Pertinent Developments in Machine Learning for Small Experiments

Christopher Tunnell

Rice University

Software and Computing for Small HEP Experiments

Nov 16, 2021

A few things you maybe hadn’t heard and I think are cool

- Interesting developments in field (see LOIs)
  - Papers
  - NEXT Energy Distance
- Efforts worth knowing:
  - NSF (and soon DOE?) Institutes
  - Extrkx, didacts, DANCE-Edu
- Small experiments as ML Testbed [remaining time]
  - Uncertainty quantification
  - Robust physics-informed architectures

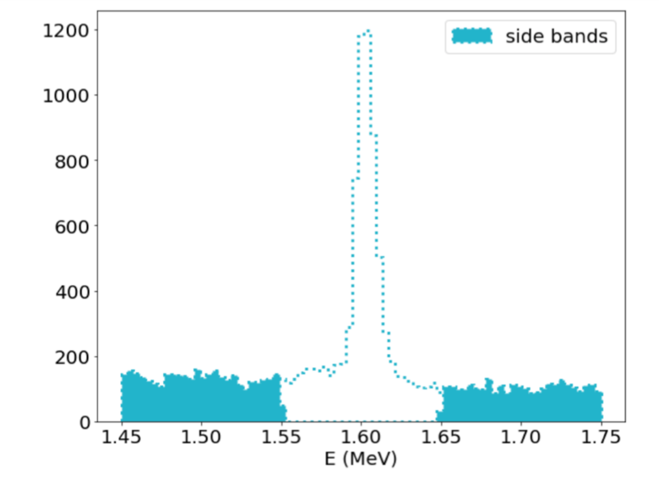
Lots of work CNN classifiers

- CNN: Major advances in computer vision and natural language processing (NLP) in past decades, so we leverage heavily computer vision a la CNN
  - Few borrow from NLP….
- Classifier: NNs were breakthrough in classification problems rather than regressions
- References
  - DUNE: Neutrino interaction classification with a convolutional neural network in the DUNE far detector (2020) [link]
  - IceCube: A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory (2021) [link]
  - Minerva: Neutral pion reconstruction using machine learning in the experiment at $\langle E_\nu \rangle$ 6 GeV [link]
  - MicroBooNE: Electromagnetic Shower Reconstruction and Energy Validation with Michel Electrons and π0 Samples for the Deep-Learning-Based Analyses in MicroBooNE **[link]**
  - EXO: [link]
  - LUX: [link]
  - *Most HEP results have ML in places within pipeline*
- Mostly medium-scale experiments since small experiments rarely have effort to publish

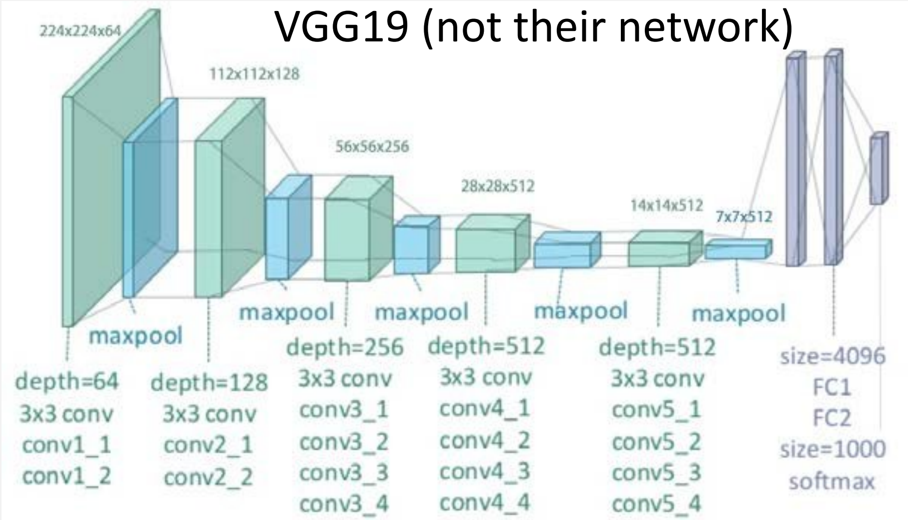
A Review on Machine Learning for Neutrino Experiments (2020) Psihas et al

Cool thing: NEXT’s Energy Distance

- NEXT is 0nuBD experiment.
  - MC and data are always different, which can make supervised learning ill-defined
  - Do data augmentation to avoid overfitting (top right)
  - How quantify overfitting? And show learning important features in data
- Procedure: Augment, then look at “Energy Distance” in side band
- <u>Small experiments are great test beds for new methods</u>

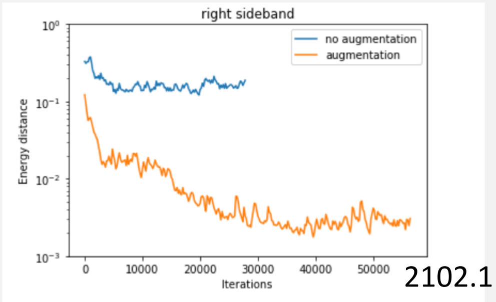


VGG19 (not their network)



$$E_{n,m}(X,Y) := 2A - B - C$$

$$A := \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\|x_i - y_j\|, B := \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|, C := \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|y_i - y_j\|$$



2009.10783

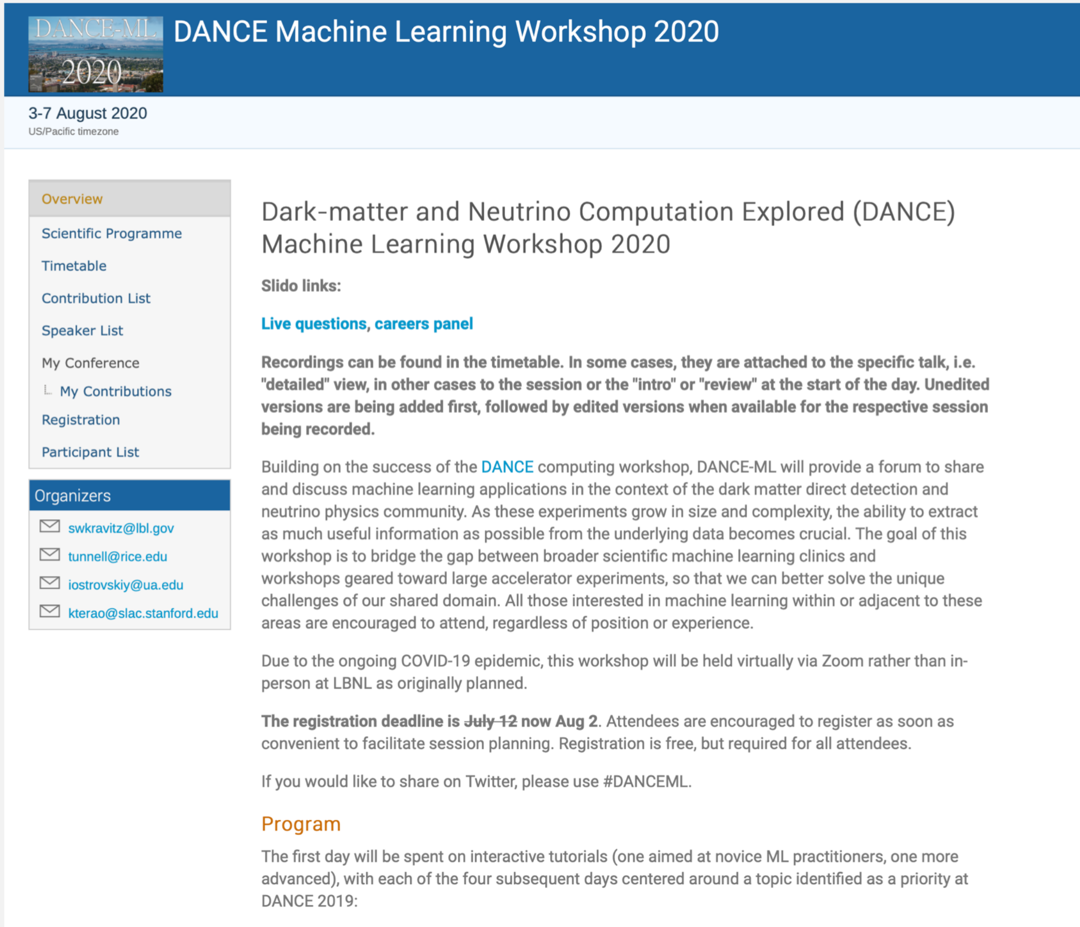
2102.11931 also interesting

Entities with relevant ML components:

- Physics + AI w/ IAIFI: Institute for Artificial Intelligence and Fundamental Interactions https://iaifi.org [mailing list]
- Fast ML w/ A3D3: Accelerated Artificial Intelligence Algorithms for Data-Driven Discover https://a3d3.ai
- Community with CLARIPHY: https://clariphy.org [google group]
- Track finding w/ Exa.TrkX: https://exatrkx.github.io/
- ML reconstruction w/ DIDACTS: https://didacts.org [output]
- HEP Software: https://iris-hep.org/
- DANCE-Edu: Fellowship and post-pandemic summer schools

DANCE-ML 2022?

- Successful DANCE-ML in 2020 (Thanks, Scott!)
- No desire for remote. Again in 22?
  - Already funded



DANCE Machine Learning Workshop 2020

3-7 August 2020

US/Pacific timezone

- Overview
- Scientific Programme
- Timetable
- Contribution List
- Speaker List
- My Conference
  - My Contributions
- Registration
- Participant List

Organizers

swkravitz@lbl.gov

tunnell@rice.edu

iostrovskiy@ua.edu

kterao@slac.stanford.edu

Dark-matter and Neutrino Computation Explored (DANCE) Machine Learning Workshop 2020

Slido links:

Live questions, careers panel

Recordings can be found in the timetable. In some cases, they are attached to the specific talk, i.e. "detailed" view, in other cases to the session or the "intro" or "review" at the start of the day. Unedited versions are being added first, followed by edited versions when available for the respective session being recorded.

Building on the success of the DANCE computing workshop, DANCE-ML will provide a forum to share and discuss machine learning applications in the context of the dark matter direct detection and neutrino physics community. As these experiments grow in size and complexity, the ability to extract as much useful information as possible from the underlying data becomes crucial. The goal of this workshop is to bridge the gap between broader scientific machine learning clinics and workshops geared toward large accelerator experiments, so that we can better solve the unique challenges of our shared domain. All those interested in machine learning within or adjacent to these areas are encouraged to attend, regardless of position or experience.

Due to the ongoing COVID-19 epidemic, this workshop will be held virtually via Zoom rather than in-person at LBNL as originally planned.

The registration deadline is ~~July 12~~ now Aug 2. Attendees are encouraged to register as soon as convenient to facilitate session planning. Registration is free, but required for all attendees.

If you would like to share on Twitter, please use #DANCEML.

Program

The first day will be spent on interactive tutorials (one aimed at novice ML practitioners, one more advanced), with each of the four subsequent days centered around a topic identified as a priority at DANCE 2019:

Probabilistic Programming

- Deep link between ML and auto. differentiation
- Machine-learning optimization of experiment
  - Articles on what this means [1] [2], great talk
- MODE Collaboration in Europe [info] [workshop]
- Key idea:
  - Auto differentiation key to recent ML advances (backprop)
  - For some function ‘f’, can see how output varies as vary parameters
  - ‘f’ can be your simulation or ML model
  - Can use to optimize experiments w/ e.g. NEST

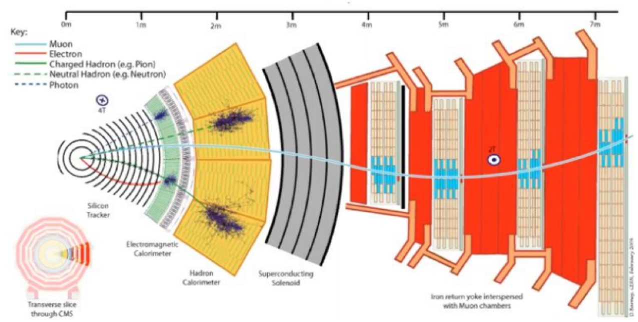
$$f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$$

automatic differentiation

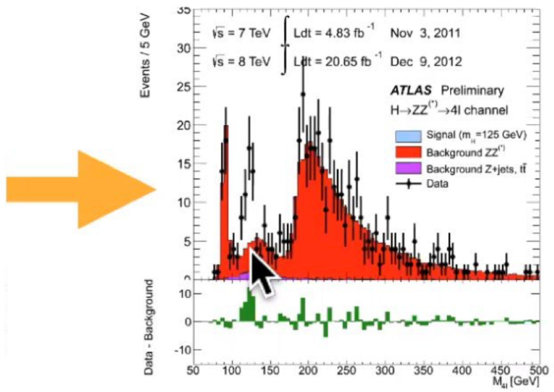
$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

Differentiable programming in particle physics

- **Differentiable analysis**
  Unify analysis pipeline by simultaneously optimizing the free parameters of an analysis with respect to the desired physics objective



- **Differentiable simulation**
  Enable efficient simulation-based inference, reducing the number of events needed by orders of magnitude



Baydin, Cranmer, Feickert, Gray, Heinrich, Held, Melo, Neubauer, Pearkes, Simpson, Smith, Stark, Thais, Vassilev, Watts. 2020. “Differentiable Programming in High-Energy Physics.” In Snowmass 2021 Letters of Interest (LOI), Division of Particles and Fields (DPF), American Physical Society. https://snowmass21.org/loi.

Two novel classes of ML problems that experiments face, but small experiments can tackle more easily

	Known	**Unknown**
Knowns	Known Knowns	Unknown Knows
Unknowns	Known Unknowns	Unknown Unknowns

Exponential advances in direct-detection dark matter reveals missing methods

Detector's extreme simplicity and sensitivity requires deep understanding of each step of pipeline, and 1% improvements worth millions of dollars

What is needed for discovery...

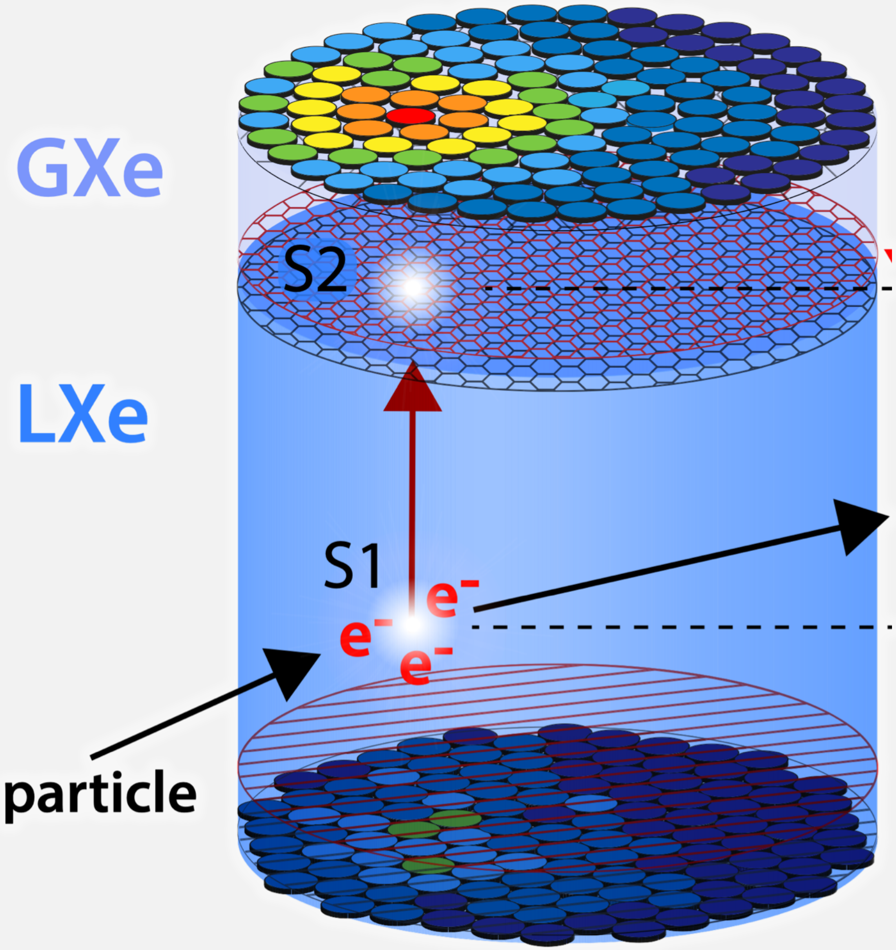
Known science
+uncertainty
+synth. data

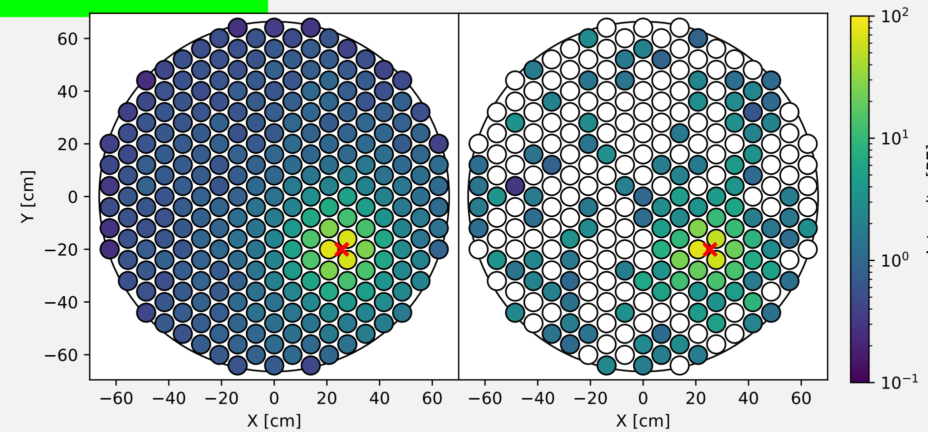
Observed
data and
~~ground truth~~

Machine Learning

Pr[X|data, science]

Exponential advances in direct-detection dark matter reveals missing methods





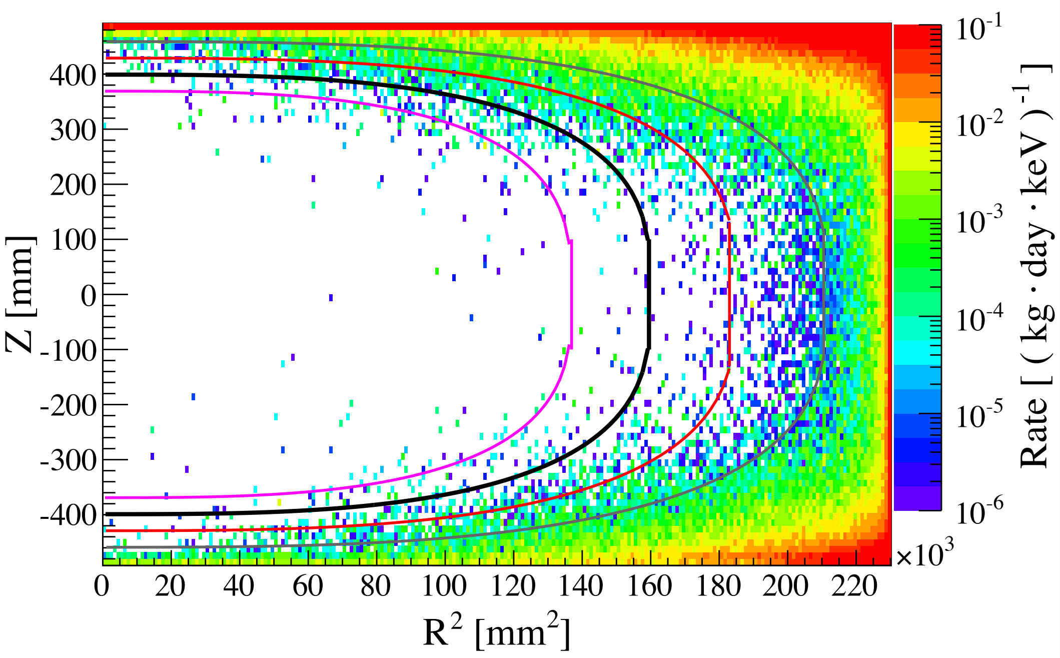
- Measurement process, this is all we do
  - Inverse problem
  - 1/100 sensor spacing
  - Poisson counting noise
  - Almost perfect forward model
- Dimensions have some meaning, not images, but no standard architectures

$$\mathcal{R}^{500} \rightarrow \mathcal{R}^{2}$$

Known Unknowns Bayesian Networks

	Known	**Unknown**
Knowns	Known Knowns	Unknown Knows Graph denoiser
Unknowns	Known Unknowns Bayesian Networks	Unknown Unknowns Physics-constrained Graph NN

Known Unknowns Bayesian Networks



1512.0750

TABLE I: Expected number of events for each background component in the fiducial mass; in the full $cS1 \in [3, 70]$ PE, $cS2_b \in [50, 8000]$ PE search region and in a reference region between the NR median and -2σ quantile in $cS2_b$. Uncertainties <0.005 events are omitted. The ER rate is unconstrained in the likelihood; for illustration, we list the best-fit values to the data in parentheses.

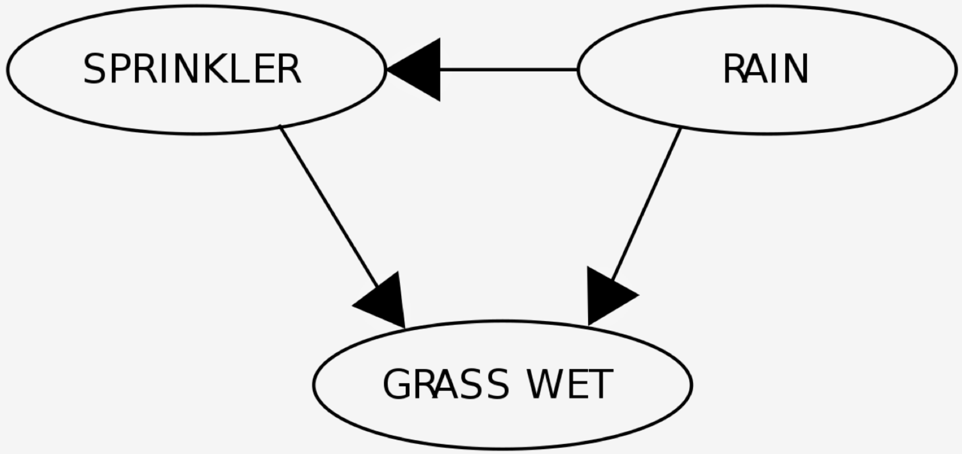
	Full	Reference
Electronic recoils (*ER*)	(62 ± 8)	$(0.26^{+0.11}_{-0.07})$
Radiogenic neutrons (*n*)	0.05 ± 0.01	0.02
CNNS (ν)	0.02	0.01
Accidental coincidences (*acc*)	0.22 ± 0.01	0.06
Wall leakage (*wall*)	0.5 ± 0.3	0.01
Anomalous (*anom*)	$0.10^{+0.10}_{-0.07}$	0.01 ± 0.01
Total background	63 ± 8	$0.36^{+0.11}_{-0.07}$
50 GeV/c^2, 10^{-46}cm^2 WIMP	1.66 ± 0.01	0.82 ± 0.06

1705.06655

w/ UDelaware, Tina Peters, Hagit Shatkay

Known Unknowns Bayesian Networks

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

Posterior!

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

For 'graphical models', these tables can be learned from full-chain simulations!

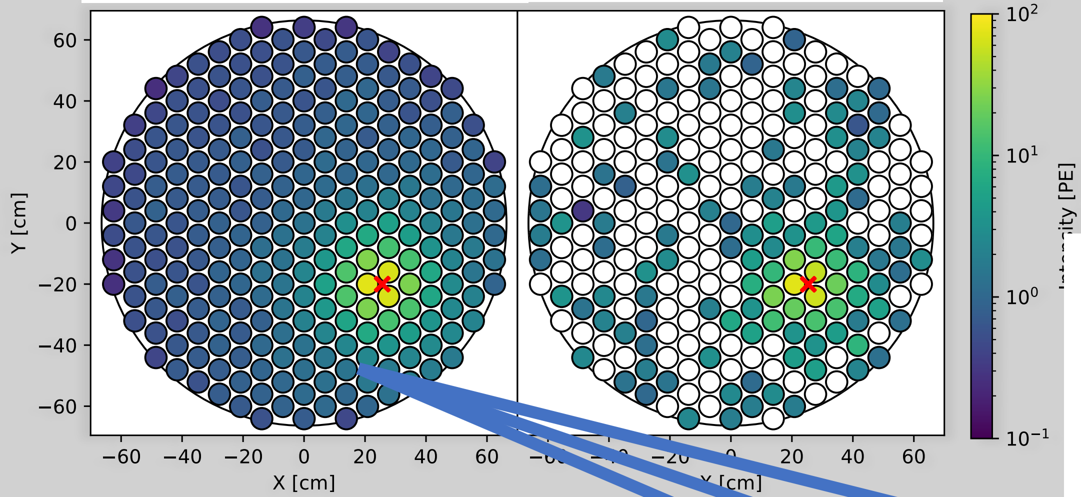
w/ UDelaware, Tina Peters, Hagit Shatkay

Known Unknowns Bayesian Networks

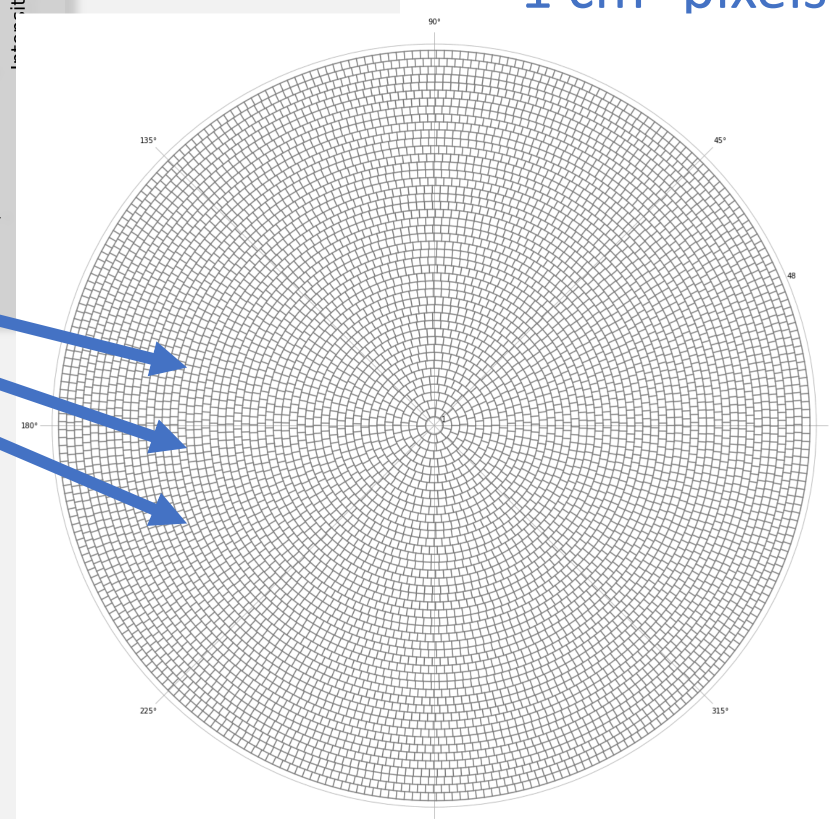
Consider localization

Sensor array

Average | Poisson fluctuations



1 cm^2 pixels



Connect each PMT to each pixel

No explicit likelihood

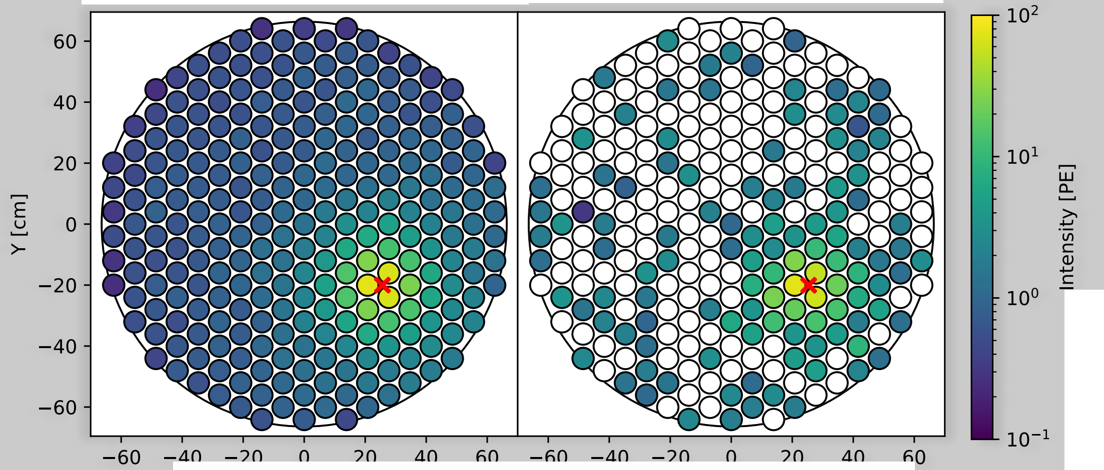
Tutorial

Known Unknowns Bayesian Networks

Sensor array

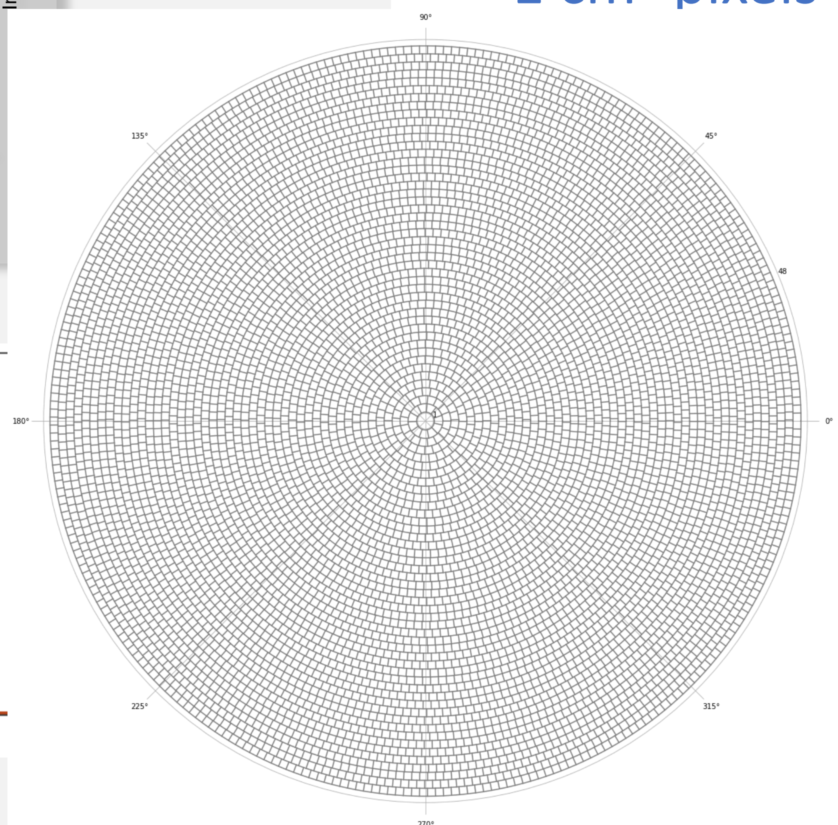
Average

Poisson fluctuations

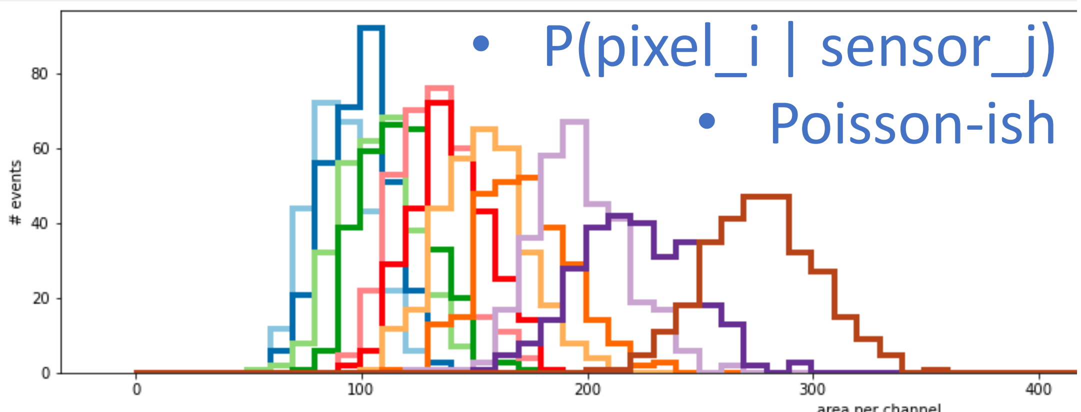


$$P(\text{PMT \#0} = x_0, \text{Position} = p) = \frac{e^{-\lambda_{p,0}} {\lambda_{p,0}}^{x_0}}{x_0!},$$

1 cm² pixels



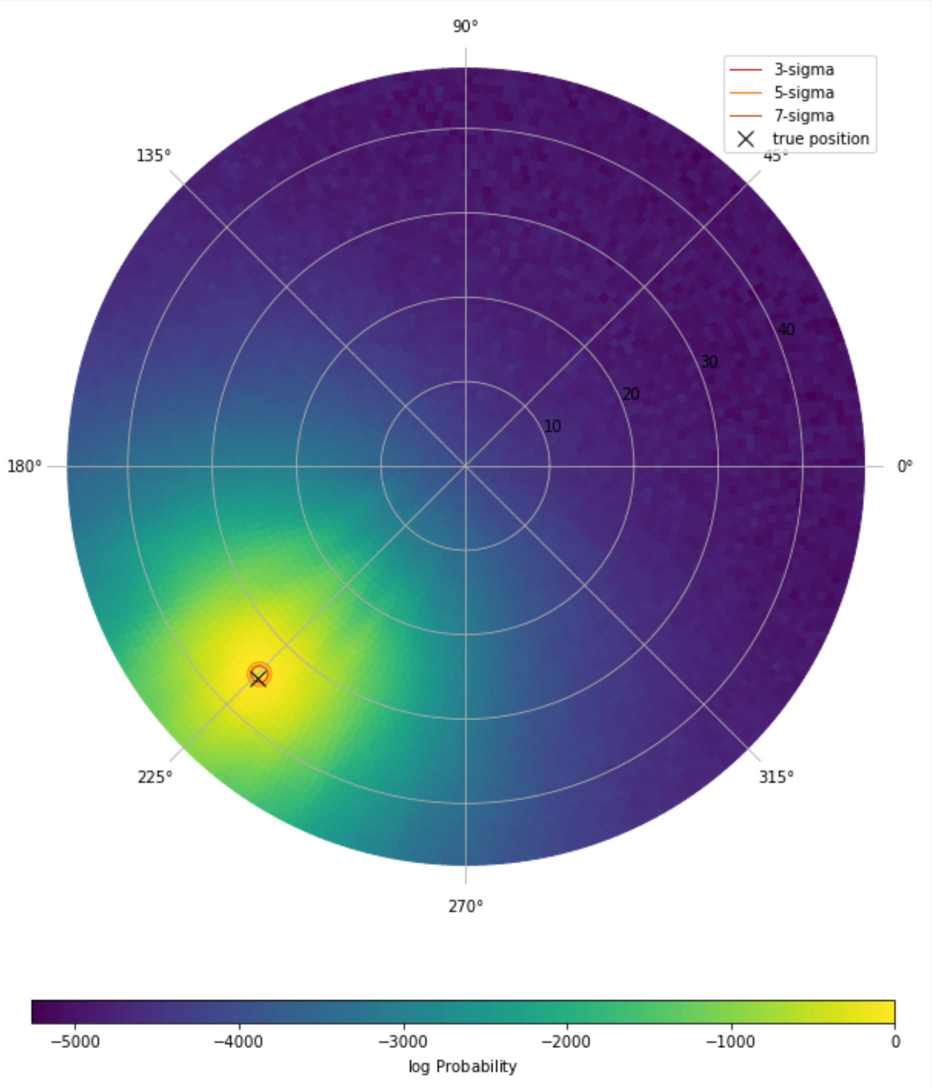
- P(pixel_i | sensor_j)
  - Poisson-ish

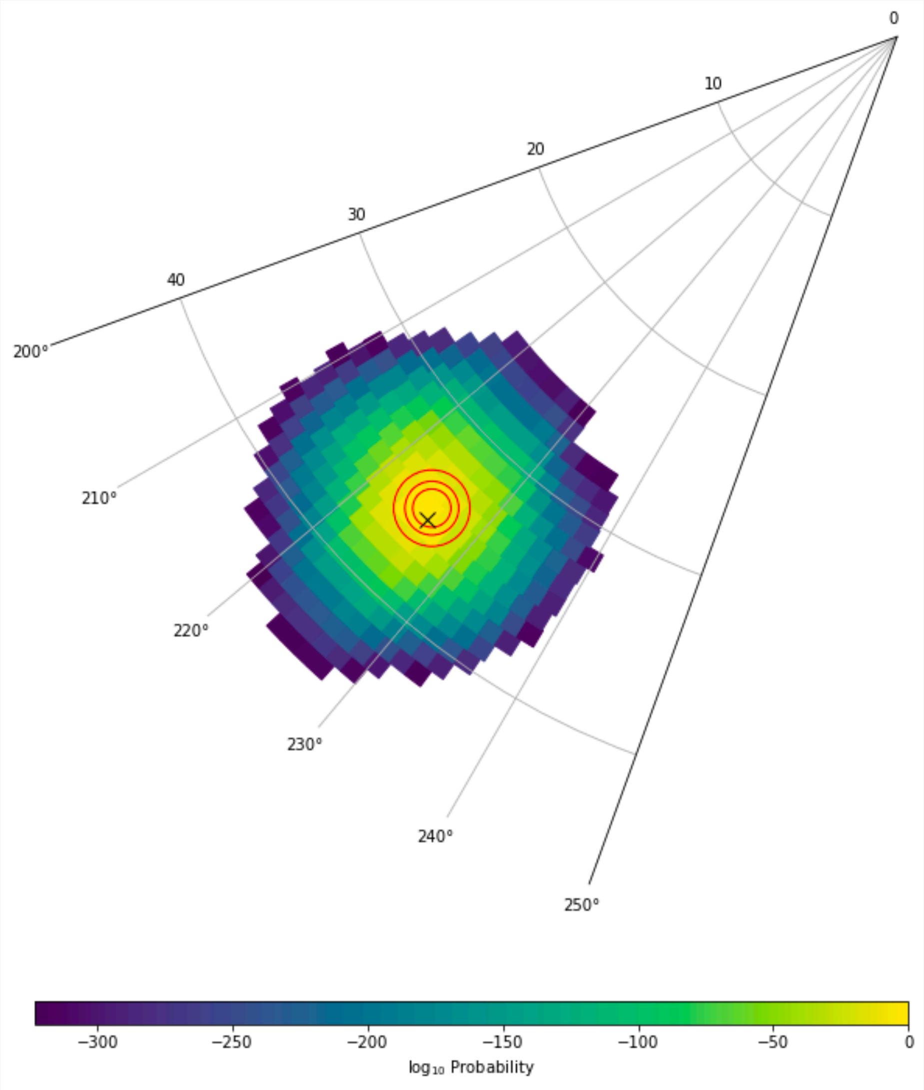


Tutorial

Known Unknowns Bayesian Networks

Per Event Uncertainty on regressed variable!!





[Tutorial]

Paper in preparation, but feel free to reach out

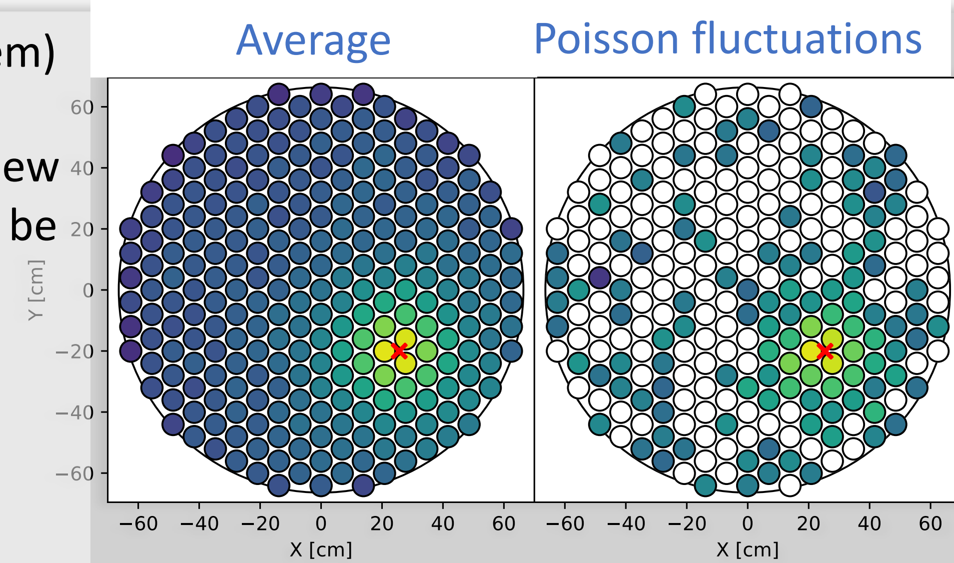
Unknown Unknowns Physics-constrained Graph NN

	Known	Unknown
Knowns	Known Knowns	Unknown Knows Graph denoiser
Unknowns	Known Unknowns Bayesian Networks	Unknown Unknowns Physics-constrained Graph Neural Net

Unknown Unknowns Physics-constrained Graph NN

What makes "reconstruction" (i.e. inverse problem) difficult?

- Every new experiment measures something new
    - Experimental physics has been shown to be 48% voodoo
    - Likelihood analysis pointless at raw-data level
- Petabytes of data, efficiency matters
- Out-of-the-box ML often fails
    - Often regression, not classifier
    - Data graphical, not square image
    - Poisson counting noise, not normal
    - No ground truth on data for supervised learning
        - MC simulations mostly work
    - ML learns model from data, scientists apply model to data
        - "We have initial theory of cat photos"
    - Interpretability / Physical constraints (energy positive)

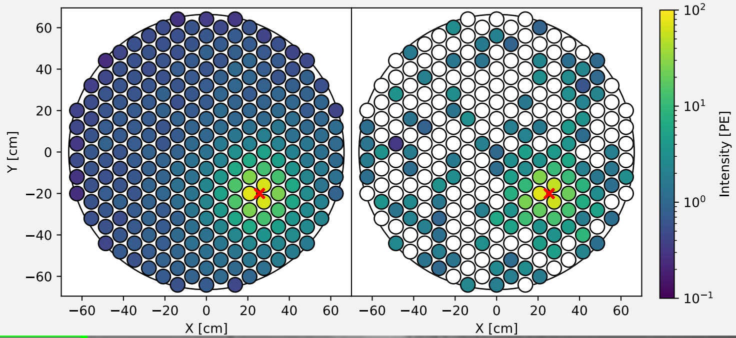


$$\mathcal{R}^{500} \rightarrow \mathcal{R}^2$$

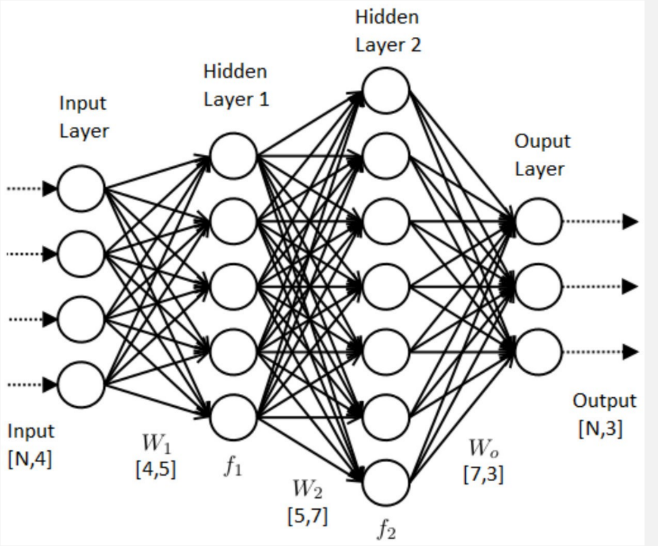
Unknown Unknowns Physics-constrained Graph NN

Consider localization problem

$$\mathcal{R}^{500} \to \mathcal{R}^2$$



Current state-of-art is MLP with elu and two hidden layers, but can't go deep learning

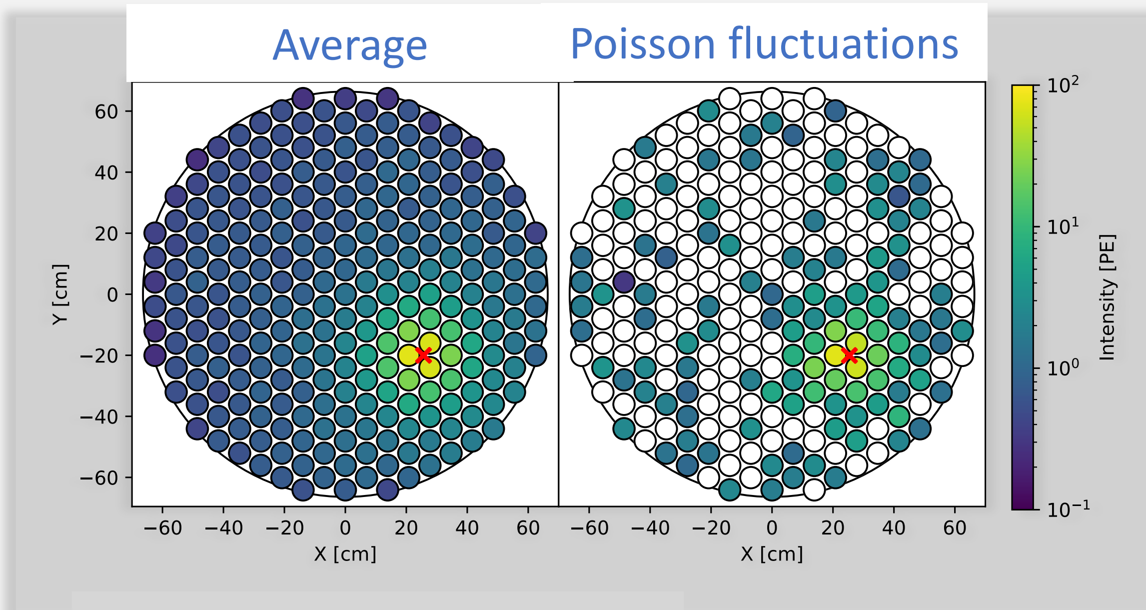


CNNs go deep by introducing locality via learned kernel to reduce complexity of layers. Above is edge detector.

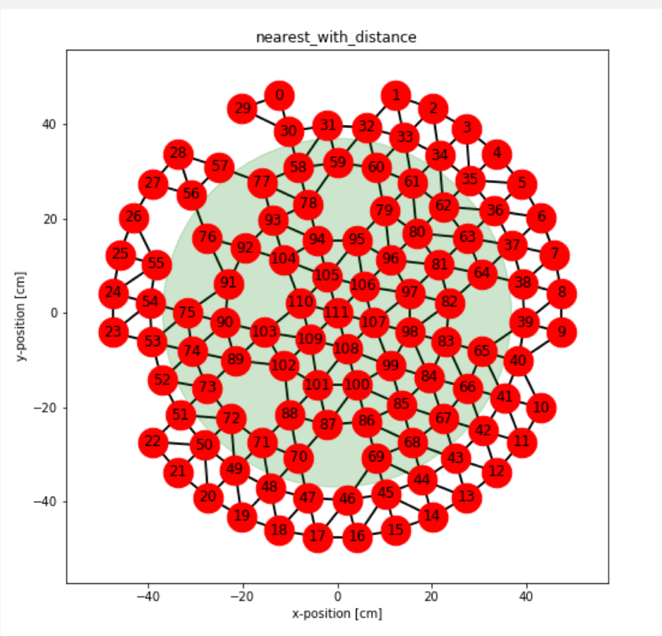
Unknown Unknowns Physics-constrained Graph NN

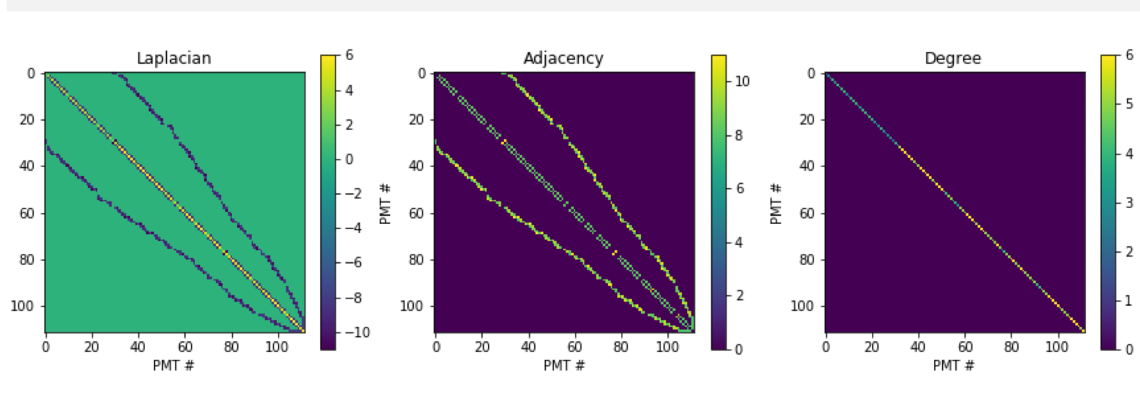
Naively:

- Try convolutional graph NN
  - Kernel hard, ~~hex~~
- but ‘convolutional’ approaches require translational invariance



$$\mathcal{R}^{500} \rightarrow \mathcal{R}^{2}$$

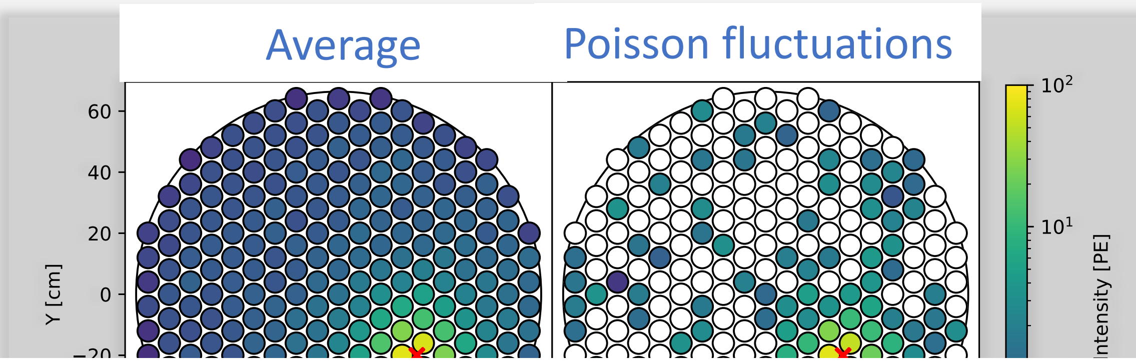


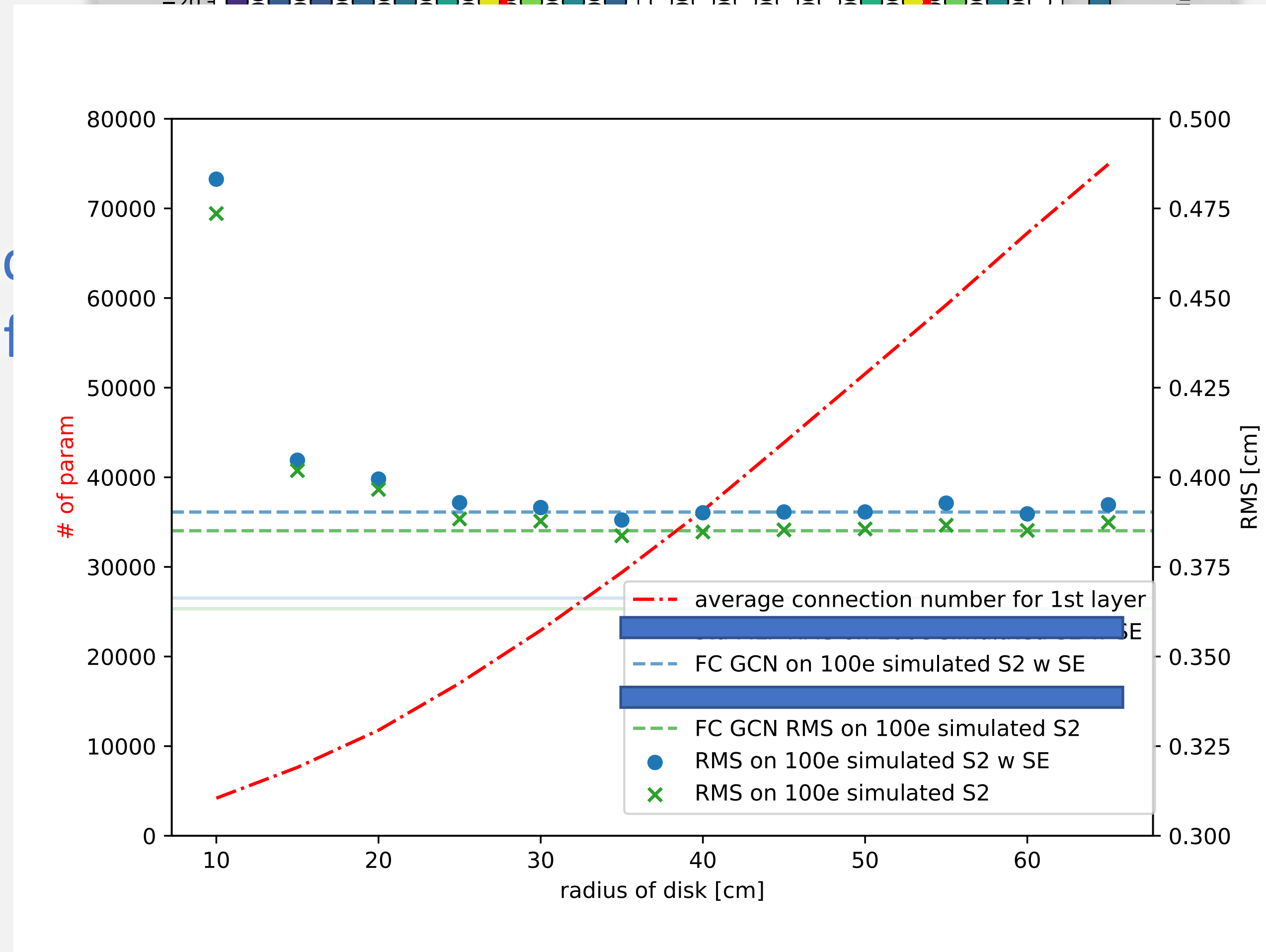


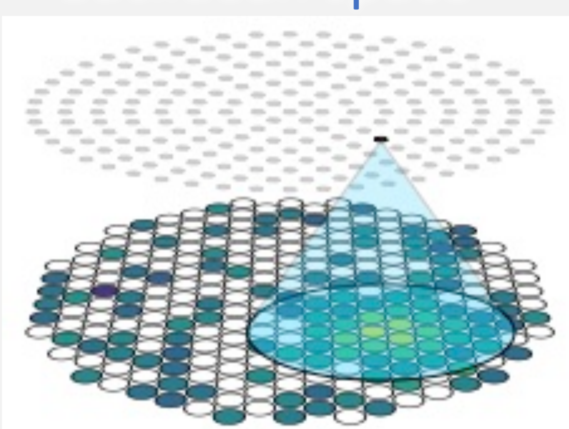
Unknown Unknowns Physics-constrained Graph NN

Naively:

- Try convolutional graph NN
  - Kernel hard, ~~hex~~
- but ‘convolutional’ approaches require translational invarian
- Treat each ‘node’ as i embedded in Euclidean space



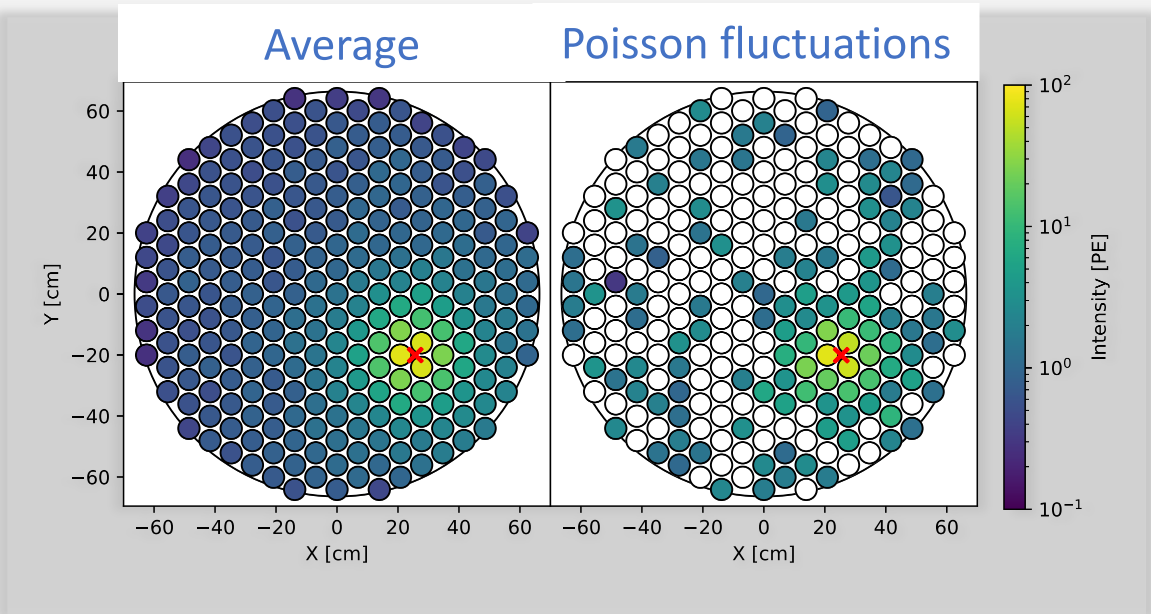


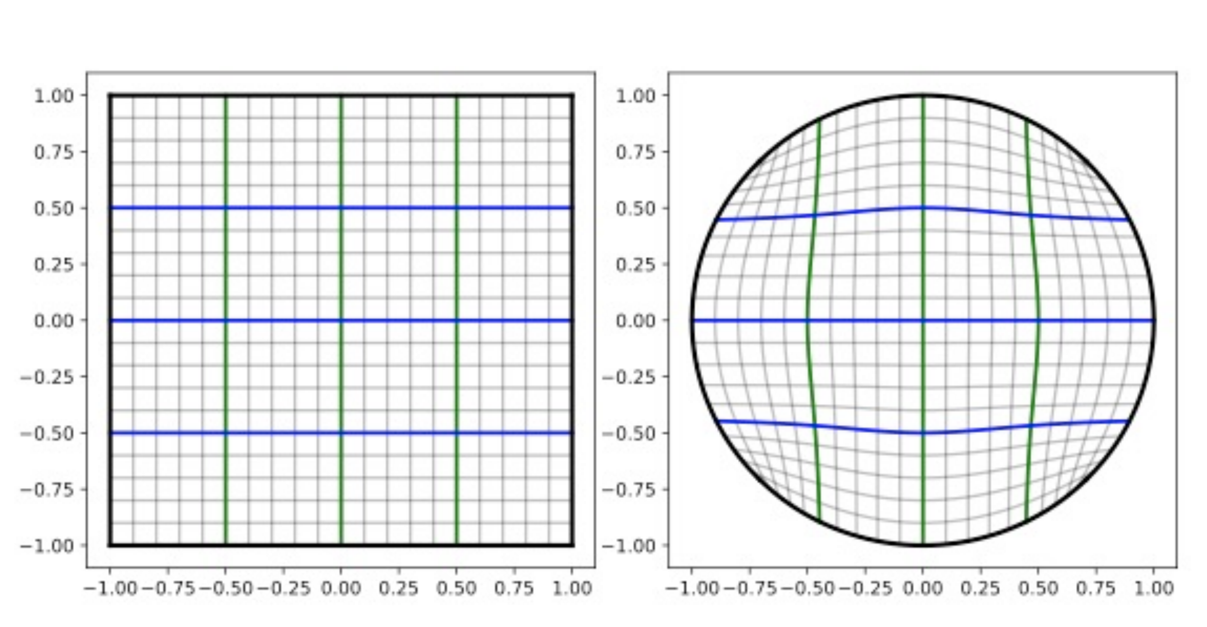


Unknown Unknowns Physics-constrained Graph NN

Physical architectural constraints

- Events cannot be outside of the detector
  - NN doesn't know
  - Corrections ill-defined
- Deeper learning allows more 'nonlinearity'
- Constrain output by tanh then transform





More publications (read later)

- Safety of Quark/Gluon Jet Classification, A. Romero, D. Whiteson, M. Fenton, J. Collado and P. Baldi, arXiv 2103.09103 (16 Mar 2021) [1 citation].
- Efficient sampling of constrained high-dimensional theoretical spaces with machine learning, J. Hollingsworth, M. Ratz, P. Tanedo and D. Whiteson, arXiv 2103.06957 (11 Mar 2021) [5 citations].
- Progress in developing a hybrid deep learning algorithm for identifying and locating primary vertices, S. Akar, G. Atluri, T. Boettcher, M. Peters, H. Schreiner et. al., arXiv 2103.04962 (08 Mar 2021).
- Learning to Isolate Muons, J. Collado, K. Bauer, E. Witkowski, T. Faucett, D. Whiteson et. al., arXiv 2102.02278 (03 Feb 2021) [1 citation].
- Foundations of a Fast, Data-Driven, Machine-Learned Simulator, J. Howard, S. Mandt, D. Whiteson and Y. Yang, arXiv 2101.08944 (21 Jan 2021) [4 citations].
- Accelerated Charged Particle Tracking with Graph Neural Networks on FPGAs, A. Heintz, V. Razavimaleki, J. Duarte, G. DeZoort, I. Ojalvo et. al., arXiv 2012.01563 (30 Nov 2020) [9 citations].
- Learning to Identify Electrons, J. Collado, J. Howard, T. Faucett, T. Tong, P. Baldi et. al., arXiv 2011.01984 (03 Nov 2020) [1 citation].
- Mapping Machine-Learned Physics into a Human-Readable Space, T. Faucett, J. Thaler and D. Whiteson, Phys.Rev.D 103 036020 (2021) (22 Oct 2020) [12 citations].
- Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks, M. Fenton, A. Shmakov, T. Ho, S. Hsu, D. Whiteson et. al., arXiv 2010.09206 (19 Oct 2020) [6 citations].
- FPGAs-as-a-Service Toolkit (FaaST), D. Rankin, J. Krupa, P. Harris, M. Acosta Flechas, B. Holzman et. al., arXiv 2010.08556 (16 Oct 2020) [4 citations].
- Sparse autoregressive models for scalable generation of sparse images in particle physics, Y. Lu, J. Collado, D. Whiteson and P. Baldi, Phys.Rev.D 103 036012 (2021) (23 Sep 2020) [4 citations].
- Bayesian Neural Networks for Fast SUSY Predictions, B. Kronheim, M. Kuchera, H. Prosper and A. Karbo, Phys.Lett.B 813 136041 (2021) (13 Jul 2020) [1 citation].
- Accelerating Deep Neural Networks for Real-time Data Selection for High-resolution Imaging Particle Detectors, Y. Jwa, G. Guglielmo, L. Carloni and G. Karagiorgi, Unknown (01 Jun 2019) [1 citation].