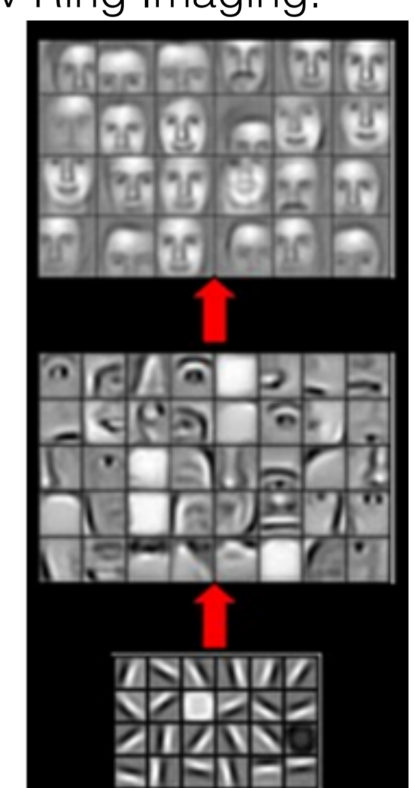


Slide from Kyle Cranmer

Feature Learning

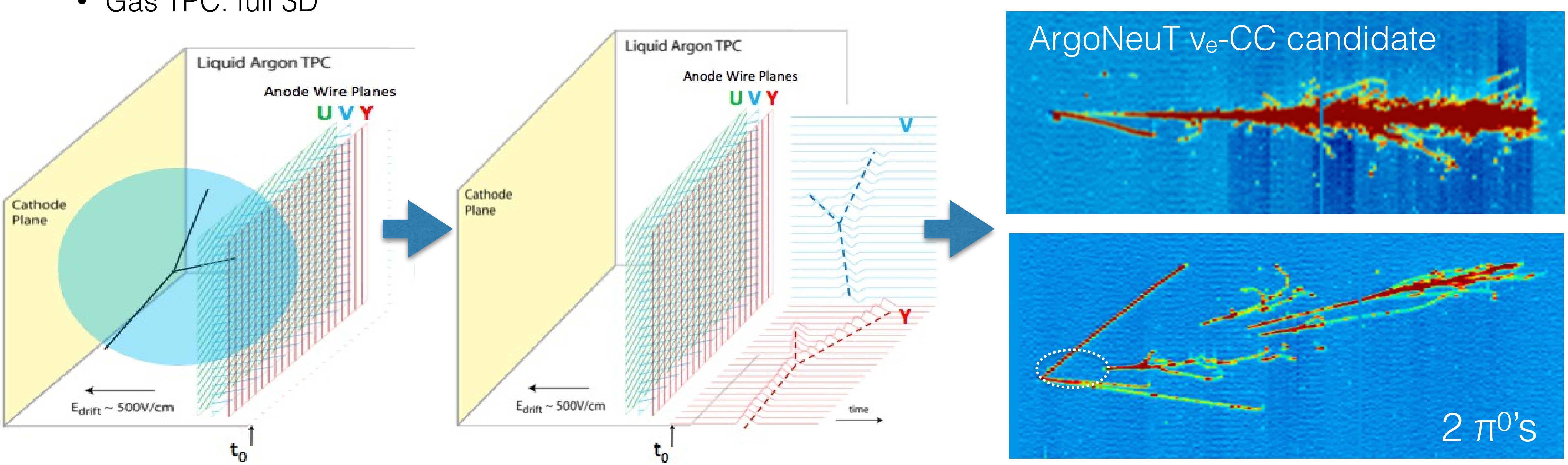
- ***Feature Engineering***: e.g. Event Reconstruction ~ Feature Extraction, Pattern Recognition, Fitting, …
- Deep Neutral Networks can ***Learn Features*** from ***raw data***.
- Example: ***Convolutional Neural Networks*** - Inspired by visual cortex
    - ***Input***: Raw data… for example 1D = Audio, 2D = Images, 3D = Video
    - ***Convolutions*** ~ learned feature detectors
    - ***Feature Maps***
    - ***Pooling*** - dimension reduction / invariance
    - ***Stack***: Deeper layers recognize higher level concepts.
- Over the past few years, CNNs have lead to ***exponential improvement / superhuman performance on Image classification*** challenges. Current best > 150 layers.
- Obvious HEP application: ***"Imaging" Detectors*** such as TPCs, High Granularity Calorimeters, or Cherenkov Ring Imaging.





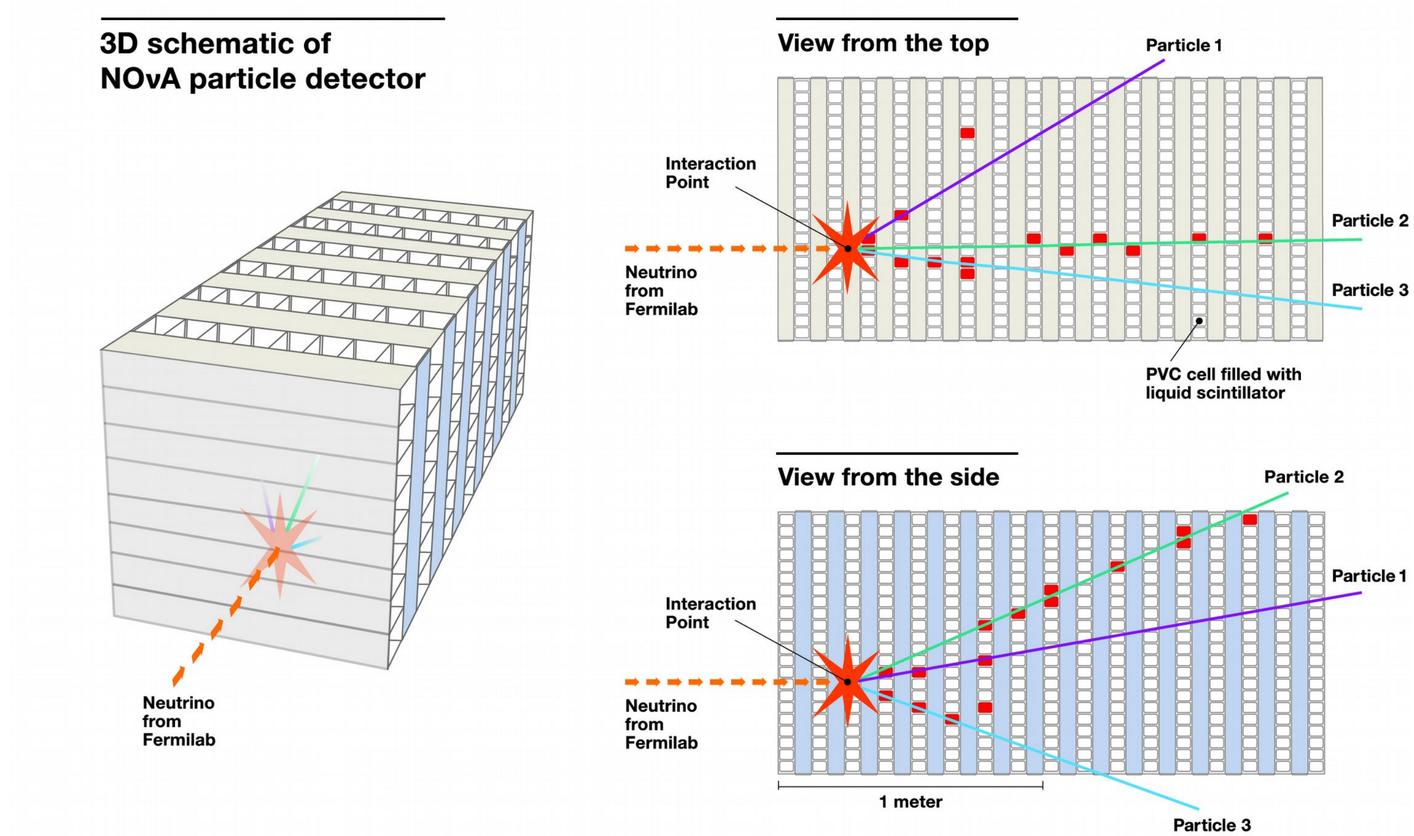
Neutrino Detectors

- ***Need large mass/volume*** to maximize chance of neutrino interaction.
- Technologies:
    - Water/Oil Cherenkov
    - Segmented Scintillators
    - ***Liquid Argon Time Projection Chamber: promises ~ 2x detection efficiency.***
        - ***Provides tracking, calorimetry, and ID all in same detector***.
        - Chosen technology for US's flagship LBNF/DUNE program.
        - Usually 2D read-out… 3D inferred.
    - Gas TPC: full 3D



Neutrino Physics

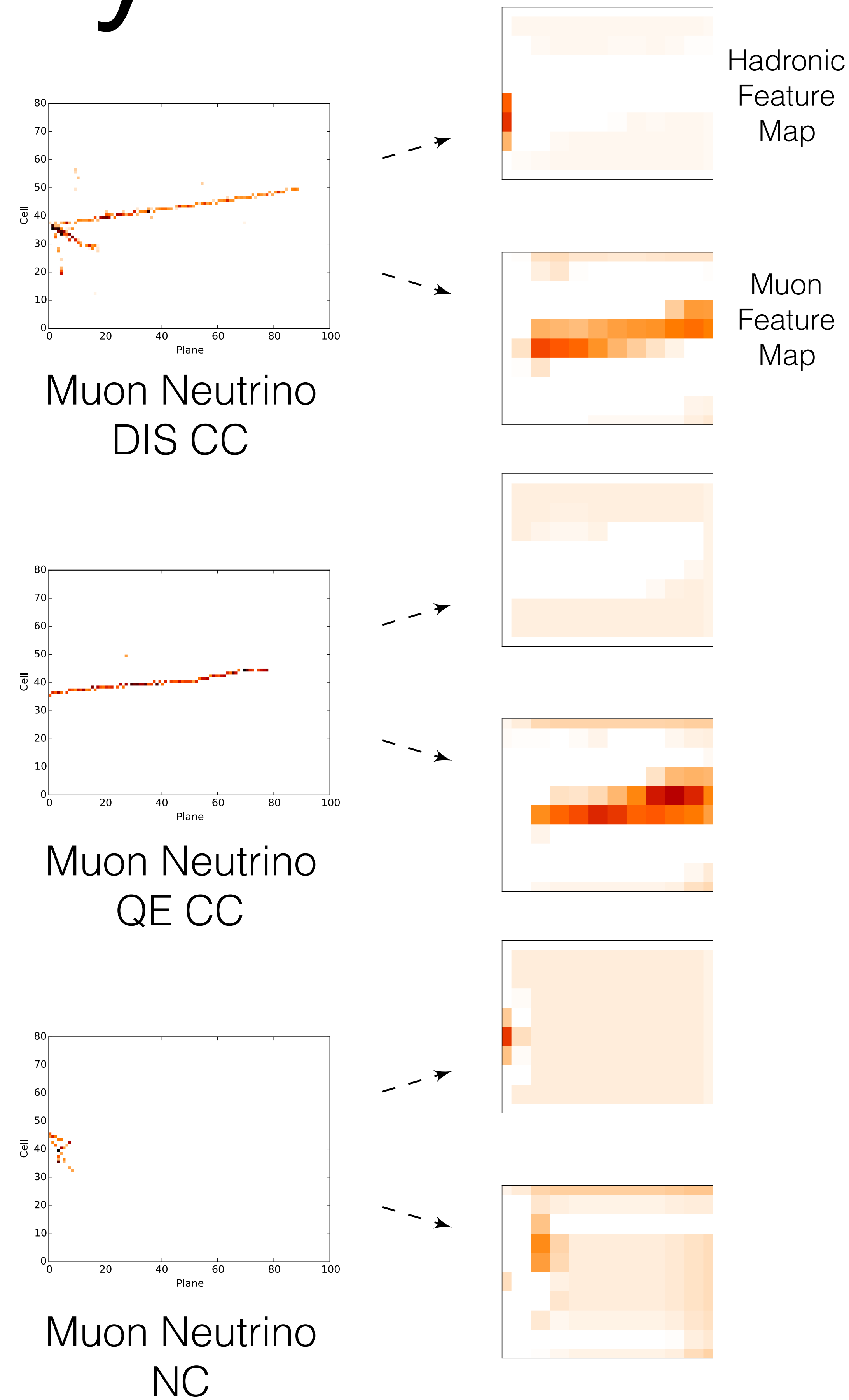
- Core Physics requires just measuring ***neutrino flavor and energy***.
- Generally clean (low multiplicity) and high granularity.
- ***First HEP CNN application: <u>Nova</u>*** using Siamese Inception CNN.



	CVN Selection Value	ν_e sig	Tot bkg	NC	ν_μ CC	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	88.4	509.0	344.8	132.1	32.1	–	14.8%
$s/\sqrt{b}$ opt	0.94	43.4	6.7	2.1	0.4	4.3	49.1%	86.6%
$s/\sqrt{s+b}$ opt	0.72	58.8	18.6	10.3	2.1	6.1	66.4%	76.0%

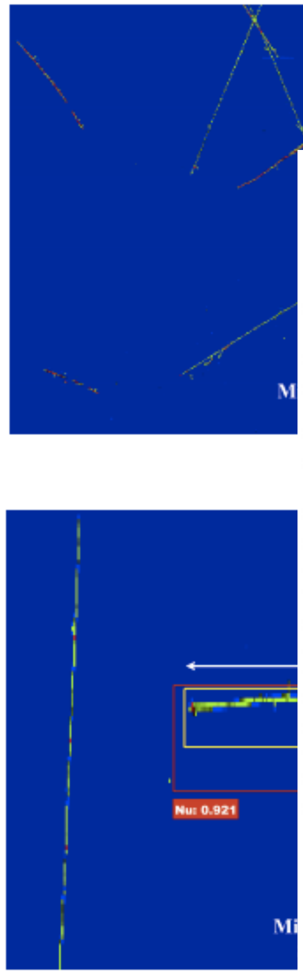
	CVN Selection Value	ν_μ sig	Tot bkg	NC	Appeared ν_e	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	355.5	1269.8	1099.7	135.7	34.4	–	21.9%
$s/\sqrt{b}$ opt	0.99	61.8	0.1	0.1	0.0	0.0	17.4%	99.9%
$s/\sqrt{s+b}$ opt	0.45	206.8	7.6	6.8	0.7	0.1	58.2%	96.4%

- ***40% Better Electron Efficiency for same background.***



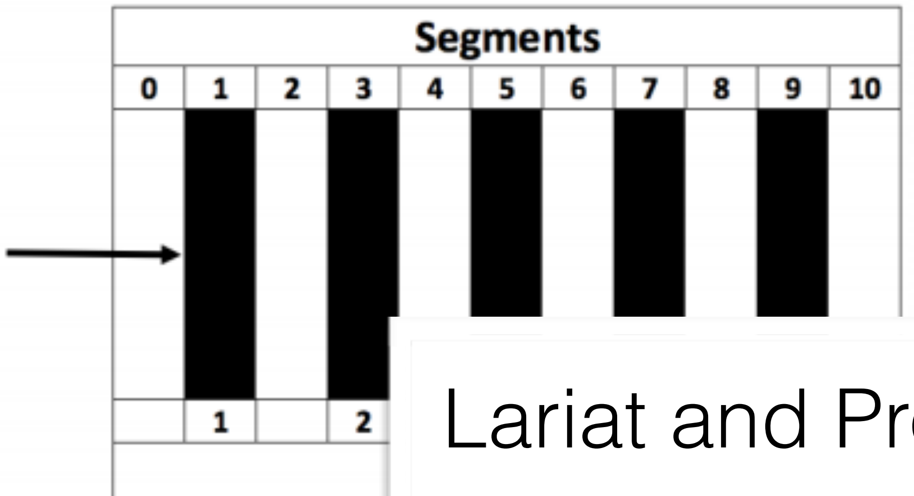
http://arxiv.org/pdf/1604.01444.pdf

MicroBooNE



JINST 12 (2017) r

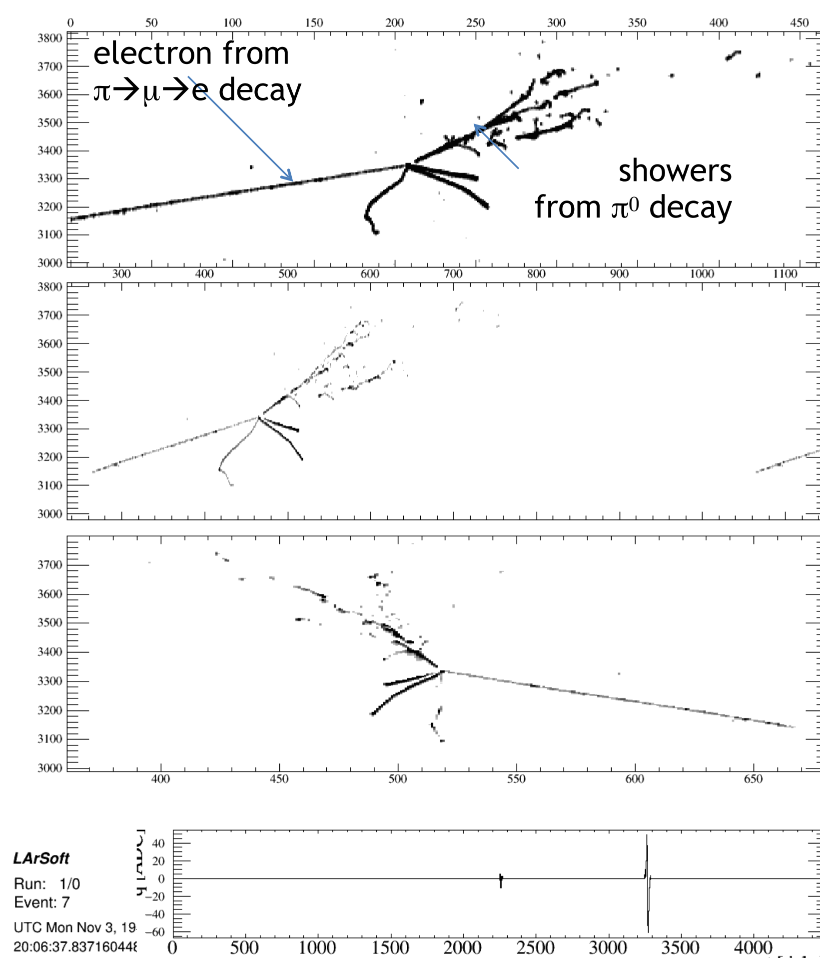
MINERvA and MENNDL



PhyStat-nu Fermila

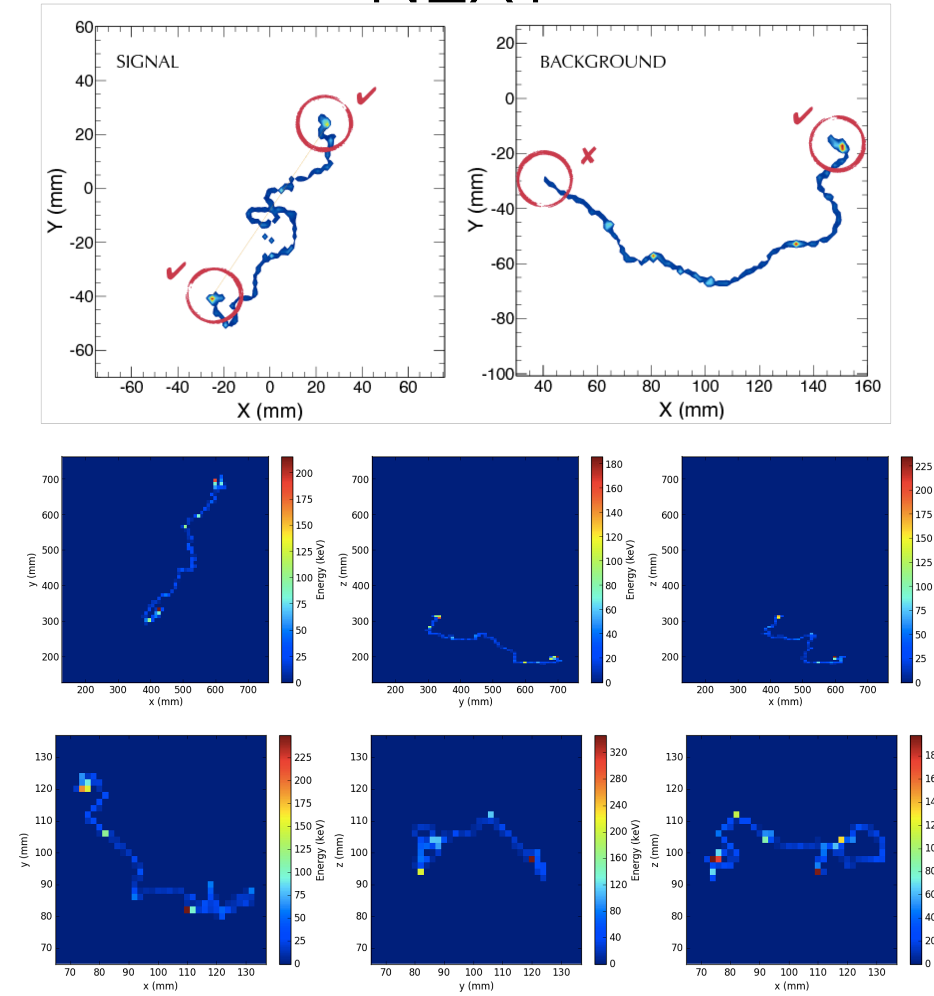
Lariat and ProtoDUNE

EM / hadronic component discrimination



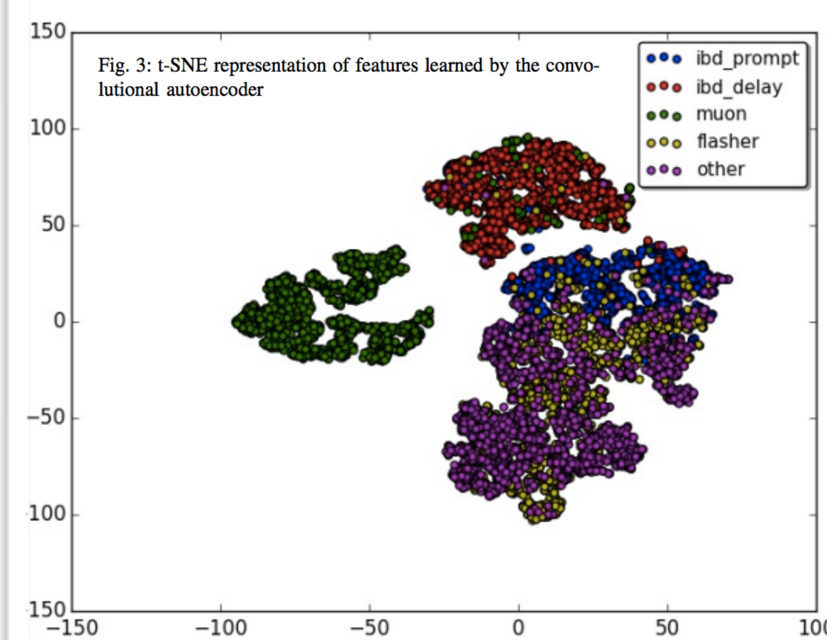
Private communication, Robert Sutlej

NEXT



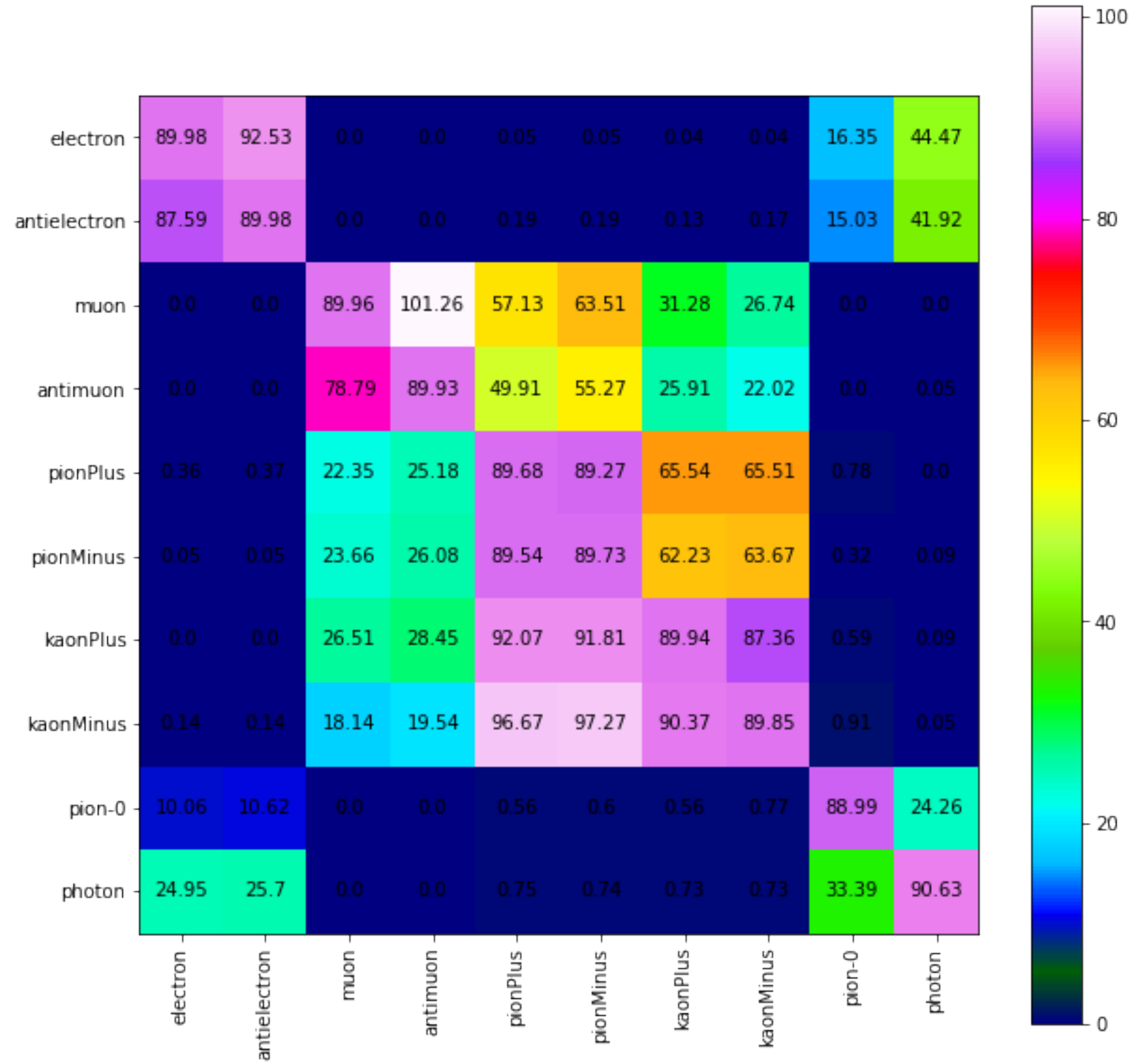
JINST 12 (2017) no.01, T01004

Daya Bay



arXiv:1601.07621

LArIAT: DNN vs Alg



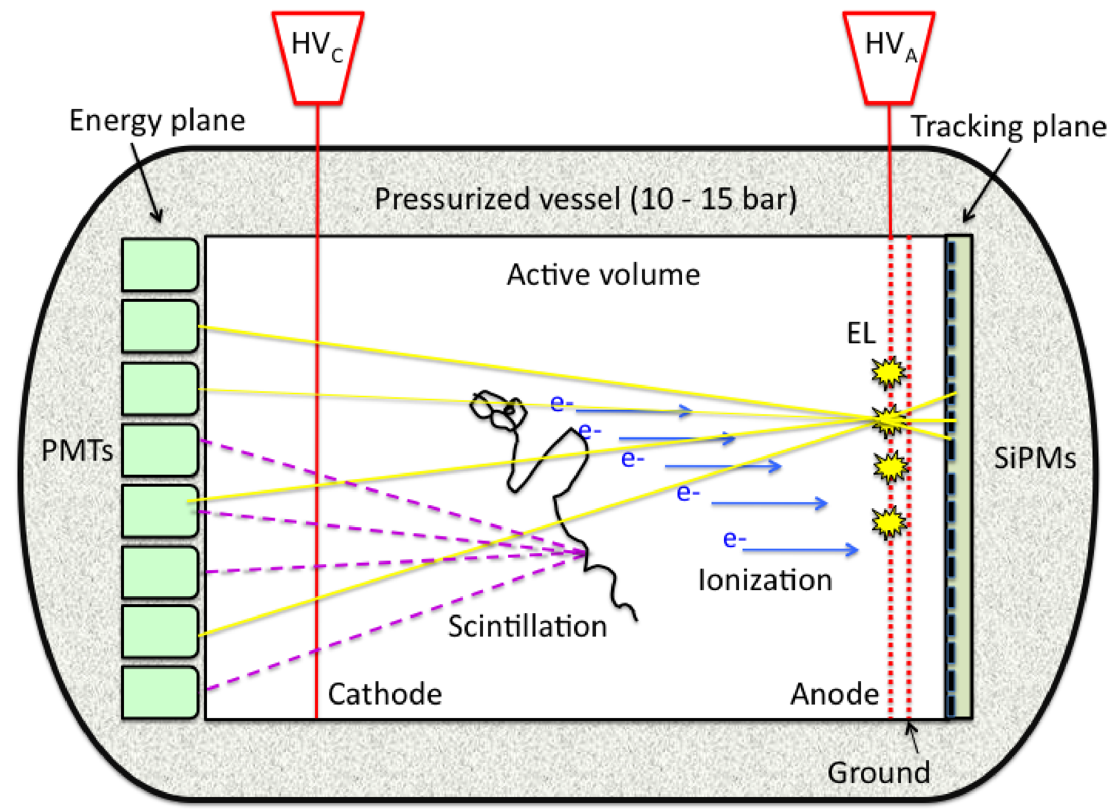
	π+	K+	μ+	e+	γ
DNN	74.42%	40.67%	6.37%	0.12%	0%
LArIAT	74.5%	68.8%	88.4%	6.8%	2.4%

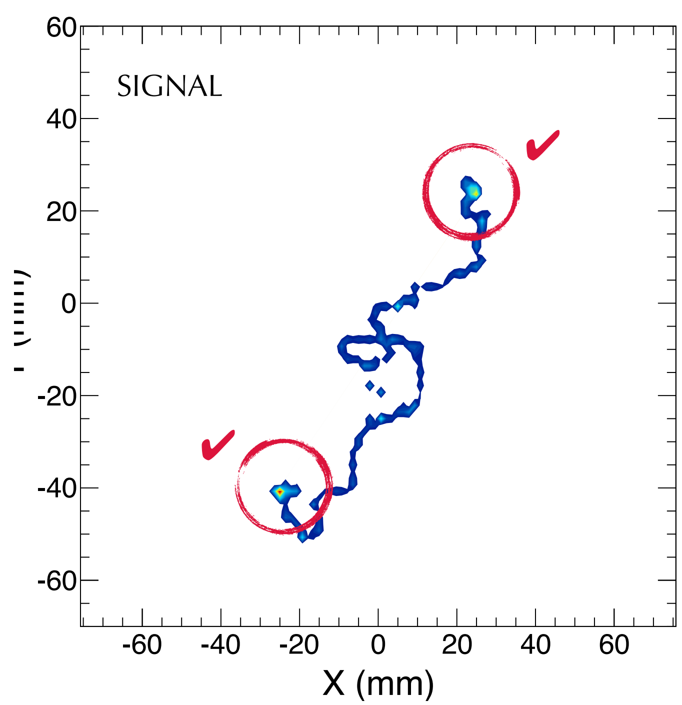
	π−	K−	μ−	e−	γ
DNN	78.68%	54.47%	13.54%	0.11%	0.25%
LArIAT	78.7%	73.4%	91.0%	7.5%	2.4%

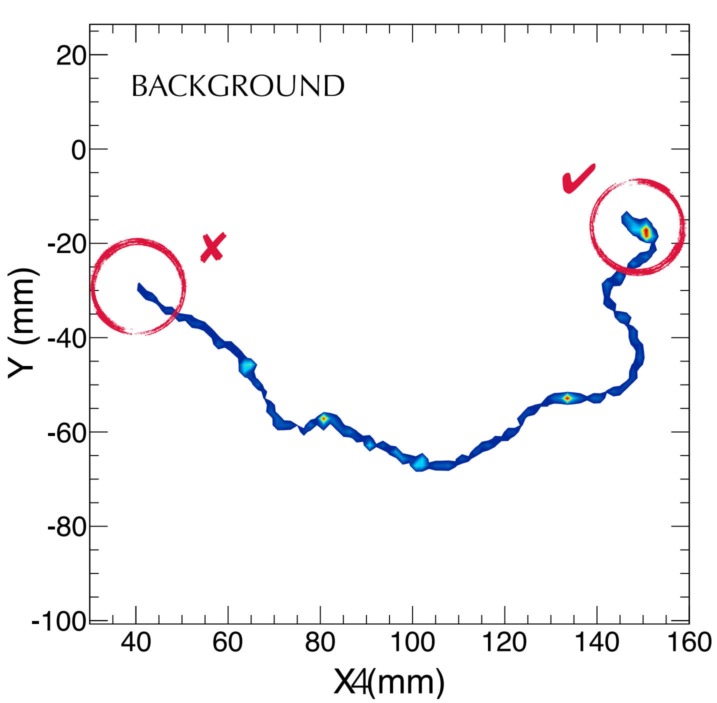
NEXT Experiment

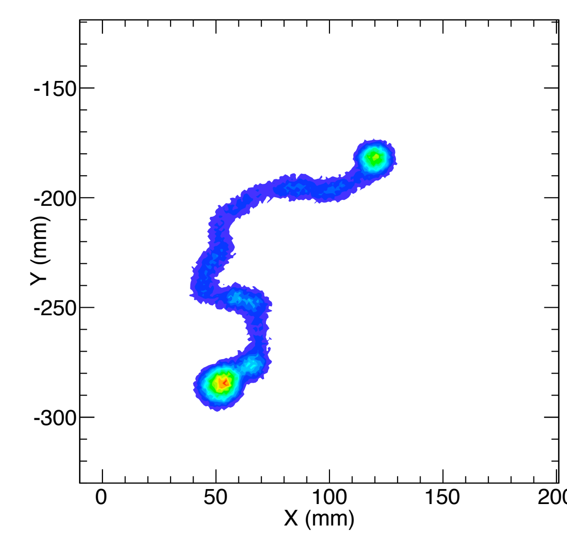
(J. Renner, J.J. Gomez, …, AF)

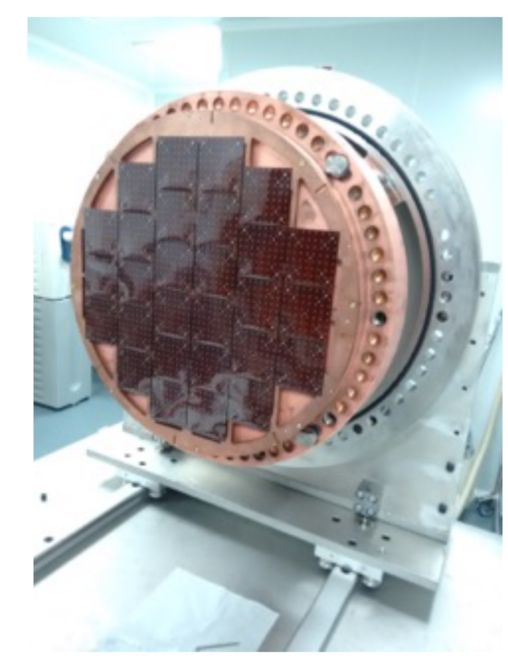
- ***Neutrinoless Double Beta Decay*** using Gas TPC/SiPMs
- Signal: 2 Electrons. Bkg: 1 Electron.
- Hard to distinguish due to ***multiple scattering***.
- ***3D readout***… candidate for 3D Conv Nets.
- Just a handful of signal events will lead to ***noble prize***
    - Can we trust a DNN at this level?







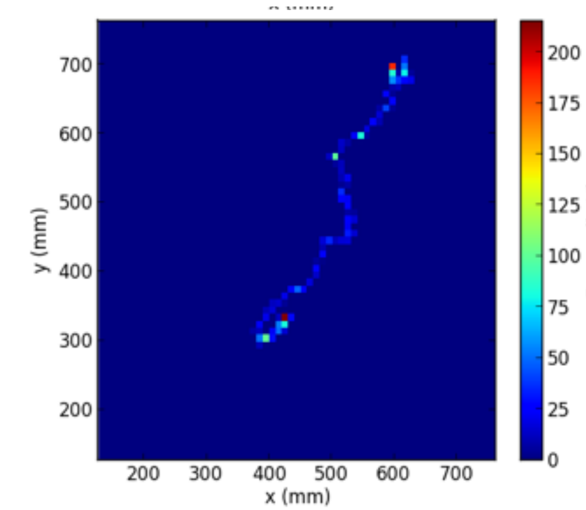


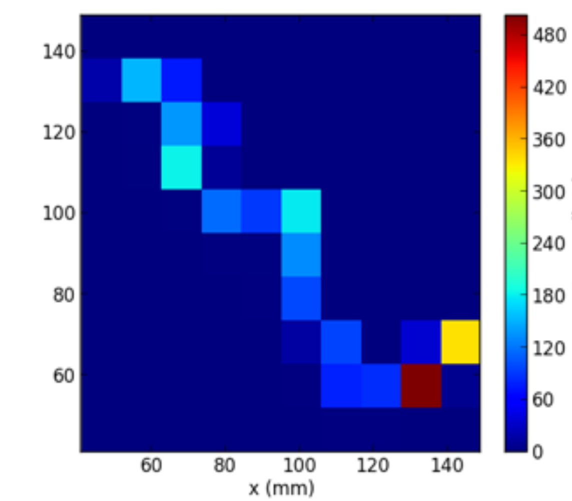


NEXT Detector Optimization

- Idea 1: use DNNs to ***optimize detector***.
  - Simulate data at different resolutions
  - Use DNN to quickly/easily assess best performance for given resolution.

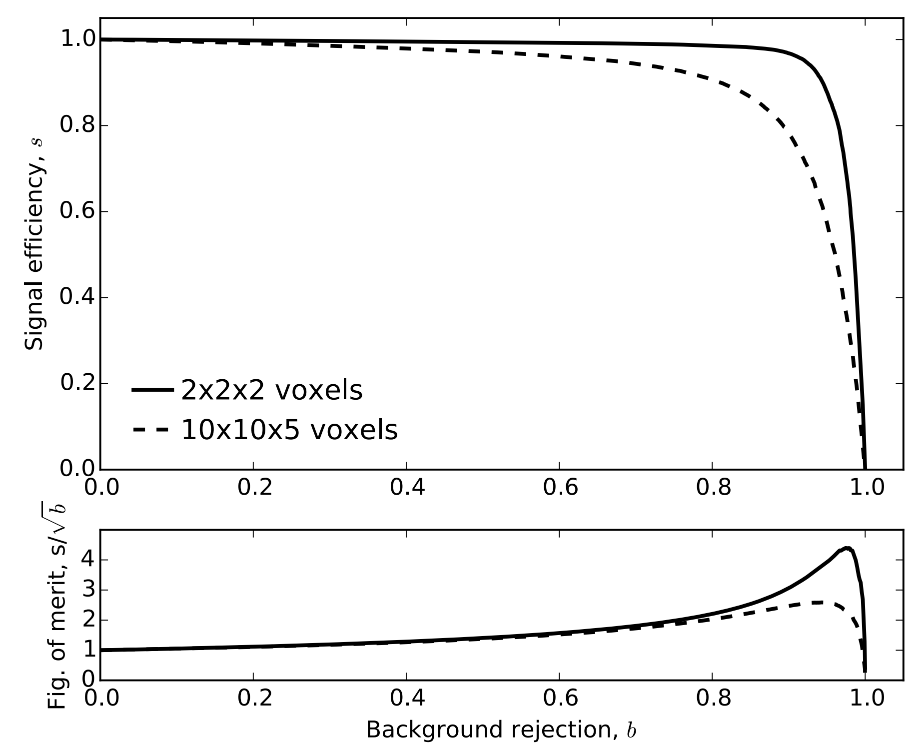
Analysis	Signal eff. (%)	B.G. accepted (%)
DNN analysis (2 x 2 x 2 voxels)	86.2	4.7
Conventional analysis (2 x 2 x 2 voxels)	86.2	7.6
DNN analysis (10 x 10 x 5 voxels)	76.6	9.4
Conventional analysis (10 x 10 x 5 voxels)	76.6	11.0





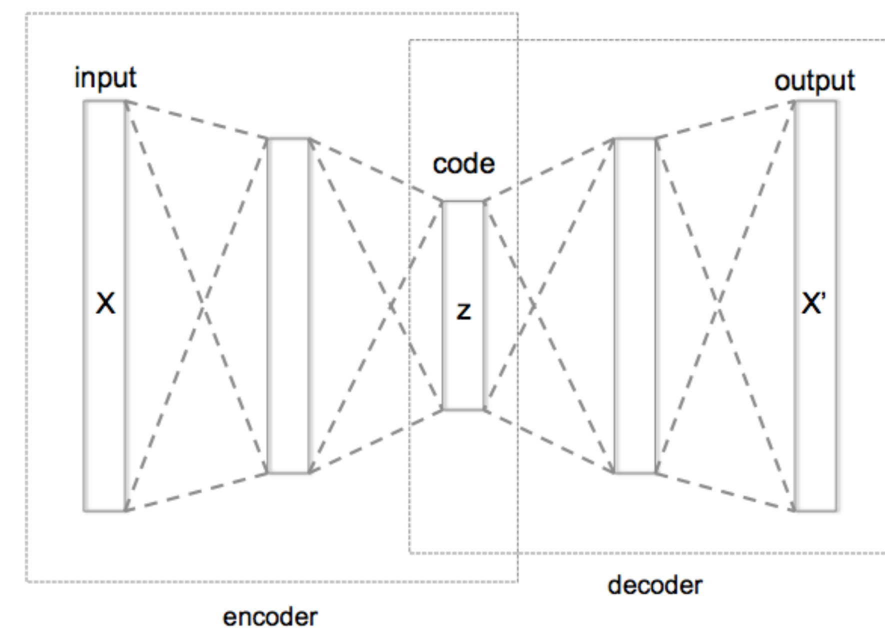
- Idea 2: ***systematically study*** the relative importance of various physics/detector effects.
  - Start with simplified simulation. Use DNN to assess performance.
  - Turn on effects one-by-one.

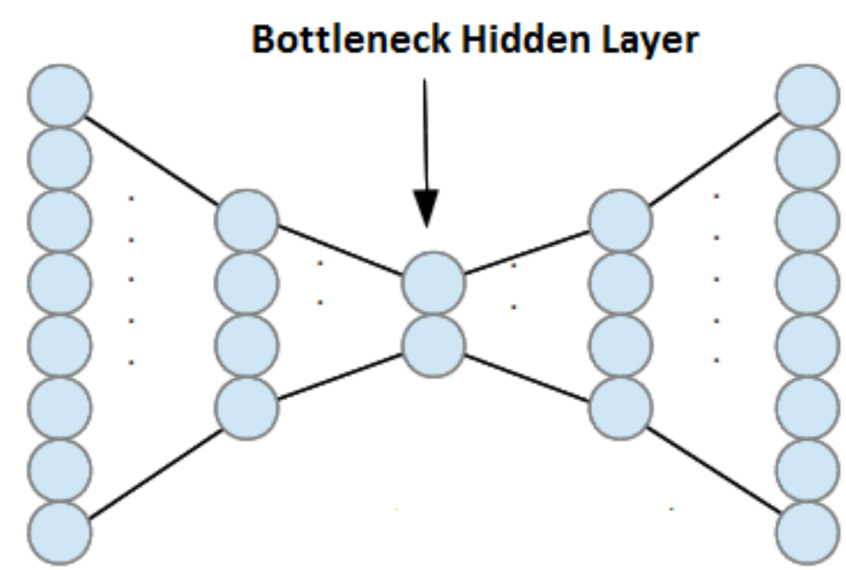
2x2x2 voxels	Run description	Avg. accuracy (%)
	Toy MC, ideal	99.8
	Toy MC, realistic $0\nu\beta\beta$ distribution	98.9
	Xe box GEANT4, no secondaries, no E-fluctuations	98.3
	Xe box GEANT4, no secondaries, no E-fluctuations, no brem.	98.3
	Toy MC, realistic $0\nu\beta\beta$ distribution, double multiple scattering	97.8
	Xe box GEANT4, no secondaries	94.6
	Xe box GEANT4, no E-fluctuations	93.0
	Xe box, no brem.	92.4
	Xe box, all physics	92.1
	NEXT-100 GEANT4	91.6
10x10x5 voxels		
	NEXT-100 GEANT4	84.5



Semi-supervised Learning

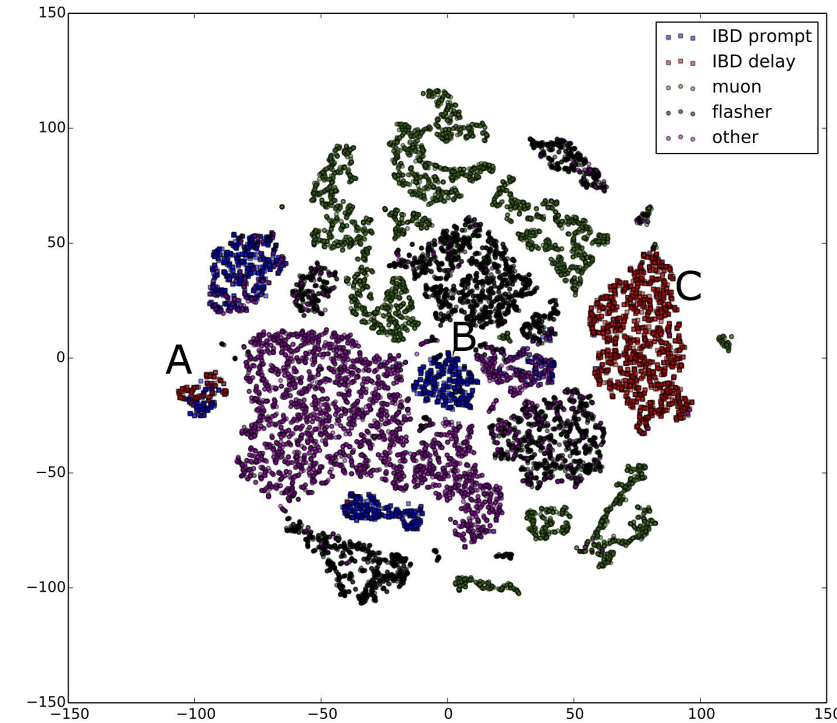
- Basic idea: Train network to ***reproduce the input***.
- Example: ***Auto-encoders***
  - ***De-noising auto-encoders***: add noise to input only.
  - ***Sparse auto-encoders***:
    - ***Sparse latent (code) representation*** can be exploited for ***Compression, Clustering, Similarity testing,*** …
  - ***Anomaly Detection***
    - Reconstruction Error
    - Outliers in latent space
  - ***Transfer Learning***
    - Small labeled training sample?
      - Train auto-encoder on large unlabeled dataset (e.g. data).
      - Train in latent space on small labeled data. (e.g. rare signal MC).
- Easily think of a dozen applications.



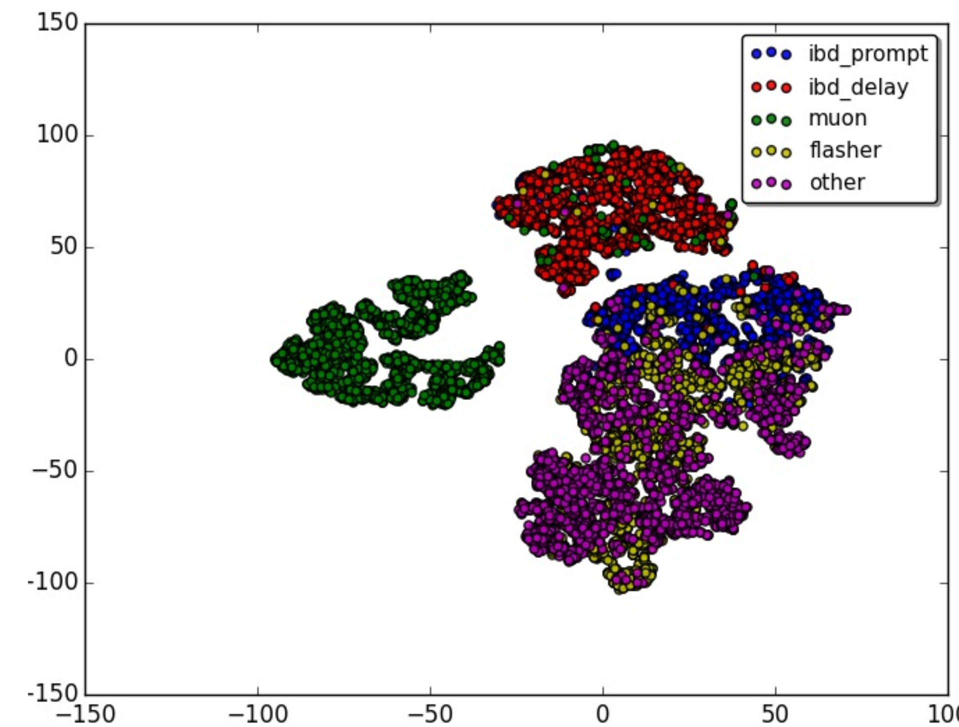


Learning Representations

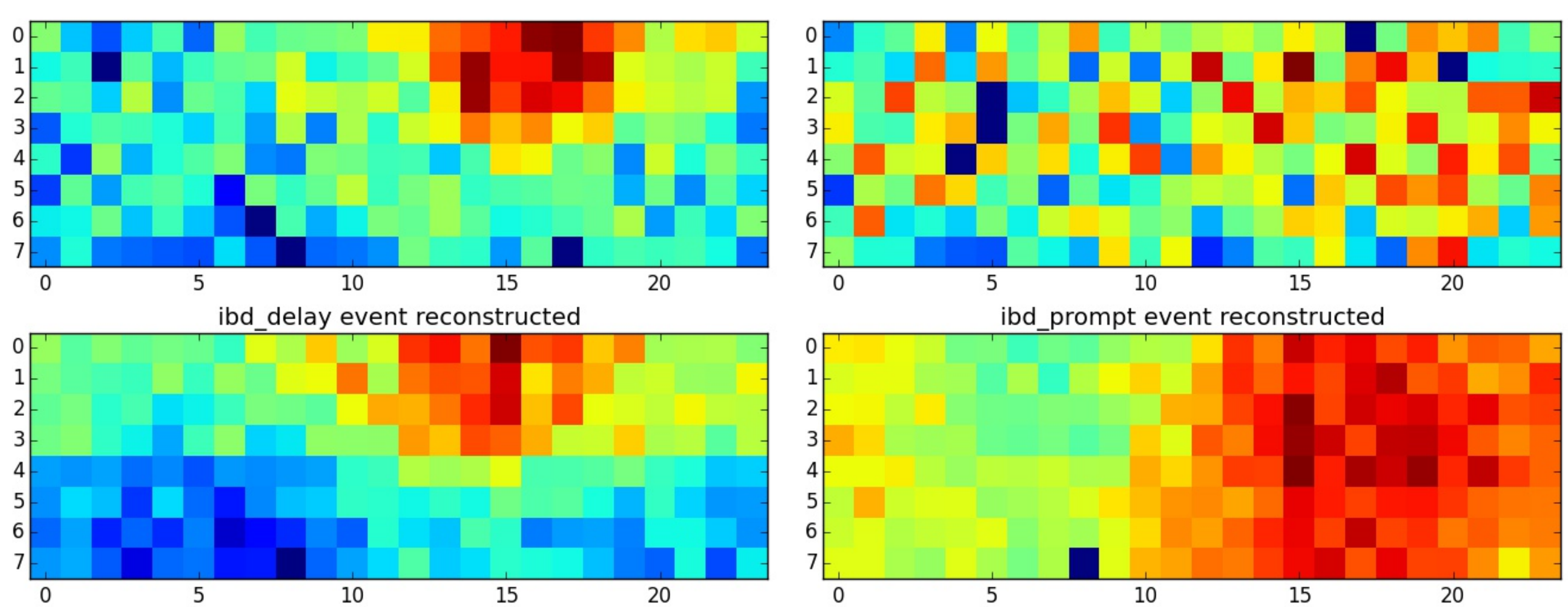
- Example: ***Daya Bay Experiment*** (*Evan Racah, et al*)
- Input: 8 x 24 PMT unrolled cylinder. ***Real Data (no simulation)***
- 2 Studies:
    - ***Supervised CNN Classifier***
        - Labels from standard analysis: Prompt/Delayed Inverse Beta Decay, Muon, Flasher, Other.
    - ***Convolutional Auto-encoder*** (semi-supervised)
        - Clearly separates muon and IBD delay ***without any physics knowledge***.
        - Potentially could have ID'ed problematic data (e.g. flashers) much earlier.



t-SNE reduction of 26-dim representation of the last fully connected layer.



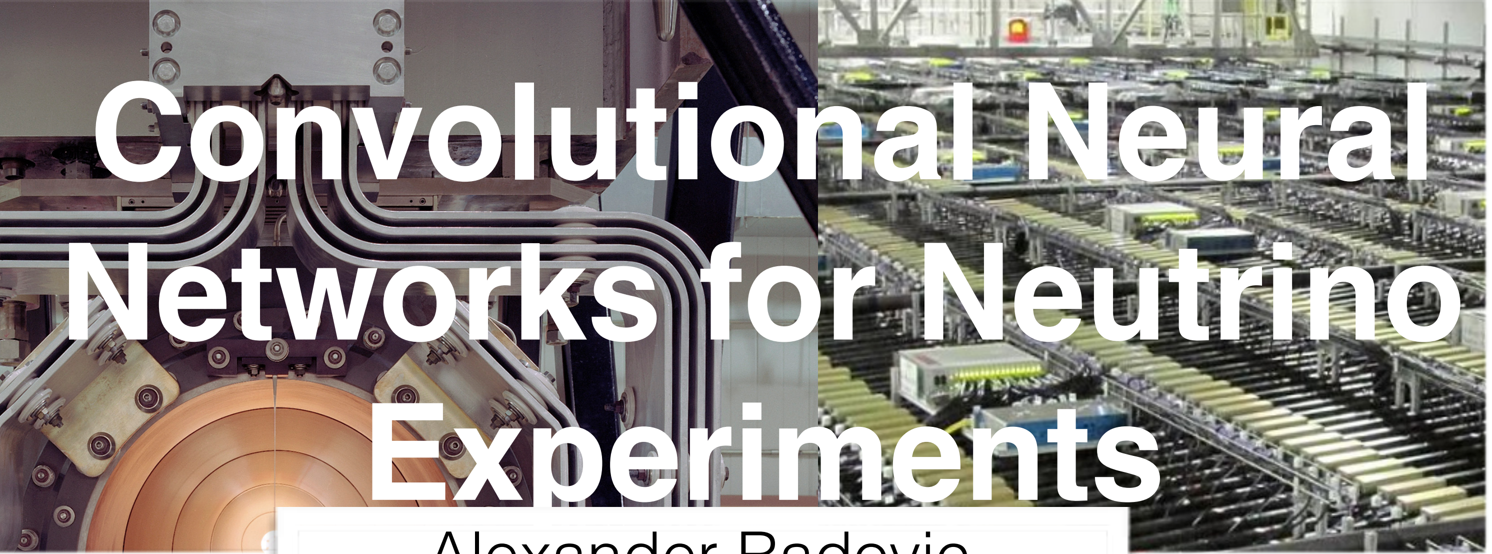
t-SNE reduction of 10 parameter latent representation.



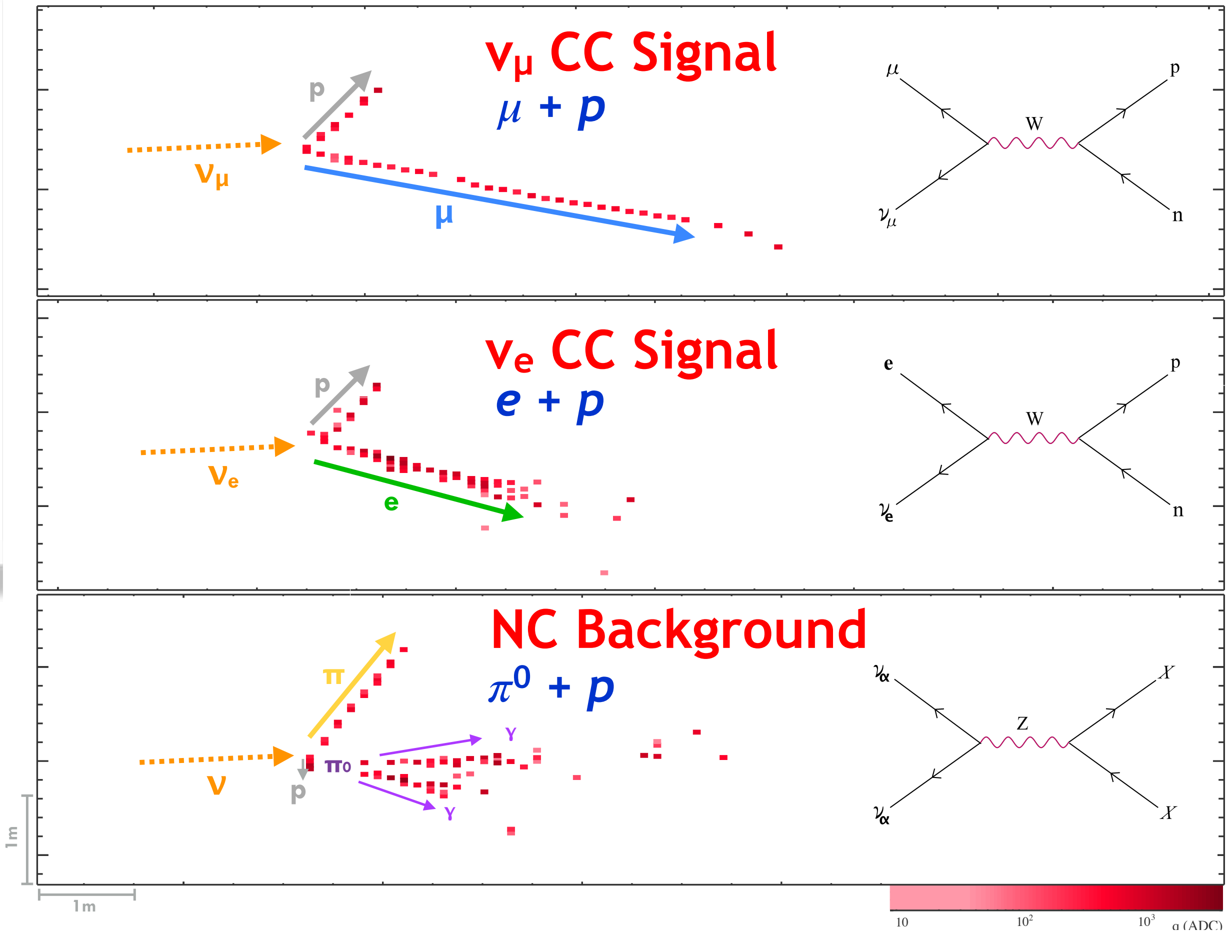
(a) Example of an "IBD delay" event

(b) Example of an "IBD prompt" event

Convolutional Neural Networks for Neutrino Experiments

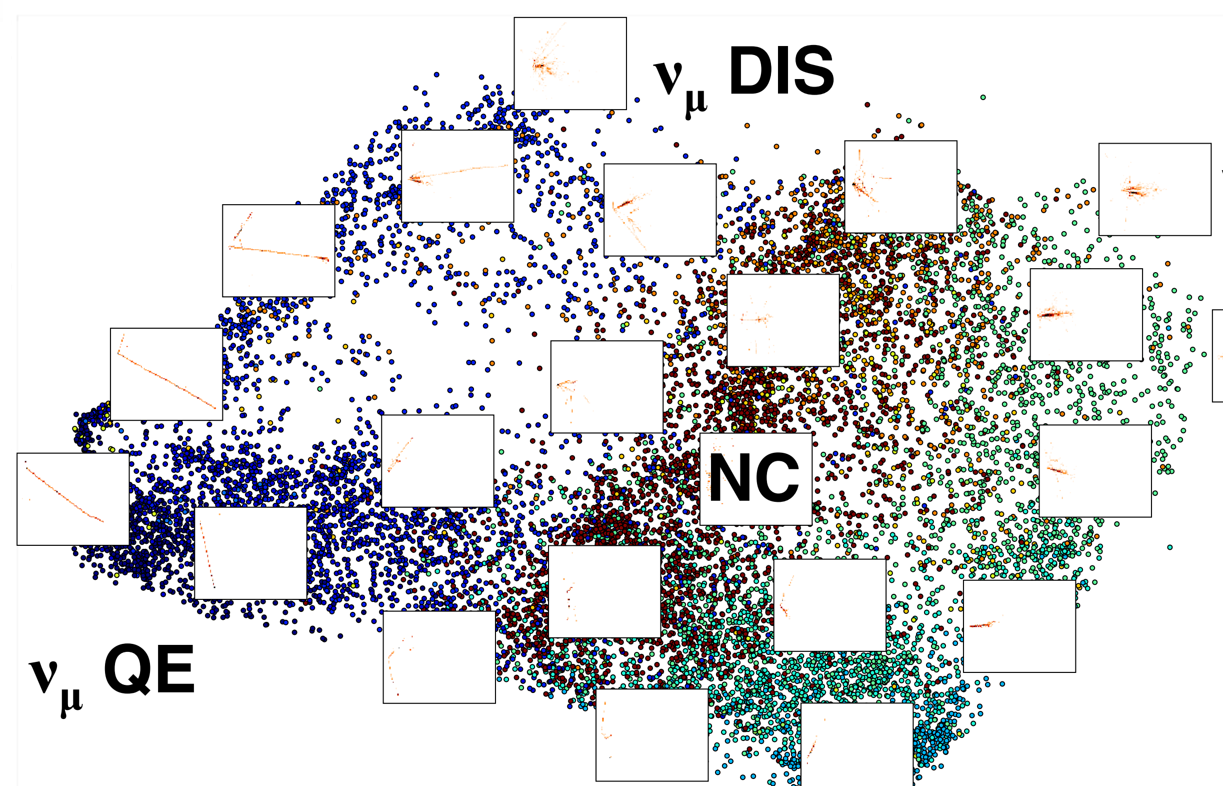


Alexander Radovic
College of William and Mary



Why Convolutional Neural Networks?

- That means that any oscillation analysis can benefit fror precise identification of the interaction in two ways:
  - Estimating the lepton flavor of the incoming neutrino.
  - Correctly identifying the type of neutrino interaction, t better estimate the neutrino energy, aka is it a quasi elastic event or a resonance event?



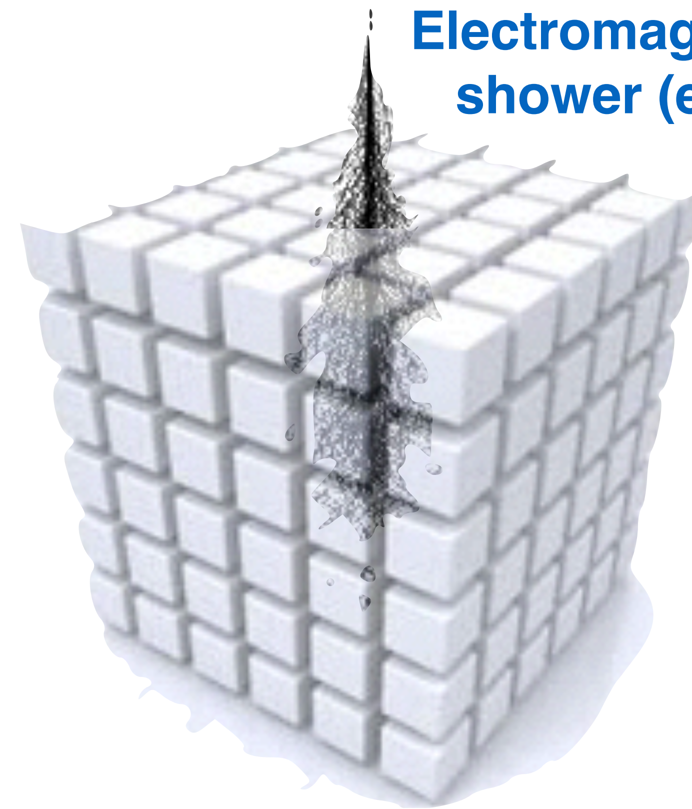
t-SNE projection of final features to 2D.

- Our detectors are also often the perfect domain:
  - Large ~uniform volumes where spatially invariant response is a benefit.
  - Usually only one or two detector systems.

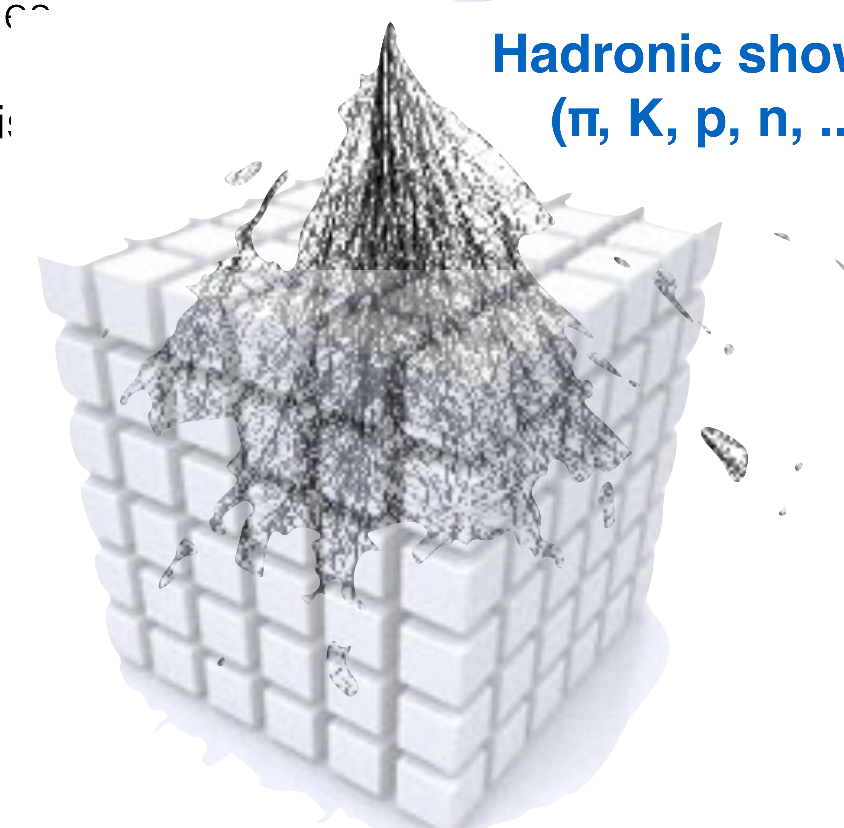
However our CNN achieves **73%** efficiency and **76%** purity on ν_e selection at the $s/\sqrt{s+b}$ optimized cut.
Equivalent to **30%** more exposure with the old PIDs.

Generative Models

- ***Likelihood Approximation relies simulation***
  - Most ***computationally expensive*** step, so any speedup has huge impact.
  - More generally, ***simulation based on data*** would be a powerful tool.
    - For example, we can build a Hadronization model purely from data.
- DNNs Generative Models enable building simulations purely from examples.
  - ***Generative Adversarial Nets*** (Goodfellow, et. al. arxiv:1406.2661). Simultaneously train 2 Networks:
      - ***Discriminator*** (D) that tries to distinguish output and real exampl
      - ***Generator*** (G) that generate the output that is difficult to distingui
  - ***Variational Auto-encoders***:
    - Learn a ***latent variable probabilistic model*** of the input dataset.
    - ***Sample latent space*** and use ***decoder to generate data***.
- ***Particle showering*** is ***slowest*** part of the micro-physics simulation…
  - Various techniques for fast showering (e.g. shower template libraries) are common.
  - DNN Generative Models are being pursued inside the experiments (K. Cranmer, G. Louppe, …) for this task…



Electromagnetic shower (e, γ)



Hadronic shower (π, K, p, n, ..)

Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis

Luke de Oliveira[a], Michela Pag

[a] *Lawrence Berkeley National Lab*

[b] *Department of Physics, Yale Un*

E-mail: lukedeoliveira@lbl.

Abstract: We provide a bridg
and simulated physical processe
Adversarial Network (GAN) arc
energy depositions from particle
the Location-Aware Generative A
from simulated high energy partic
span over many orders of magnit
jet mass, n-subjettiness, etc.). W
of image quality and validity of
a base for further explorations of

CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

Michela Paganini[a,b], Luke de Oliveira[a], and Benjamin Nachman[a]

[a] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA, 94720, USA*

[b] *Department of Physics, Yale University, New Haven, CT 06520, USA*

E-mail: michela.paganini@yale.edu, lukedeoliveira@lbl.gov, bnachman@cern.ch

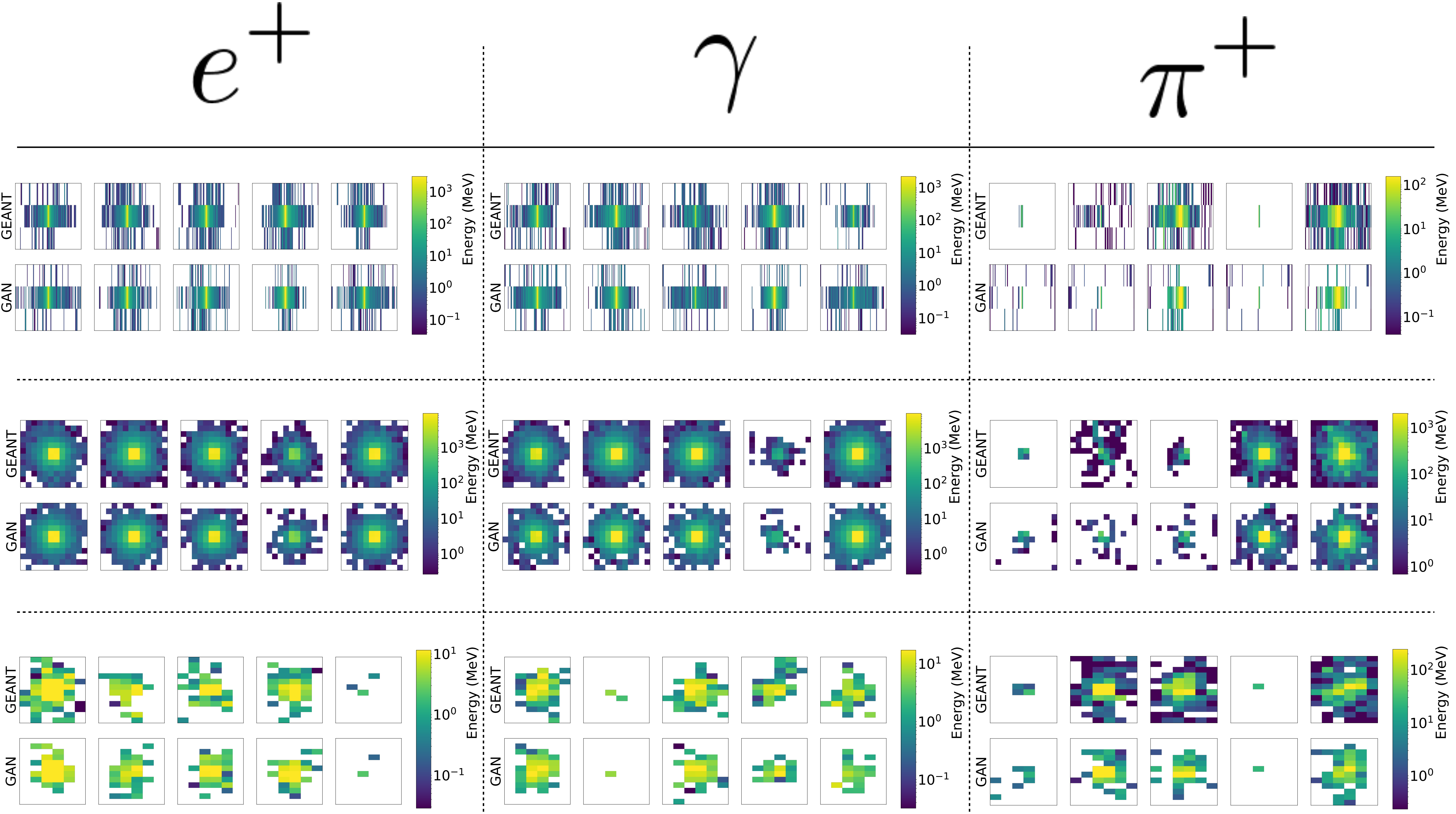
Abstract: Simulation is a key component of physics analysis in particle physics and nuclear physics. The most computationally expensive simulation step is the detailed modeling of particle showers inside calorimeters. Full detector simulations are too slow to meet the growing demands resulting from large quantities of data; current fast simulations are not precise enough to serve the entire physics program. Therefore, we introduce CaloGAN, a new fast simulation based on generative adversarial neural networks (GANs). We apply the CaloGAN to model electromagnetic showers in a longitudinally segmented calorimeter. This represents a significant stepping stone toward a full neural network-based detector simulation that could save significant computing time and enable many analyses now and in the future. In particular, the CaloGAN achieves speedup factors comparable to or better than existing fast simulation techniques on CPU ($100\times$-$1000\times$) and even faster on GPU (up to $\sim 10^5\times$)) and has the capability of faithfully reproducing many aspects of key shower shape variables for a variety of particle types.

https://arxiv.org/pdf/1701.05927.pdf

https://arxiv.org/pdf/1705.02355.pdf

Qualitative Performance (2)

Yale



Timing

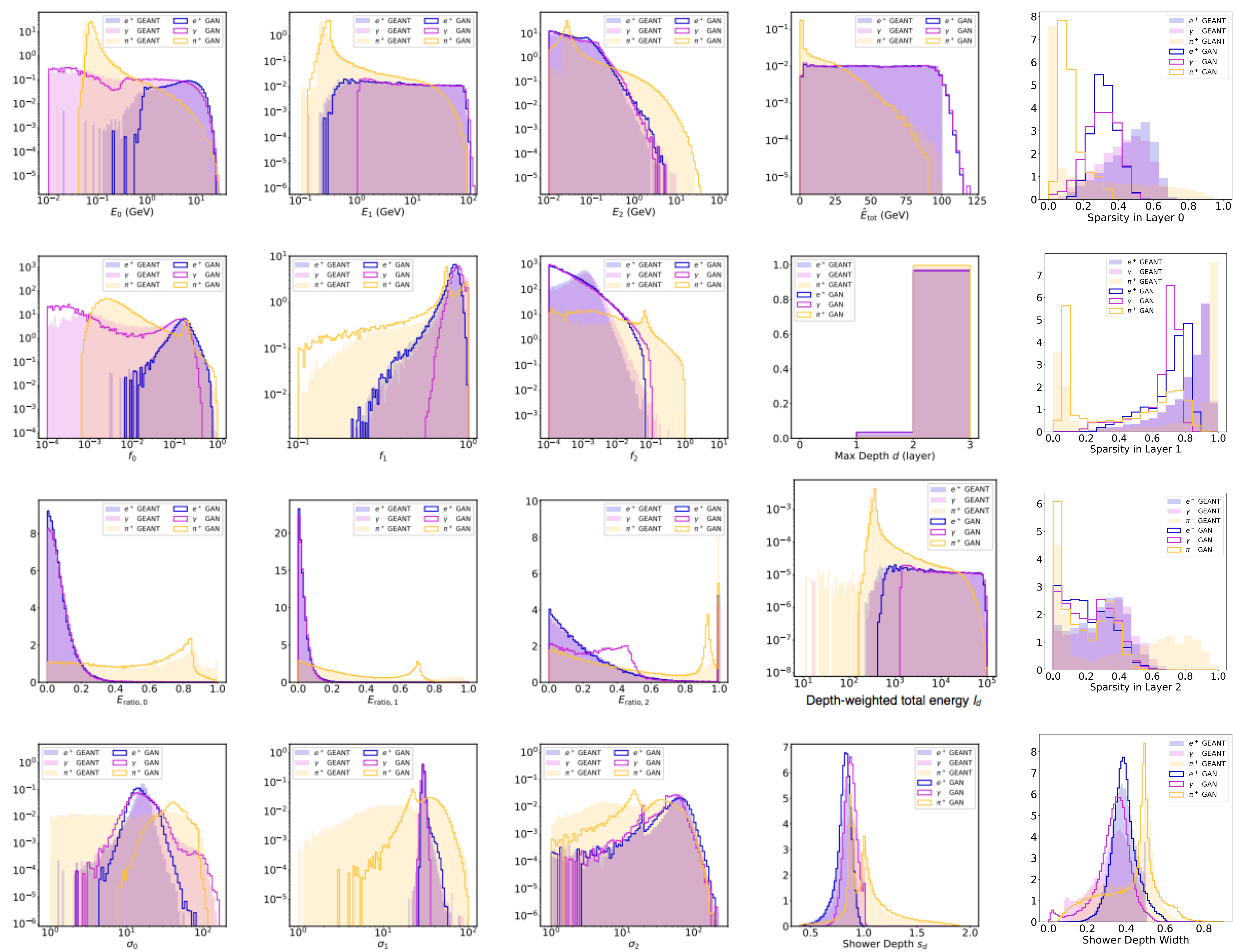
M. Paganini et al., 1705.02355

Generation Method	Hardware	Batch Size	milliseconds/shower
GEANT4	CPU	N/A	1772 ←
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012 ←

See also S. Vallecorsa et al. (GeantV), C. Guthrie et al. (NYU), W. Wei et al. (LCD dataset group), D. Salamani et al. (Geneva), D. Rousseau et al. (Orsay), L. de Oliveira et al. (Berkeley)

Shower Shapes

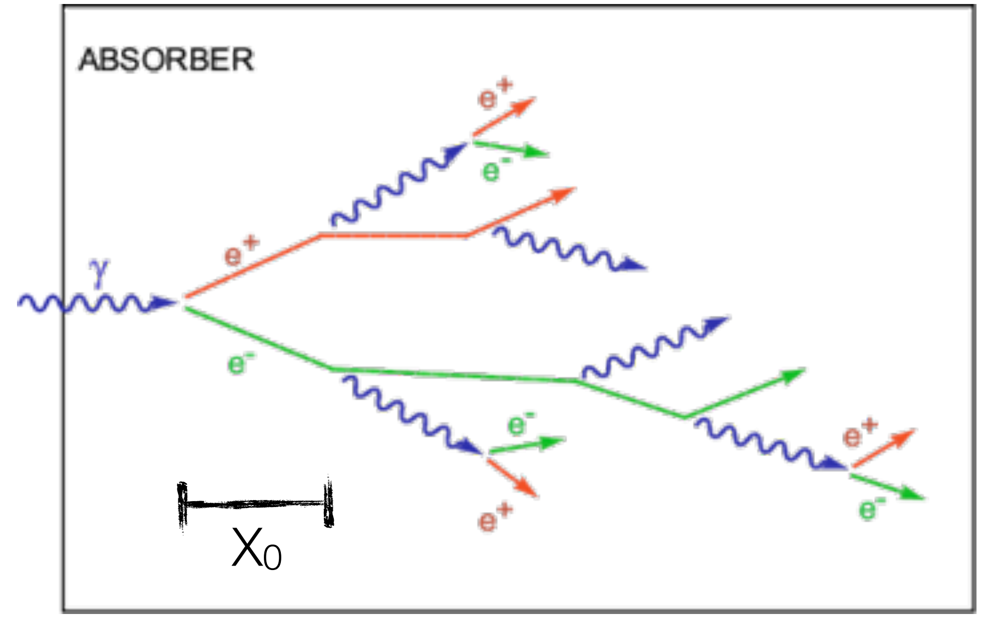
Check: does the LAGAN recover the true data distribution as projected onto a set of meaningful 1D manifolds?

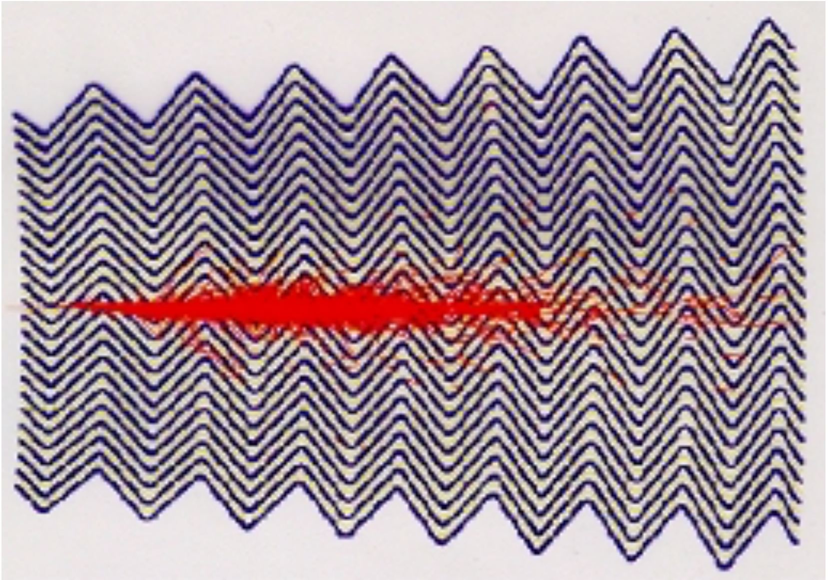


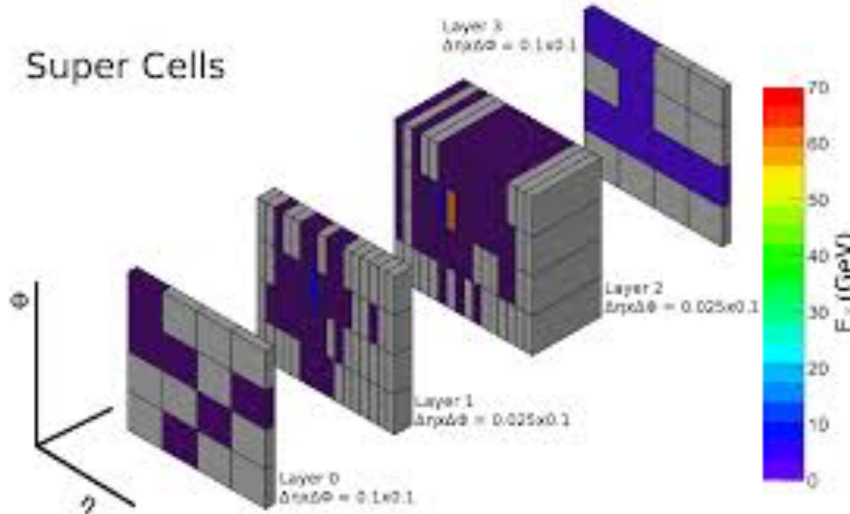
Calorimetry with Deep Learning

How do we “see” particles?

- Particles deposit their energy in a ***stochastic process*** know as ***“showering”***, secondary particles, that in turn also shower.
  - Number of secondary particles ~ Energy of initial particle.
  - Energy resolution improves with energy: $\sigma(E) / E = a/\sqrt{E} \oplus b/E \oplus c$.
    - a = sampling, b = noise, c = leakage.
  - Density and Shape of shower characteristic of type of particle.
- ***Electromagnetic calorimeter***: Low Z medium
  - ***Light particles***: electrons, photons, $\pi^0 \rightarrow \gamma\gamma$ interact with electrons in medium
- ***Hadronic calorimeters***: High Z medium
  - ***Heavy particles***: Hadrons (particles with quarks, e.g. charged pions/protons, neutrons, or jets of such particles)
    - Punch through low Z.
    - Produce secondaries through strong interactions with the nucleus in medium.
    - Unlike EM interactions, not all energy is observed.

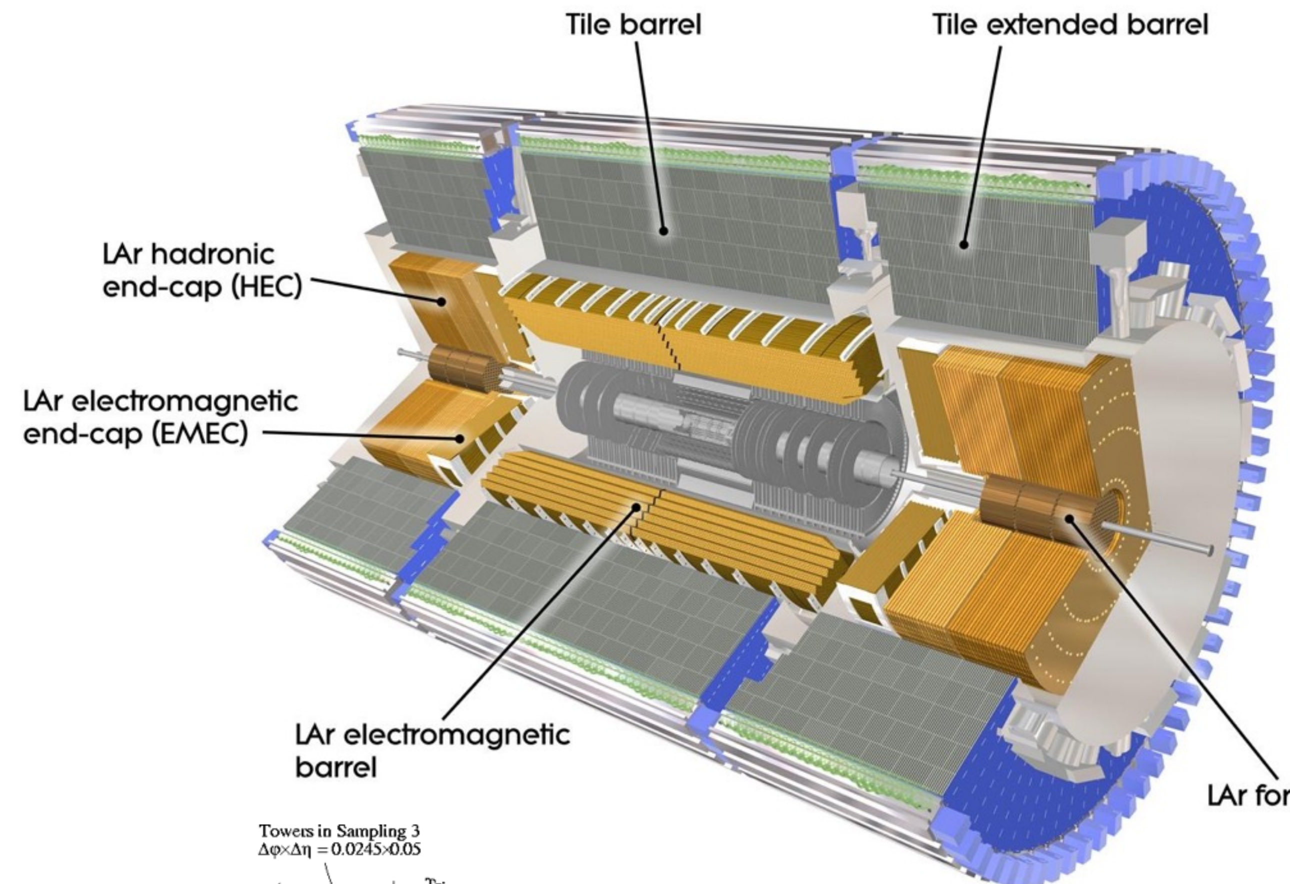


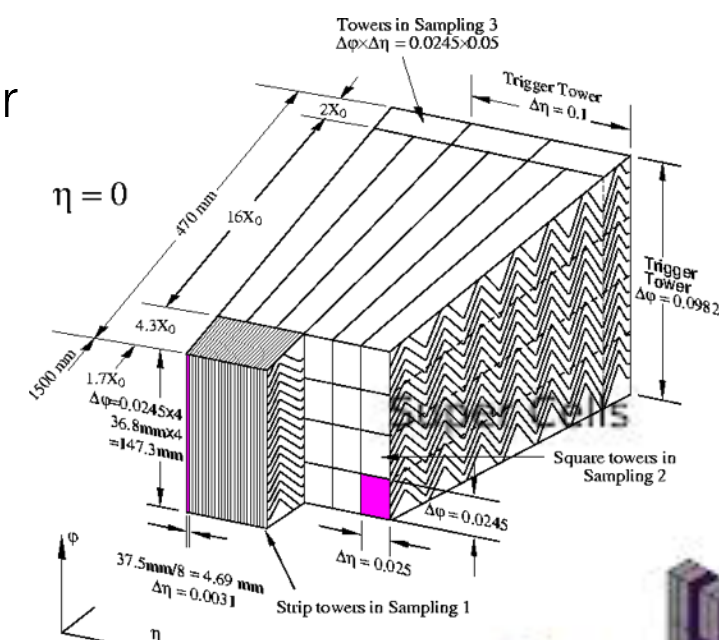


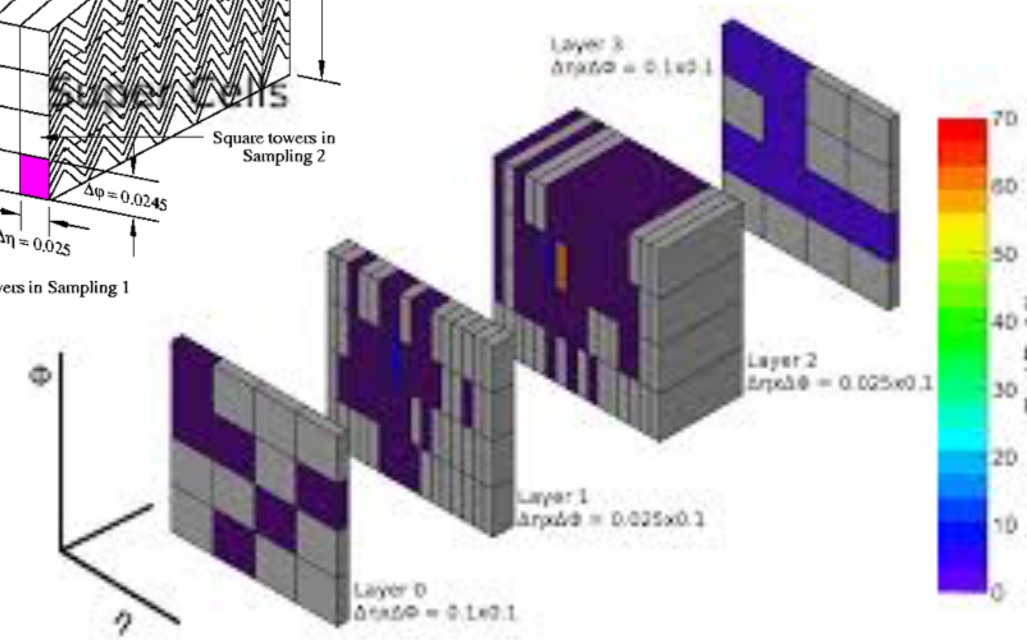


ATLAS Calorimeter

- ***Ideally suited*** for "***imaging***" ~ 64 x 36 x 7 3D Image
    - 200K Calorimeter cells measure energy deposits.
    - Interesting Challenges: non-uniform granularity, cylindrical geometry.
- ***High impact***:
    - ***Improve Identification and energy resolution*** make the peaks stand out.
    - Turn DNN into generative model for ***fast shower simulation***.
- ***High potential***: we don't use all information so room for improvement
    - *e/gamma:* take full advantage of the high granularity and accordion structure
    - *hadronic calibration:* take full advantage of longitudinal sampling and other handles
    - *particle flow:* correlate with tracks (and vertex) for hadronic calibration, taus, jet-tagging, boosted objects…
- ***Problem: Private Data…***

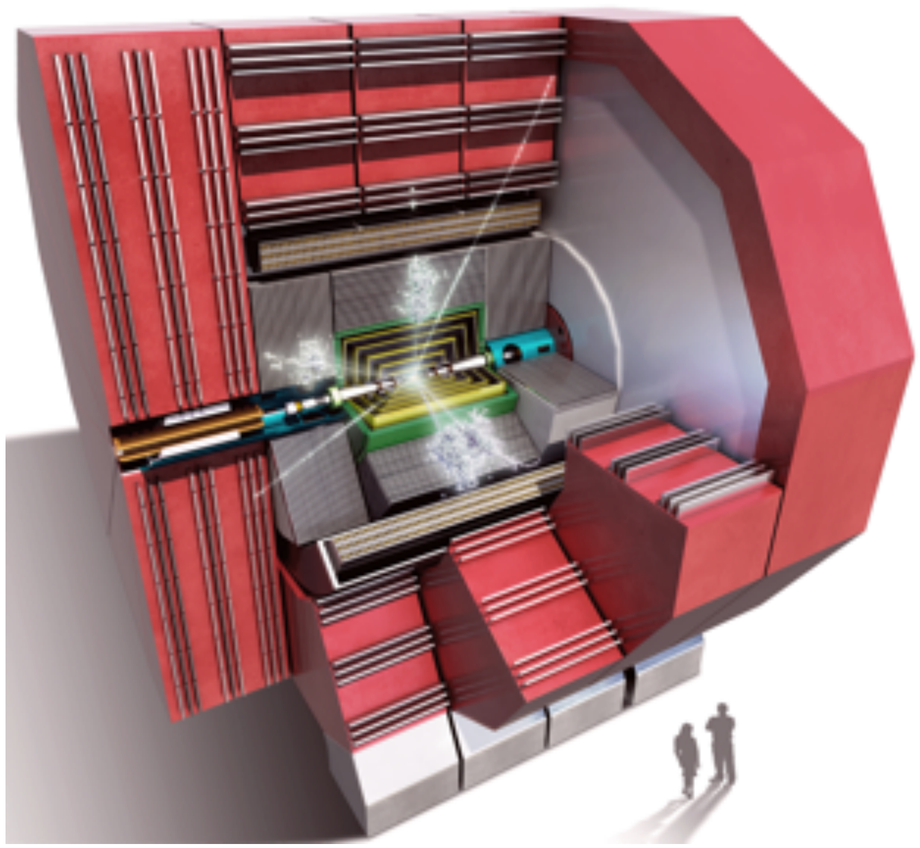


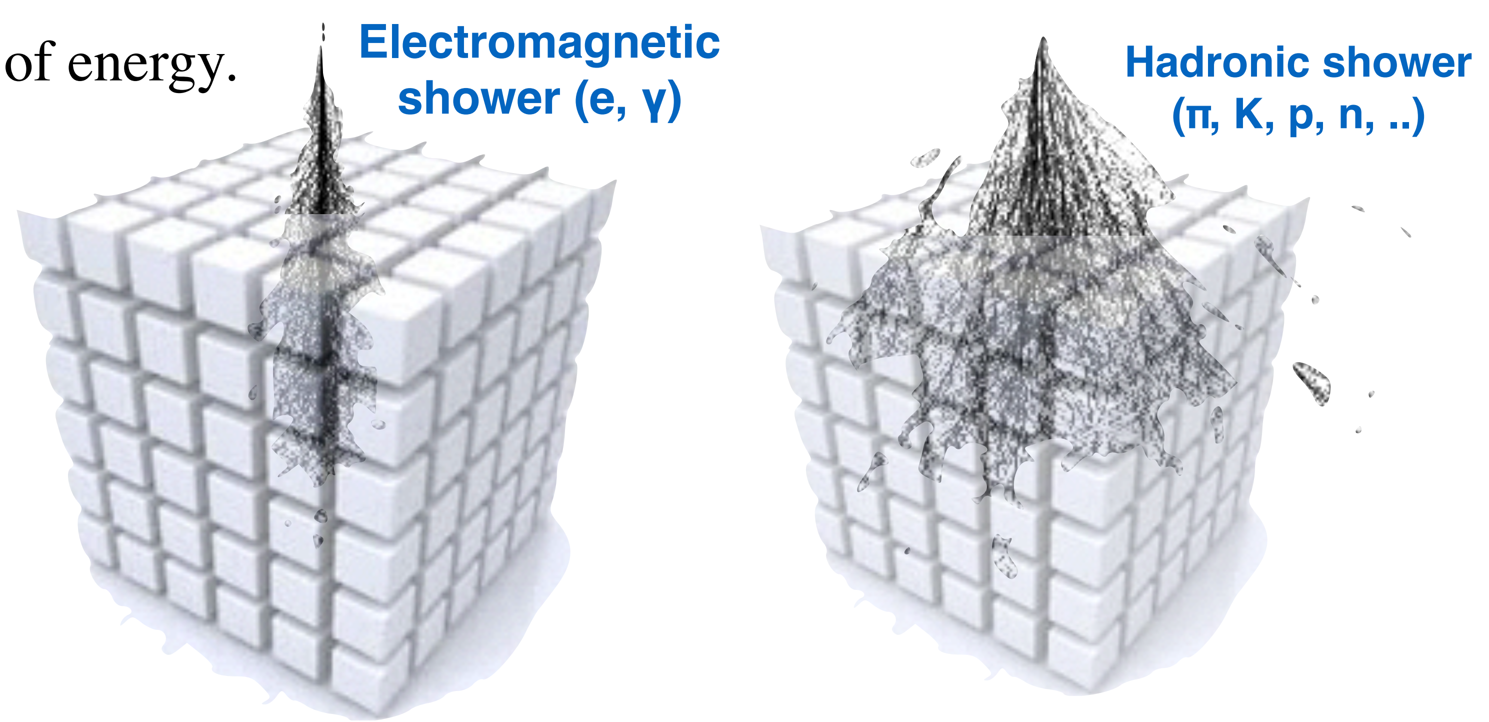




Calorimeter Dataset

- CLIC is a proposed CERN project for a linear accelerator of electrons and positrons to TeV energies (~ LHC for protons)
  - LCD is a detector concept.
  - Not a real experiment yet, so we could simulate data and make it public.
- The LCD calorimeter is an array of absorber material and silicon sensors comprising the most granular calorimeter design available
  - Data is essentially a 3D image
- With at effective eta/phi resolution of 0.003x0.003, we can down sample to get ~ ATLAS granularity: 0.025x0.1 (pre-sampler) to 0.2x0.1 Tile D.
- ***Data**:* 1 million single e, γ, $\pi^{+/-}$, π^{0}. 10-500 GeV of energy.





Calorimetry with Deep Learning: Particle Classification, Energy Regression, and Simulation for High-Energy Physics

Federico Carminati, Gulrukh Khattak, Maurizio Pierini
CERN

Amir Farbin
Univ. of Texas Arlington

Benjamin Hooberman, Wei Wei, and Matt Zhang
Univ. of Illinois at Urbana-Champaign

Vitória Barin Pacela
Univ. of Helsinki
California Institute of Technology

Sofia Vallecorsafac
Gangneung-Wonju National Univ.

Maria Spiropulu and Jean-Roch Vlimant
California Institute of Technology

Abstract

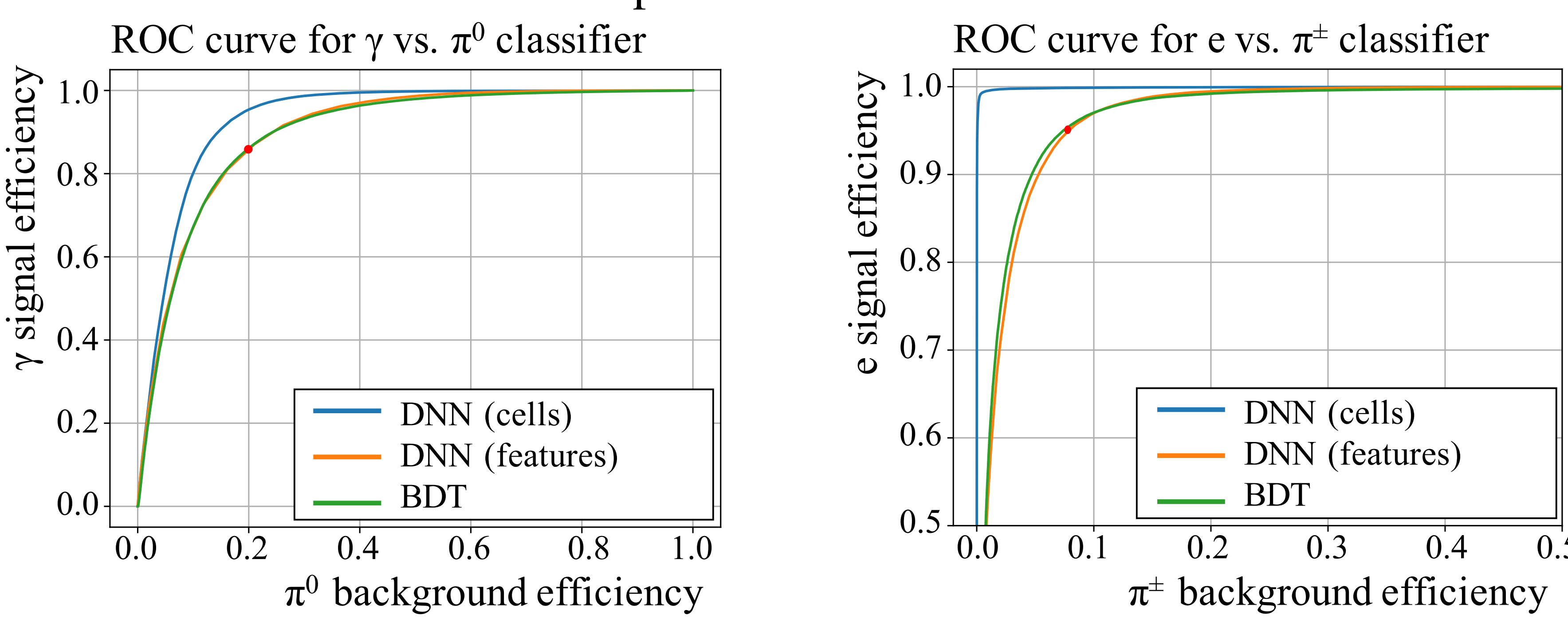
We present studies of the application of Deep Neural Networks and Convolutional Neural Networks for the classification, energy regression, and simulation of particles produced in high-energy particle collisions.We train cell-based Neural Nets that provide significant improvement in performance for particle classification and energy regression compared to feature-based Neural Nets and Boosted Decision Trees, and Generative Adversarial Networks that provide reasonable modeling of several but not all shower features.

1. e/γ Particle Identification (Classification)

- Photon/lepton ID requires factor ~10000 jet rejection
- Jet like photon/lepton classification tasks:
  - *Task 1:* Electrons vs Electromagnetic $\pi^{+/-}$ (HCAL/ECAL Energy < 0.025)
  - *Task 2:* Photons vs Merging π^0 (2γ opening angel < 0.01 rad)
- *Comparison*:
  - ***Feature based*** BDT and DNN
  - ***Cell-based*** DNN (fully connected).
- Significant Improvement with cell-based DNNs.

	γ **vs.** π^0				e **vs.** π			
Model	**acc.**	**AUC**	$\Delta\epsilon_{\text{sig}}$	ΔR_{bkg}	**acc.**	**AUC**	$\Delta\epsilon_{\text{sig}}$	ΔR_{bkg}
BDT	83.1%	89.8%	-	-	93.8%	98.0%	-	-
DNN (features)	82.8%	90.2%	0.9%	0.95	93.6%	98.0%	-0.1%	0.95
DNN (cells)	87.2%	93.5%	9.4%	1.63	99.4%	99.9%	4.9%	151

Table 1: Performance parameters for BDT and DNN classifiers.



2. *Energy Calibration (Regression)*

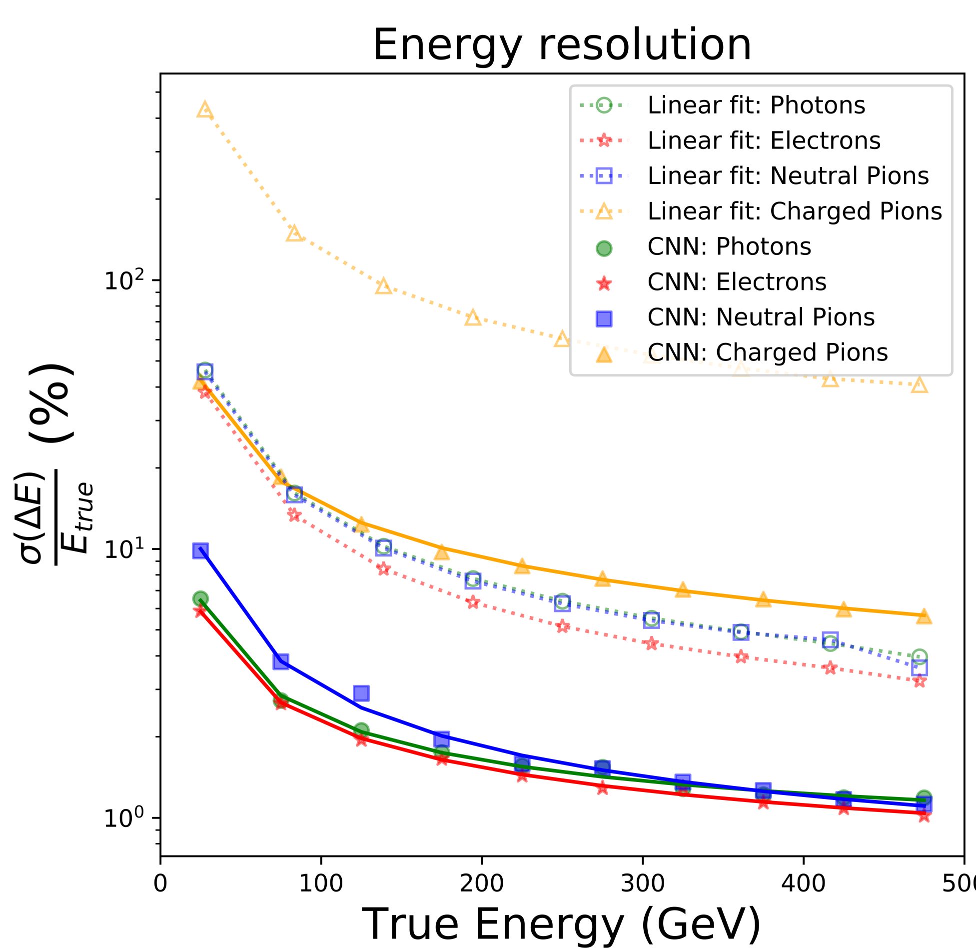
- Energy ***resolution improves with energy***:
  - $\sigma(E) / E = a/\sqrt{E} \oplus b/E \oplus c$.
  - a = sampling, b = noise, c = leakage.
- *Comparison*:
  - ***Simple calibration***: Sum energies (no noise) and scale.
  - ***CNN calibration***: Cells → Particle energy
- Significant Improvement with CNN

Simple Linear Model

Particle Type	a	b	c
Photons	55.5	1.85	1245
Electrons	42.3	1.51	1037
Neutral pions	55.3	1.71	1222
Charged pions	442	25	11706

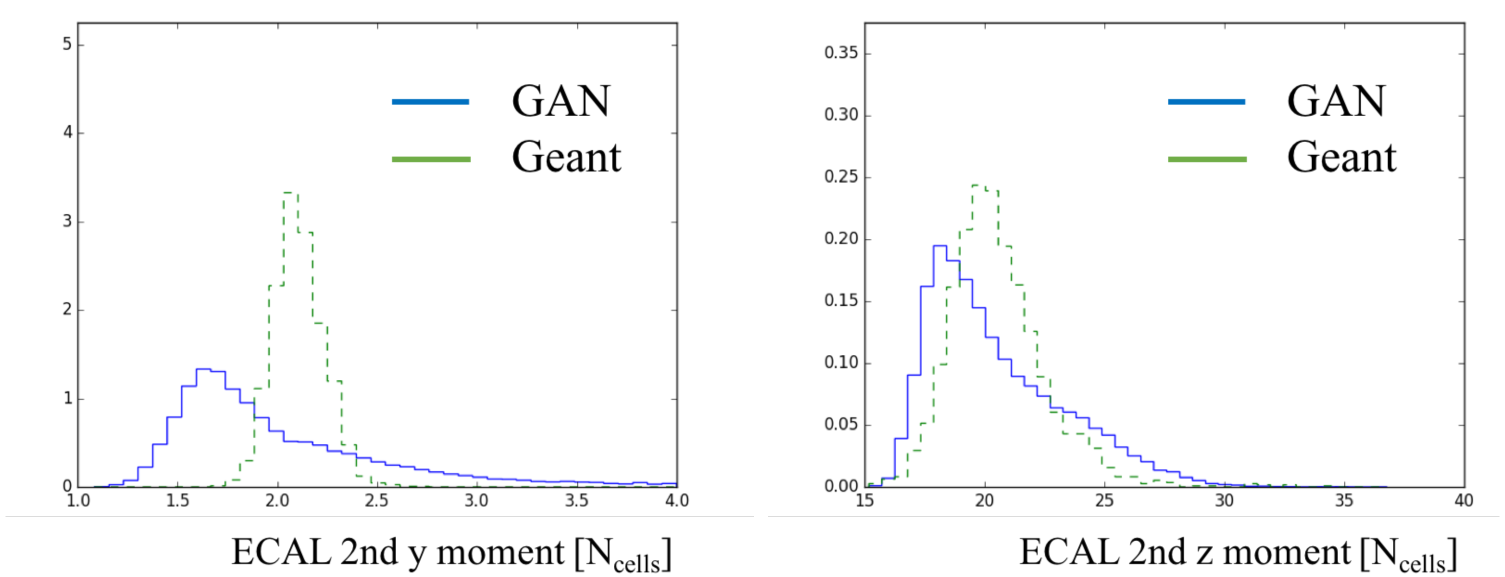
CNN Model

Particle Type	a	b	c
Photons	18.3	0.75	131
Electrons	18.7	0.574	111
Neutral pions	19.3	0.45	231
Charged pions	114	1.02	893



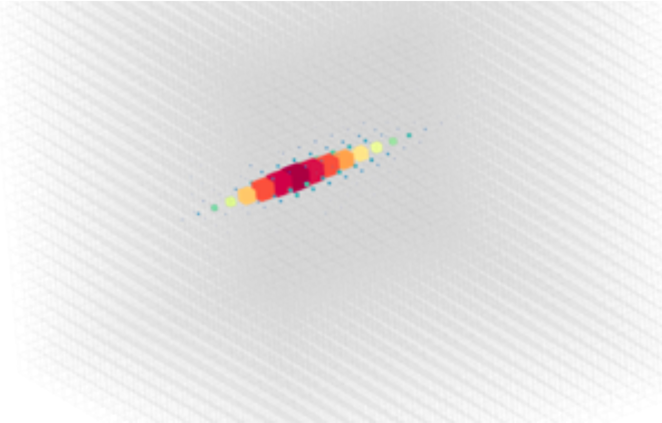
3. Simulation (Generative Model)

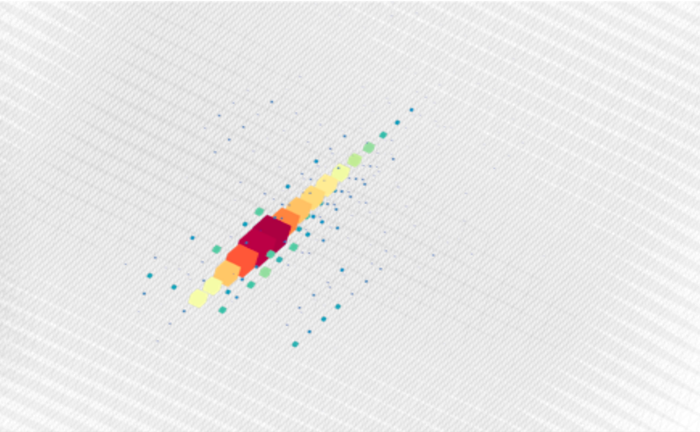
- Physics measurements typically require extremely detailed and precise simulation,
  - Software packages (e.g. Geant4) simulated the well understood ***micro-physics*** governing the interaction of particles with matter.
  - Generally very CPU intensive
    - *Example*: ATLAS experiment uses half of the experiment's computing resources for simulation.
- *Task:* ***CNN GAN conditioned on particle energy***
  - Accelerate simulation by many orders of magnitude.
- Promising start… but not yet faithfully reproducing all commonly used features extracted from generated images.

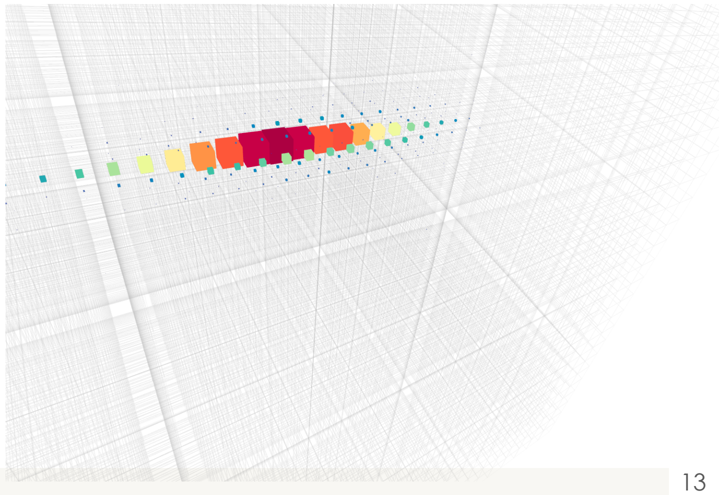


Some images

- Slice energy spectrum
- Start with photons & electrons

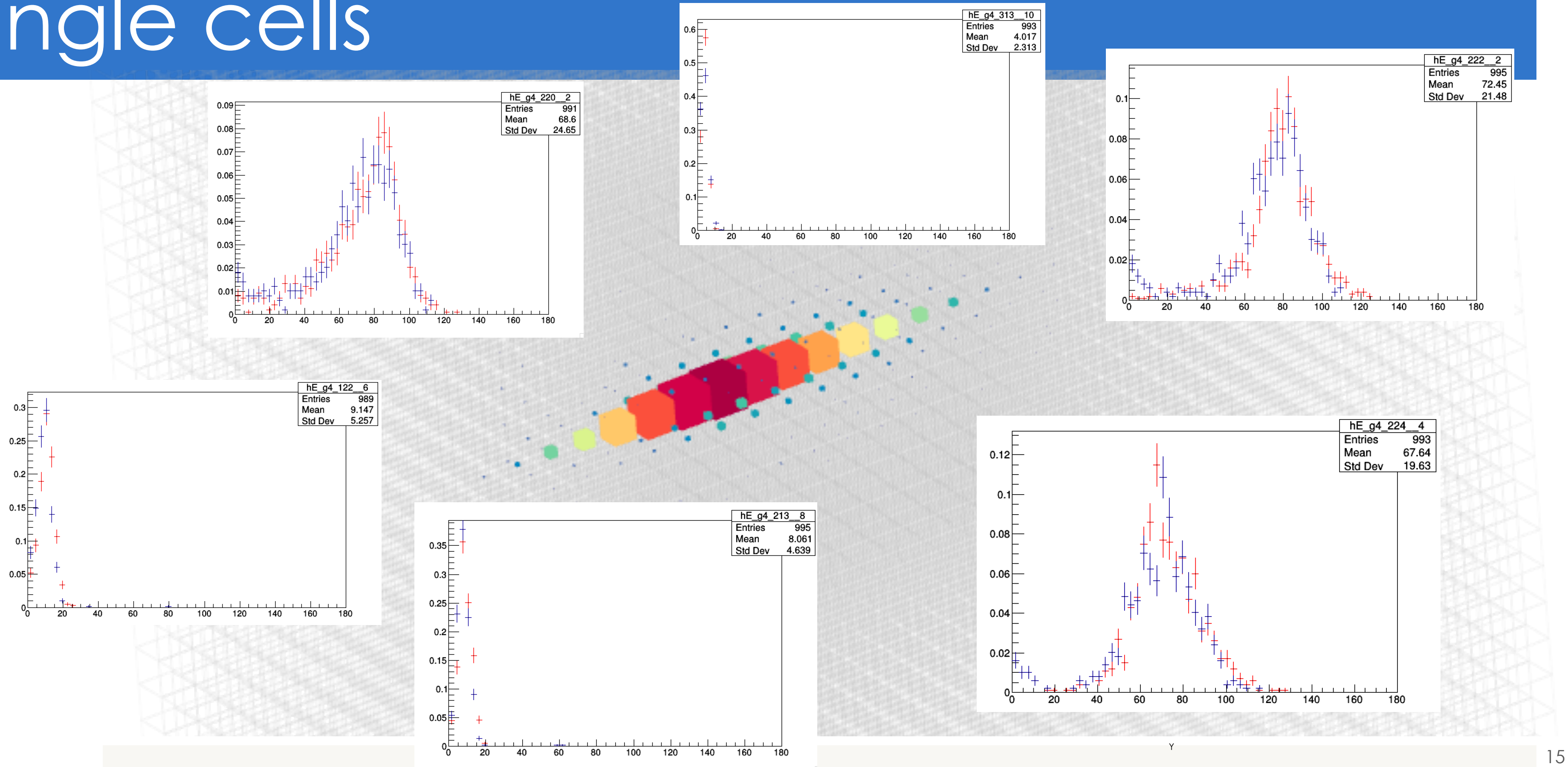






Preliminary

Single cells



Jet Physics with Deep Learning

Modern Machine Learning

for Classification, Regression, and Generation in Jet Physics

BERKELEY LAB

Benjamin Nachman

Lawrence Berkeley National Laboratory

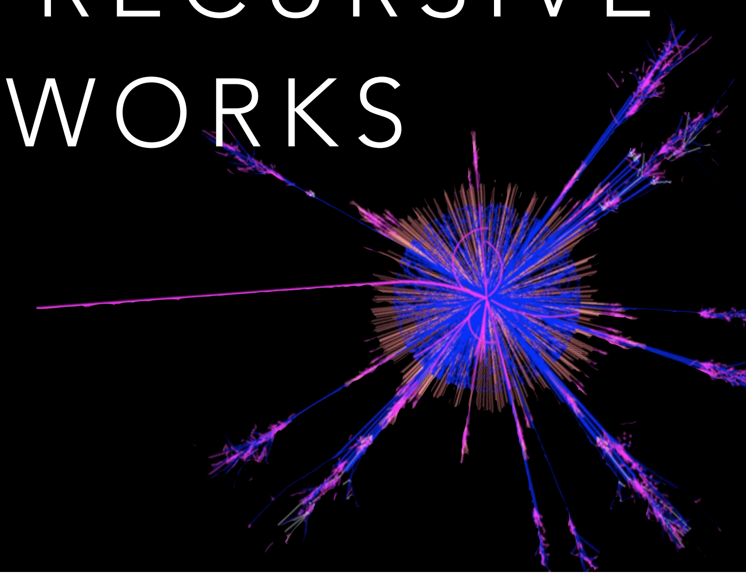
CERN Data Science Seminar, Novem

Next slides stolen from Ben Nachmans and Kyle Cranmer's Excellent CERN Seminars

Center for Cosmology and Particle Physics

ARXIV:1702.00748

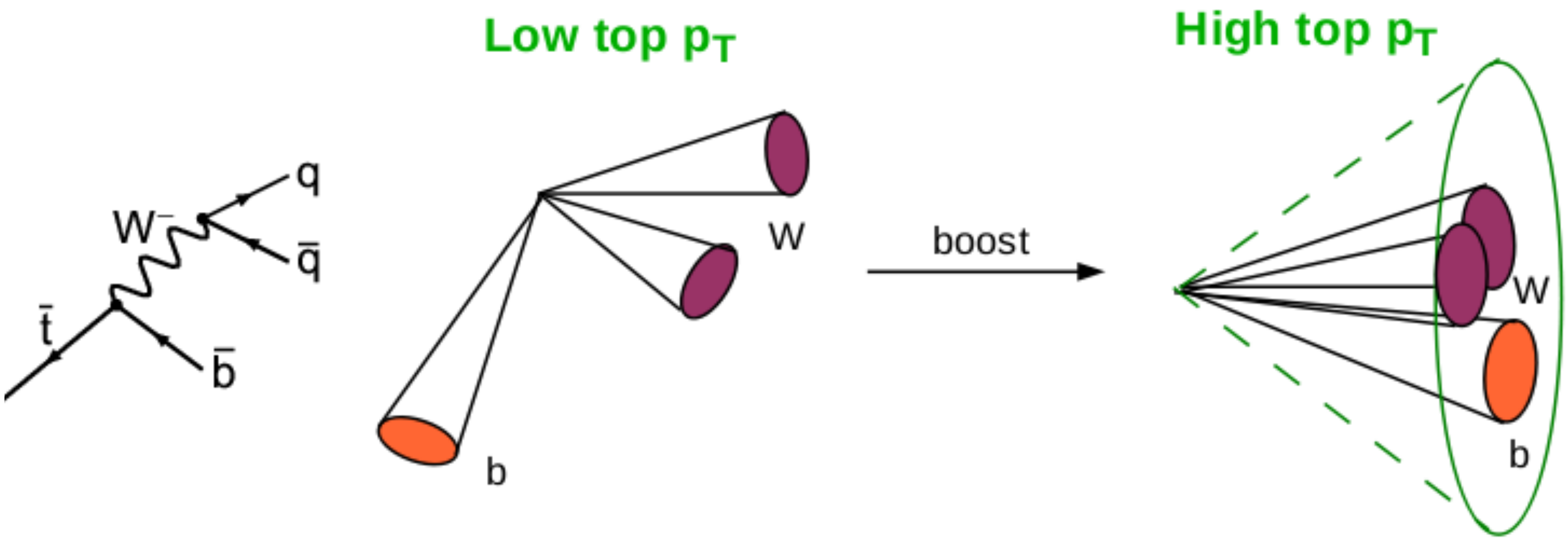
QCD-AWARE RECURSIVE NEURAL NETWORKS



@KyleCranmer
New York University
Department of Physics
Center for Data Science

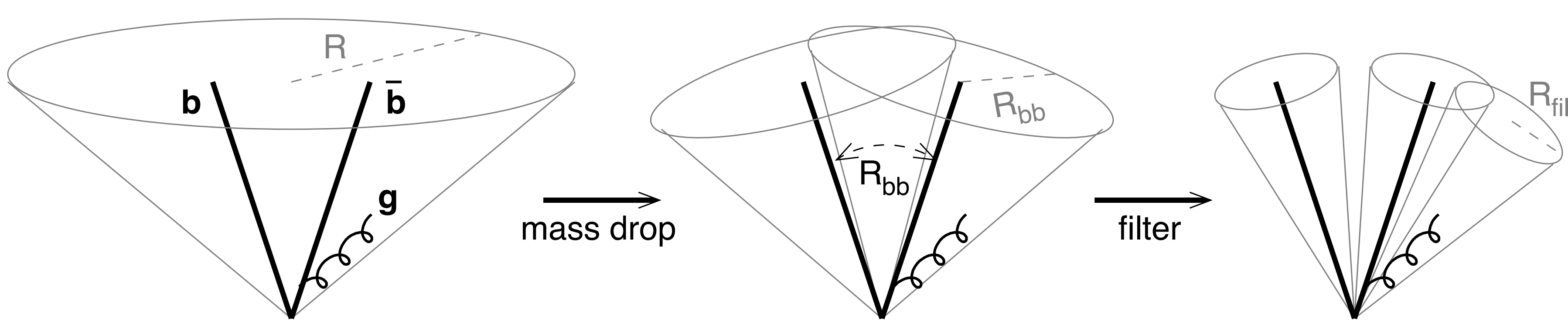
with:
Gilles Louppe
Kyunghyun Cho
Joan Bruna
Cyril Becot

Many scenarios for physics Beyond the Standard Model include highly boosted W, Z, H bosons or top quarks



Identifying these rests on subtle substructure inside jets

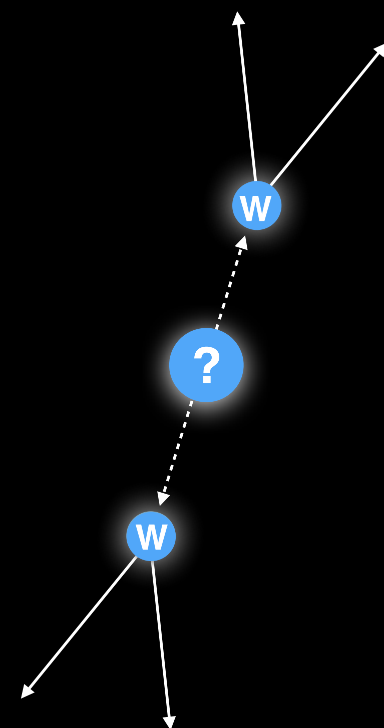
- an enormous number of theoretical effort in developing observables and techniques to tag jets like this



Goal: Find W jets in an enormous sea of generic q/g jets

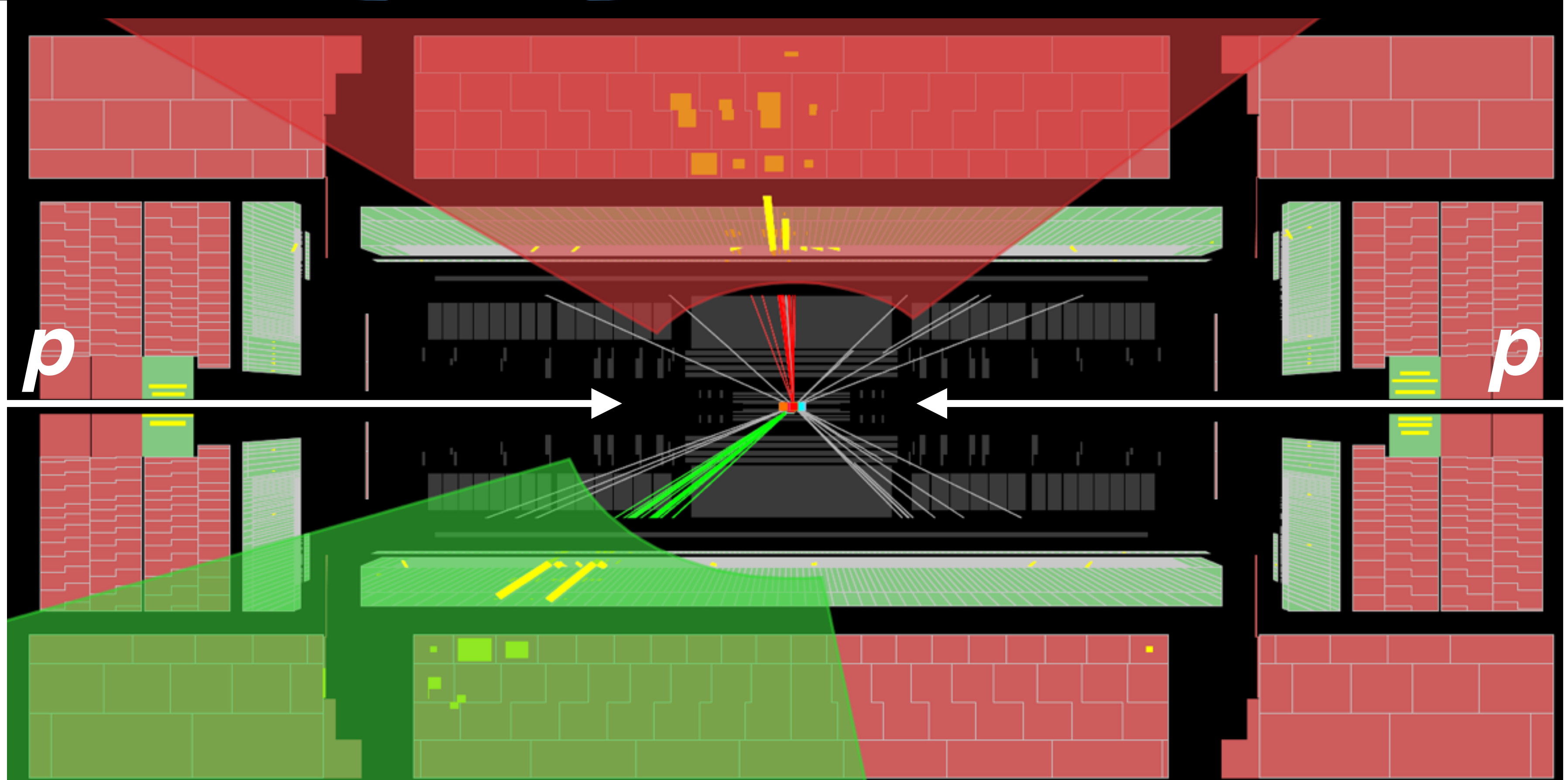
These jets have a non-trivial structure!

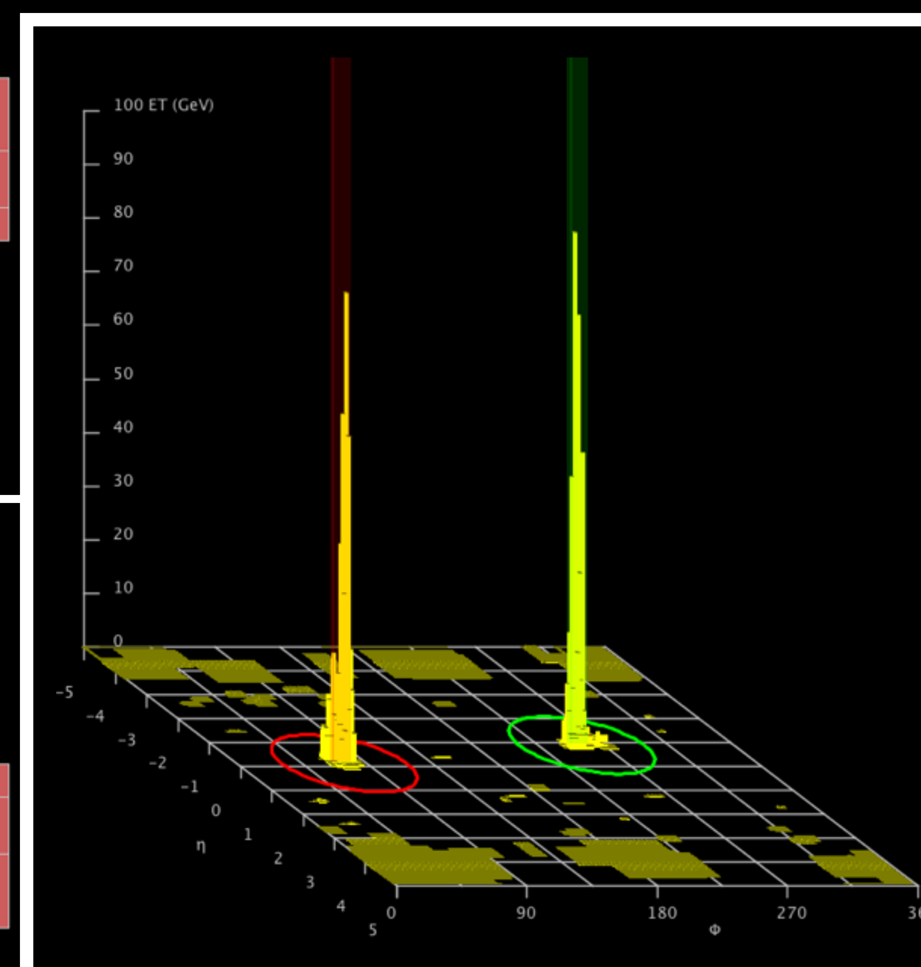
W bosons are naturally boosted if they result from the decay of something even heavier



Searching for new particles decaying into boosted W bosons requires **looking at the radiation pattern inside jets**

like a digital image!

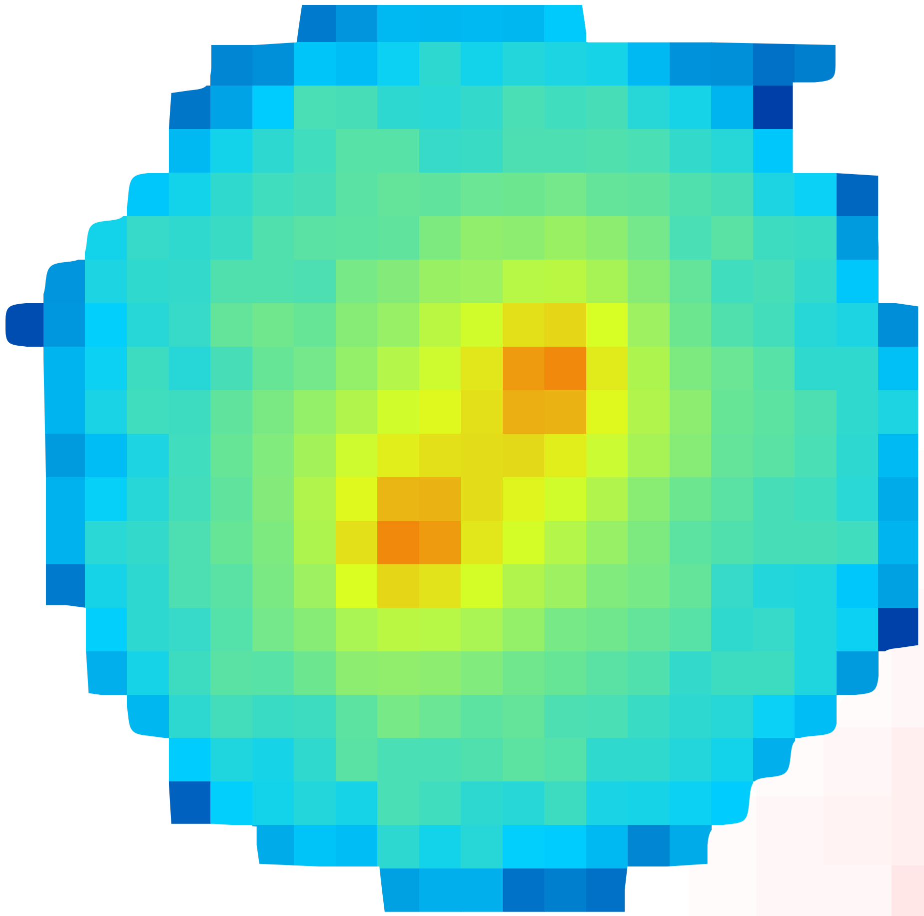




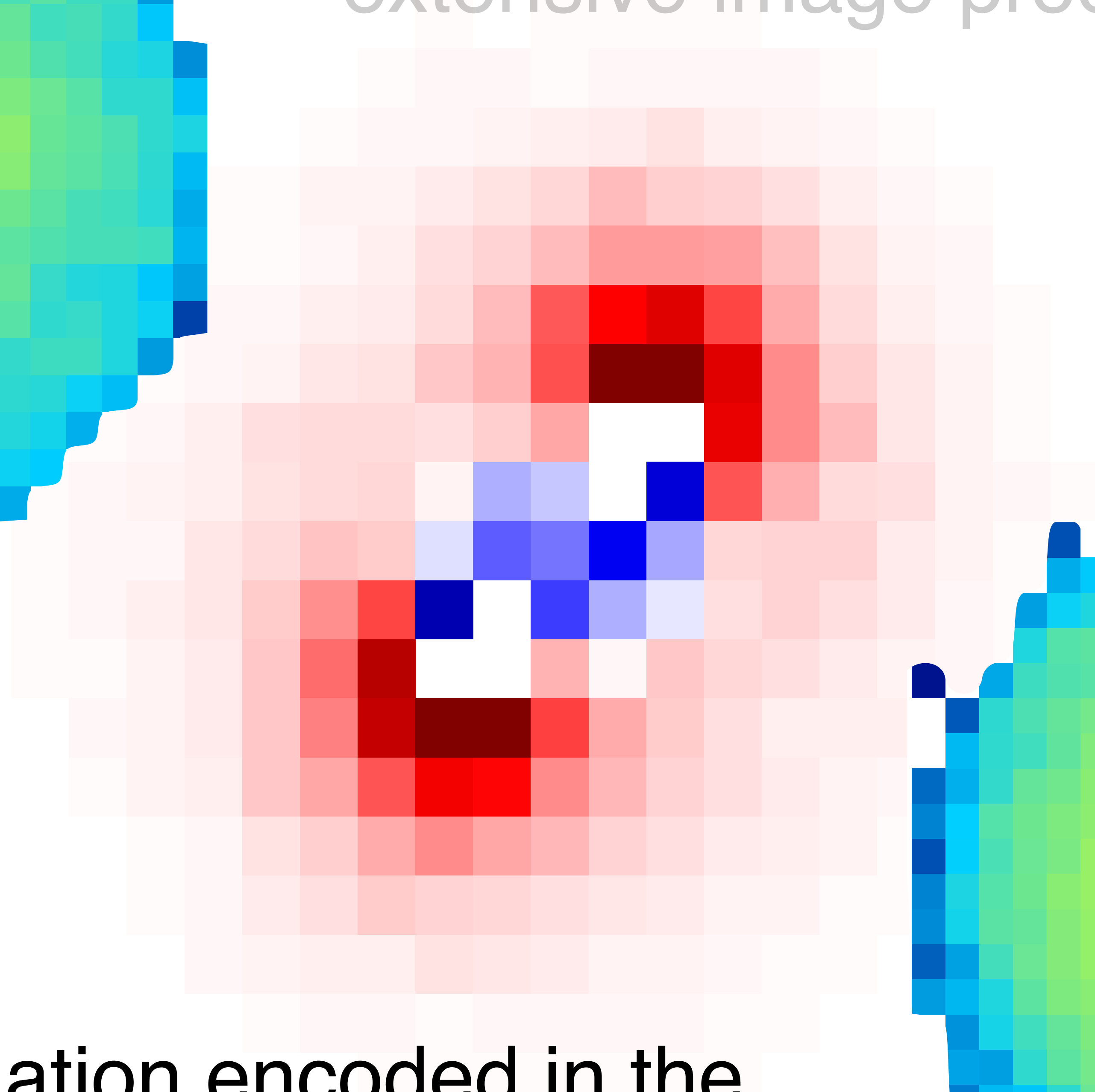
Why images?

Can directly visualize physics

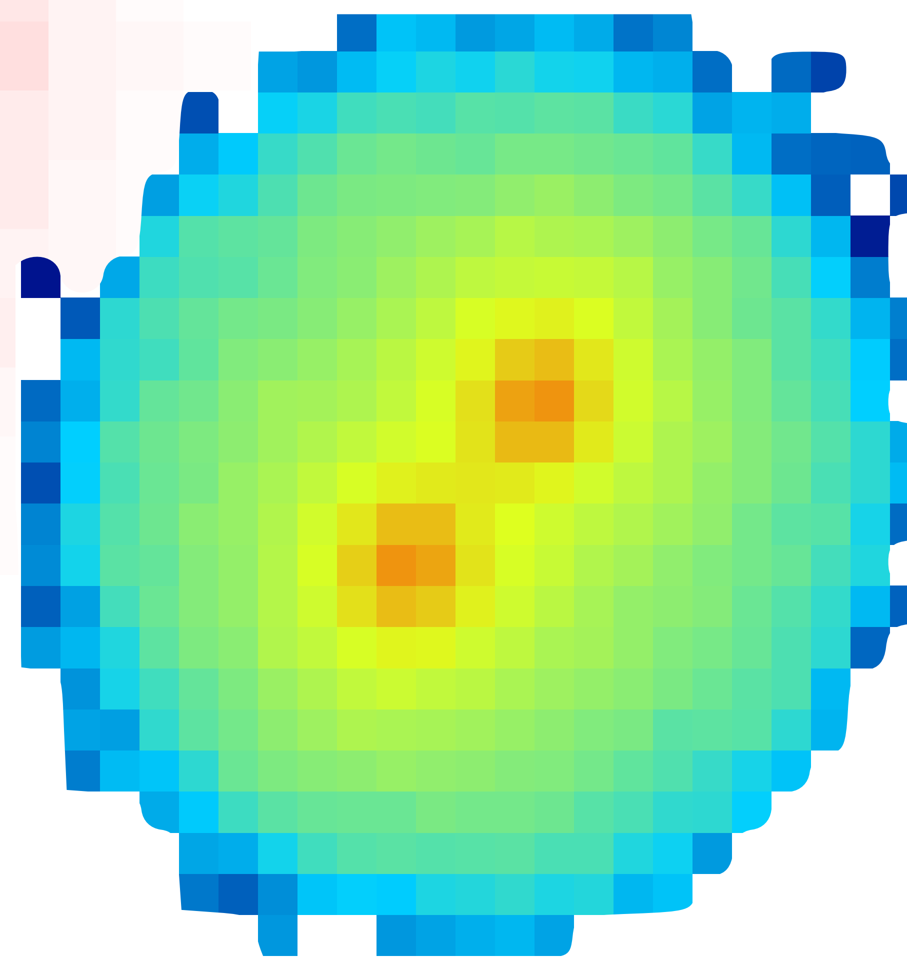
and we can benefit from the extensive image processing literature



W ⇢ $q\bar{q}$



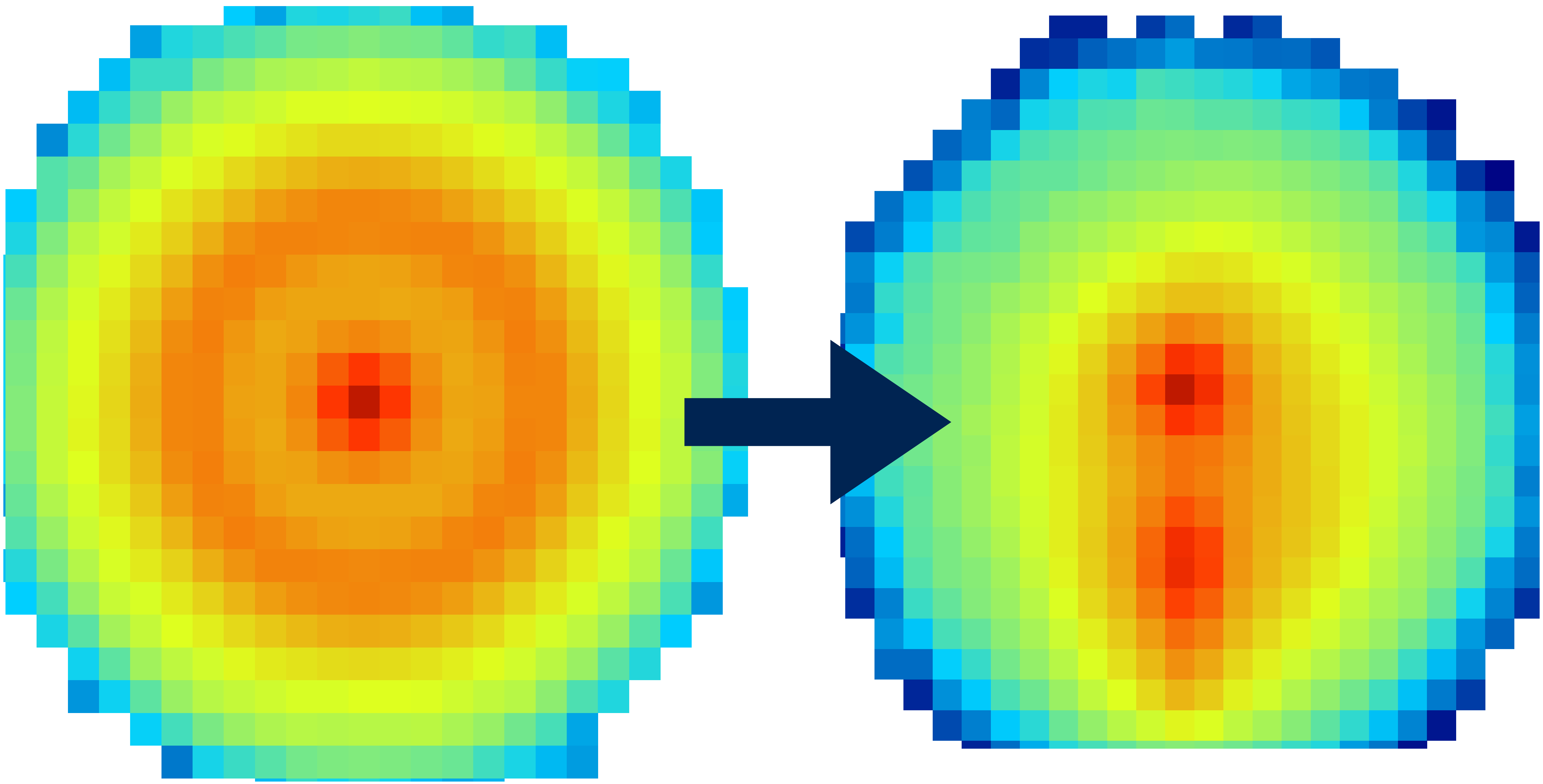
g ⇢ $q\bar{q}$



there is information encoded in the physical distance between pixels

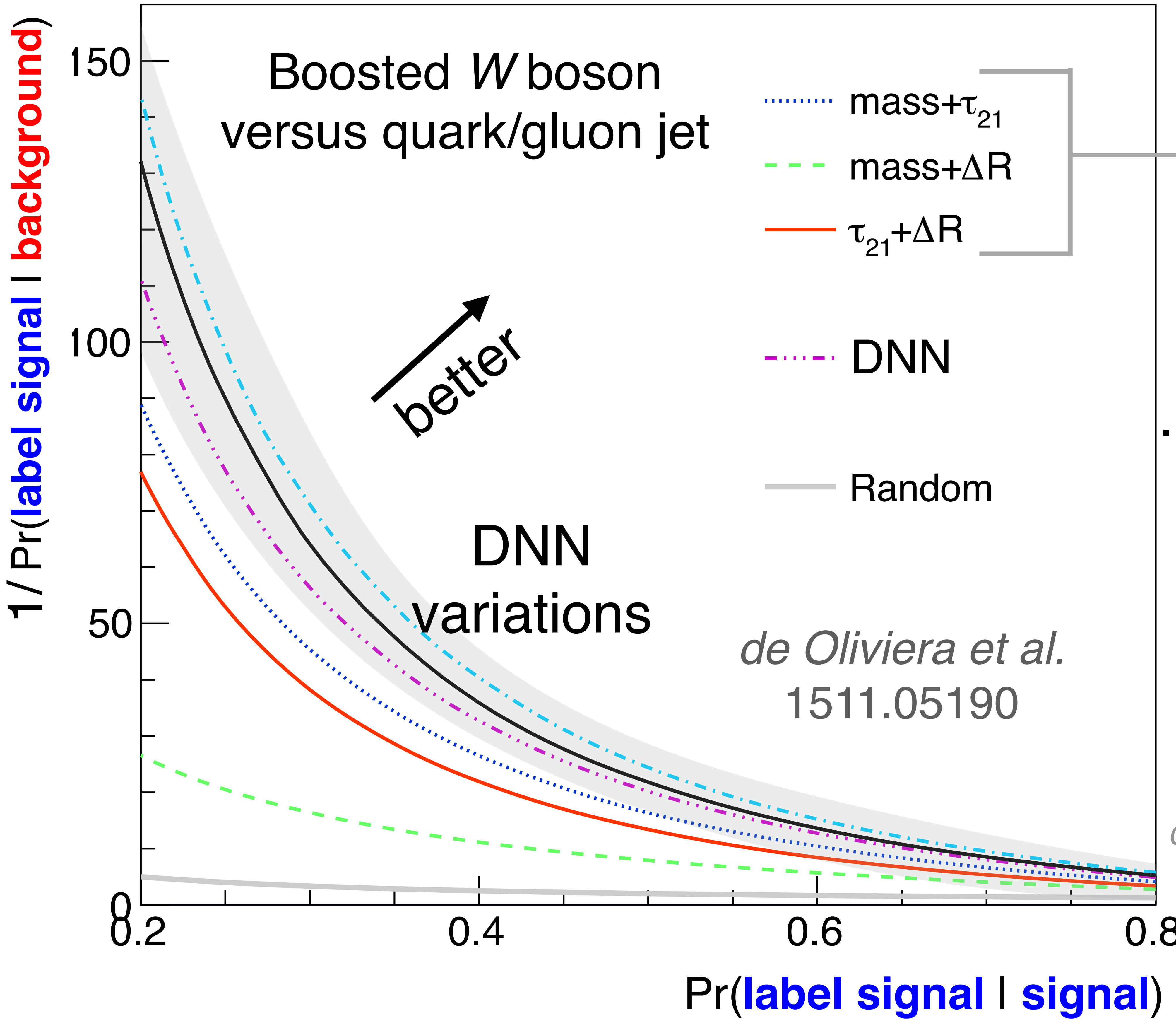
Pre-processing & spacetime symmetries

One of the first typical steps is pre-processing



Can help to learn faster & smarter; but must be careful!

Modern Deep NN's for Classification



mass, τ_{21}, ΔR are all simple functions of the image

…what the DNN is learning is active R&D!

See also

L. Almeida et al. 1501.05968

Baldi et al. 1603.09349

J. Barnard et al. 1609.00607

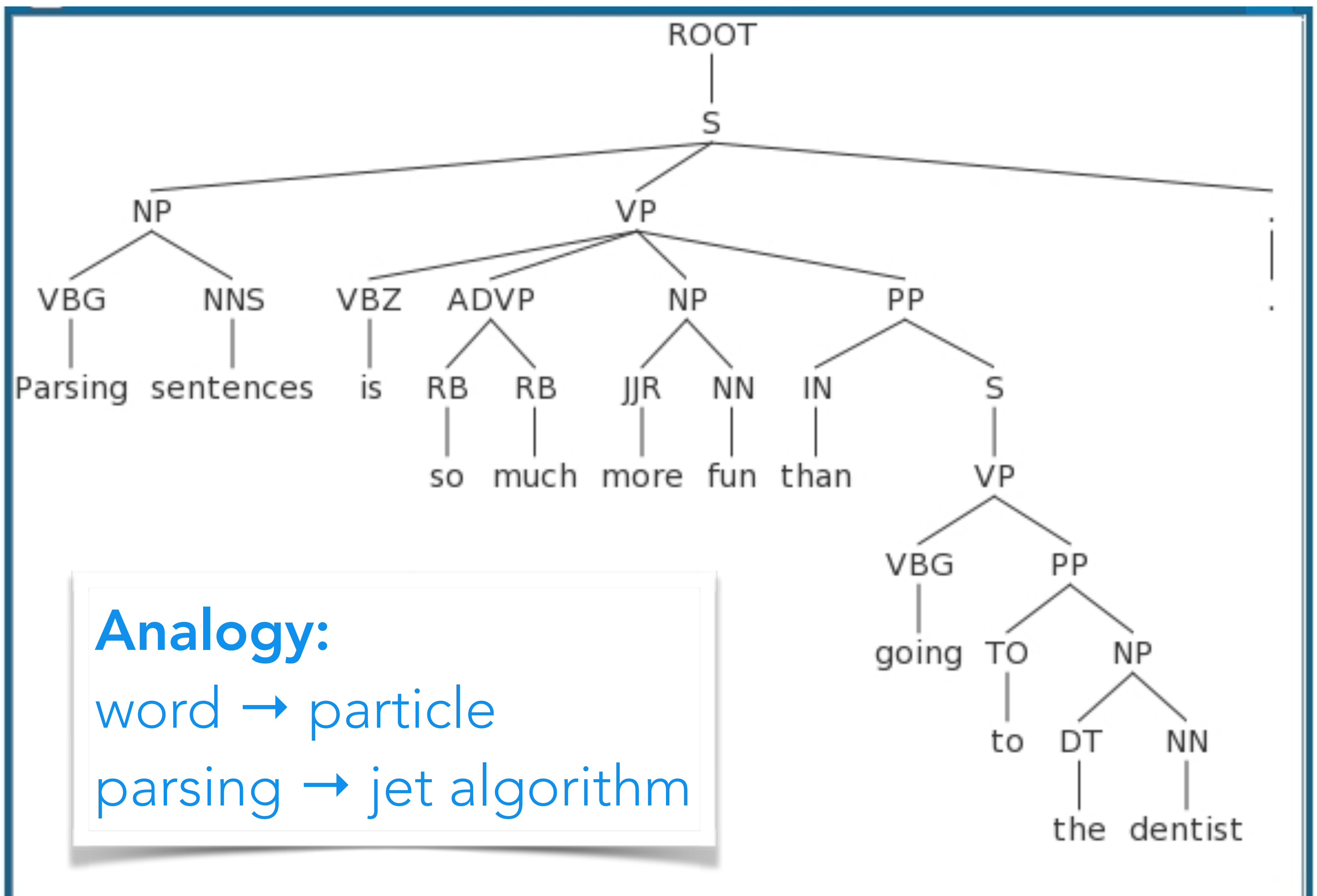
P. Komiske et al. 1612.01551

G. Kasieczka et al. 1701.08784

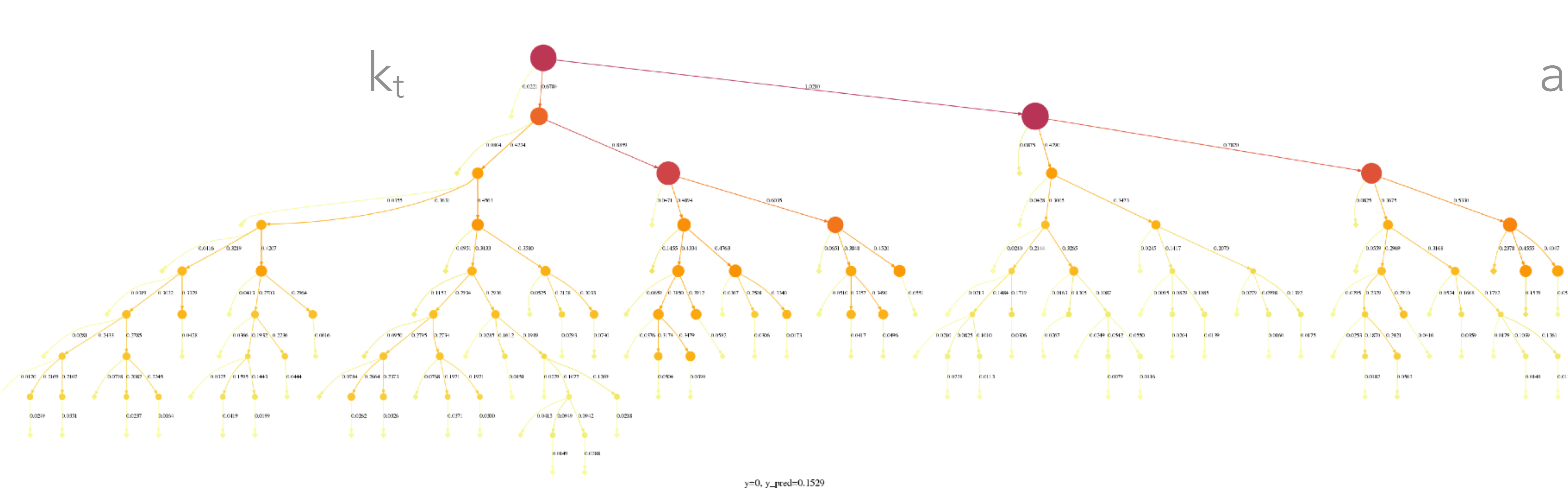
W. Bhimji et al. 1711.03573

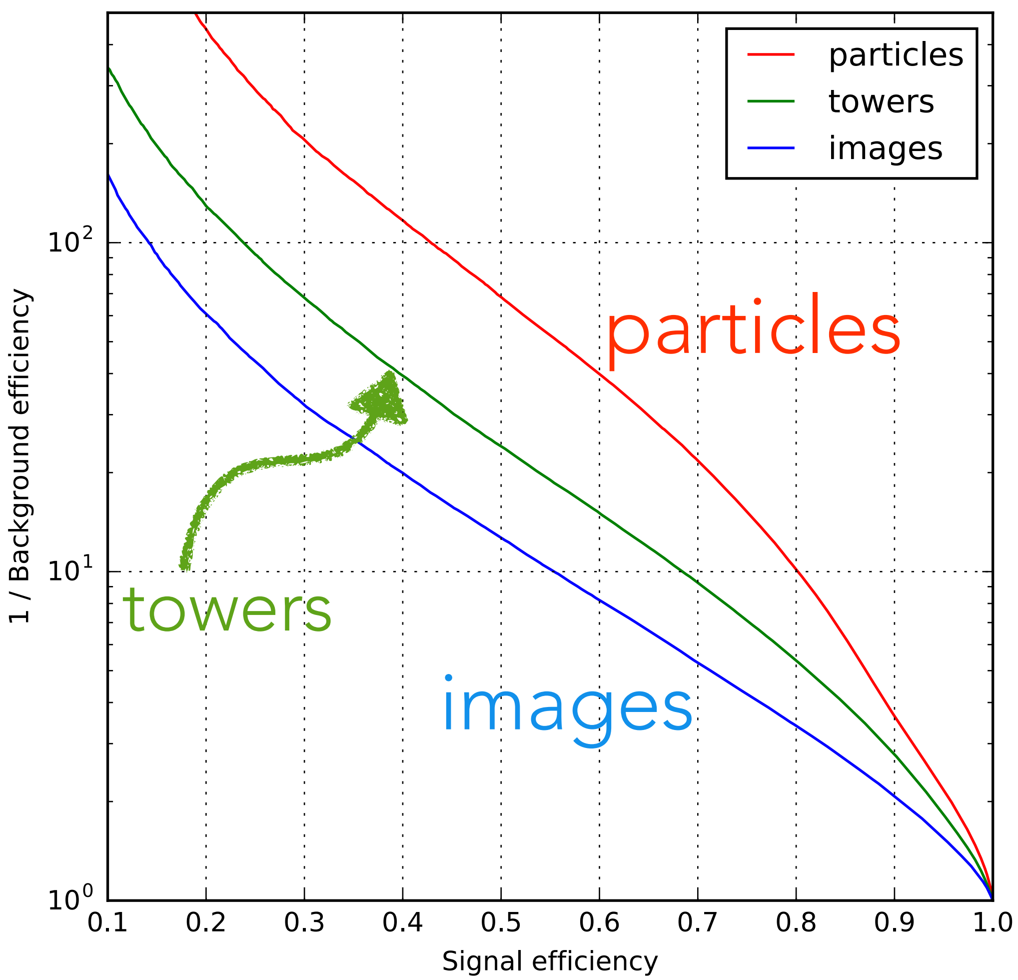
Recursive Neural Networks showing great performance for Natural Language Processing tasks

- neural network's topology given by parsing of sentence!



QCD-INSPIRED RECURSIVE NEURAL NETWORKS





- W-jet tagging example using data from Dawe, et al arXiv:1609.00607
- down-sampling by projecting into images looses information
- RNN needs much less data to train!

Neural Message Passing for Jet Physics

Isaac Henrion, Johann Brehmer, Joan Bruna, Kyunghun Cho, Kyle Cranmer
Center for Data Science
New York University
New York, NY 10012
{henrion*, johann.brehmer, bruna, kyunghyun, kyle.cranmer*}@nyu.edu

Gilles Louppe
Department of Computer Science
University of Liège
Belgium
g.louppe@ulg.ac.be

Gaspar Rochette
Department of Computer Science
École Normale Supérieure
Paris, France
gaspar.rochette@ens.fr

Abstract

Supervised learning has incredible potential for particle physics, and one application that has received a great deal of attention involves collimated sprays of particles called jets. Recent progress for jet physics has leveraged machine learning techniques based on computer vision and natural language processing. In this work, we consider message passing on a graph where the nodes are the particles in a jet. We design variants of a message-passing neural network (MPNN); (1) with a learnable adjacency matrix, (2) with a learnable symmetric adjacency matrix, and (3) with a set2set aggregated hidden state and MPNN with an identity adjacency matrix. We compare these against the previously proposed recursive neural network with a fixed tree structure and show that the MPNN with a learnable adjacency matrix and two message-passing iterations outperforms all the others.

Table 1: Summary of classification performance for several approaches.

Network	Iterations	ROC AUC	$R_{\epsilon=50\%}$
RecNN-k_t (without gating) [10]	1	0.9185 ± 0.0006	68.3 ± 1.8
RecNN-k_t (with gating) [10]	1	0.9195 ± 0.0009	74.3 ± 2.4
RecNN-desc-p_T (without gating) [10]	1	0.9189 ± 0.0009	70.4 ± 3.6
RecNN-desc-p_T (with gating) [10]	1	0.9212 ± 0.0005	$\mathbf{83.3} \pm \mathbf{3.1}$
RelNet	1	0.9161 ± 0.0029	67.69 ± 6.80
MPNN (directed)	1	0.9196 ± 0.0015	89.35 ± 3.54
MPNN (directed)	2	0.9223 ± 0.0008	$\mathbf{98.26} \pm \mathbf{4.28}$
MPNN (directed)	3	0.9188 ± 0.0031	85.93 ± 8.50
MPNN (undirected)	1	0.9193 ± 0.0015	86.41 ± 3.80
MPNN (undirected)	2	0.8949 ± 0.1004	$\mathbf{97.27} \pm \mathbf{5.02}$
MPNN (undirected)	3	0.9185 ± 0.0036	84.53 ± 8.64
MPNN (set, directed)	1	0.9189 ± 0.0017	88.23 ± 4.53
MPNN (set, directed)	2	0.9191 ± 0.0046	87.46 ± 14.14
MPNN (set, directed)	3	0.9176 ± 0.0049	$\mathbf{88.33} \pm \mathbf{9.84}$
MPNN (set, undirected)	1	0.9196 ± 0.0014	85.65 ± 4.48
MPNN (set, undirected)	2	0.9220 ± 0.0007	$\mathbf{94.70} \pm \mathbf{2.95}$
MPNN (set, undirected)	3	0.9158 ± 0.0054	75.94 ± 12.54
MPNN (id)	1	0.9169 ± 0.0013	74.75 ± 2.65
MPNN (id)	2	0.9162 ± 0.0020	74.41 ± 3.50
MPNN (id)	3	0.9158 ± 0.0029	74.51 ± 5.20

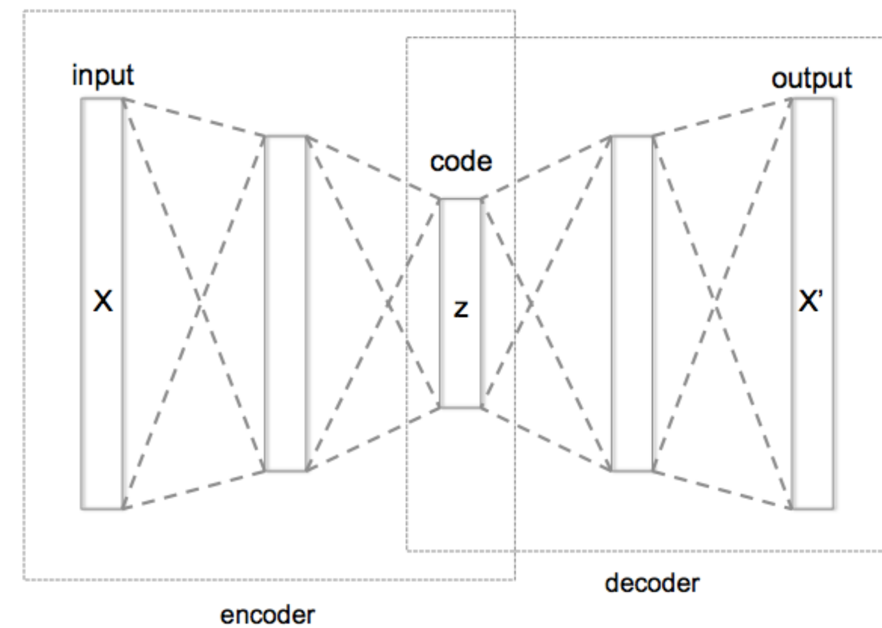
➔ Making your data into an image isn't always the best idea.

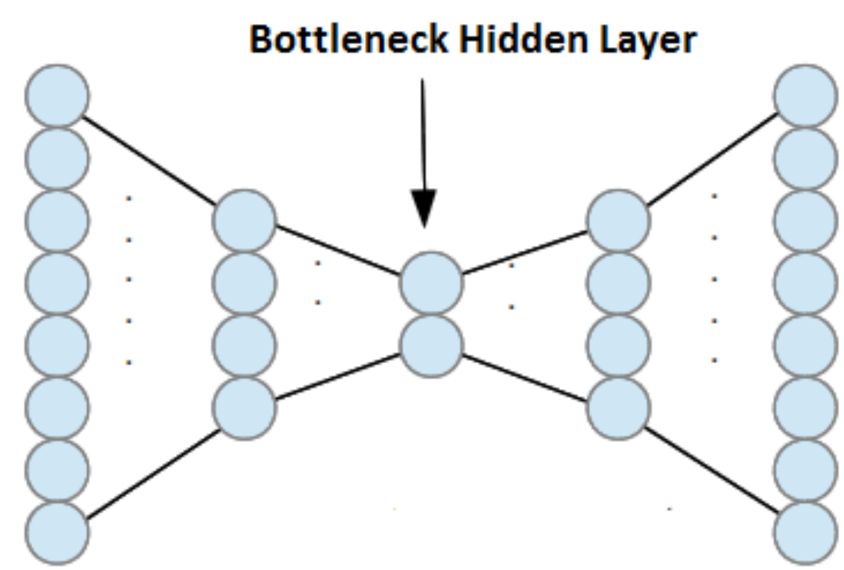
Final Remarks

- Deep Learning can help get the most out of a given accelerator, detector, and data set.
- Deep Learning can help design better experiments.
- Deep Learning may help address HEP problems:
    - US's flagship project, DUNE, and other LArTPC experiments need help with automatic reconstruction. They are ideally suited for DNNs.
    - Computing for HL-LHC will be prohibitively expensive unless we find some clever techniques.
- Over the past couple of years many DL solutions have been demonstrated, often with toys…
- Over the next few years:
    - Bring them into our experiments and make them realistic
    - Target physics measurements where DL can have significant impact
    - Move DL to production and make DL mainstream
- Deep Learning will fundamentally change how scientific computing is done…

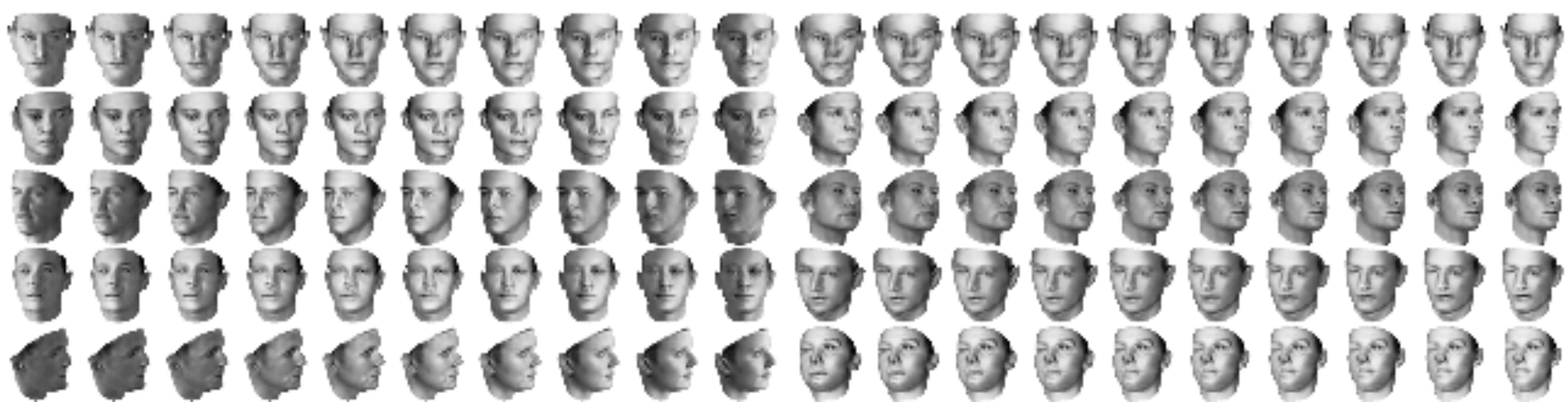
Semi-supervised Learning

- Basic idea: Train network to ***reproduce the input***.
- Example: ***Auto-encoders***
  - ***De-noising auto-encoders***: add noise to input only.
  - ***Sparse auto-encoders***:
    - ***Sparse latent (code) representation*** can be exploited for ***Compression, Clustering, Similarity testing,*** …
  - ***Anomaly Detection***
    - Reconstruction Error
    - Outliers in latent space
  - ***Transfer Learning***
    - Small labeled training sample?
      - Train auto-encoder on large unlabeled dataset (e.g. data).
      - Train in latent space on small labeled data. (e.g. rare signal MC).
- Easily think of a dozen applications.





(a) Azimuth (pose) (b) Elevation



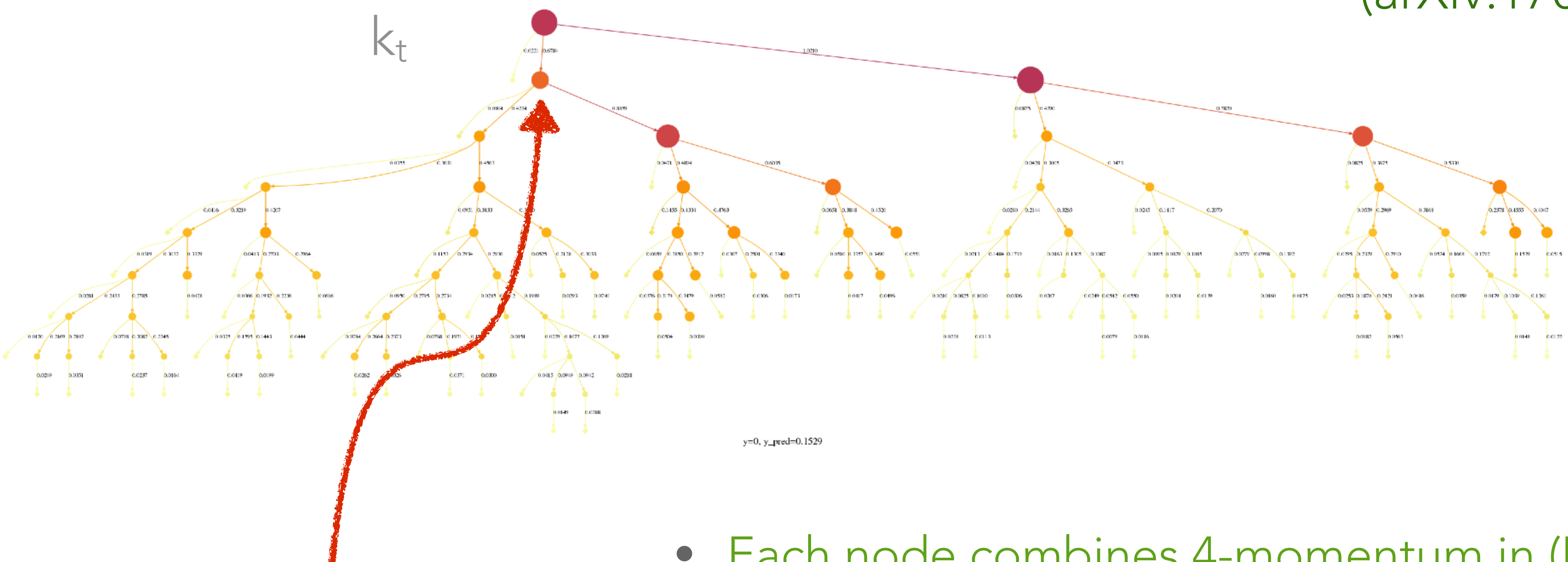
(c) Lighting (d) Wide or Narrow

Figure 3: **Manipulating latent codes on 3D Faces:** We show the effect of the learned continuous latent factors on the outputs as their values vary from -1 to 1. In (a), we show that one of the continuous latent codes consistently captures the azimuth of the face across different shapes; in (b), the continuous code captures elevation; in (c), the continuous code captures the orientation of lighting; and finally in (d), the continuous code learns to interpolate between wide and narrow faces while preserving other visual features. For each factor, we present the representation that most resembles prior supervised results [7] out of 5 random runs to provide direct comparison.

https://arxiv.org/pdf/1606.03657.pdf

QCD-INSPIRED RECURSIVE NEURAL NETWORKS

(arXiv:1702.00748)



$$\mathbf{h}_k^{\text{jet}} = \begin{cases} \mathbf{u}_k & \text{if } k \text{ is a leaf} \\ \mathbf{z}_H \odot \tilde{\mathbf{h}}_k^{\text{jet}} + \mathbf{z}_L \odot \mathbf{h}_{k_L}^{\text{jet}} + & \text{otherwise} \\ \hookrightarrow \mathbf{z}_R \odot \mathbf{h}_{k_R}^{\text{jet}} + \mathbf{z}_N \odot \mathbf{u}_k & \end{cases}$$

$$\mathbf{u}_k = \sigma\left(W_u g(\mathbf{o}_k) + b_u\right)$$

$$\mathbf{o}_k = \begin{cases} \mathbf{v}_{i(k)} & \text{if } k \text{ is a leaf} \\ \mathbf{o}_{k_L} + \mathbf{o}_{k_R} & \text{otherwise} \end{cases}$$

$$\tilde{\mathbf{h}}_k^{\text{jet}} = \sigma\left(W_{\tilde{h}} \begin{bmatrix} \mathbf{r}_L \odot \mathbf{h}_{k_L}^{\text{jet}} \\ \mathbf{r}_R \odot \mathbf{h}_{k_R}^{\text{jet}} \\ \mathbf{r}_N \odot \mathbf{u}_k \end{bmatrix} + b_{\tilde{h}} \right)$$

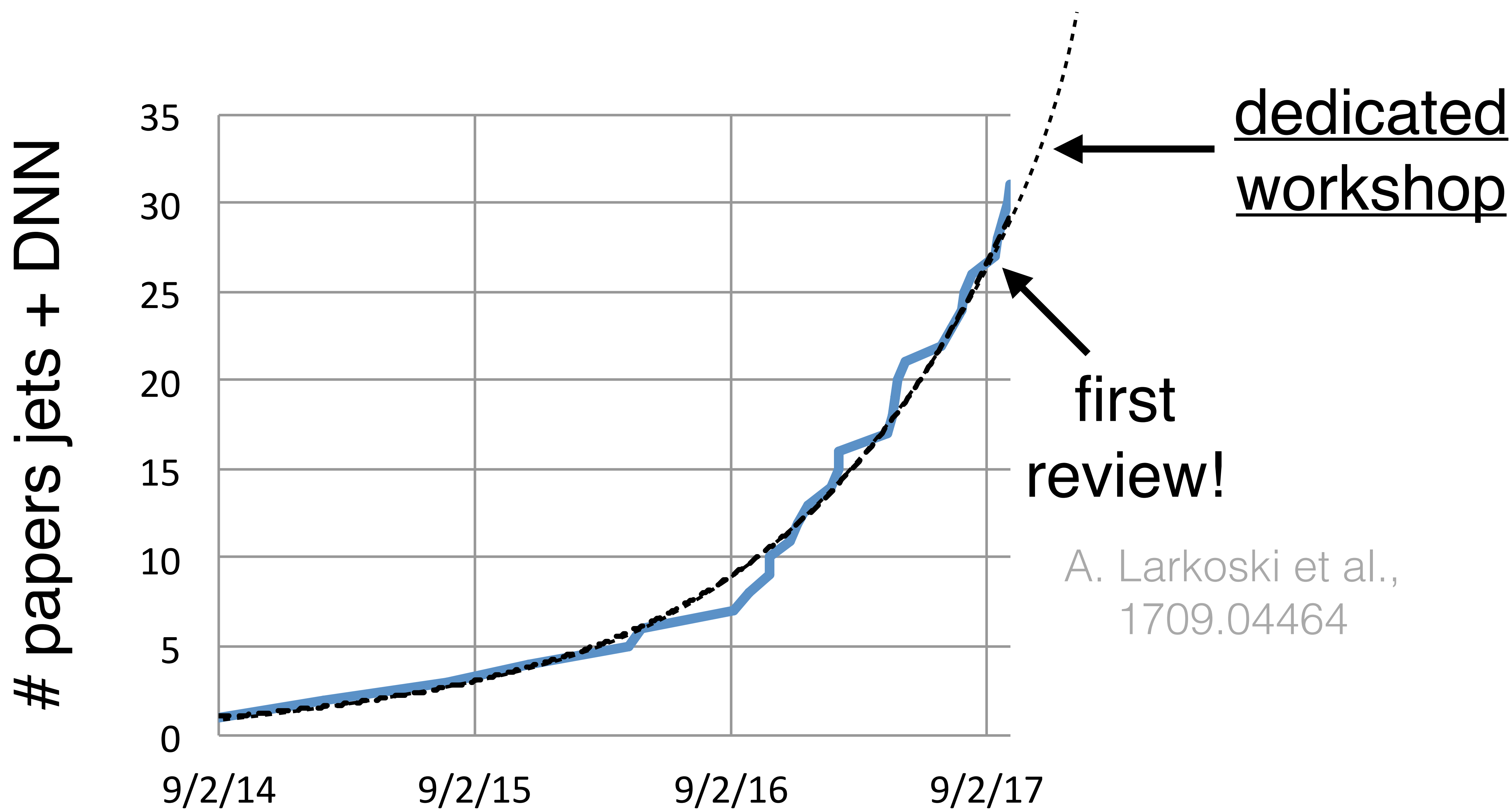
$$\begin{bmatrix} \mathbf{z}_H \\ \mathbf{z}_L \\ \mathbf{z}_R \\ \mathbf{z}_N \end{bmatrix} = \text{softmax}\left(W_z \begin{bmatrix} \tilde{\mathbf{h}}_k^{\text{jet}} \\ \mathbf{h}_{k_L}^{\text{jet}} \\ \mathbf{h}_{k_R}^{\text{jet}} \\ \mathbf{u}_k \end{bmatrix} + b_z \right)$$

$$\begin{bmatrix} \mathbf{r}_L \\ \mathbf{r}_R \\ \mathbf{r}_N \end{bmatrix} = \text{sigmoid}\left(W_r \begin{bmatrix} \mathbf{h}_{k_L}^{\text{jet}} \\ \mathbf{h}_{k_R}^{\text{jet}} \\ \mathbf{u}_k \end{bmatrix} + b_r \right)$$

- Each node combines 4-momentum in (E-scheme recombination of o_k) and a non-linear transformation of hidden state of children h_{kL}, $h_{kR} \in \mathbb{R}^{40}$
- Recursively applied (shared weights, Markov)
- "gating" allows for weighting of information of L/R children and for to flow directly along one branch

Exciting New Directions

So far only scratches the surface
…this is a very active field of research!



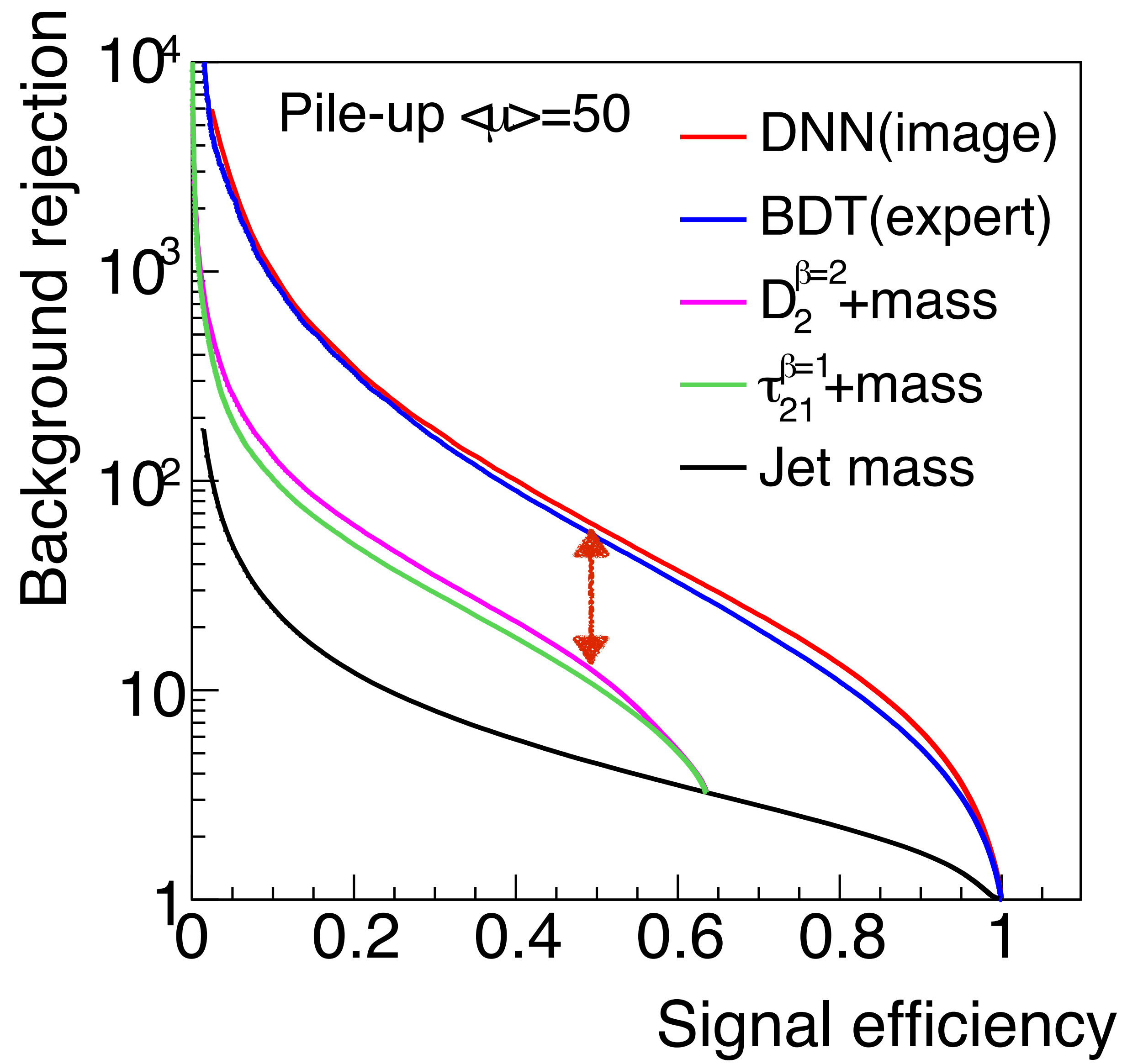
A. Larkoski et al., 1709.04464

Oliveira, et. al arXiv:1511.05190
Whiteson, et al arXiv:1603.09349
Barnard, et al arXiv:1609.00607

While the DNN shows a significant improvement with respect to the jet mass combined with single theory inspired variable (eg. τ_{21}, D_2), only a small improvement with respect to a BDT using several theory-inspired variables

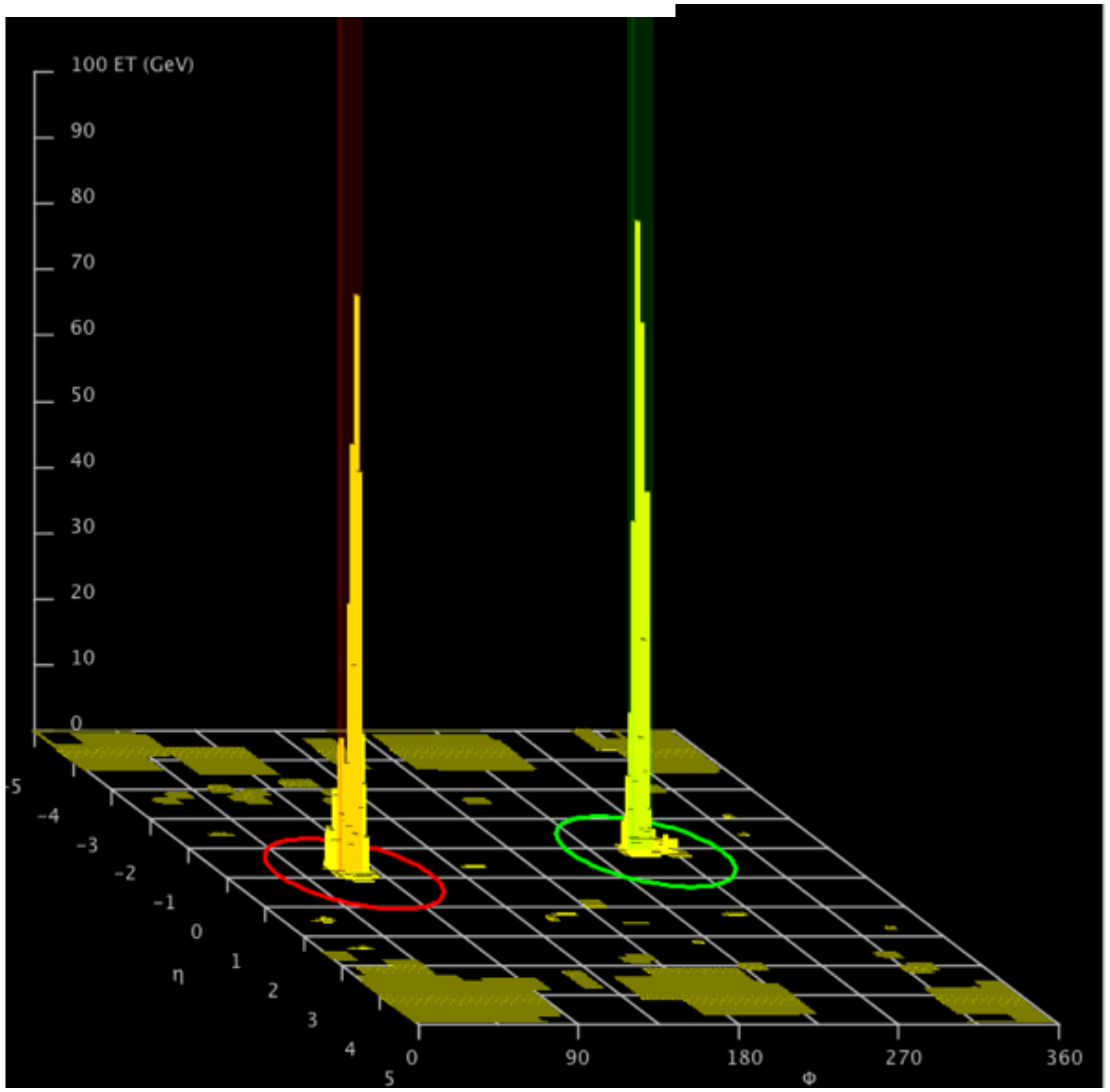
Other Problems:

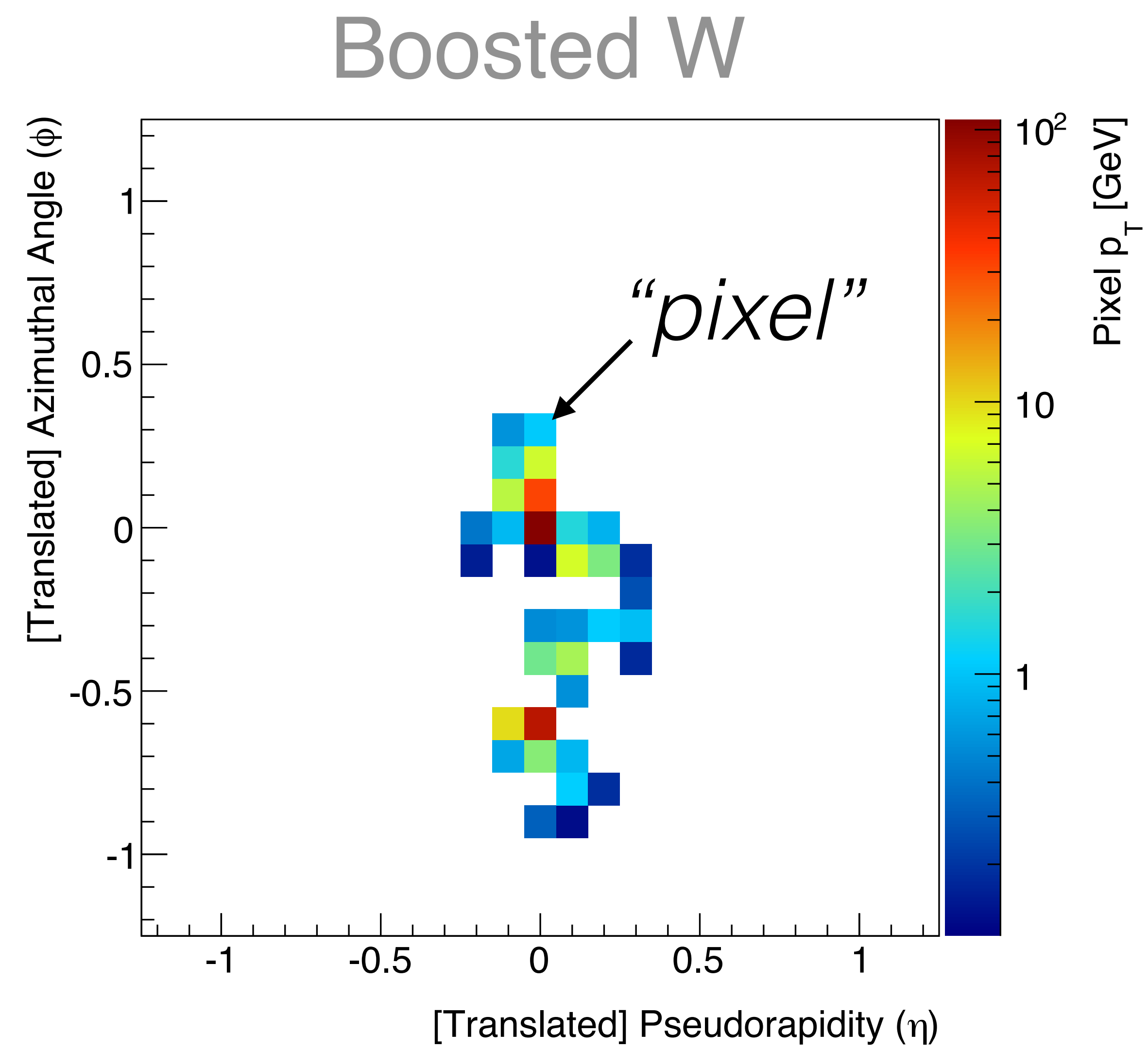
- image-based approach not easily generalized to non-uniform calorimeters
- not easy to extend to tracks, projecting into towers looses information
- theory inspired variables work on set of 4-vectors & have important theoretical properties



the Jet Image

J. Cogan et al. JHEP 02 (2015) 118





Credit: Peter G Trimming (Wikipedia)

no smooth edges, clear features, low occupancy (number of hit pixels)

One of the most useful physics-inspired features is the ***jet mass***