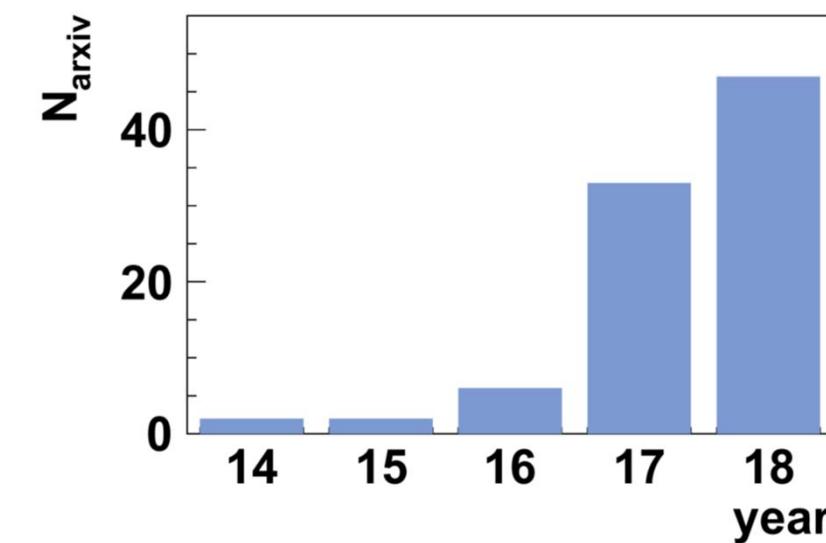


# Machine Learning

deep learning in particle  
& astroparticle physics



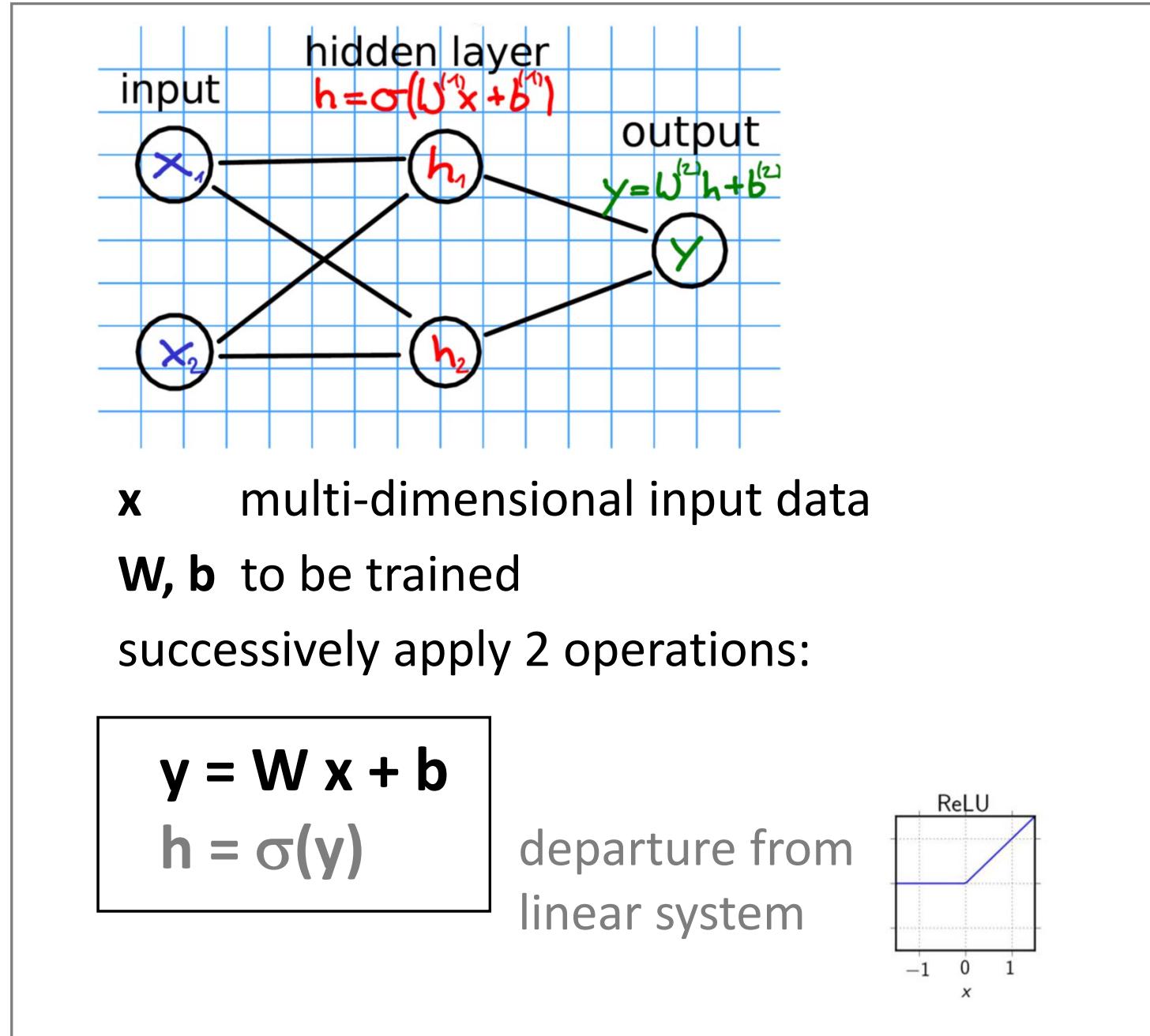
- (1) Deep Learning has arrived in all areas of astroparticle physics
- (2) Developments in field of machine learning
- (3) Aim at applications of machine learning in physics in standard curriculum
- (4) Workshops, Coordination, Funding

# Message 1

Deep Learning has arrived in all areas of astroparticle physics

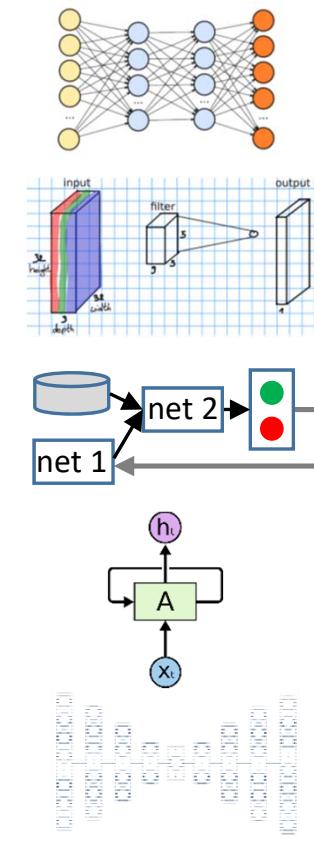
# Deep Learning:

*Neural networks with ‘many’ hidden layers*



## Architectures

- Fully connected
- Convolutional
- Adversarial
- Recurrent
- Autoencoder



## Improved set of tools

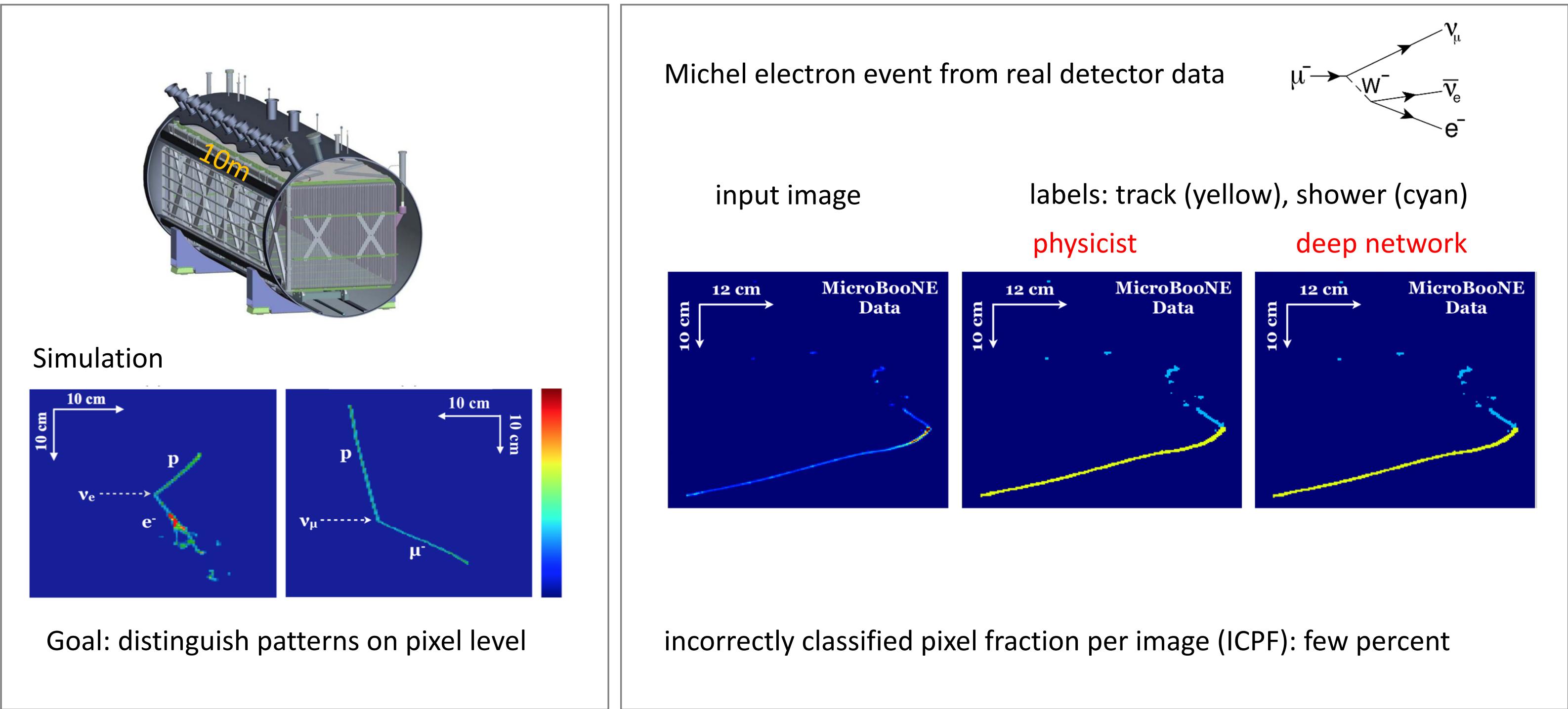
Train millions of parameters by:

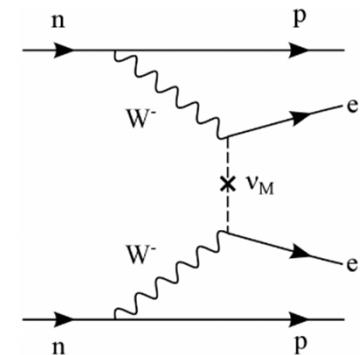
- Data preprocessing
- Normalization etc

## Computing

- Graphics Processing Unit (GPU)
- Software Libraries

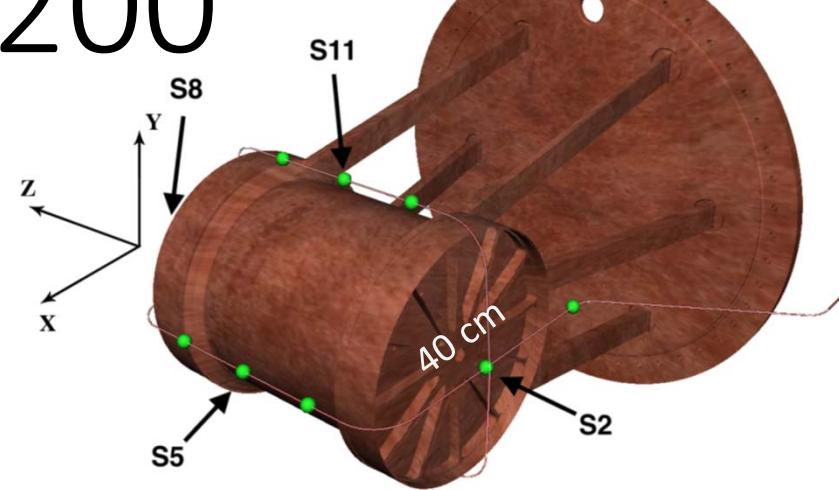
# Neutrino detection in LAr TPC: MicroBooNE





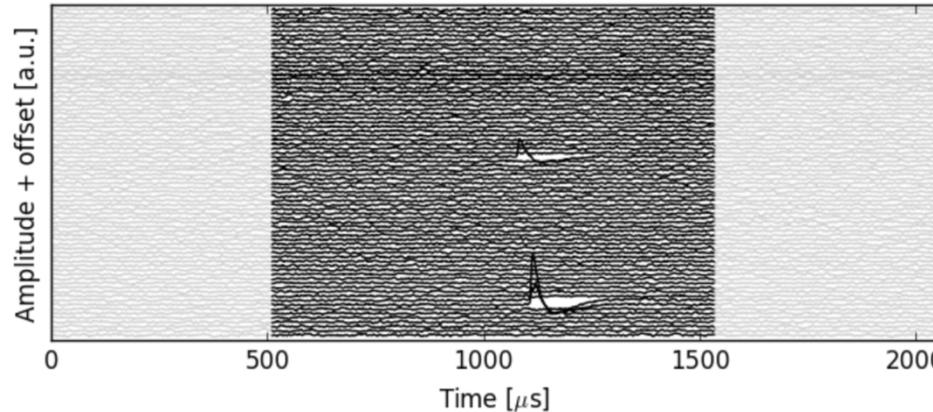
# Search for Neutrinoless Double Beta Decay

EXO-200



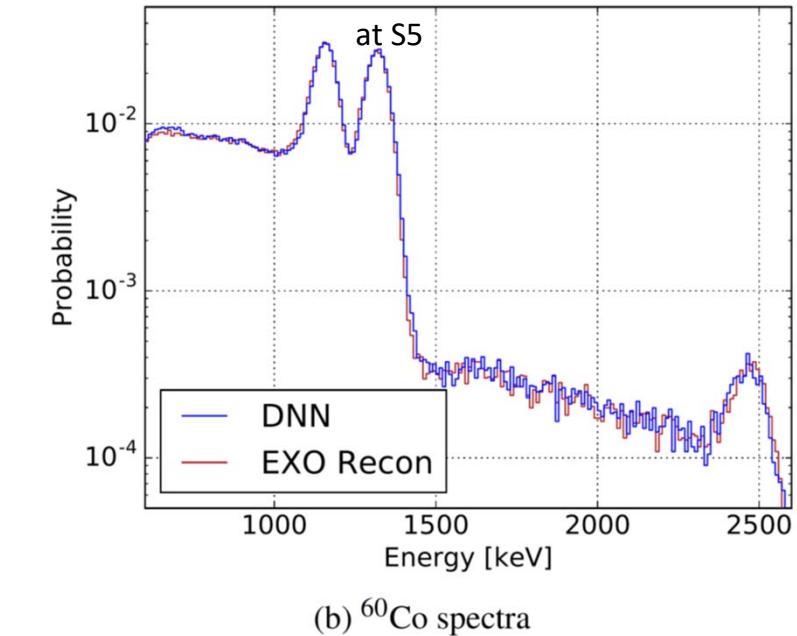
Double Time Projection Chamber, liquid Xenon  
Charge on wires, scintillation light in APD's

Signals on wires

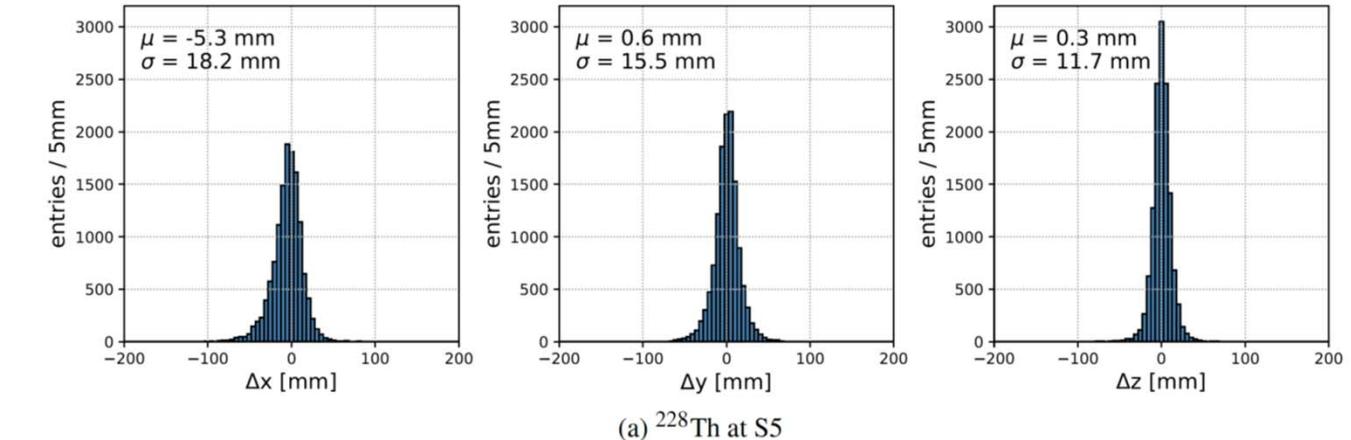


Energy reconstruction:

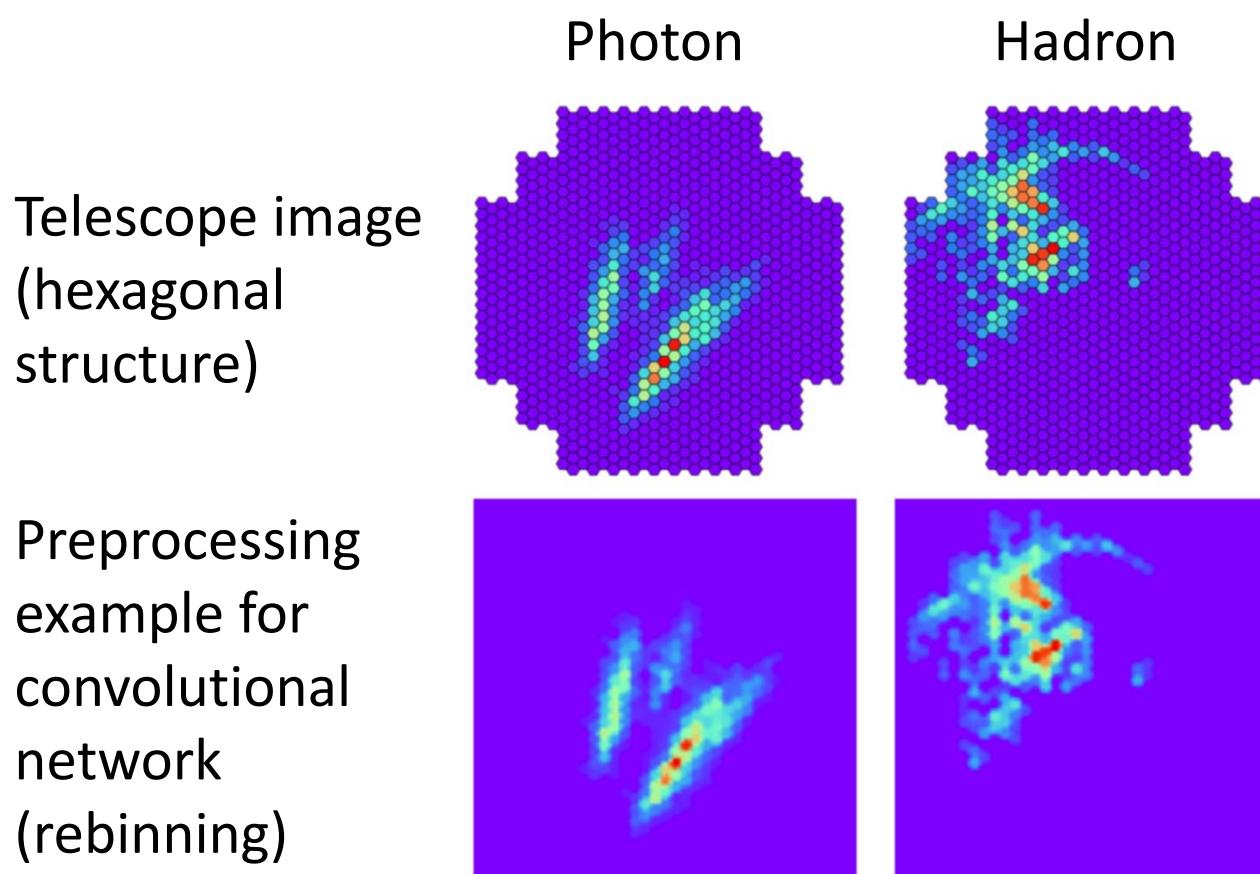
Deep neural network  
with slightly better  
energy resolution



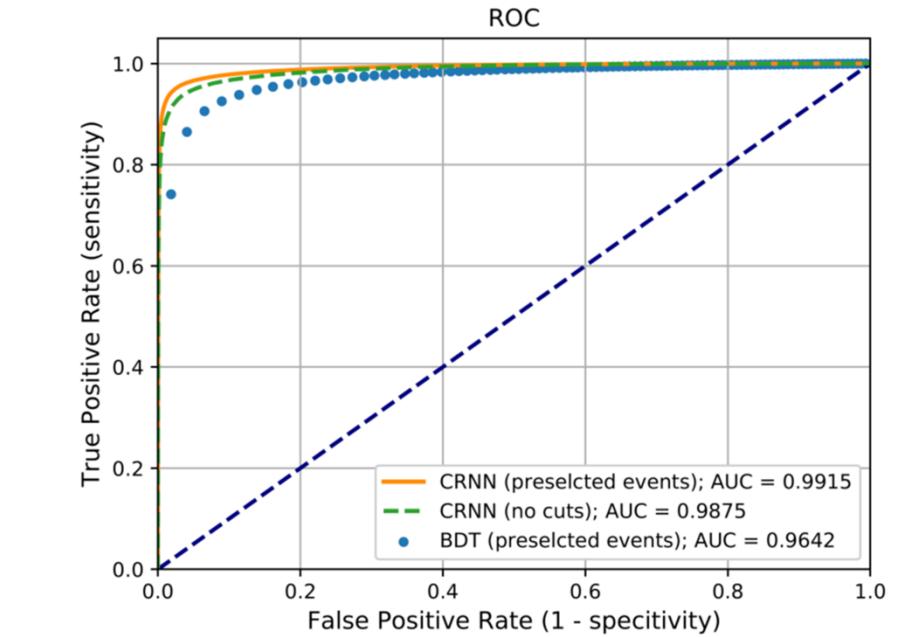
Position resolution:



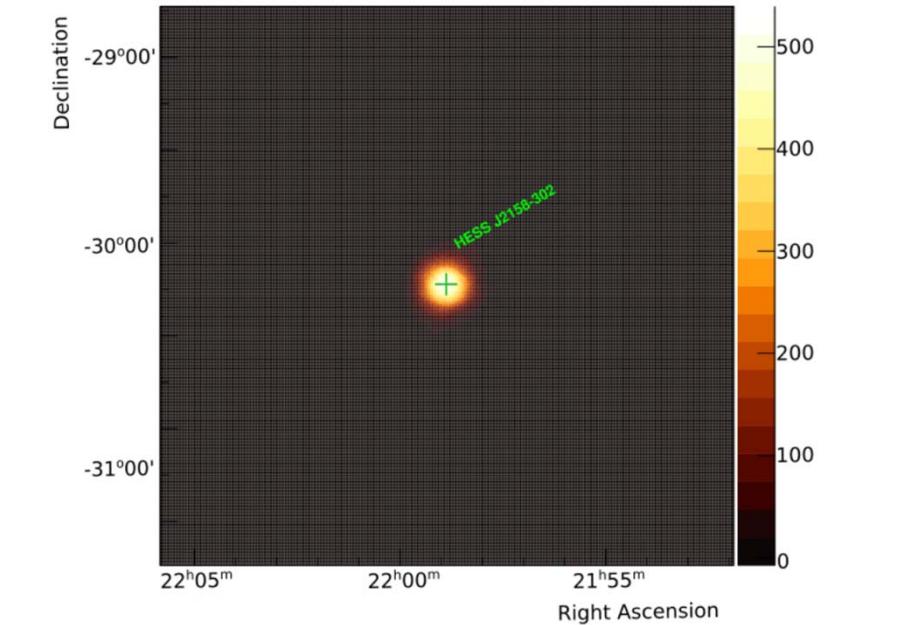
# Imaging Atmospheric Cherenkov Telescopes



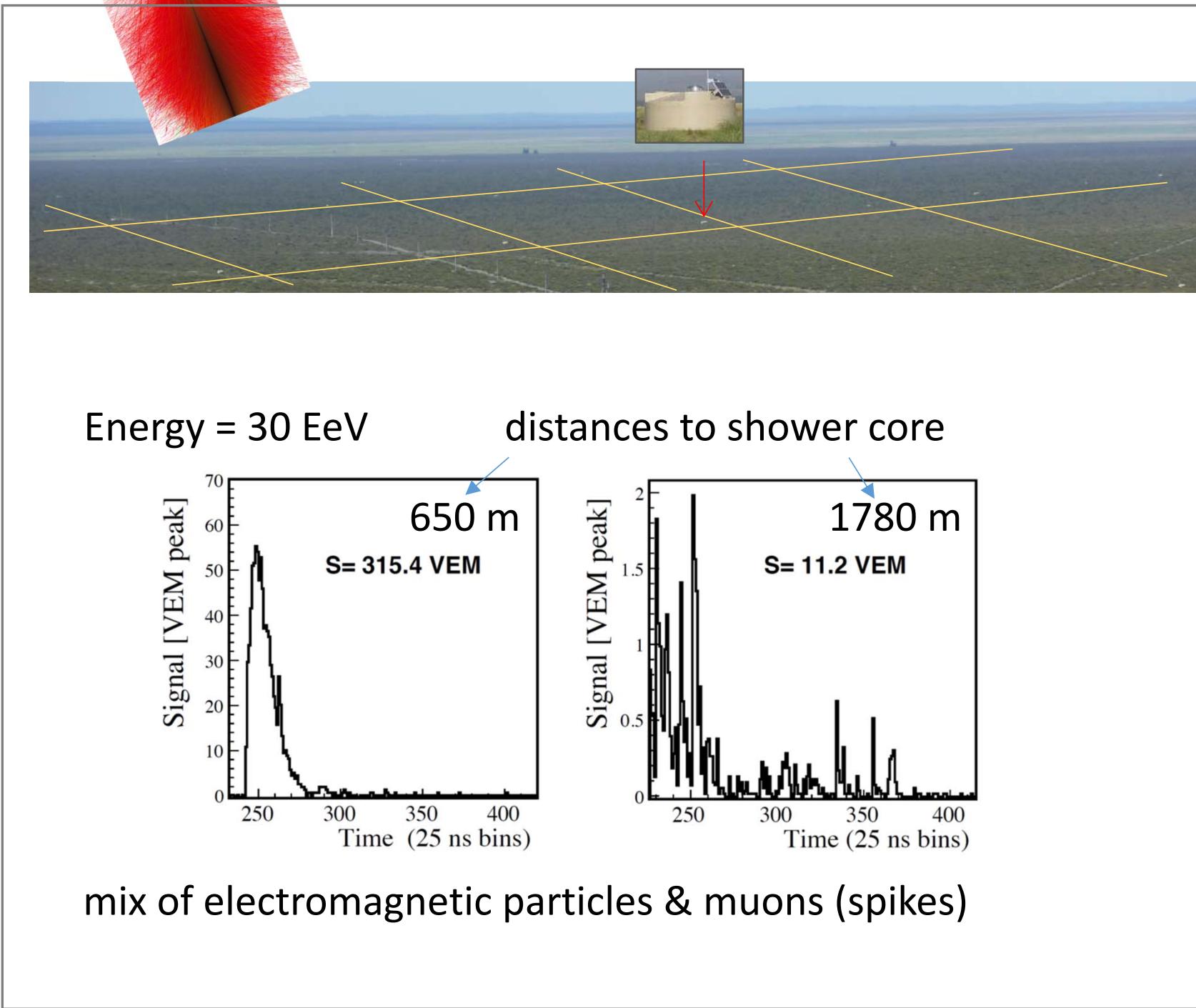
Simulation:  
deep network better  
to reject background  
compared to Boosted  
Decision Tree



Data:  
excess events observed in  
direction of PKS 2155-304  
for one flare observation

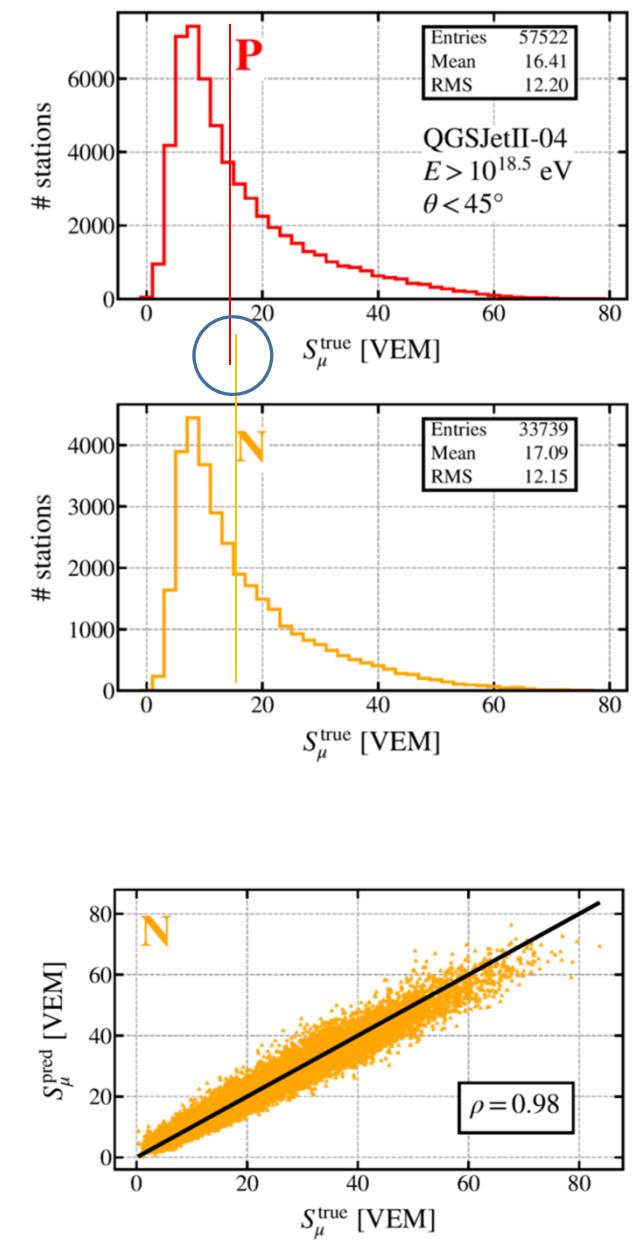


# Cosmic nuclei: muon fraction in air showers



Simulation:  
Small differences  
in single station  
signals for different  
nuclei

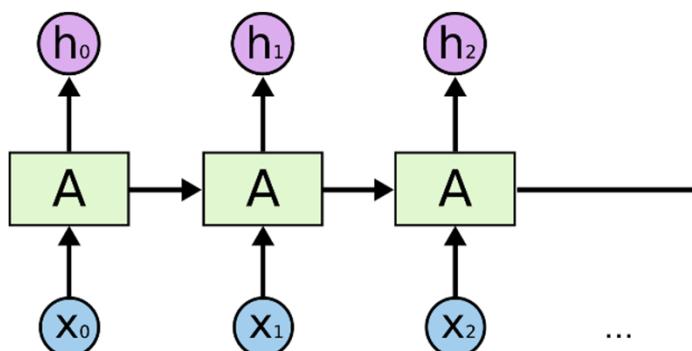
Resolution 2.5 VEM  
(vertical equivalent  
muons)



# Denoising Gravitational Waves with Recurrent Denoising Autoencoder

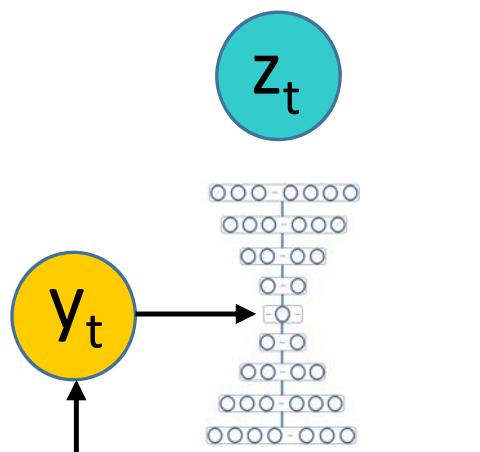


**Recurrent Network**

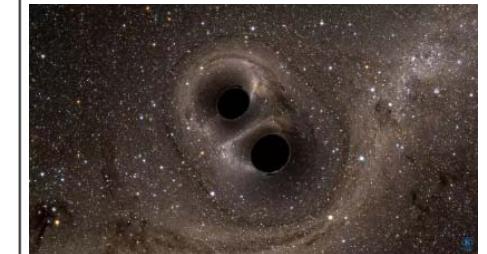
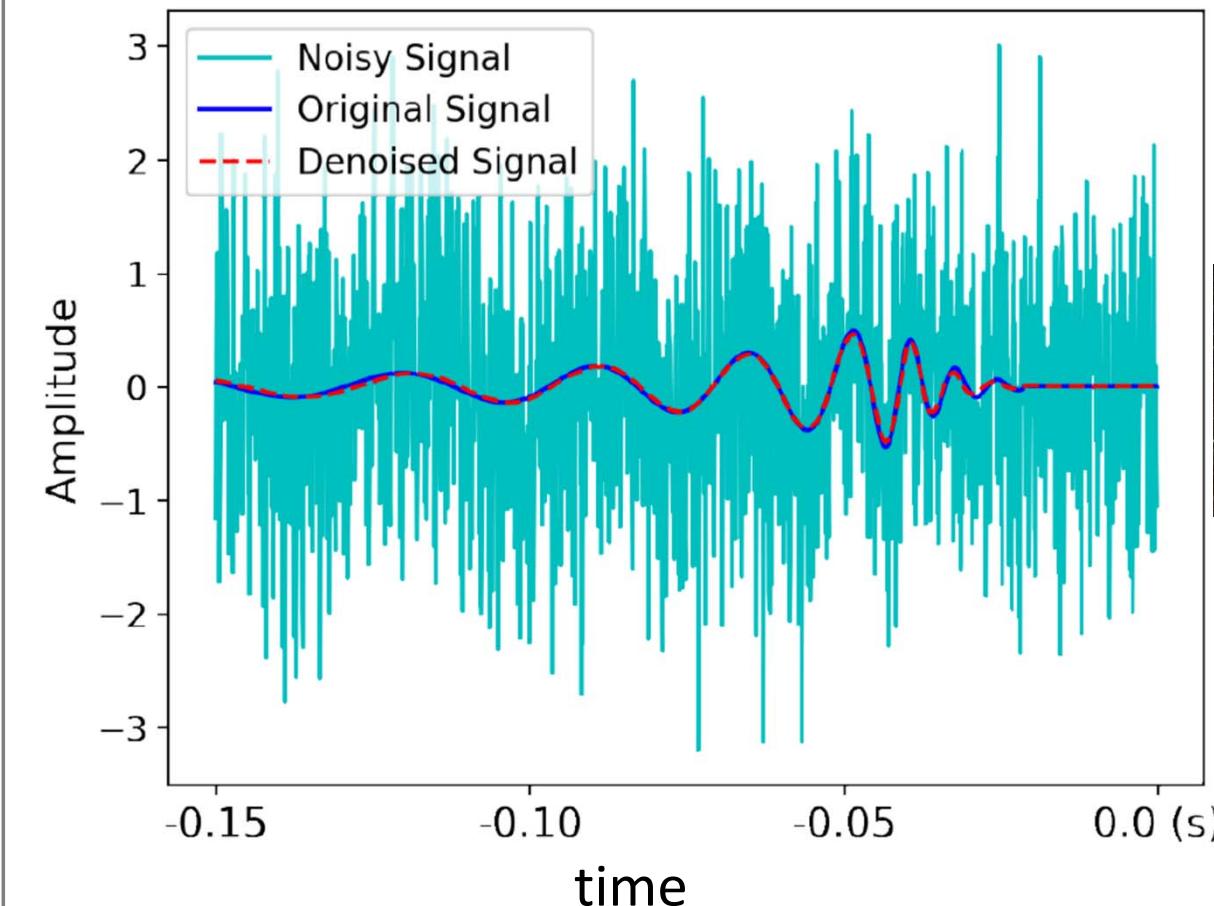


many-to-one network

**Autoencoder Network**



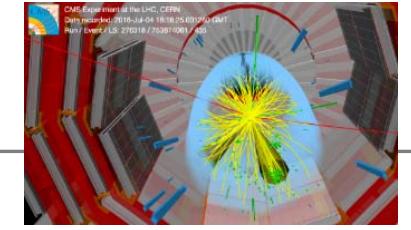
**Gravitational wave (simulated)**



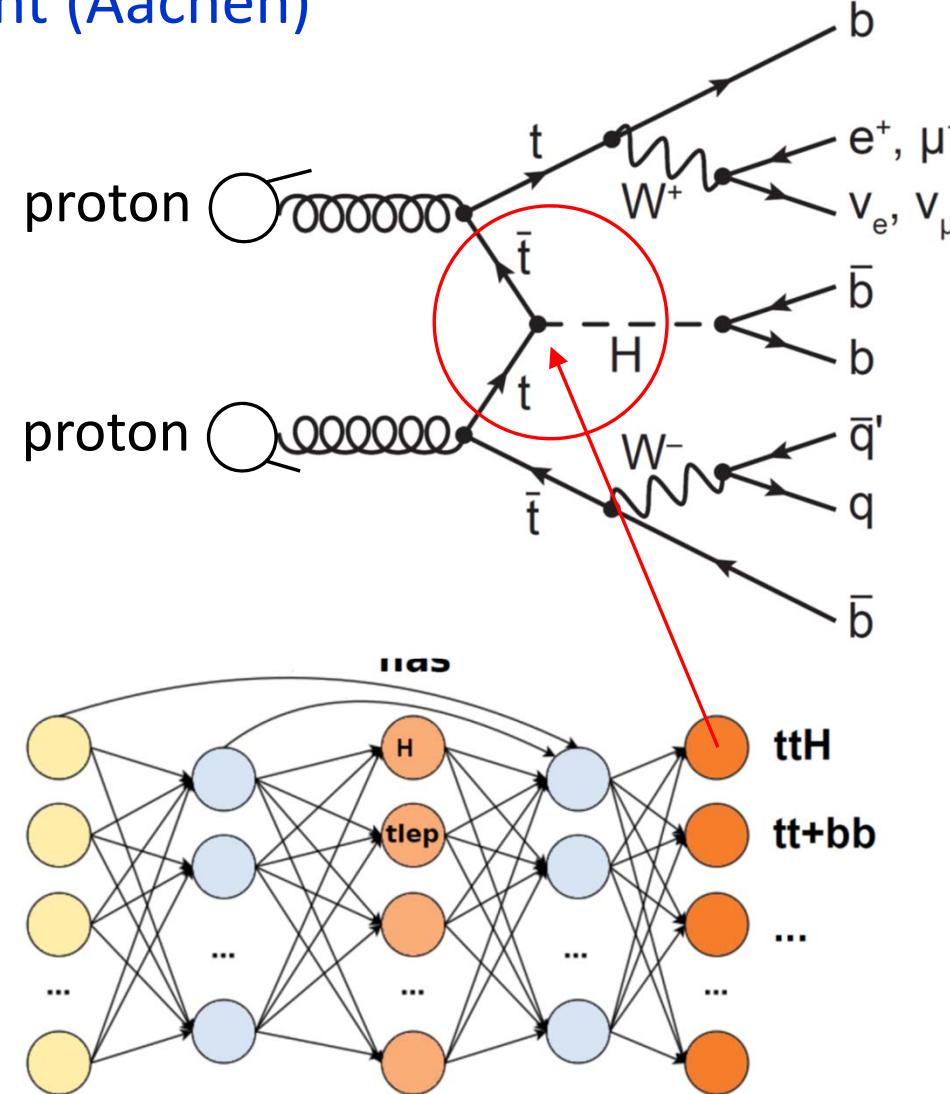
binary black hole merger

Excellent recovery of original signal

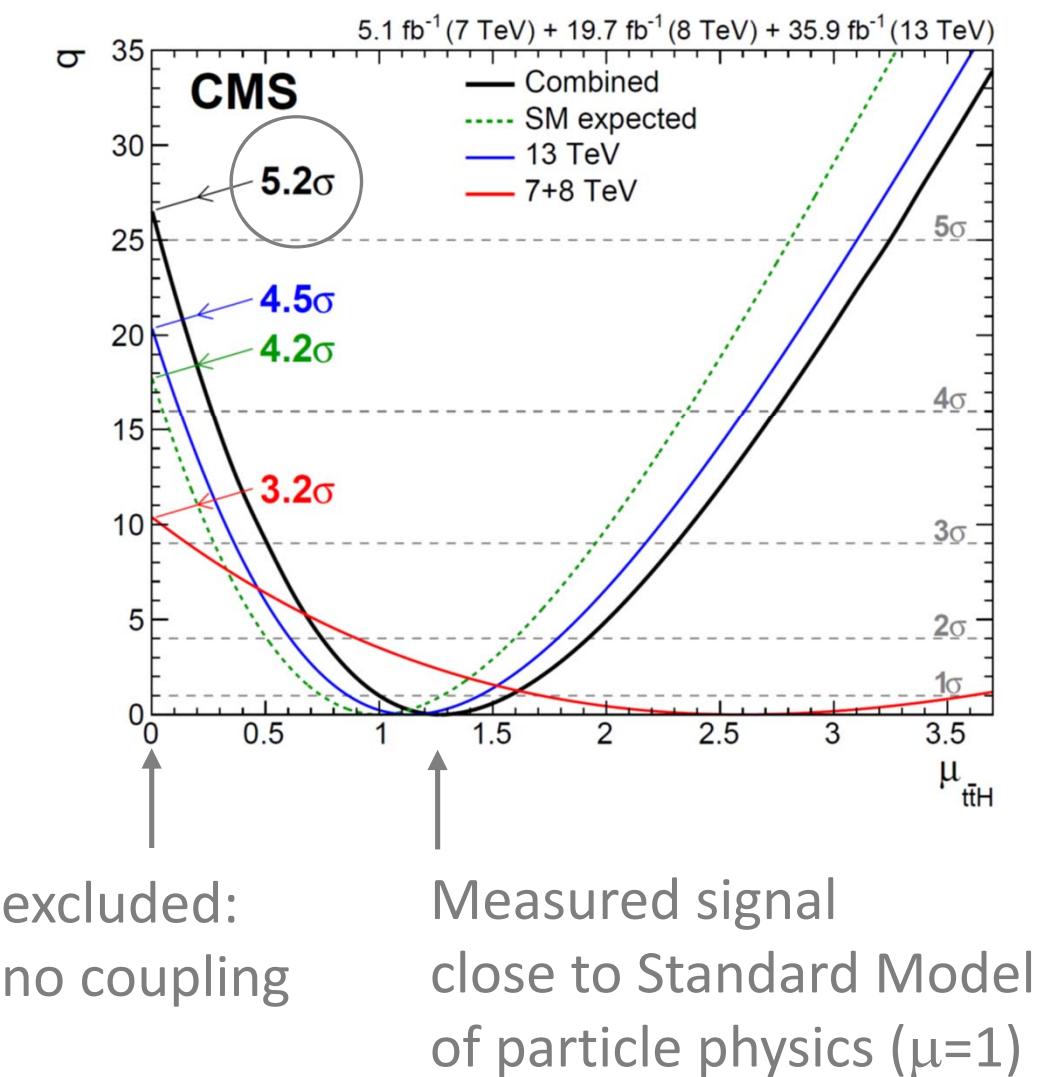
# LHC: Coupling Top-Quark – Higgs Boson



Deep Learning predicts physics process for each event (Aachen)

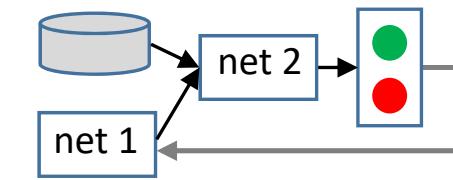


Observation of ttH production

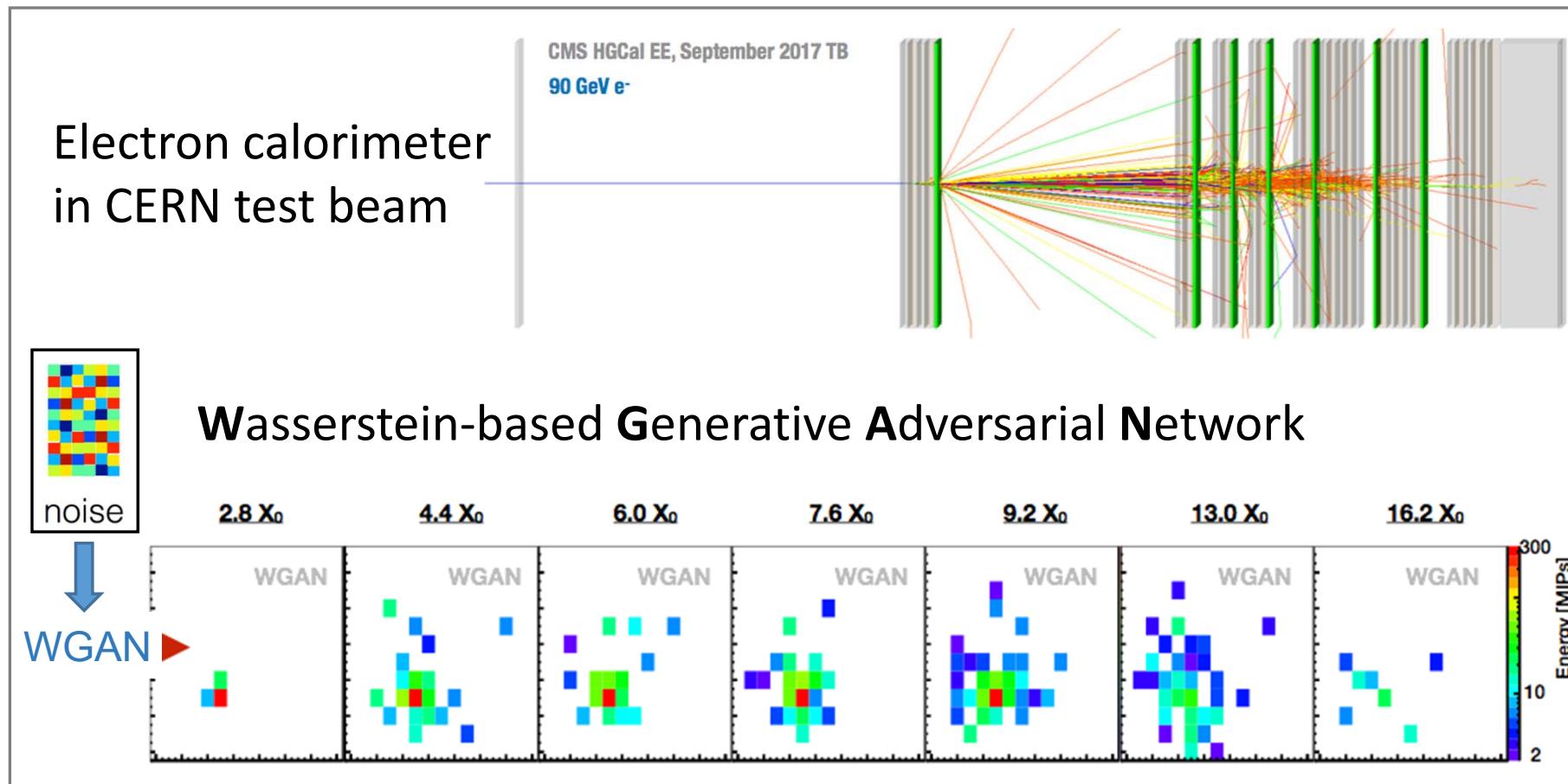


***Deep Learning: already standard in particle physics***

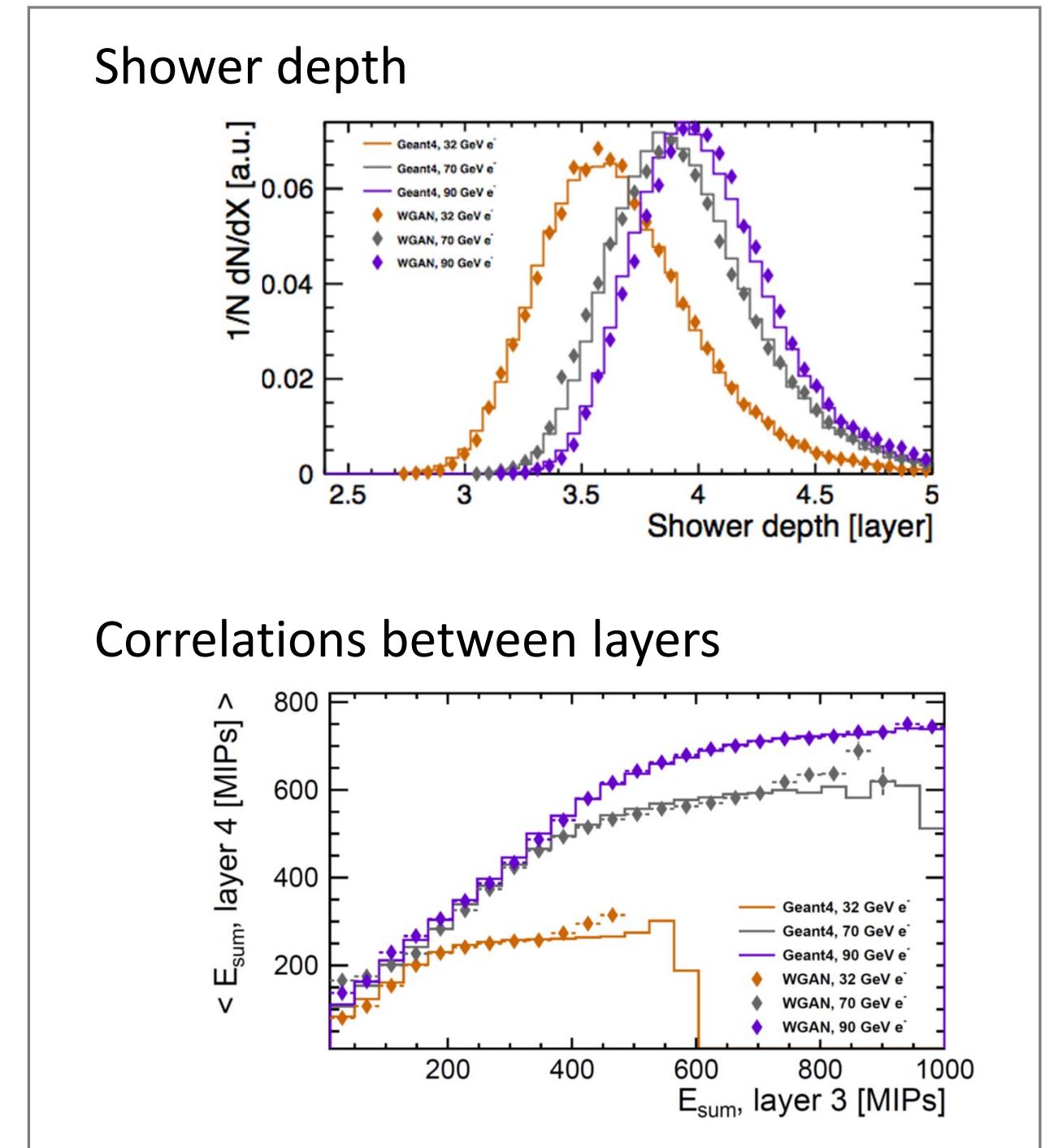
# Calorimeter simulations: WGAN



M.E., J. Glombitza, T. Quast  
arxiv 1807.01954



Generation Method	Hardware	milliseconds/shower	Goal: use measured data to train WGAN
GEANT4	CPU	<b>2000</b>	
WGAN	CPU	52	Air Shower simulation?
	GPU	<b>0.3</b>	



# Message 2

Variants and further developments in the field of machine learning

# Information Field Theory

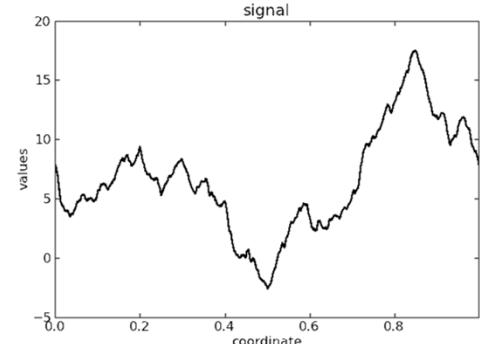
Torsten Enßlin, arXiv:1804.03350  
<http://www.mpa-garching.mpg.de/ift/>

Bayesian method to fuse multiple information sources,  
learning from a single data set

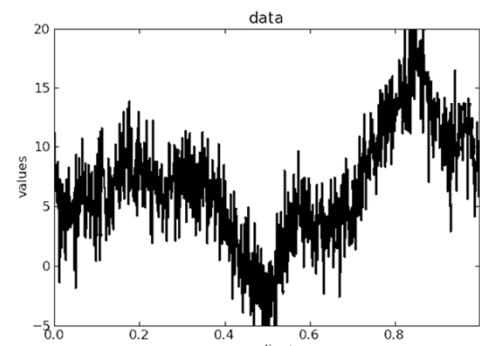
$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s) \mathcal{P}(s)}{\mathcal{P}(d)} = \frac{1}{Z(d)} e^{-H(d,s)}$$

$s$  = signal  
 $d$  = data

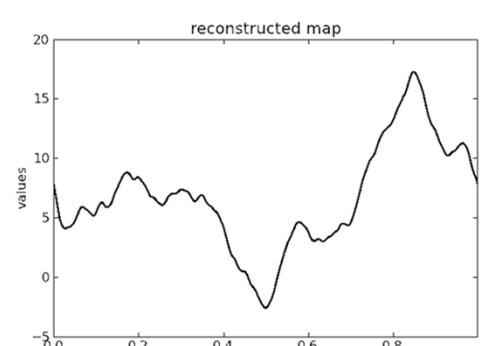
Simulated  
signal



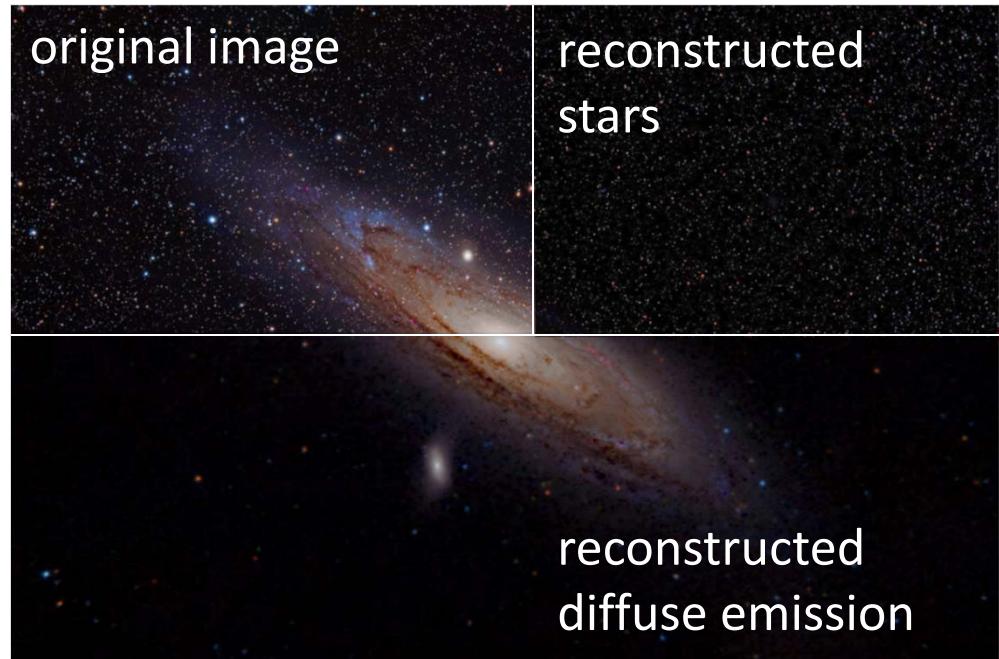
Simulated  
data



Reconstructed  
signal



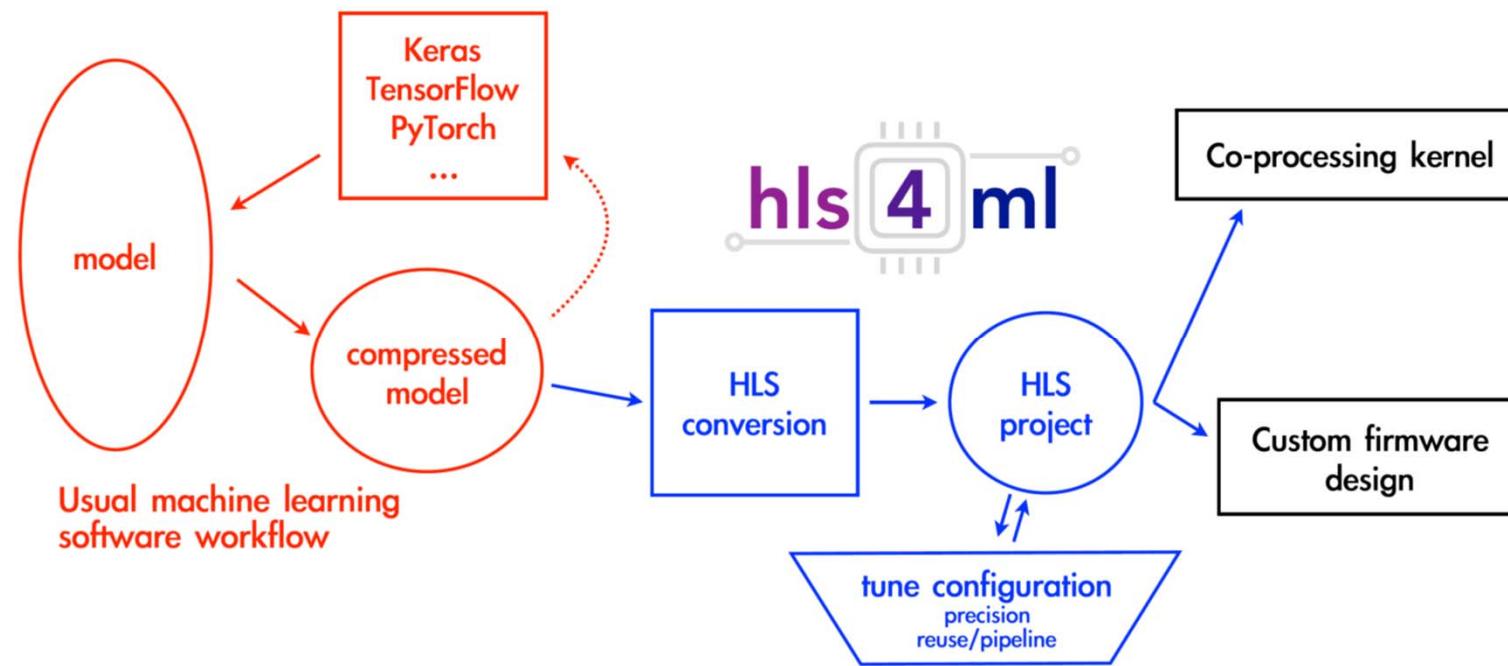
Separate contribution to Hubble's Andromeda galaxy



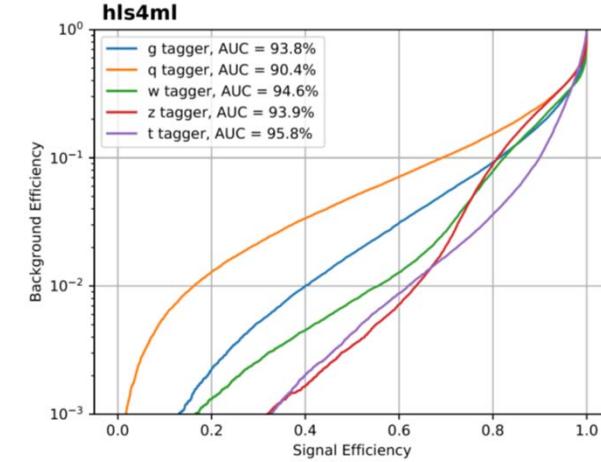
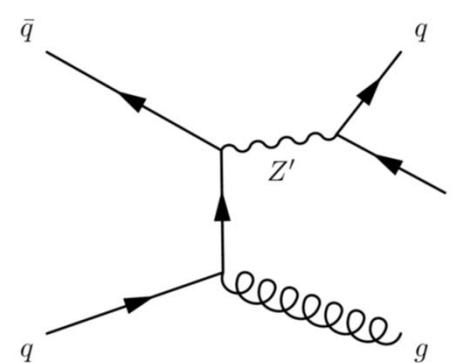
Hope for better understanding of Galactic Magnetic Field

F Boulanger, T Ensslin, A Fletcher, P Girichides, S Hackstein, M Havercorn, J Hoerandel, T Jaffe, J Jasche, M Kachelriess, K Kotera, C Pfrommer, J Rachen, L Rodrigues, B Ruiz-Granados, A Seta, A Shukurov, G Sigl, T Steininger, V Vacca, E van der Velden, A van Vliet, J Wang  
*IMAGINE: A comprehensive view of the interstellar medium, Galactic magnetic fields and cosmic rays*, arXiv:1805.02496

# Deep neural networks in FPGAs



Distinguish jets from quarks, gluons, ‘fat’ jets



Remarkable

- Fully connected neural network to identify jets with 4389 parameters
- Implemented in FPGA using network compression & reduced precision
- Latency of inference 75–150 ns with clock frequency 200 MHz → LHC

Trigger

# Message 3

Physics students should receive a training offer for applications of machine learning in physics in the standard curriculum

# „Broschüre“

Draft Version 1.0,  
15-Juli-2018



Arbeitskreis Physik, moderne  
Informationstechnologie und  
Künstliche Intelligenz der DPG

## Ausbildung für Physikstudierende im Erkenntnisgewinn durch moderne datengetriebene Methoden: Vertiefungsrichtung Deep Learning

Autoren: Martin Erdmann für das AKPIK Vorstandsgremium

Die Entwicklung neuer Technologien ist entscheidend für Erkenntnisgewinne in Grundlagen- und angewandten Wissenschaften. Fortschritte bei modernen Datenanalyseverfahren kommt dabei besondere Bedeutung zu: Sie sind der Schlüssel, um neue Erkenntnisse aus Messdaten zu schöpfen und können im besten Fall entscheidenden Mehrwert aus Messapparaturen erwirken. Zahlreiche Neu- und Weiterentwicklungen im Bereich von Analyseverfahren finden kontinuierlich statt, so dass hier eine Momentaufnahme erfasst wird.

In der Physik haben Datenanalysen schon immer zentrale Bedeutung. Im Vergleich mit bisherigen Analyseverfahren schöpfen moderne Verfahren allerdings die Daten weitgehender aus, was Erkenntnisprozesse zumindest beschleunigt oder sogar erst ermöglicht.

**Die Grundlagen für den Erkenntnisgewinn durch aktuelle datengetriebene Methoden sollen daher im Physikstudium stärker verankert und in entsprechenden Kursen gelehrt werden.**

Aktuell werden in großer Anwendungsbreite Verfahren der Deep Learning Technologie erprobt, die vielfach herausragende und neue Möglichkeiten dieser Technologie für

wissenschaftliche Anwendungen zeigen (siehe unten: Forschungsbeispiele). Ein verstärkender Faktor für den Einsatz in der Physik sind die Entwicklung vergleichsweise komfortabler Softwarebibliotheken durch Großkonzerne, die als Open Source Programme zur Verfügung gestellt werden.

Für die kommenden Jahre wird daher erwartet, dass die Verfügbarkeit, Anwenderfreundlichkeit sowie die Anwendungsbreite der Analyseverfahren mit Deep Learning eine gute Grundlage für die vertiefte Ausbildung von Physikstudierenden im Bereich der Erkenntnisgewinne durch moderne datengetriebene Methoden darstellt (siehe unten Kursbeispiel). Voraussetzung für entsprechende Kurse sind selbstverständlich solide Grundkenntnisse in statistischen Verfahren zur Datenanalyse, die im Bachelorstudium der Physik gelehrt werden.

Fertigkeiten und Kenntnisse in modernen Methoden der Datenanalyse haben über die wissenschaftliche Forschung hinaus Bedeutung. Die Mehrzahl der Physik-Studierenden wird eine Stelle in der freien Wirtschaft annehmen, so dass auch hier aktuelle Datenanalysemethoden zum Einsatz kommen werden, um neu aufkommende Fragestellungen in der wirtschaftlich und gesellschaftlich fortschreitenden Digitalisierung professionell zu beantworten.

1

KAT + AKPIK: M.E., U. Katz, T. Enßlin, A. Hamm,  
K. Mannheim, V. Markl, K. Morik, ...



## Inhalt

1. Ausbildungskonzept: Maschinelles Lernen für Physikforschung
2. Beispiele Anwendungen von Deep Learning Konzepten in Physik
3. Empfehlungen

- *Erweiterung des Kursangebots:* Verwendung und Entwicklung von Verfahren des maschinellen Lernens, insbesondere von Deep Learning Verfahren. Das Angebot soll explizit Physik-nahe Anwendungen enthalten.
- Einrichtung und Pflege von *Grafikprozessoren* an Rechenzentren der Universitäten für Kursteilnehmer und Forscher mit einer Softwareumgebung, die für Erstellung, Training und Evaluation von tiefen Netzwerken geeignet ist.
- Einrichtung und Pflege einer *Ausbildungsplattform* für Austausch von Kursmaterialien für Vorlesungen und Übungen. Marktplatz für Netzwerkarchitekturen und vortrainierte Netzwerkparameter, Möglichkeiten des Benutzeroberflächenaustauschs über Forum, Sammlung und Zurverfügungstellung geeigneter Datensätze.

**„Wir wollen erreichen, dass Physikstudierende im Standardcurriculum ihres Studiums ein Ausbildungsangebot in der Anwendung maschinellen Lernens in der Physik erhalten“ (Uli Katz)**

→ Konferenz der Fachbereiche Physik

Martin Erdmann, RWTH Aachen

# Message 4

Workshops, Coordination, Funding situation

## „Innovative Digitale Technologien für die Erforschung von Universum & Materie“

Aachen, Bonn, Erlangen-Nürnberg, Frankfurt, Freiburg, Göttingen, Hamburg, Karlsruhe, Mainz, München, Wuppertal,  
assoziiert: CERN, DESY, GridKa, GSI, Jülich, Münster

### Hardware & Computing

- A)** Entwicklung von Technologien zur Nutzung von heterogenen Ressourcen
- B)** Anwendung und Test von virtualisierten Software-Komponenten im Umfeld heterogener Computing-Ressourcen

### Software & Analyse

- C)** Deep Learning, Erkenntnisgewinn durch fundierte datengetriebene Methoden
- D)** Ereignisrekonstruktion, parallele Algorithmen

**Backbone:** Ausbildung wissenschaftlicher Nachwuchs, gemeinsame Plattform

*BMBF: positiv evaluiert, einige Gruppen haben bereits Zuwendungsbescheide 10/2018-9/2021*

# Perspektiven für HEP Software und Computing in D, Wuppertal 27./28.9.2018

13:00	<b>Einleitung / Strategiepapier / Rahmenbedingungen</b>  <b>Gemeinsames Strategiepaper:</b> <b>KET, KAT, KHuK, ... (Bonn 4.-5.10.18, by invitation)</b>  <i>Bergische Universität Wuppertal</i>	Volker GUELZOW et al.	13:00 - 14:00
14:00	<b>Übersicht Verbundantrag</b>  <i>Bergische Universität Wuppertal</i>	Thomas KUHR	14:00 - 14:20
	<b>Aktuelle Themen und Trends in Computing und Software - CHEP Highlights</b>  <i>Bergische Universität Wuppertal</i>	Dr. Thomas KRESS	14:20 - 14:40
	<b>Kaffeepause</b>		
15:00	 <i>Bergische Universität Wuppertal</i>		14:40 - 15:10
	<b>NFDI</b>  <i>Bergische Universität Wuppertal</i>	Christian ZEITNITZ	15:10 - 15:30
	<b>Zeit für Diskussionen</b>  <i>Bergische Universität Wuppertal</i>		15:30 - 16:00
16:00	<b>HEP Software Foundation und gemeinsame Projekte</b>  <i>Bergische Universität Wuppertal</i>	Dr. Benedikt HEGNER	16:00 - 16:15
	<b>Softwareentwicklung in Deutschland am Beispiel Belle II</b>  <i>Bergische Universität Wuppertal</i>	Nils BRAUN	16:15 - 16:30
	<b>Erfolgreiche Zusammenarbeit in den Experimenten (Gentner-Programm und vergleichbar)</b>  <i>Bergische Universität Wuppertal</i>	Dr. Markus ELSING	
	<b>Big data analytics</b>  <i>Bergische Universität Wuppertal</i>	Prof. Martin ERDMANN	16:45 - 17:00
17:00	<b>Diskussionen, Schlussfolgerungen</b>  <i>Bergische Universität Wuppertal</i>		17:00 - 18:00
18:00			

09:00	<b>Entwicklung der Computing Modelle und WLCG Strategie</b>  <i>Bergische Universität Wuppertal</i>	Andreas PETZOLD	09:00 - 09:20
	<b>Vorstellung der deutschen Sites</b>  <i>Bergische Universität Wuppertal</i>	Dr. Torsten HARENBERG et al.	09:20 - 09:40
	<b>"Auf dem Weg zu schlankeren Sites"</b>  <i>Bergische Universität Wuppertal</i>	Dr. Johannes ELMSHEUSER	09:40 - 10:00
10:00	<b>Zur zukünftige Aufgabenverteilung Uni / HGF</b>  <i>Bergische Universität Wuppertal</i>	Christian ZEITNITZ	10:00 - 10:10
	<b>Diskussionen, Schlussfolgerunge</b>  <i>Bergische Universität Wuppertal</i>		10:10 - 11:00
11:00	<b>Kaffee</b>  <i>Bergische Universität Wuppertal</i>		11:00 - 11:30
	<b>Integration von Clouds in Grid</b>  <i>Bergische Universität Wuppertal</i>	Prof. Günter QUAST	11:30 - 11:50
	<b>DOMA und Entwicklungen in Richtung Data Lakes (inkl. dCache, caching,...)</b>  <i>Bergische Universität Wuppertal</i>	Dr. Paul MILLAR	11:50 - 12:10
12:00	<b>Analyse-Infrastrukturen und Frameworks für massiv parallele Analyse</b>  <i>Bergische Universität Wuppertal</i>	Dr. Muhammad AL-TURANY	12:10 - 12:30
	<b>Diskussionen, Schlussfolgerungen</b>  <i>Bergische Universität Wuppertal</i>		12:30 - 13:30
13:00	  <i>Bergische Universität Wuppertal</i>		
14:00	<b>Abschlussdiskussion</b>  <i>Bergische Universität Wuppertal</i>		13:30 - 14:30

# Big Data Science in Astroparticle Research (III.)

## Big Data Science in Astroparticle Research - HAP Workshop

18-20 February 2019  
RWTH Aachen University SuperC  
Europe/Berlin timezone

Overview  
Timetable  
Organisation  
Social Program  
Accommodation  
Venue & Travel  
Restaurants  
Participant List  
Map of Aachen  
Poster

contact  
[hap-workshop@physik.rwth-aachen.de](mailto:hap-workshop@physik.rwth-aachen.de)  
[0241 8027330](tel:02418027330)



Starts 18 Feb 2019, 13:00  
Ends 20 Feb 2019, 15:00  
Europe/Berlin



RWTH Aachen University SuperC  
RWTH Aachen University  
Templergraben 57,  
52062 Aachen  
phone: 0241 8090801



Andreas Haungs  
Martin Erdmann

Location

Martin Erdmann, RWTH Aachen

**Aachen, 18.-20. Feb-2019**

We invitation your contributions

- Progress in deep learning applications
- Data management and data centers
- Software
- Analysis preservation
- Platforms for algorithms & network architectures
- Education material

Keynote speakers

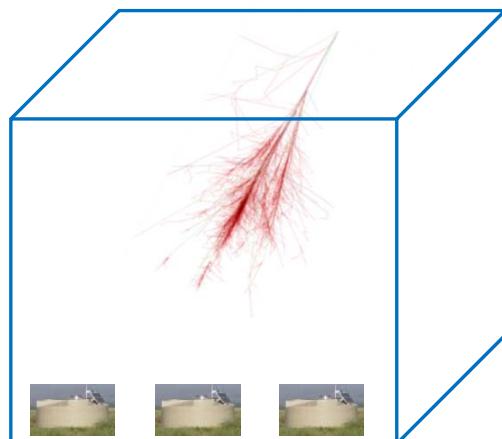
Hands-on tutorial deep networks

# Take away

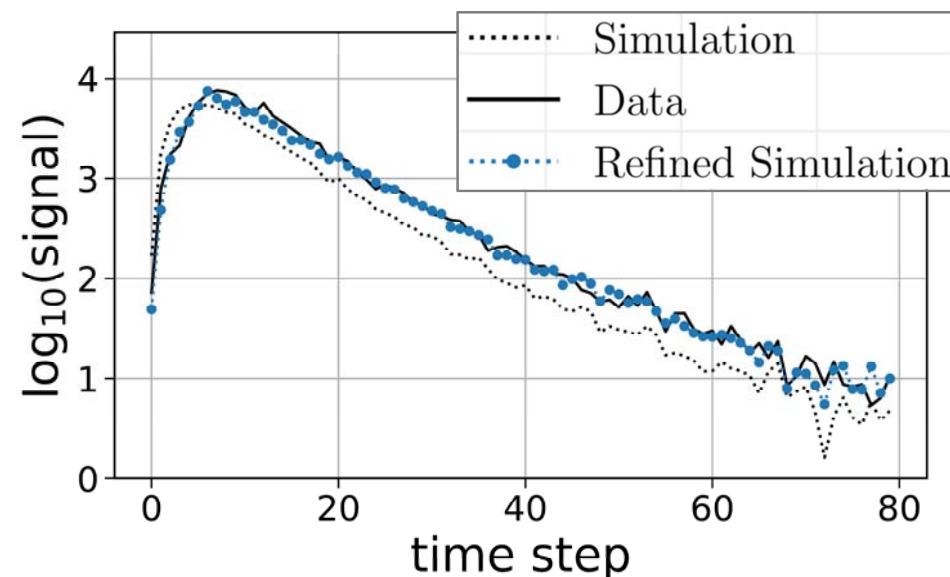
- (1) Deep Learning has arrived in all areas of astroparticle physics
- (2) Deep Learning already standard at LHC particle physics
- (3) Information Field Theory: include prior knowledge to guide parameter search
- (4) FPGA can bring machine learning close to sensors
- (5) Education: applications of machine learning in physics in standard curriculum
- (6) Community efforts ongoing & visible: DPG AKPIK, Workshops, funded projects, applications
  - Strong desire in KAT-KET-KHuK Community for joint & interdisciplinary developments:
  - Needs structure to "home" the activities (similar to Allianz, FSP,...) for coordinated efforts

# For network training: adapt simulation to data

Long-standing problem  
 $\rightarrow$  Refine simulated traces  
 $\text{data} \neq \text{simulation}$

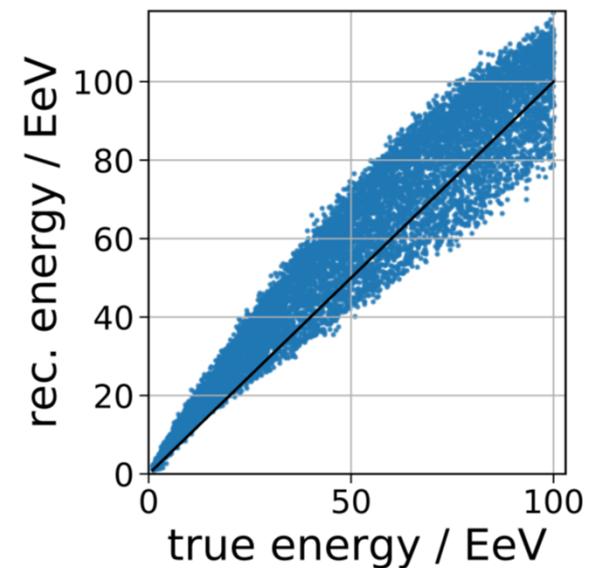


Refine simulated traces

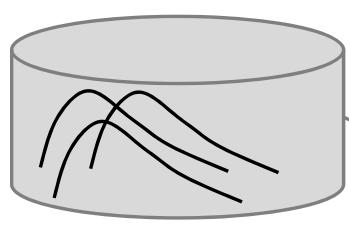


Training: badly simulated traces

Deep Network  
 net-3: cosmic  
 ray energy from  
 'data' traces

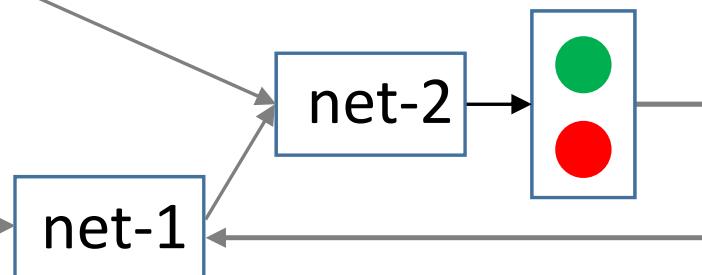


'data' traces  
 (simulated)

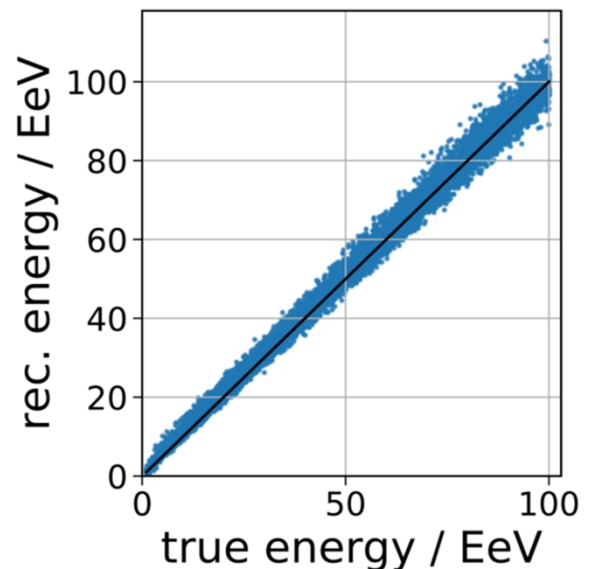


'badly simulated'  
 trace

Adversarial training  
 Wasserstein distance



Training: refined simulation  
*improved  
 energy  
 resolution*



# LHC CMS jet flavor tagging

