The FNAL Machine Intelligence Group

Gabriel N. Perdue
FNAL PAC
2017/November/16

Overview

- What is machine intelligence?
- How will we work at Fermilab?
- Near-term projects in generative models for fast simulation
- Training, workshops, and education
- Quantum machine learning and simulation
- Other miscellaneous projects and plans
- Conclusions

What is machine intelligence?

Baron Schwartz
@xaprb

Follow

When you're fundraising, it's AI
When you're hiring, it's ML
When you're implementing, it's linear regression
When you're debugging, it's printf()

12:52 AM - 15 Nov 2017

3,138 Retweets **7,439** Likes

55 3.1K 7.4K

What is machine intelligence, and what do we do?

- Machine learning plus artificial intelligence.
  - We focus on the application of advanced algorithms and architectures.
  - Learning algorithms, adaptive control, GPUs, quantum computing, etc.
- We envision a dual role in service and research.
  - Service: workshops, tutorials, community building (e.g. journal club), seminar series (with the specific aims of educating the lab community and bringing individuals to the lab that we feel have a high chance of forming partnerships), and projects that impact multiple experiments or groups in support of the lab's mission.
    - For example: fast simulation with generative models as part of the GeantV toolkit.
  - Research: domain science by members of the group with the goal of sharing and spreading techniques and lessons learned.
  - Ultimately the goal is to accelerate the process to better scientific results.
- New group (~September)

Who are we?

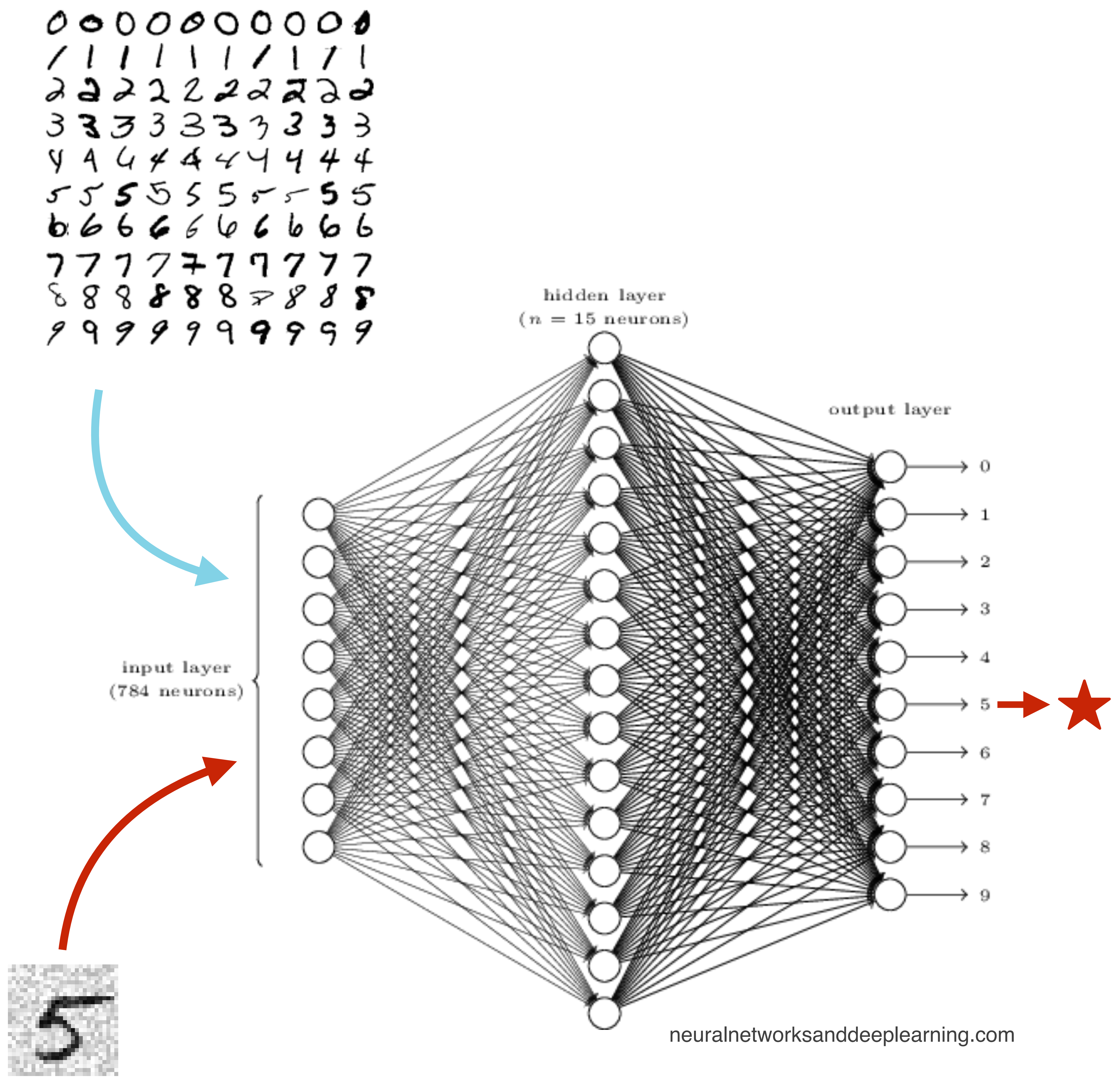
- SCD core:
  - Gabriel Perdue, Group Leader (MINERvA, DUNE, GENIE)
  - Brian Nord (DES, LSST)
  - Aristeidis Tsaris (NOvA)
  - Looking to hire an RA, to be 50% MIG, 50% cosmology, supervised by B. Nord.
- Lab community (FNAL and University) we interact with:
  - Alex Himmel (NOvA and DUNE) - working together on optical photon simulations with generative networks in DUNE
  - Alex Radovic (NOvA and DUNE) - working on the seminar series and data/code infrastructure in DUNE
  - Paddy Fox (Theory) - working on the seminar series
  - Kiel Howe (Theory) - roundtable discussion group
  - These people and several others have also contributed to the journal club.
- Looking for more active connections!

What is machine learning?

- Machine learning is a way to write programs using data.
- There are many approaches to machine learning:
  - Formal approaches are based on statistical modeling.
  - There is a wide variety of algorithms covering a wide spectrum of complexity.
- *Deep learning* is a subfield of machine learning based on models that employ hierarchical representations to solve problems - for example, using a many-layered neural network. It has had very impressive success in a number of problem domains recently:
  - Image analysis
  - Sequence to sequence models
  - Reinforcement learning
- Machine learning is not new. It is newly *successful* owing to the combination of larger, more accessible datasets and fast-enough computing. These two factors led to a Cambrian explosion of algorithmic innovation.

What can we do with machine learning?

- Two broad approaches: supervised and unsupervised learning.
- Unsupervised learning involves feeding data to an algorithm with no objective function defining performance.
  - The canonical example is clustering - searching for structure. Often useful for *visualization*.
  - Unsupervised learning is (probably) closer to how the brain learns and we don't know how to exploit it yet.
- Supervised learning involves running an algorithm over data with an objective function and a mechanism for updating the algorithm to iteratively improve the performance.
  - Note that we must define the input data and the objective function. Some algorithms attempt to learn these inputs and outputs for other algorithms, but the process here is fundamentally not *creative*.
  - In principle, we could write an algorithm by hand to accomplish the same goal, but we may find it very challenging to achieve the same scaling performance and accuracy.
  - It is a good rule of thumb in machine learning that larger datasets provide more marginal improvement in an algorithm than most algorithmic innovation. Another way of saying this is that it is much easier to solve many problems with a learning algorithm and a sufficiently large dataset than it is to derive a solution by hand.



MNIST figures from "Hands on Machine Learning with Scikit Learn and TensorFlow", from O'Reilly Media

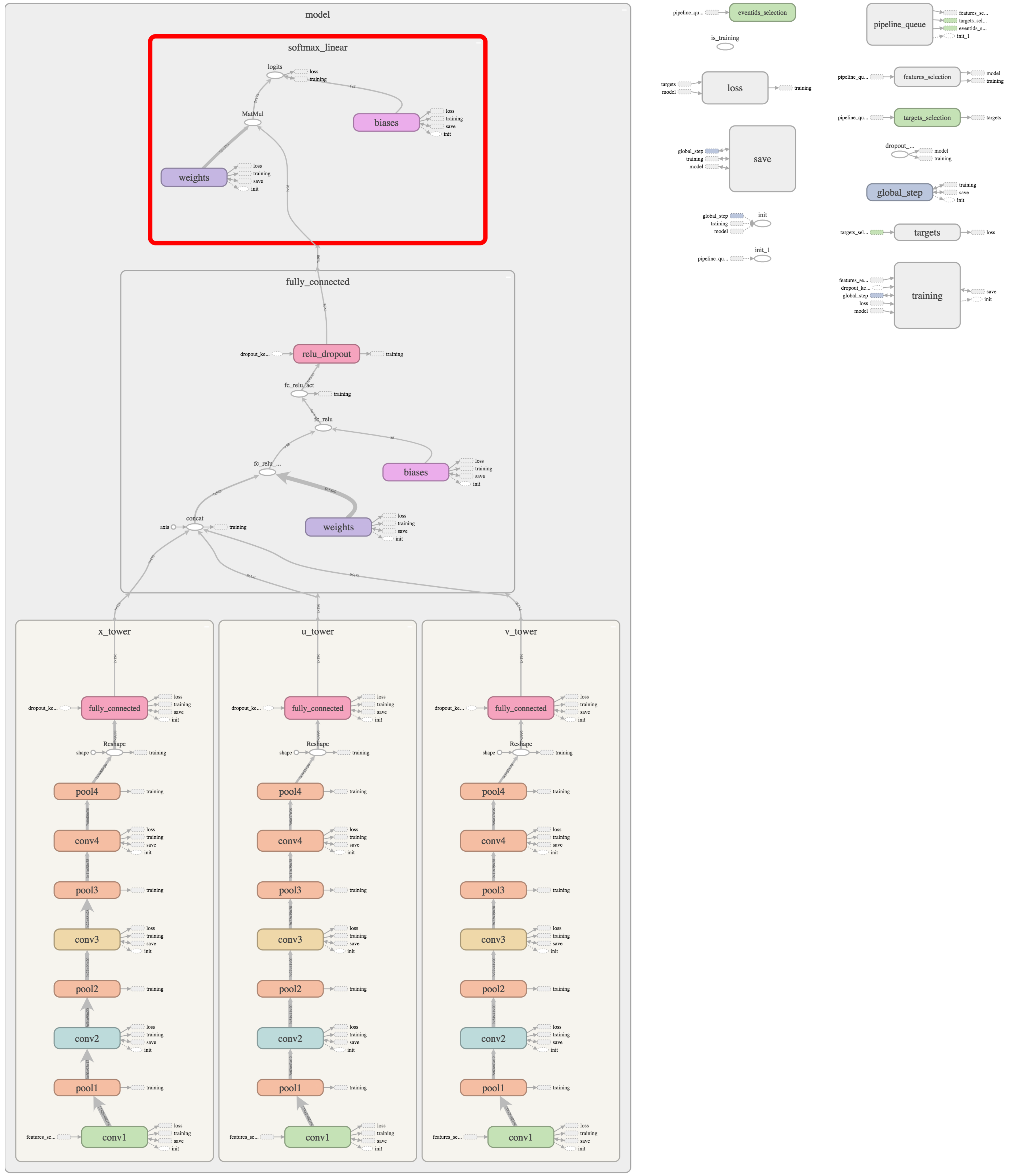
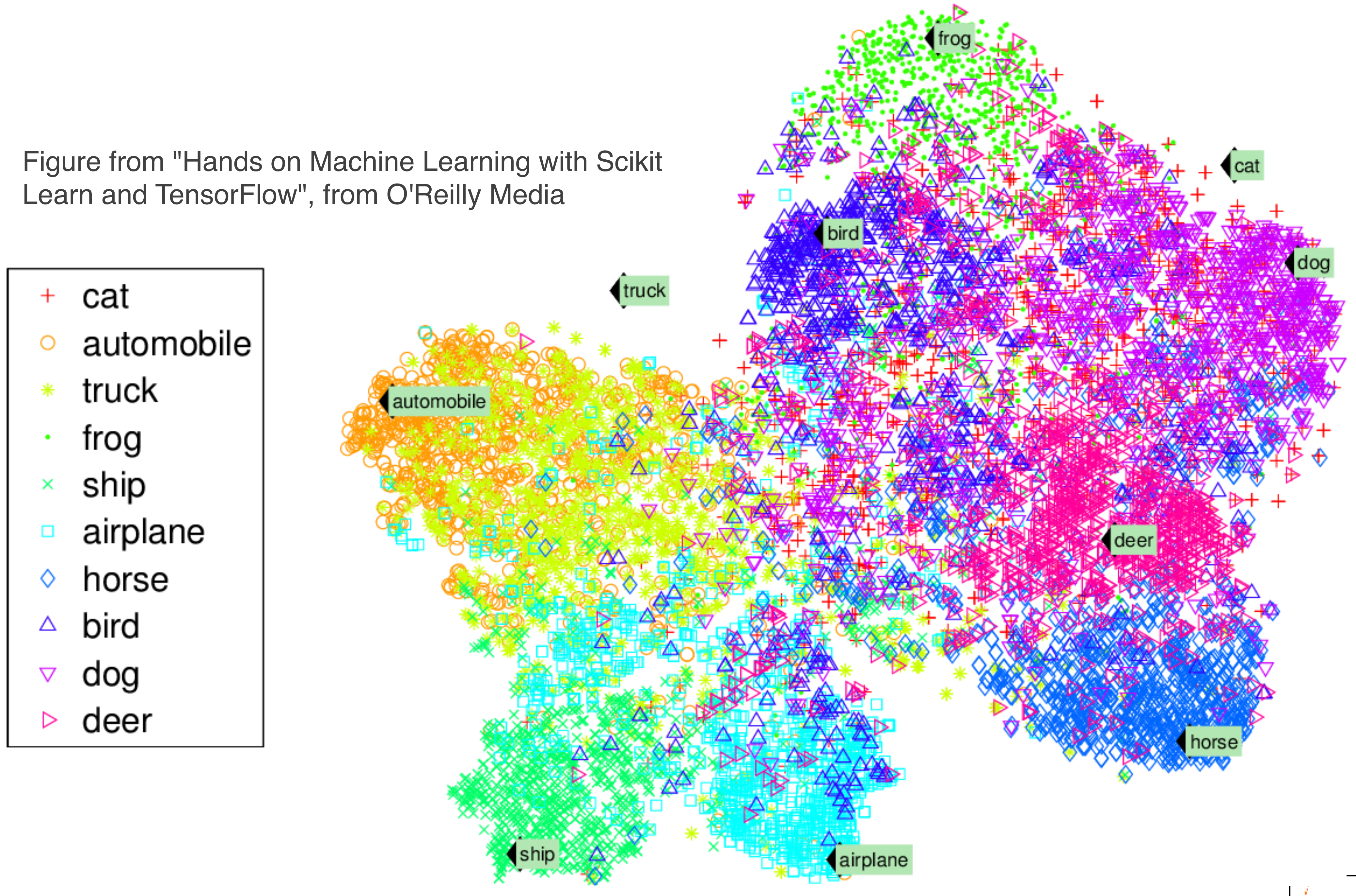
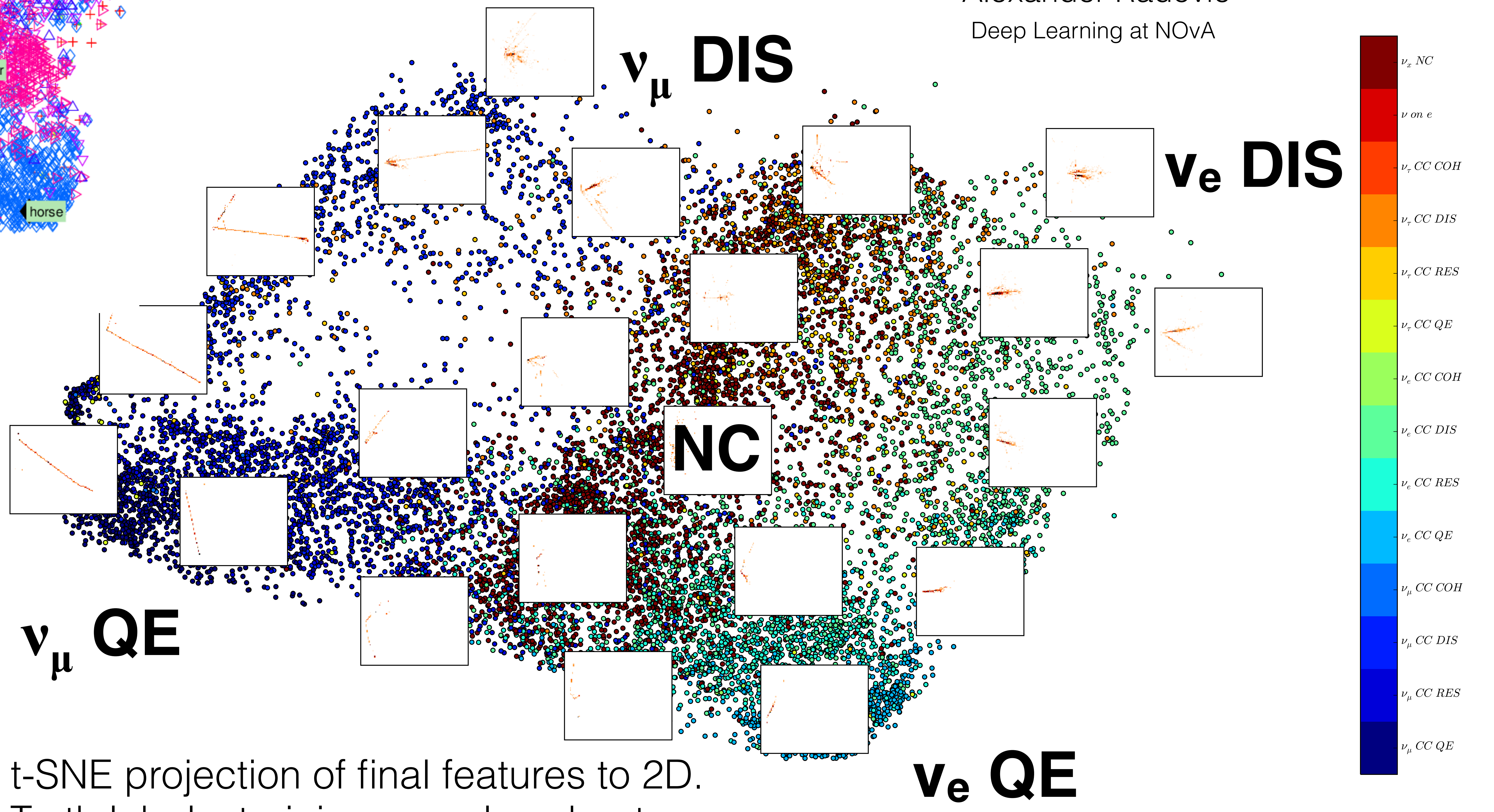


Figure from "Hands on Machine Learning with Scikit Learn and TensorFlow", from O'Reilly Media



Alexander Radovic

Deep Learning at NOvA



t-SNE projection of final features to 2D.
Truth labels, training sample subset.

Fermilab

Grounding expectations

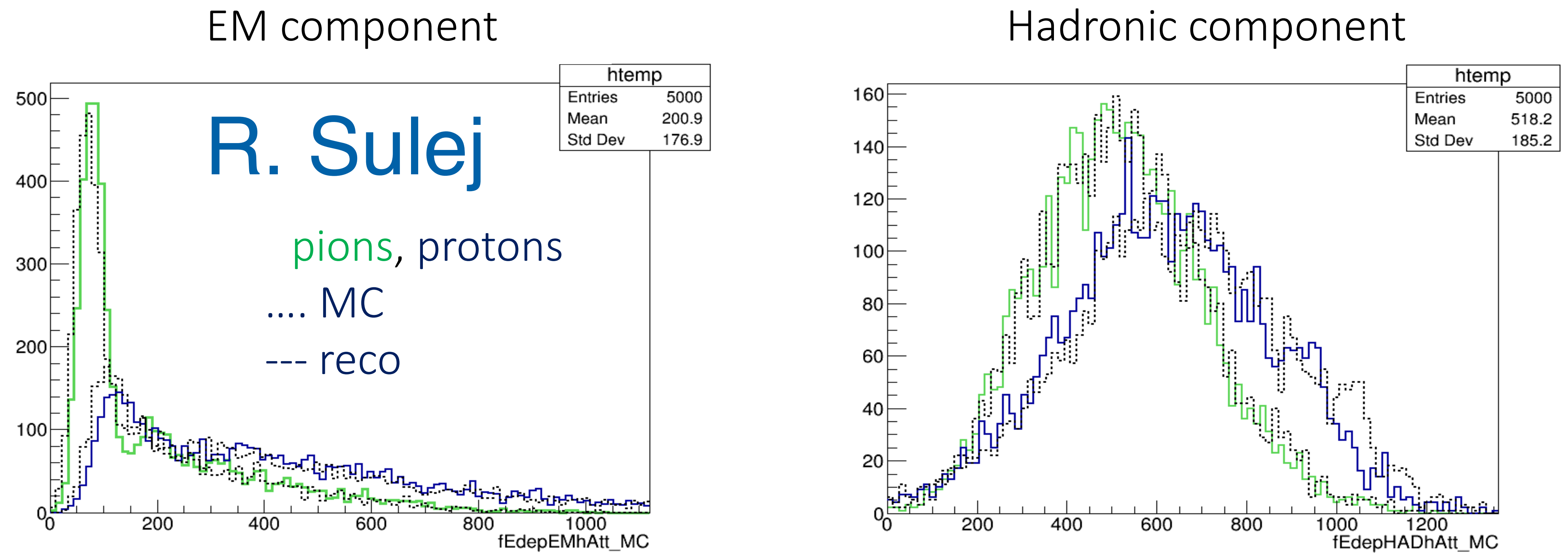
- We (humans) tend to overestimate the impact of technology in the short term and underestimate the impact in the long term.
- Throwing a machine learning algorithm at your data is not going to improve your experiment!
- It is very likely many of the “killer apps” that we’ll be using 10 years from now, though, are things we haven’t thought of yet.
- People initially adapt new tools as solutions to current problems, but eventually we will apply these techniques to questions that are predicated on the existence of fast, scalable learning tools.
- For our purposes, “AI” is meant in the sense it is most broadly used today: automated tools for decisions and analysis (“intelligence”) at scale.

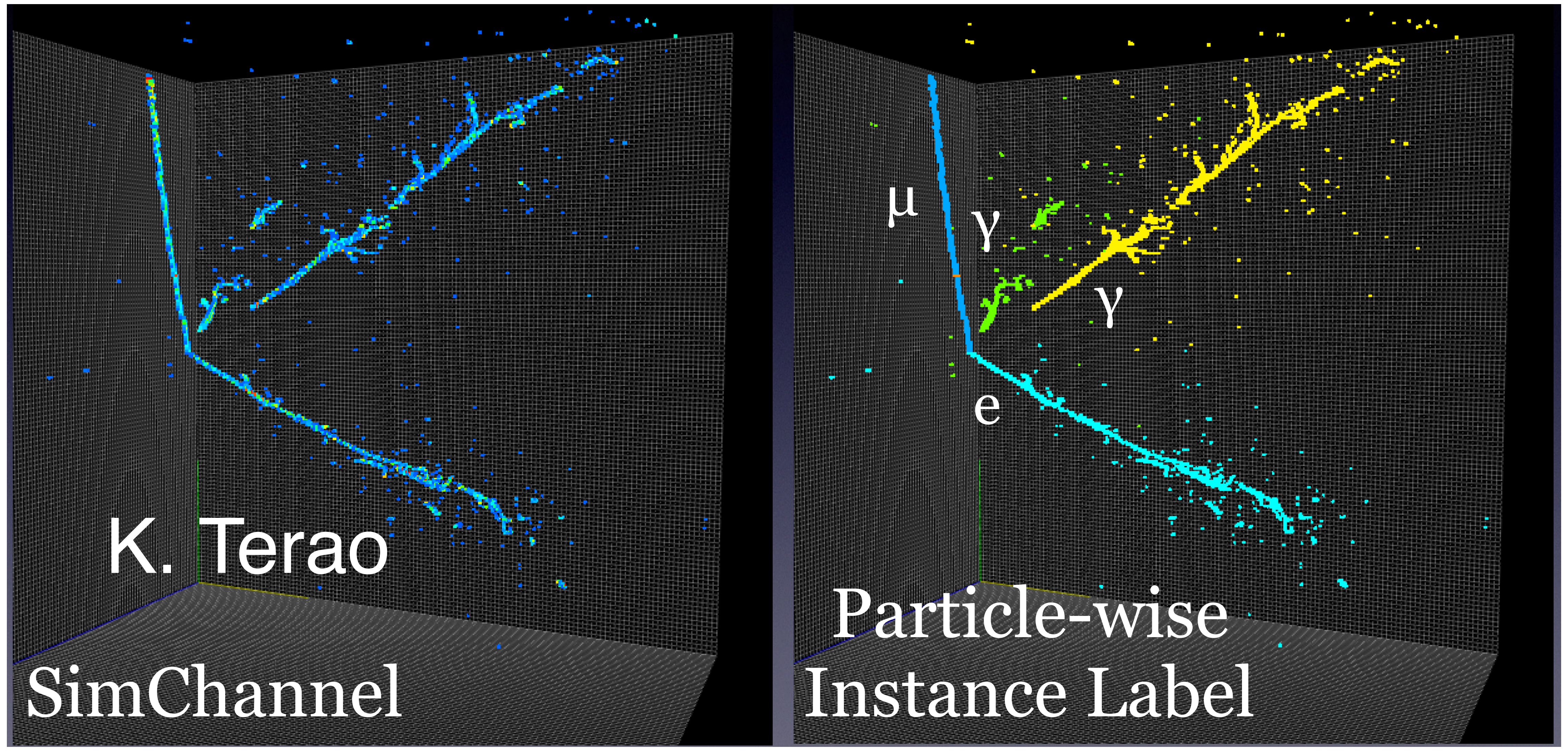
Great results in the experiments

- There is a lot of fantastic work going on in the experiments.
- Our goal is to facilitate communication between different groups, identify needs and opportunities for R&D, seed some of the needed R&D, and help to build environments and workflows that support machine learning efforts.
  - Interface to SCD frameworks, advocate, and inform SCD about new developments.
  - Develop and support common-purpose workflows.
  - Advice and consulting on tools and facilities (e.g. HPC centers).

https://indico.fnal.gov/event/15722/

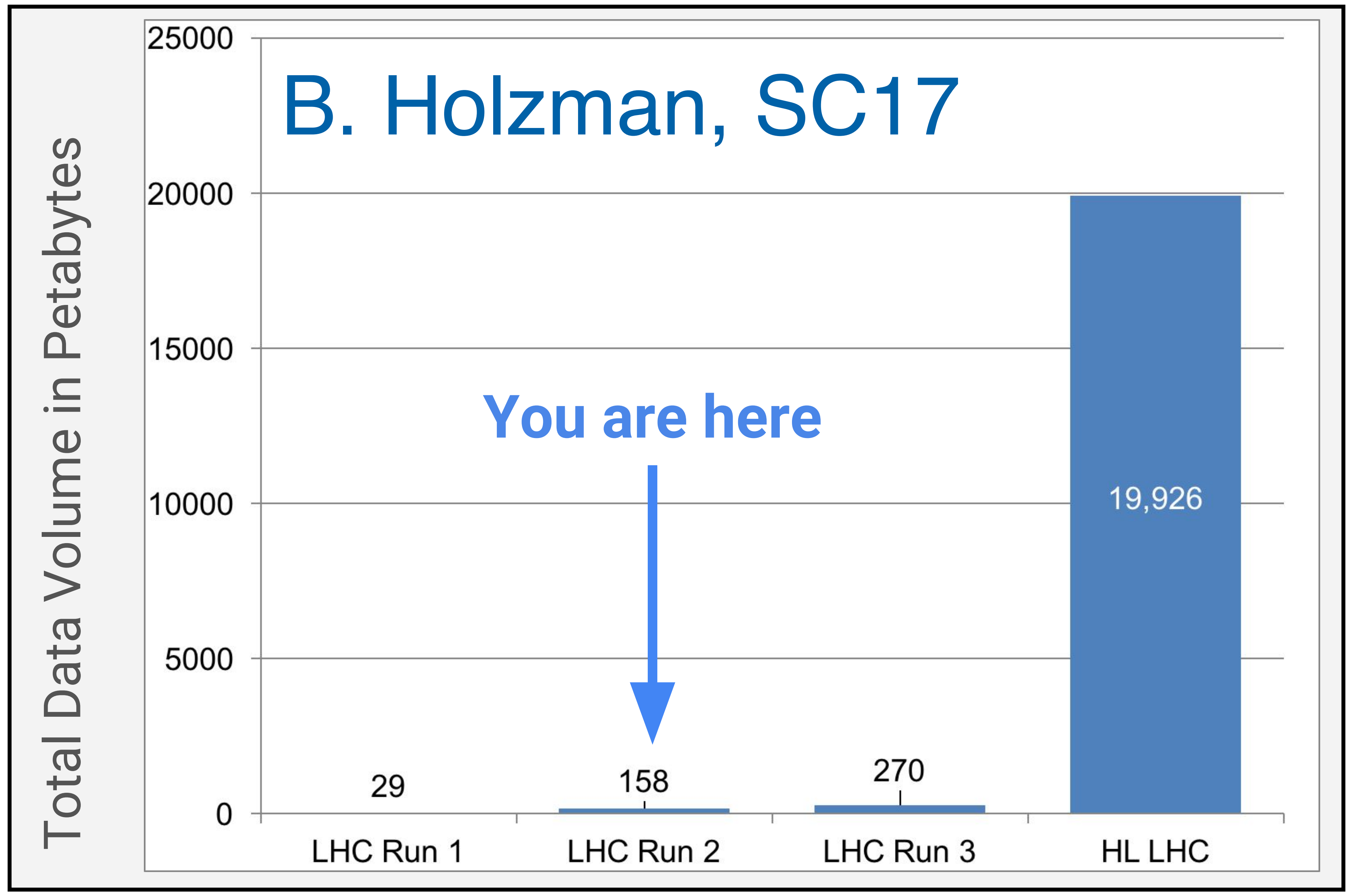
EM/track in ProtoDUNE: CNN combined with reconstruction





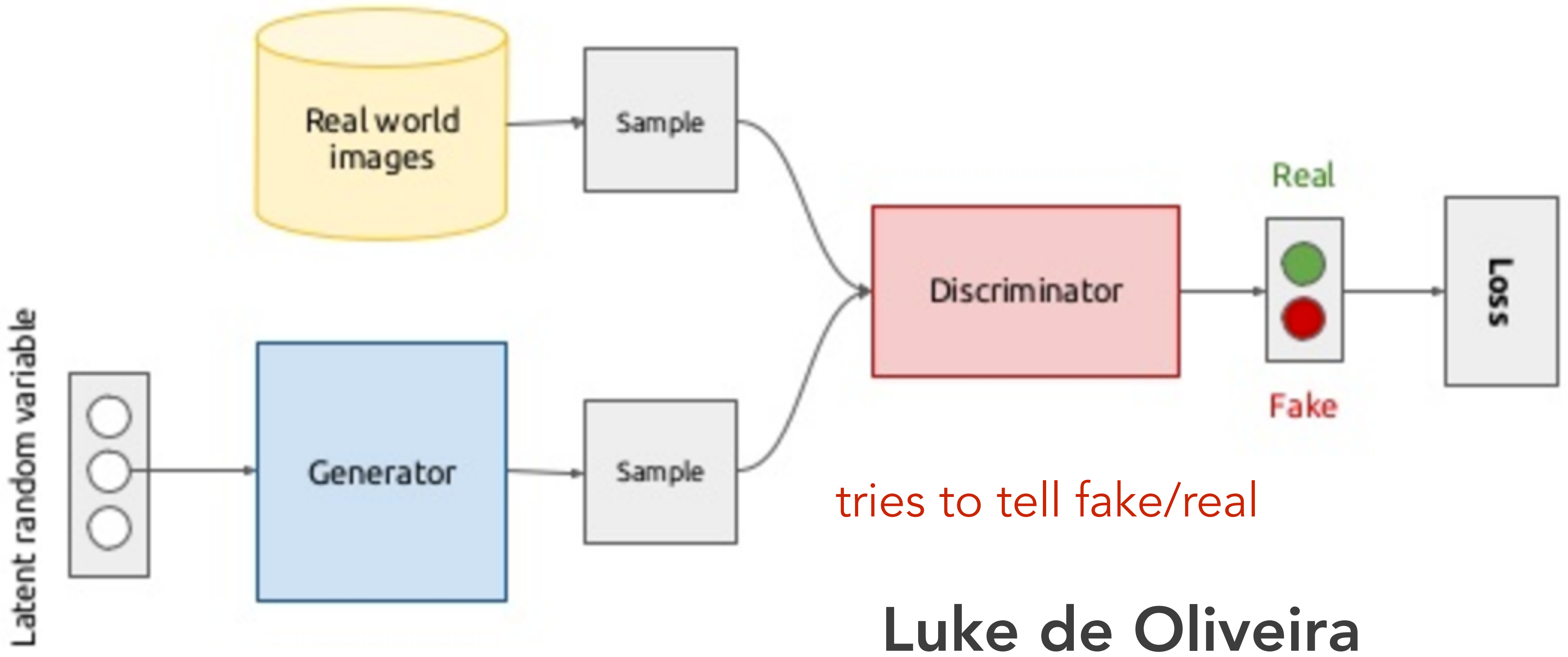
Generative models for simulation

- Future simulation needs (e.g., HL-LHC) appear likely to outstrip even optimistic resource projections.
  - Requires creative, "outside the box" thinking.
- Shower libraries face problems rooted in incompleteness and heavy data access.
- Generative models offer a potentially incredible speed-up along with better flexibility by modeling very complex distributions.
- The MIG is joining an effort in the GeantV collaboration to deliver a fast simulation program based on deep generative models.
  - Clearly benefits both the Intensity and Energy Frontiers, and lessons learned should benefit cosmological simulations as well.



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

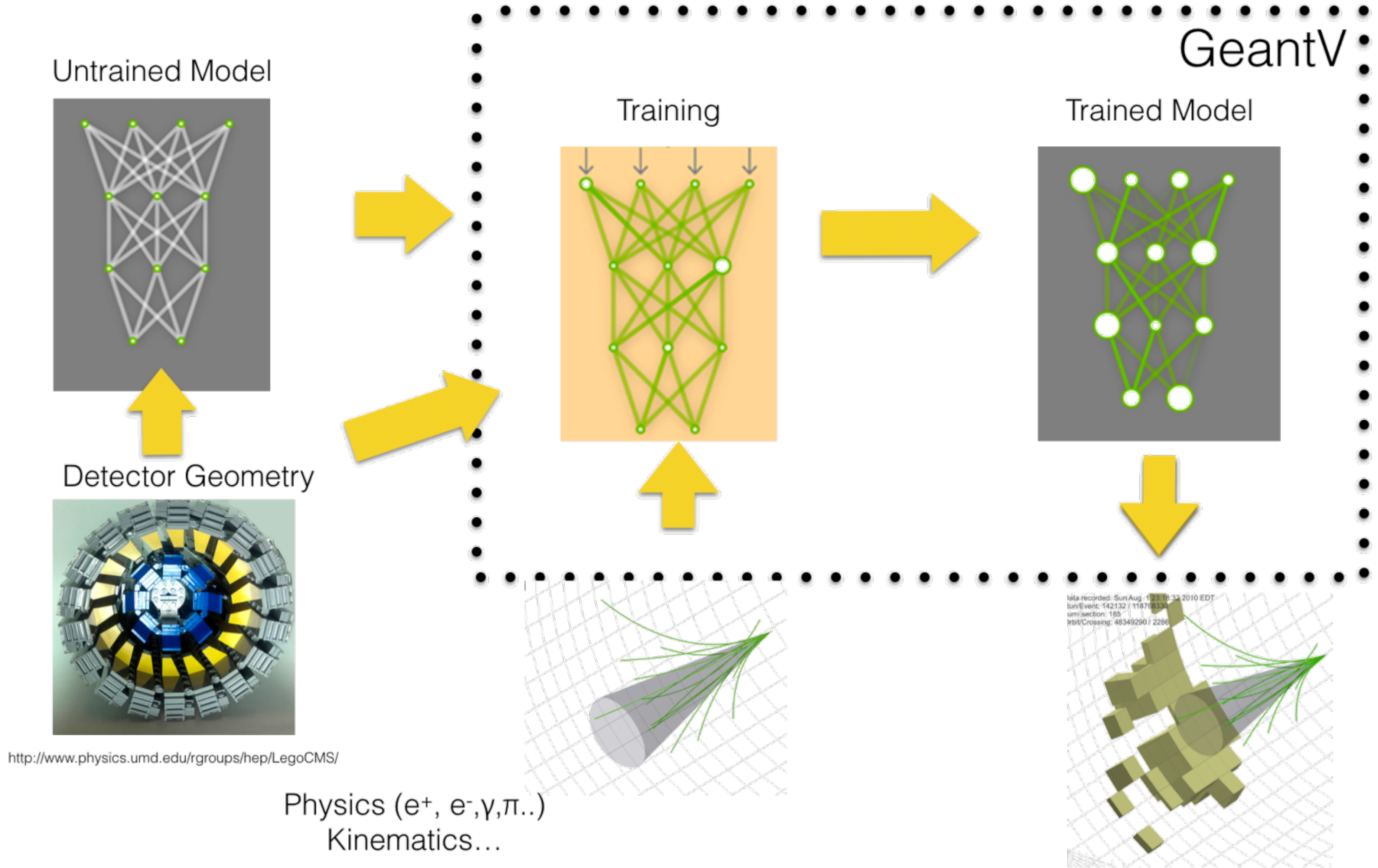
Michela Paganini*, Luke de Oliveira, Ben Nachman DS@HEP 2017



tries to tell fake/real

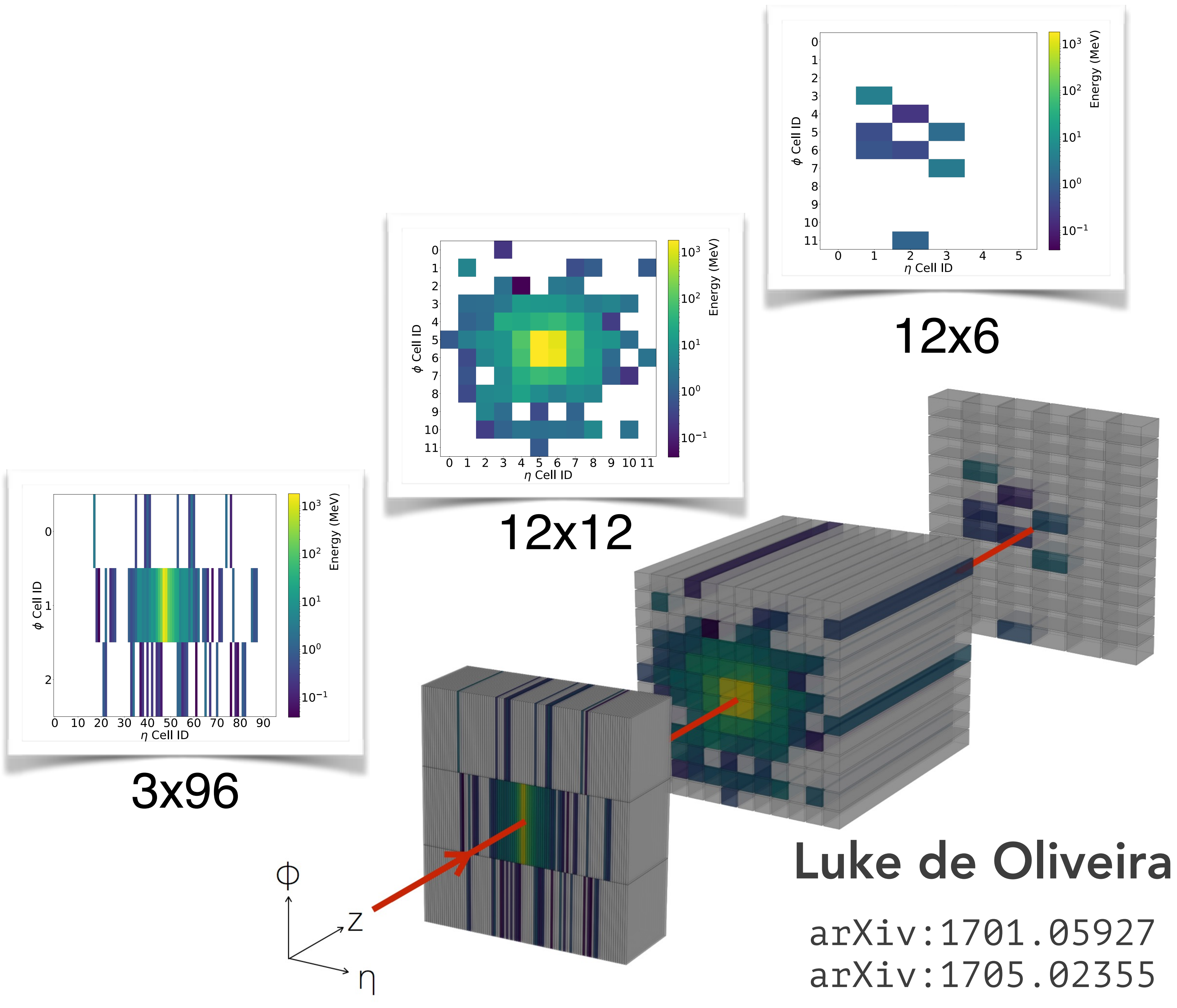
Luke de Oliveira

tries to produce real looking samples



Sofia Vallecorsa*
for the GeantV project

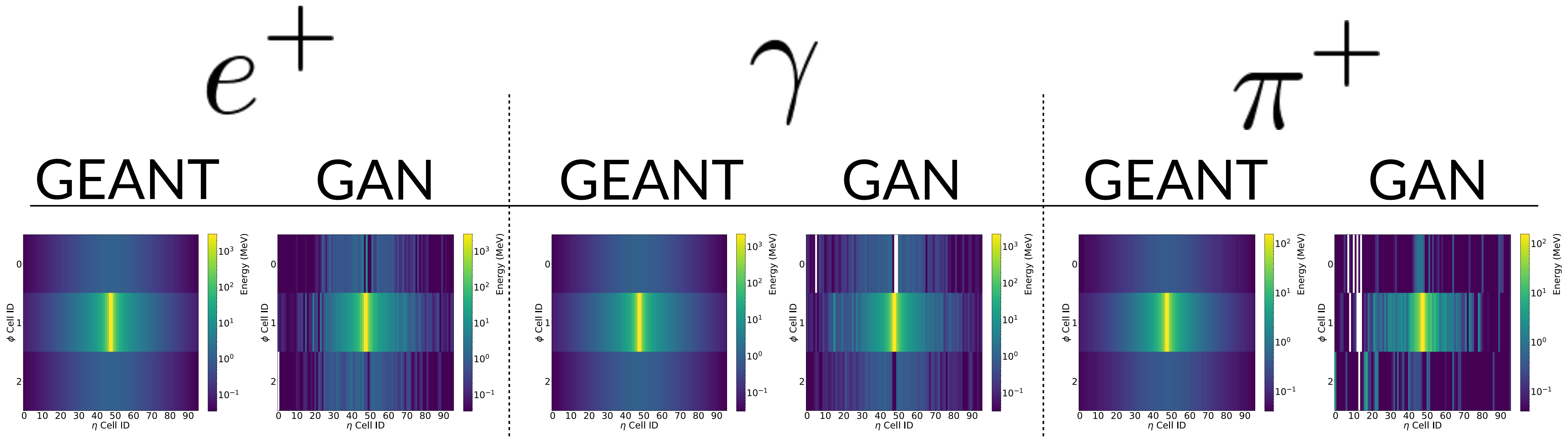
ACAT 2017



Luke de Oliveira

arXiv:1701.05927
arXiv:1705.02355

ACAT 2017



Michela Paganini*, Luke de Oliveira, Ben Nachman DS@HEP 2017

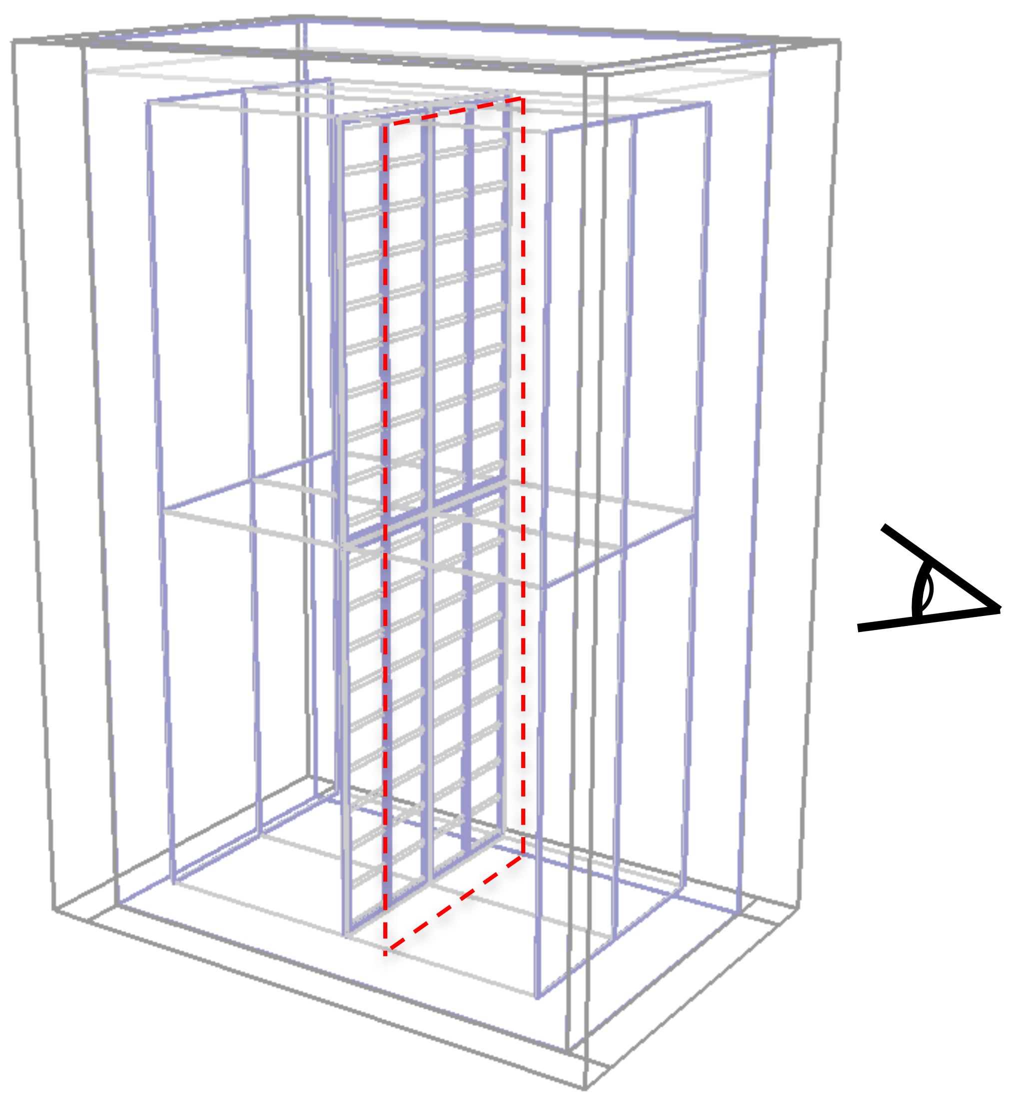
Fermilab

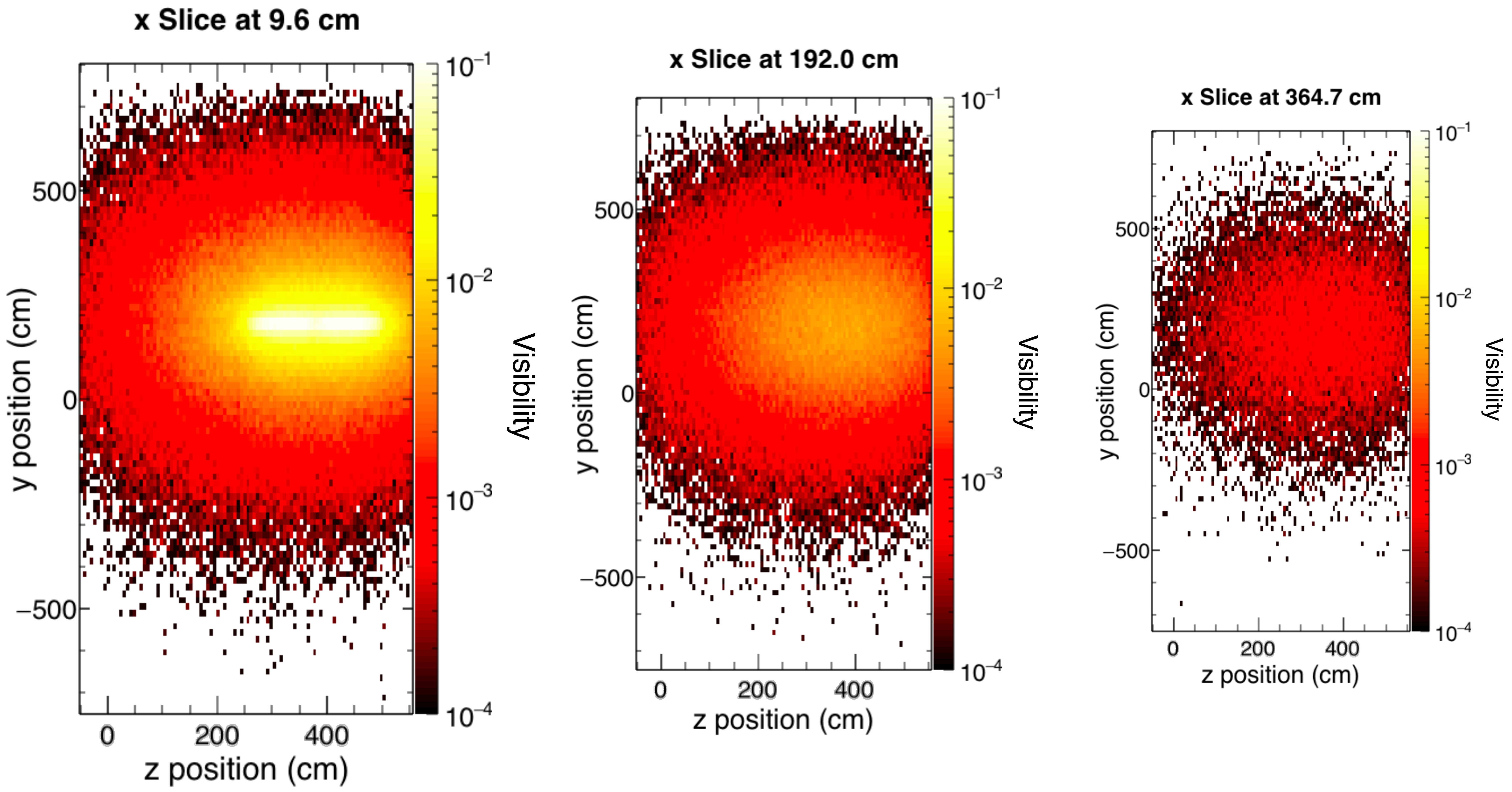
Optical photons at DUNE - partner with GeantV

- Optical photon simulation is prohibitively expensive for use in the standard simulation chain.
- The current approach to solving this issue is library look-up. But, this requires a lot of memory and access to very large table files.
- Can we improve the situation with a generative model?
- Potentially fewer concerns about leakage and splash back type effects. However, many different “vantage points” to simulate. Can the model learn to extrapolate between subsets of the voxels during training?
- We are tackling this problem in a partnership with the GeantV group.
  - Avoid replicating work by bootstrapping from their efforts so far.
  - Make important contributions to a framework and project that serves all of HEP.
  - Apply the techniques on novel datasets that support FNAL's flagship experiment.

Simulating Optical Transport

- At right, a 2D slice from the Photon Library for a single photon detector.





7

Education: journal club, roundtable, and seminars

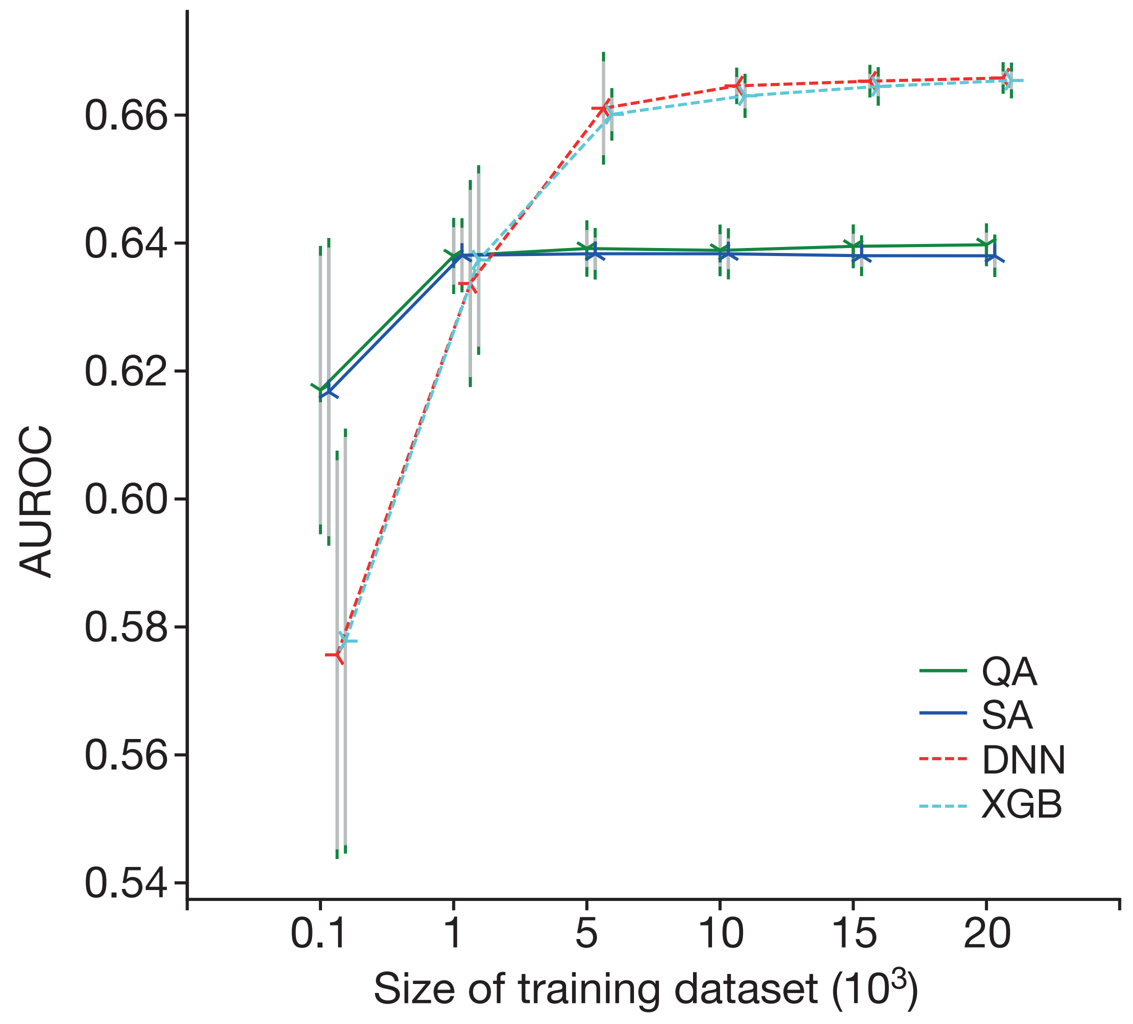
- We run a monthly meeting where we discuss papers chosen by presenters, usually with some prior discussion in the HEPMachineLearning Slack channel.
  - One or two individuals volunteer to present a paper - might be "foundational" or very recent and technical (sometimes we have both in one meeting).
- We recently began a new ~monthly community-driven effort to discuss technical problems in an informal setting.
  - The goal is to discuss technical implementation details with a wide audience and with a very low bar for participation - we don't ask people to prepare presentations, etc.
- We also run a ~monthly seminar series bringing academic and industry-based researchers in deep learning and other advanced computing techniques to the laboratory.
  - We place a special emphasis on individuals we think are likely to collaborate with groups at the lab.

Seminar success stories

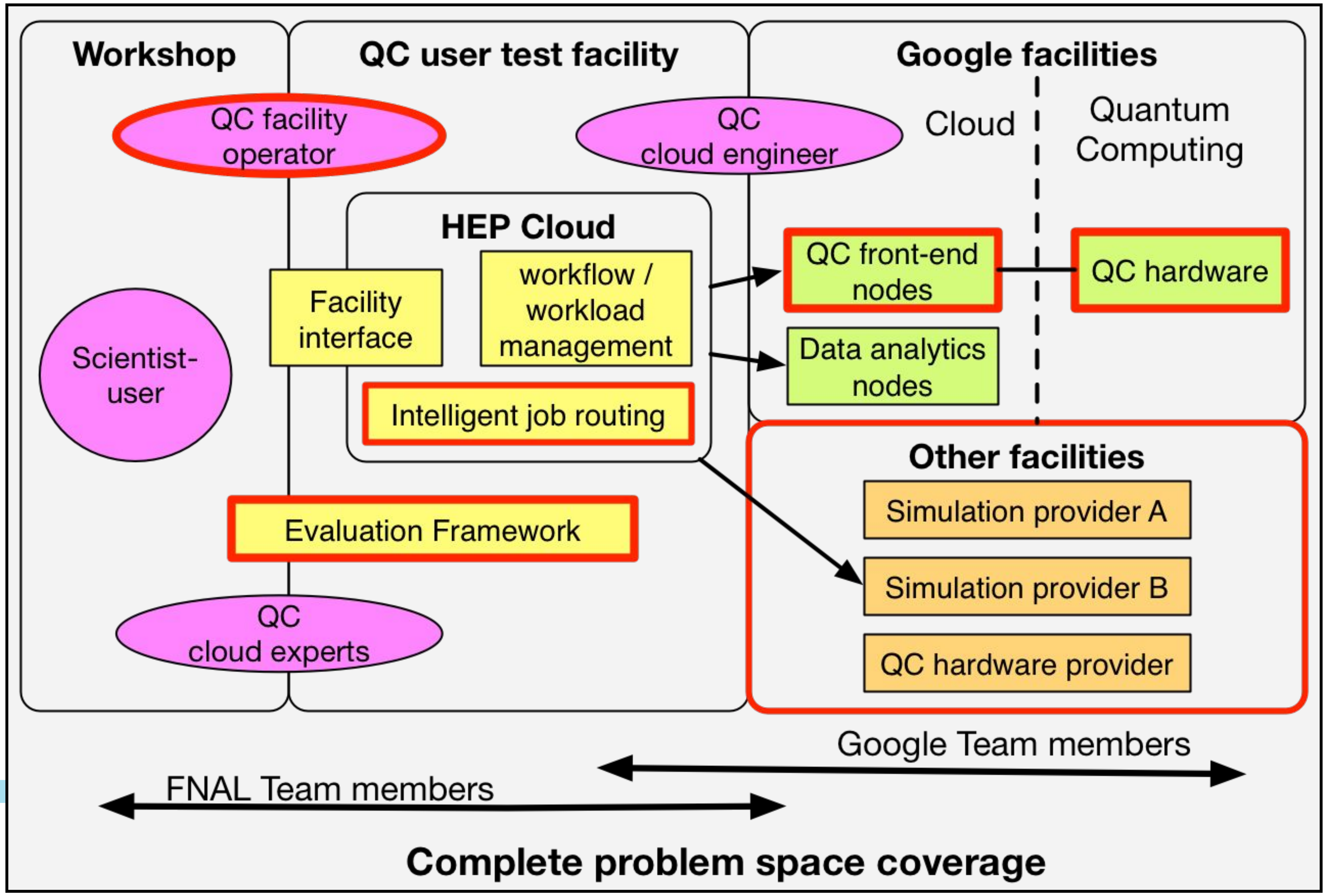
- Catherine Schuman, ORNL
  - Spoke about developments in neuromophic computing and demonstrated neural spiking networks trained by evolutionary algorithms that are competitive with deep neural networks with a tiny fraction of the number of neurons and power consumption.
  - Now working with MINERvA on classification/localization problems, and HEPTrkx on neuromorphic approaches to track finding (possible alternative to FPGA based hardware for triggering at LHC, etc.).
- Abhinav Vishnu, PNNL
  - Spoke about scaling and performance of deep learning algorithms on HPC facilities, has worked a bit with DES and MINERvA (we are all looking for grant/funding application opportunities).
- Nicholas Rubin, Rigetti Computing
  - Spoke about quantum simulation and hybrid quantum-classical algorithms.
  - In discussions about possible collaborative ventures.

Quantum machine learning

- Recently, researchers demonstrated some success using quantum annealing to build a binary signal/background classifier on a HEP dataset (Mott etc always, Nature vol 550, 375 (2017)).
  - Powerful proof-of-principle inspiration to begin investing intellectual effort.
- We are integrating into the lab's quantum information initiative to explore new applications with industrial partners (Lockheed Martin, Google).



Mott et al.



Quantum simulation

- Simulating quantum systems is likely one of the first applications of quantum computing that will yield meaningful performance advantages over classical computing.
- While quantum simulation is not necessarily a machine learning problem (although there are obvious applications for various flavors of Boltzmann Machines to function both as classifiers and as generative models), working on these problems helps us to learn about how to work with quantum computers.
- We are partnering with Los Alamos (J. Carlson, A. Roggero, S. Pastore) and U. South Carolina (A. Baroni) to develop a neutrino-nucleus scattering problem using a Hamiltonian with many realistic features.
- We plan on deploying the algorithm on Google hardware this coming Spring.

Other considerations/projects

- Looking at reinforcement learning for accelerator controls.
  - Potentially large cost savings.
  - Experience in adaptive control useful on other projects (grid computing, building control, etc.).
  - Trying to build team skill set for this work.
- More public exposure/outreach.
  - ML is very hot right now.
  - We are doing interesting, big science with ML at the lab (Higgs physics, new physics searches at the LHC, CP violation searches at the long baseline neutrino program, sterile neutrino searches at the short baseline program).
  - We should be presenting at industry events (e.g. AWS re:invent, O'Reilly AI, etc.) to raise the profile of the lab. This will help with ML recruiting, help provide exposure to industry for postdocs and students, etc.

Conclusions

- The Machine Intelligence Group aims to support efforts in machine learning and advanced algorithms across the entire lab community through improved communication, education, advocacy, and direct involvement in broadly impactful projects.
- Our goal is to be a center for activities that cut across "frontier boundaries" and leverage the resources at our disposal by plugging into common frameworks and mechanisms whenever possible.
- We are a bridge to the outside world of research in these topics, and bring interesting potential collaborators to the lab to interface with lab staff and users.
- We are always looking to partner on interesting new projects (e.g., adaptive controls via reinforcement learning as a HEPCloud decision engine, etc.).

Thank you for listening!

Core SCD group

- Gabriel Perdue, Associate Scientist, SCD, Group Leader (MINERvA, DUNE, GENIE)
  - 0.4 FTE
  - Also leads the FNAL GENIE group
- Brian Nord, Associate Scientist, SCD (DES, LSST)
  - 0.25 FTE
  - Joint appointment at the University of Chicago with a focus on scientific communication
- Aristeidis Tsaris, Research Associate, SCD (NOvA)
  - 0.2 FTE
  - Protecting a focused, coherent research program aimed at success at the next level is our primary focus when deploying effort.
- Looking to hire an additional RA, to be 50% MIG, 50% cosmology, supervised by B. Nord.

Education: workshops and tutorials (planning underway)

- We plan on executing or facilitating a series of tutorials and workshops focused on using a developing deep learning applications.
- Our group has broad expertise across a variety of technologies:
  - TensorFlow
  - Theano
  - Keras as a frontend (backends include TensorFlow, Theano, and new libraries like PlaidML)
  - Caffe
  - Scikit-learn
  - XGBoost
- Additionally, SCD is well-equipped to provide VMs for tutorial software environemnts.