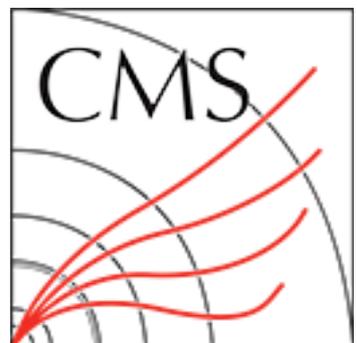# Exploring end-to-end deep learning solutions for event classification at CMS

Michael Andrews[1,2], Manfred Paulini[1,2], Sergei Gleyzer[1,3], Barnabas Poczos[4]
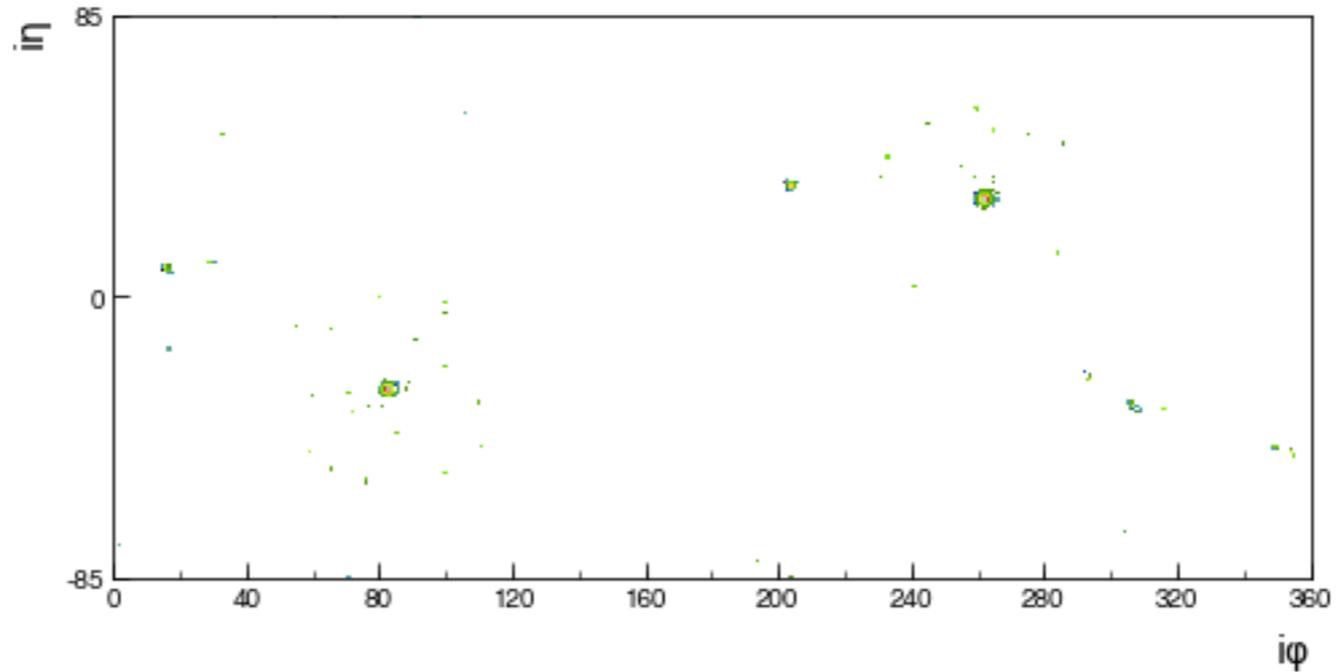on behalf of the CMS experiment

[1]CMS, [2]Carnegie Mellon University-Physics, [3]University of Florida, [4]Carnegie Mellon University-ML

ACAT 2017, 2017-AUG-22

Motivation
Typical
Event
Classification

**Sensor
Data**

Physics Reconstruction—lossy?

	Energy	p_T	R9	$\sigma_{i\eta i\eta}$	...
1	31.6	20.3	0.85	0.79	
2	45.2	37.1	0.92	0.83	
..					

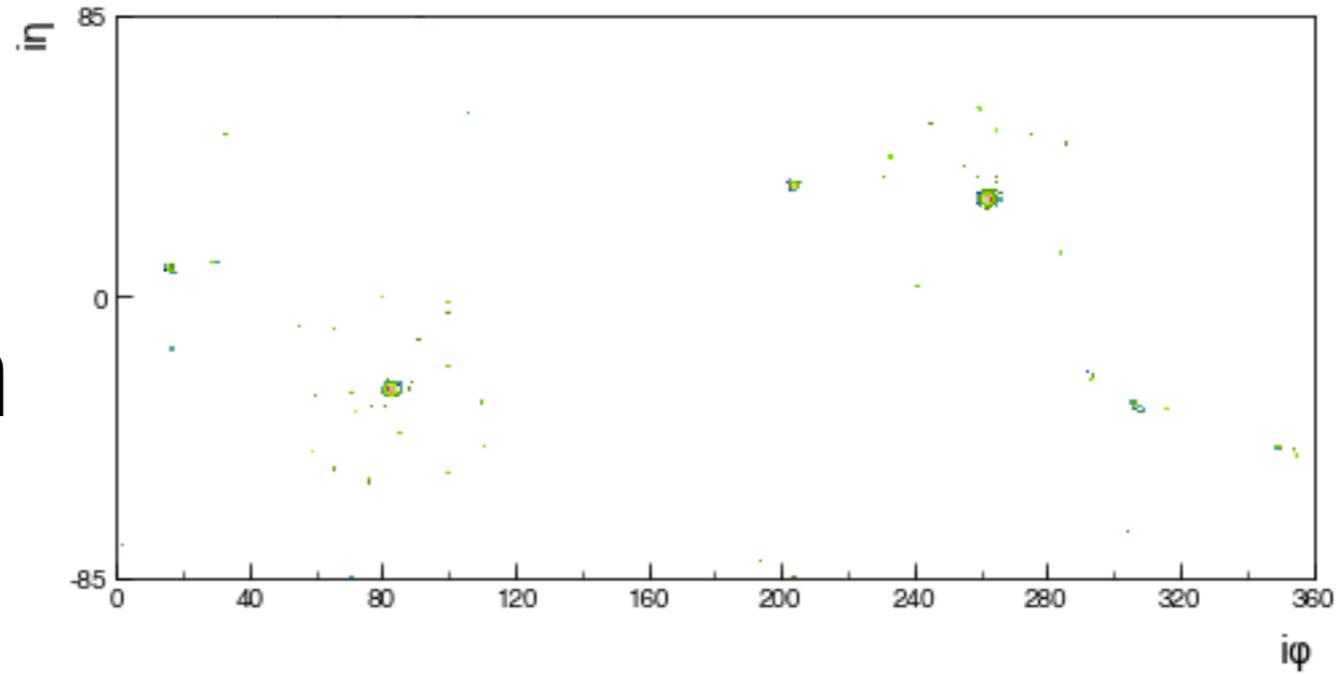**Kinematic
Data**

Classification

Class X
(e.g. Signal or Bkgnd)
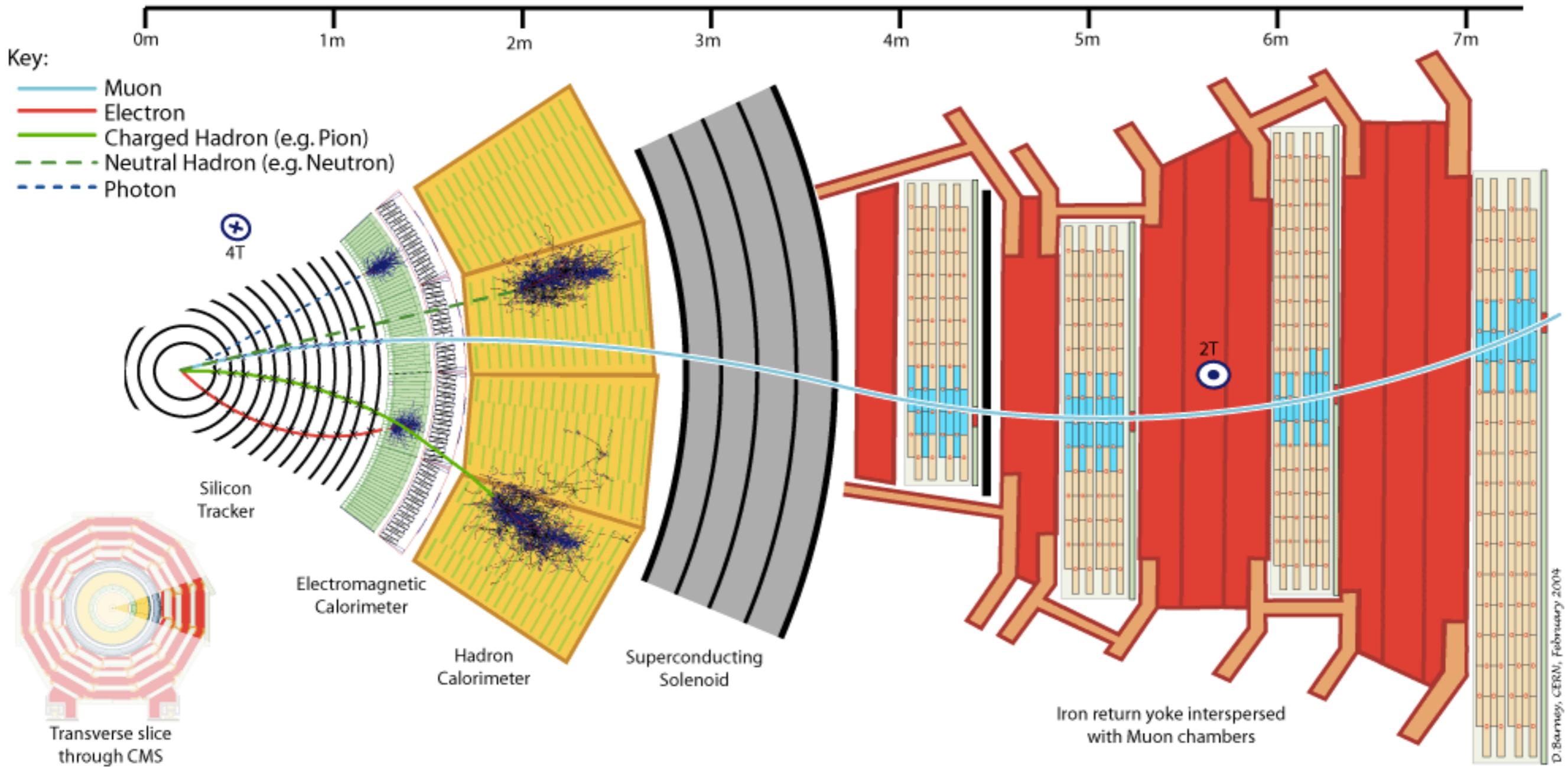
**Class
Output**

2

End-to-end
Event Classification

Sensor Data

End-to-end Classification

Class X
(e.g. Signal or Bkgnd)

Class Output

CMS

Key:
- Muon
- Electron
- Charged Hadron (e.g. Pion)
- Neutral Hadron (e.g. Neutron)
- Photon

4T

2T

Silicon Tracker

Electromagnetic Calorimeter

Hadron Calorimeter

Superconducting Solenoid

Iron return yoke interspersed with Muon chambers

Transverse slice through CMS

D. Barney, CERN, February 2004

4
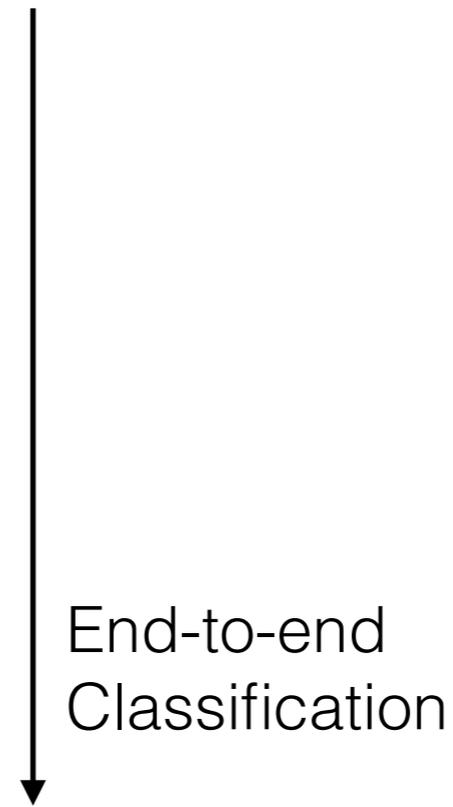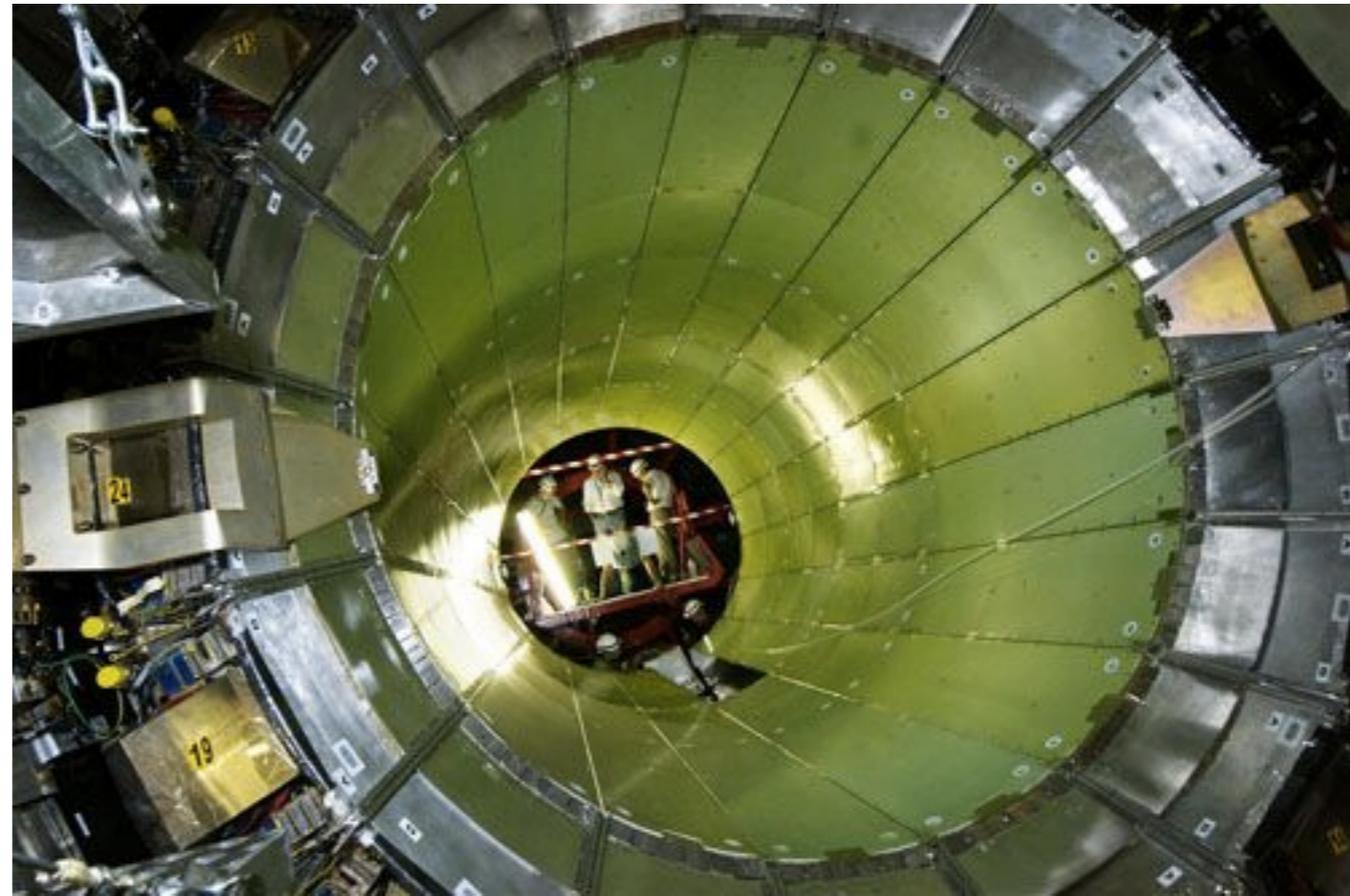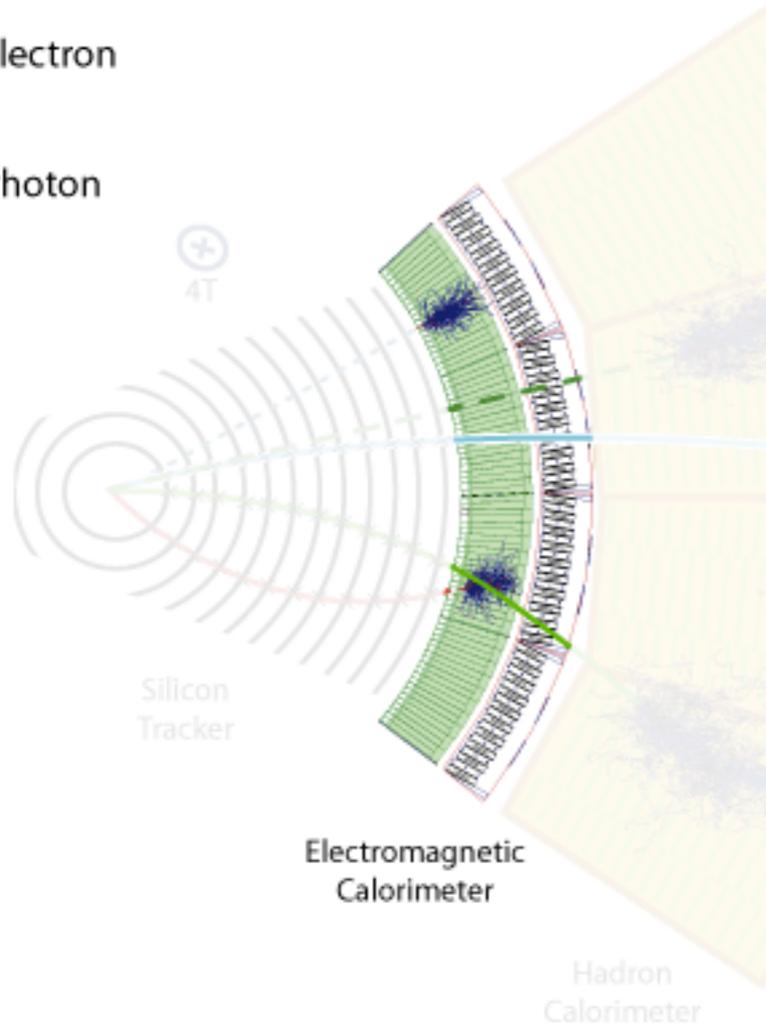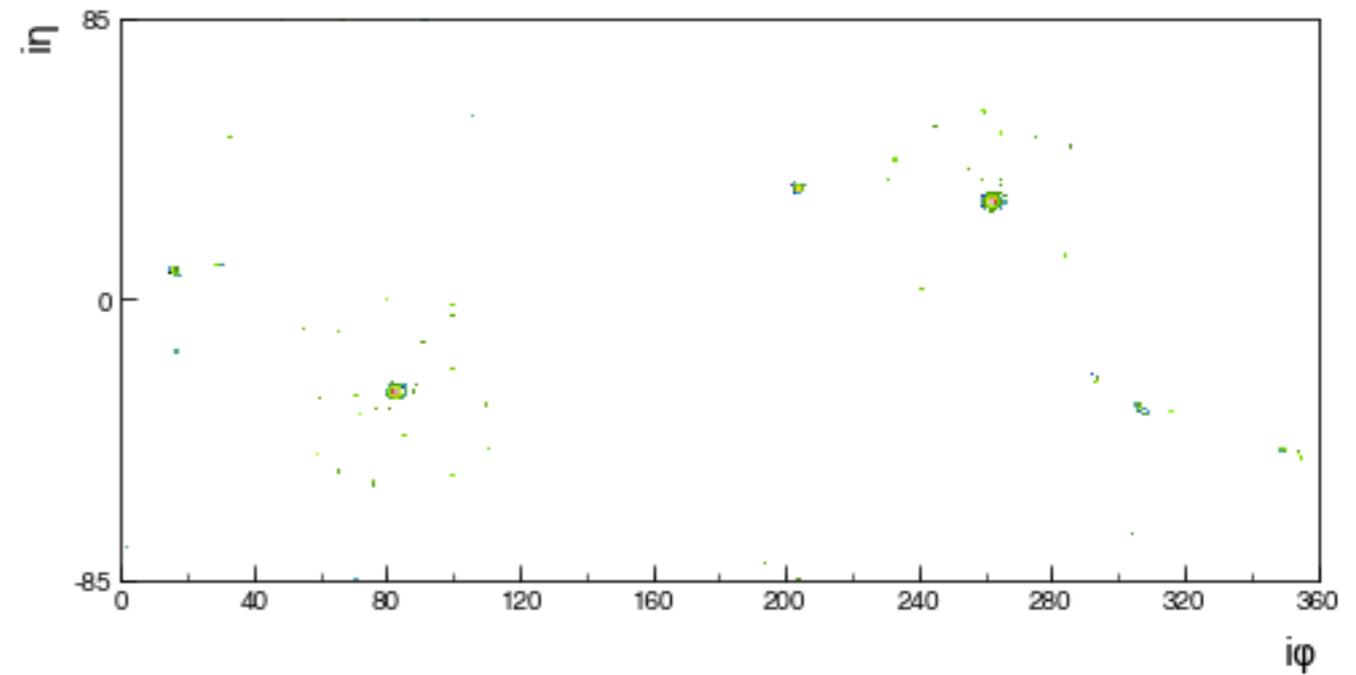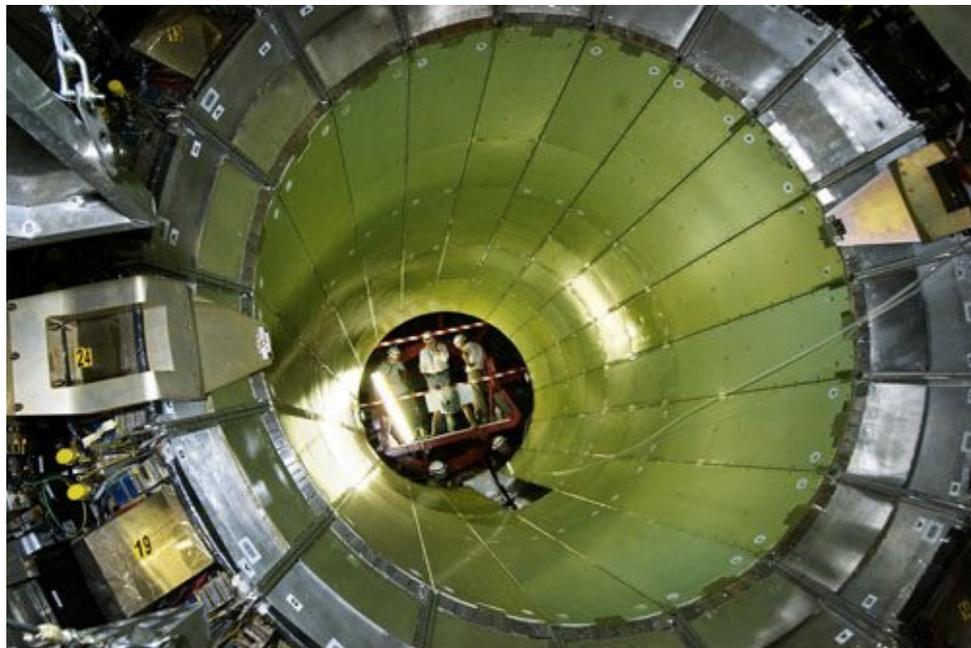
ECAL

Key:
—— Electron
- - - - Photon

Silicon
Tracker

Electromagnetic
Calorimeter

Hadron
Calorimeter

ECAL Longitudinal Section

Barrel ECAL (EB)

$\eta = 1.479$

$\eta = 1.653$

$\eta = 2.6$

$\eta = 3.0$

Preshower (ES)

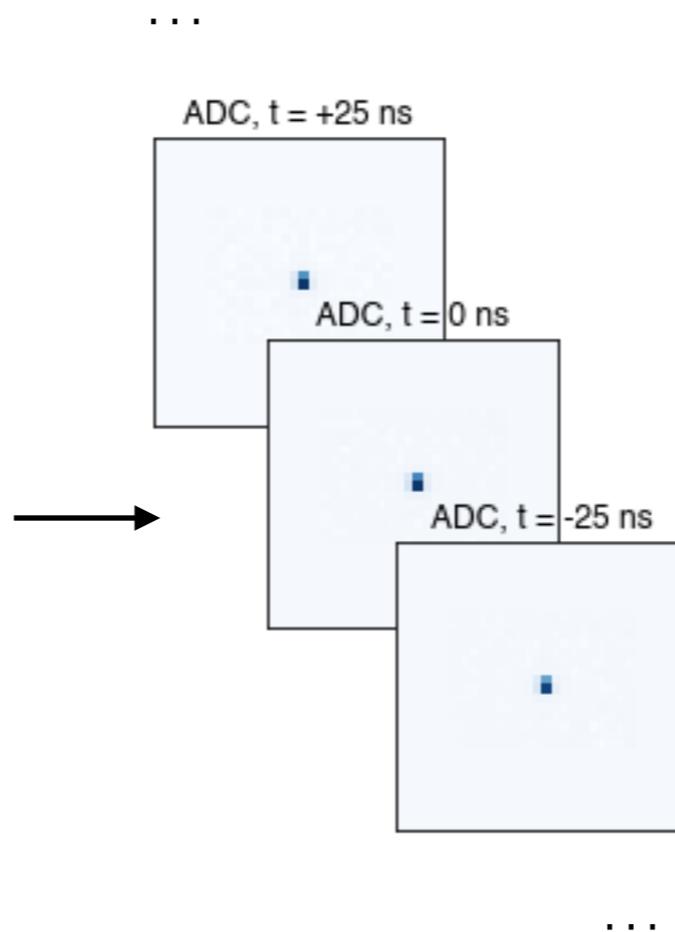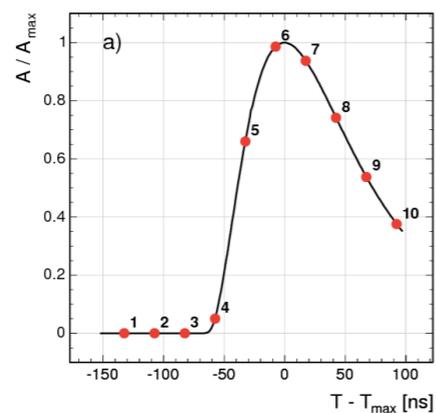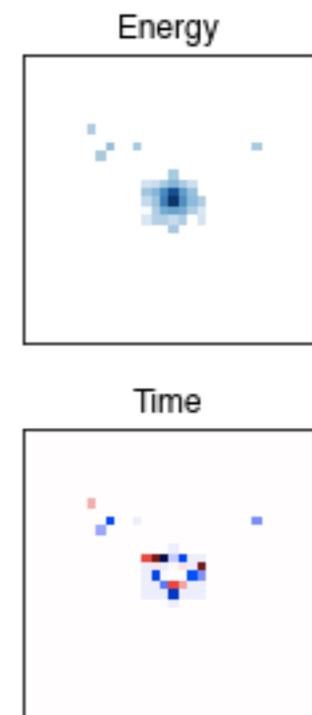Endcap ECAL (EE)

y

z
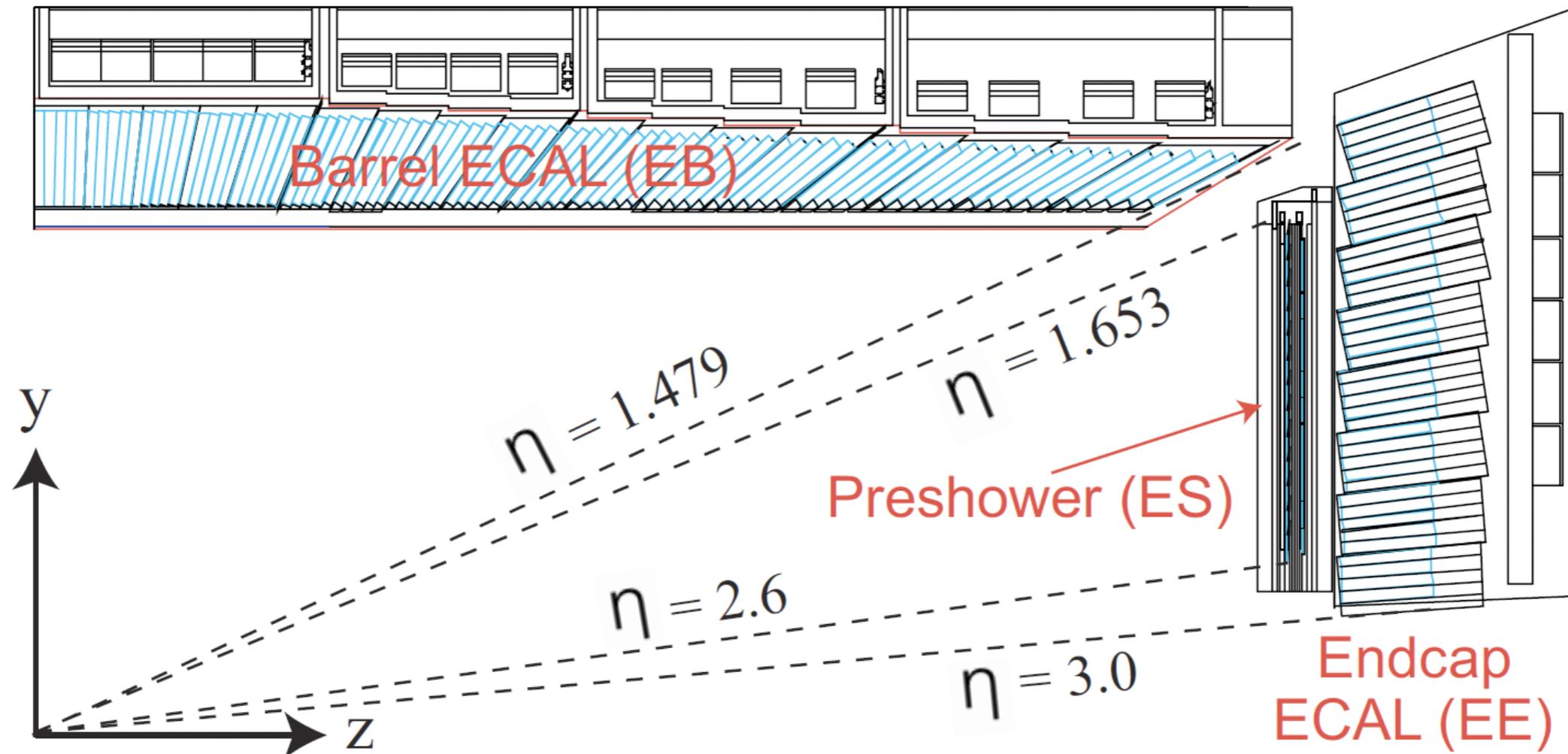
ECAL Graphical Representation

ECAL Hit Reconstruction

Scale: 1 pixel = 1 crystal

Scintillating Crystal

Signal Pulse

Digitized Hit ("DIGI")

**Reconstructed Hit** ("Rec Hit or RH")

Related Work

‣ Various Neutrino experiments have been successfully using Convolutional Neural Nets (CNNs) for some time now *(see DS@HEP 2017, IML Workshop 2017)*

‣ Several applications of jet "images" involving projecting 4-momenta onto image grids
(see DS@HEP 2017, IML Workshop 2017)

‣ Applications involving 4-momenta a la NLP
(see Kyle Cranmer et al., Jean-Roch Vlimant et al.)

‣ Applications involving EGSuperCluster Rechits for photons vs fakes (*see Andre Holzner et al.)*

Object Classification

Object Classification: e^- vs γ

‣ **Simulated SingleEG particle gun:** *e^- vs γ*

‣ $p_{T,gen}$ = 50 GeV, Flat $|\eta_{gen}|$ < 1.4

‣ No pileup or underlying event

‣ *Reminder:* ECAL not designed to distinguish between the two!

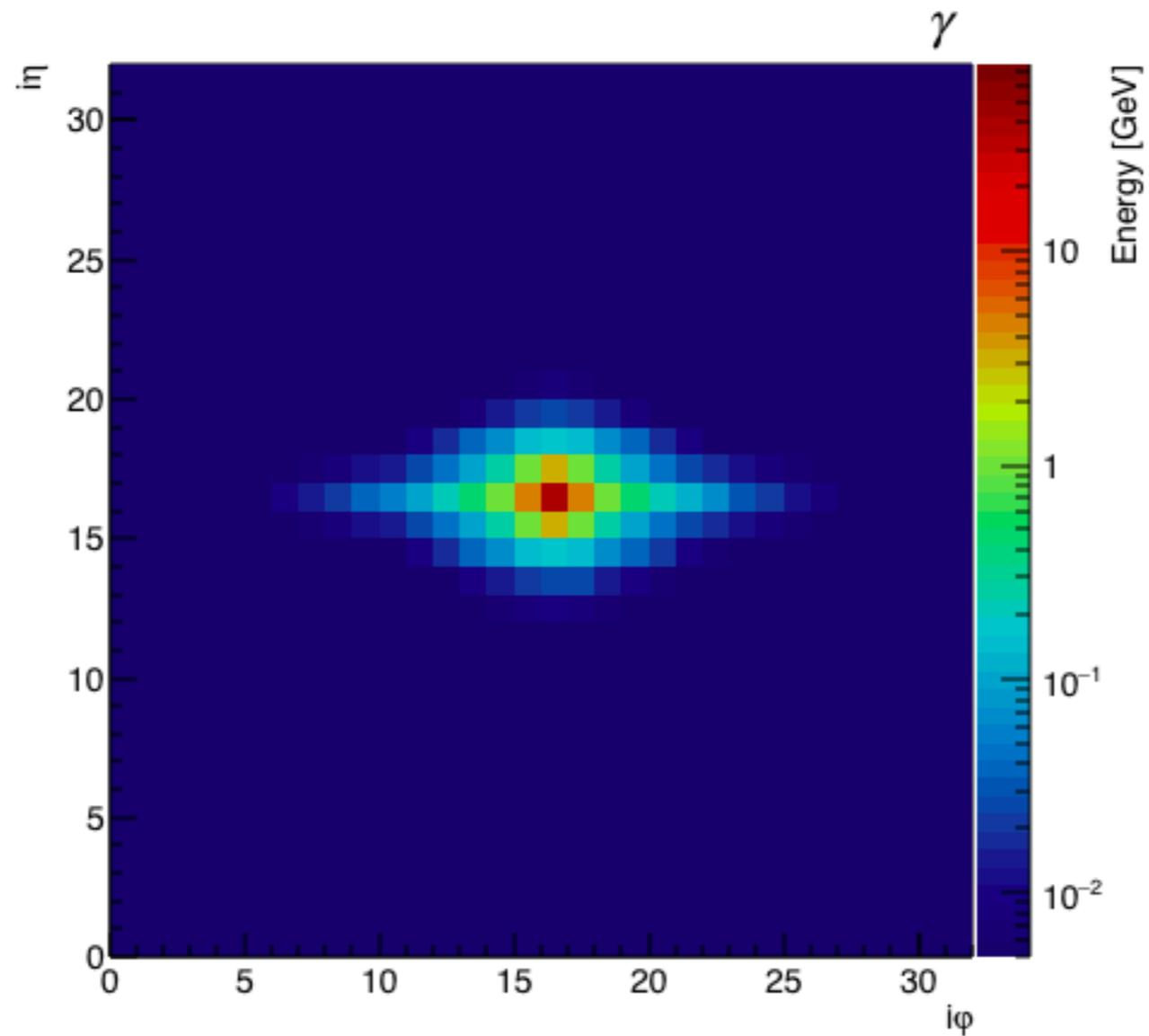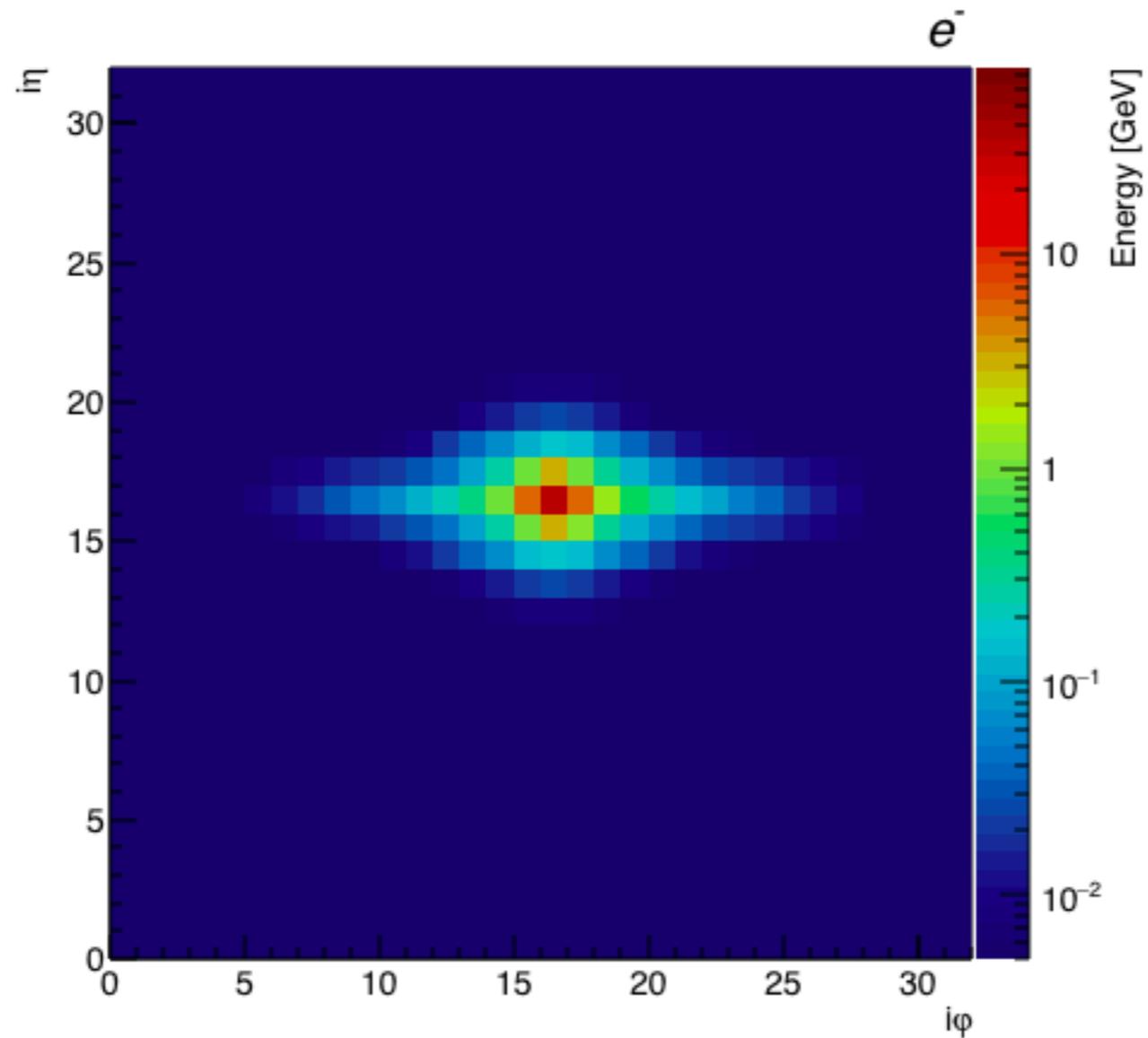‣ Take 32 x 32 grid of crystals centered on the shower seed

Shower Crops

Crop around shower seed
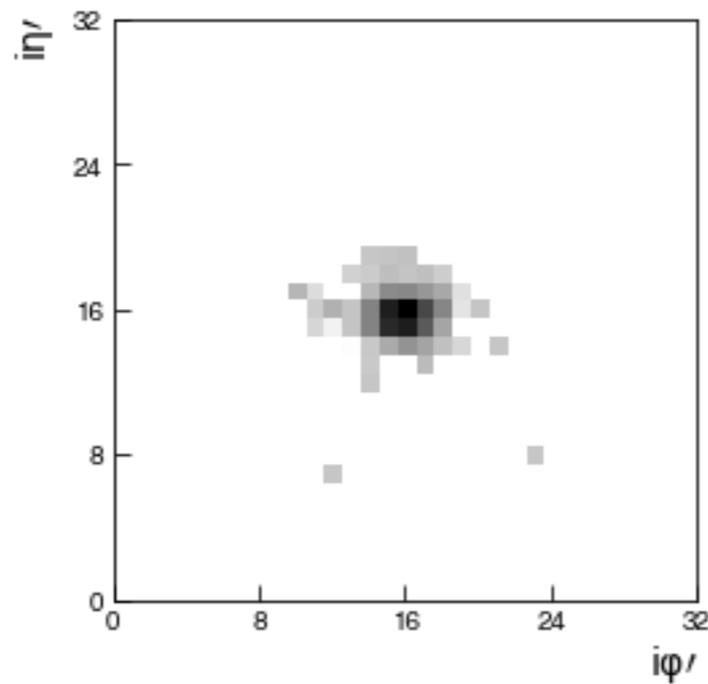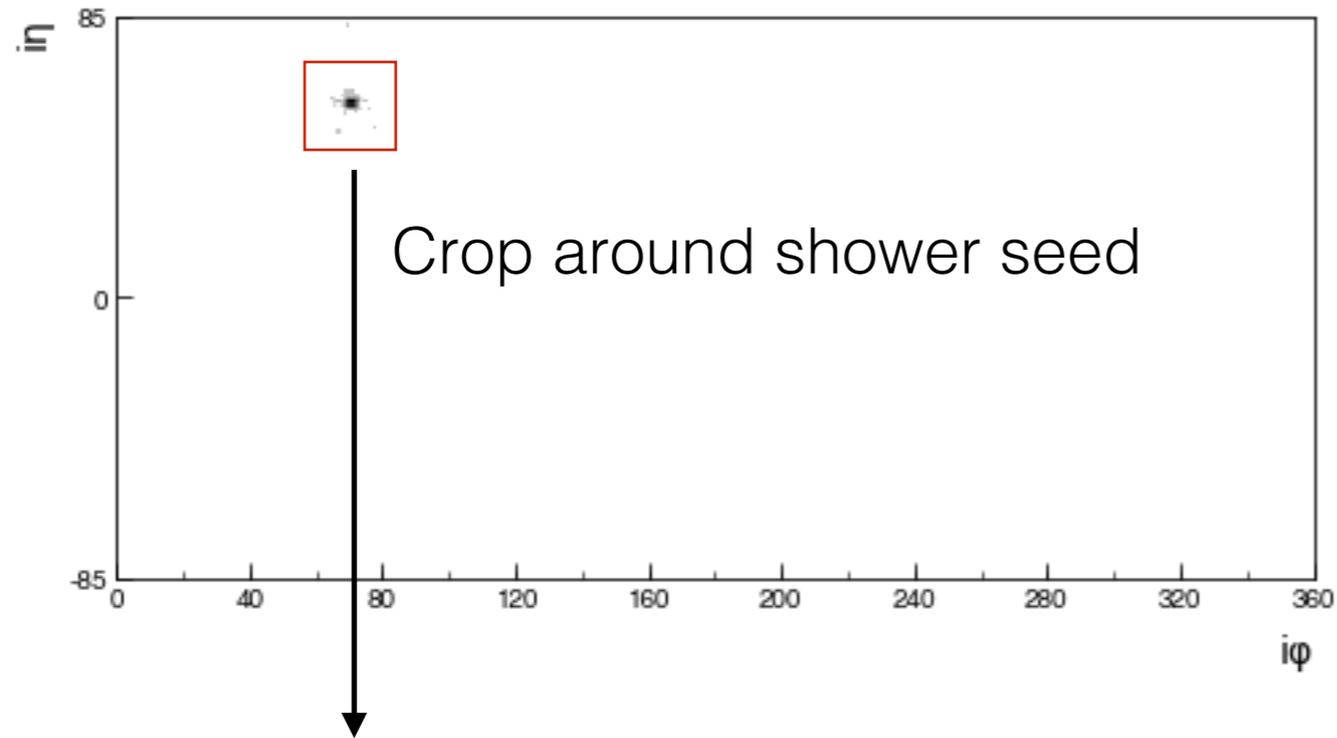
Classify

Photon or Electron
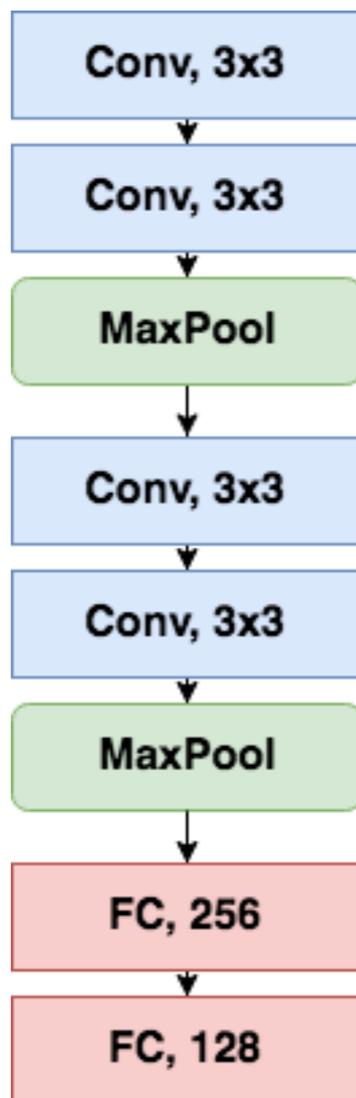
Object Classification: e⁻ vs γ

Average photon shower

Object Classification: e⁻ vs γ

Average electron shower
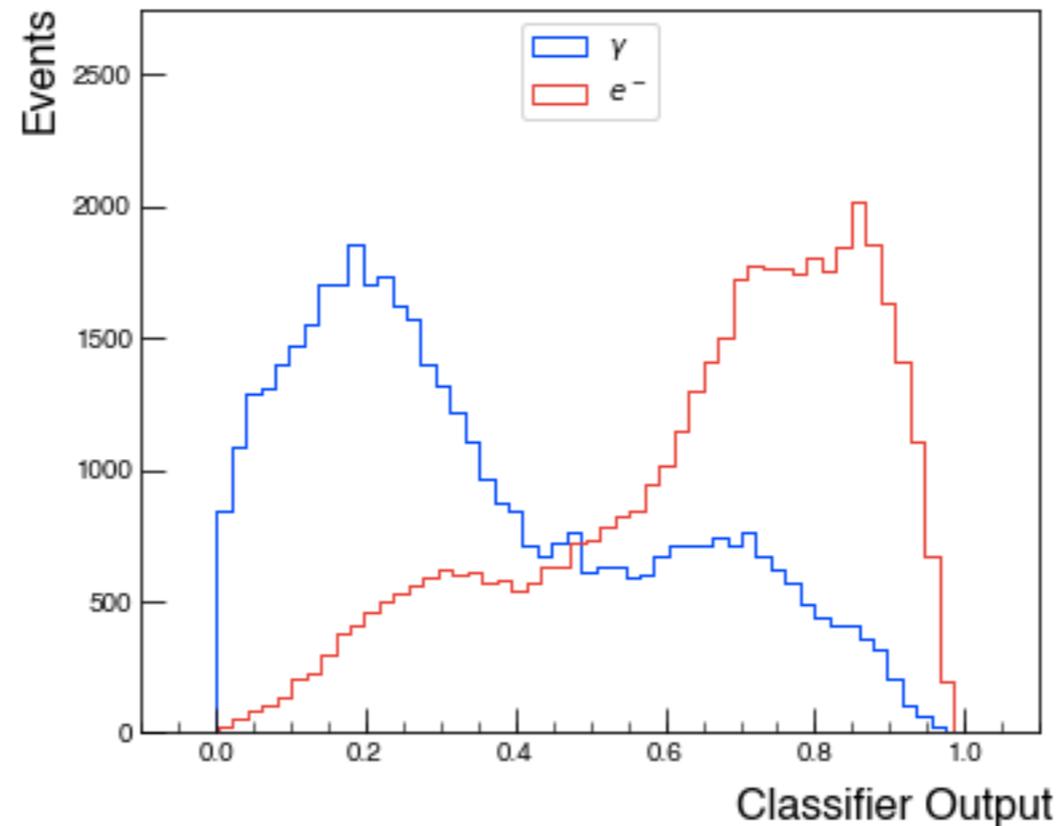
Object Classification: e⁻ vs γ

VGG

- ‣ **Baseline CNN:** VGG
- ‣ **Activation:** ReLU
- ‣ **Loss Function:** Cross-entropy
- ‣ **Learning Rate (LR):** 1.e-3
- ‣ **LR decay:** 0.5, patience=2
- ‣ **Total samples:** 498k
- ‣ **Train:Validation:Test split**
  320k **:** 89.6k **:** 88.4k
- ‣ **Figure of Merit:** Area under the curve of Receiver Operating Characteristic (ROC AUC)

Hyperparameter tuning

‣ **Preprocessing:** linear, log, pedestal subtraction

‣ **Architecture:**

    ‣ **Convolutional NN (CNN):** VGG, Inception, ResNet

    ‣ **Fully-Connected NN (FC):** 2-, 3-, 6-hidden layers, 256 nodes each

    ‣ **LSTM** (for DIGI): TimeDist(CNN)→LSTM, LSTM(CNN)

‣ **Inputs:** energy, (energy, time), (DIGI), (energy, time, DIGI)*

‣ ***Concatenation:** @input, @convolutional output, @FC output

Object Classification: e⁻ vs γ

| Network Type Inputs	Fully-Connected Flattened DIGI	CNN energy	CNN+LSTM DIGI
Test ROC AUC	0.770	0.807	0.799

Event Classification

Event Classification

‣ **Scalability?** Does single object classification scale to event classification?

‣ **Shower correlations?** Can we exploit correlations between different objects?

Residual Net CNNs

‣ **VGG does not scale well with image size**
  ‣ Deep VGGs suffer from *degradation* problem
  ‣ Fully-connected layers ⇒ too many weights, use smaller LR

‣ **Residual nets are highly scalable** *[arXiv:1512.03385v1]*
  ‣ Overcome *degradation* problem by using *identity shortcuts*
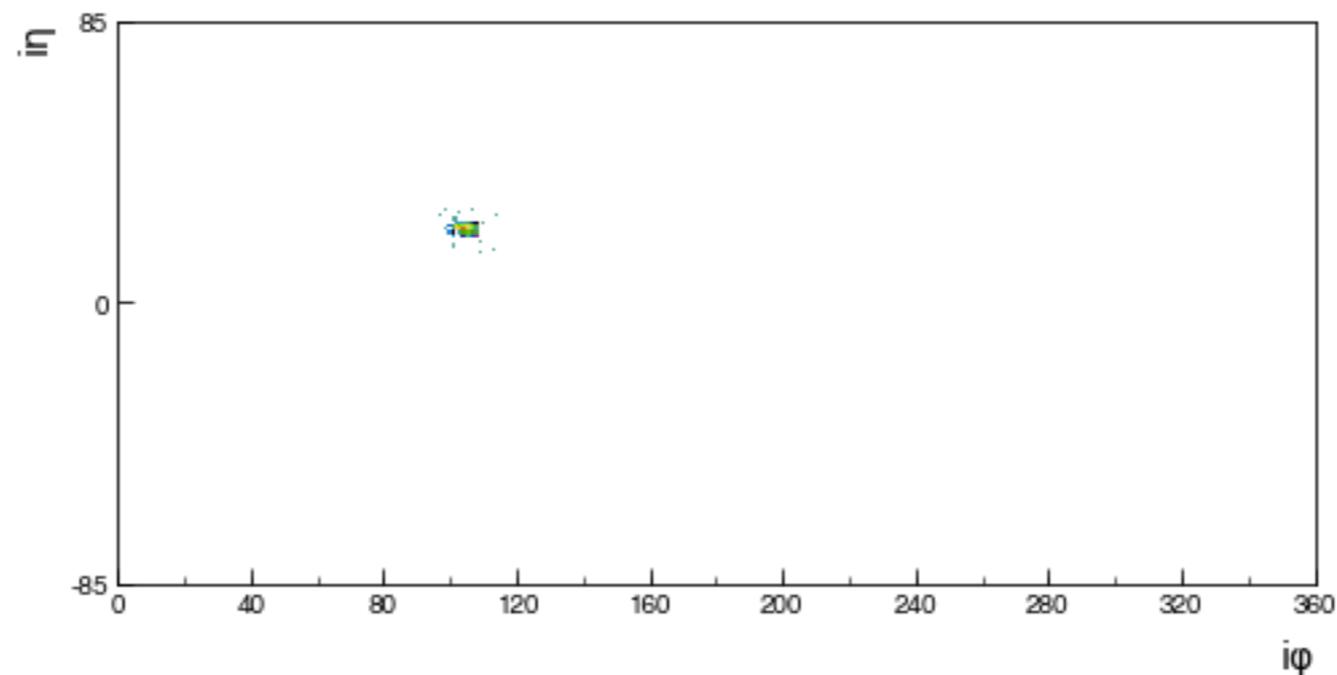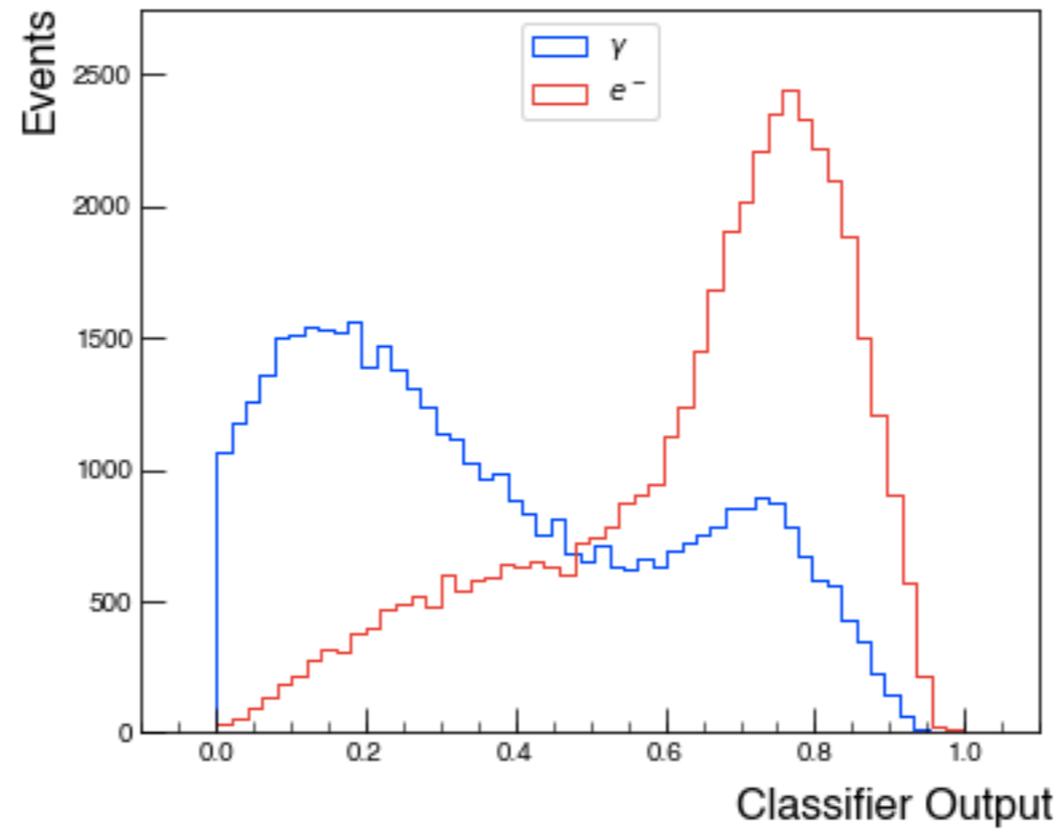 ‣ GlobalMaxPooling scales with number of feature maps

Residual Block

Conv, 3x3

ReLU

Conv, 3x3

ReLU
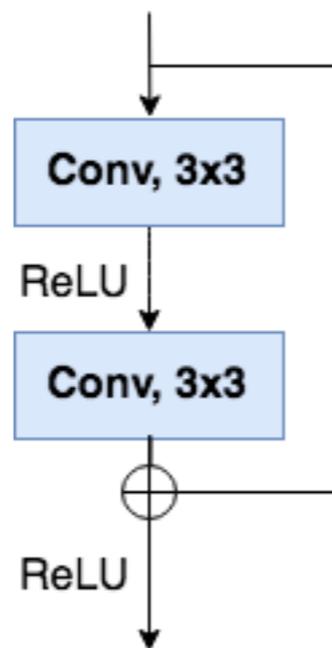
…+ GlobalMaxPooling

Scalability?

‣ **Simulated SingleEG particle gun:** e^- vs γ

‣ $p_{T,gen}$ = 50 GeV, Flat $|\eta_{gen}|$ < 1.4
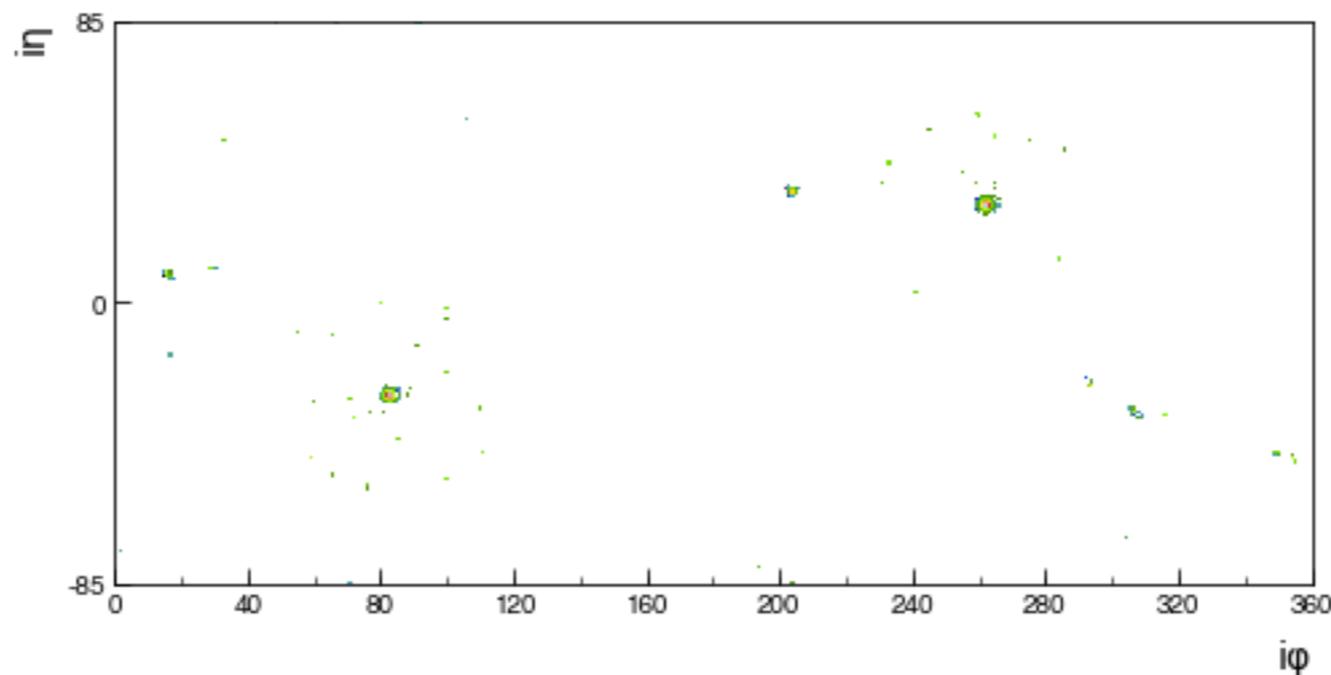
‣ No pileup or underlying event

e⁻ vs γ

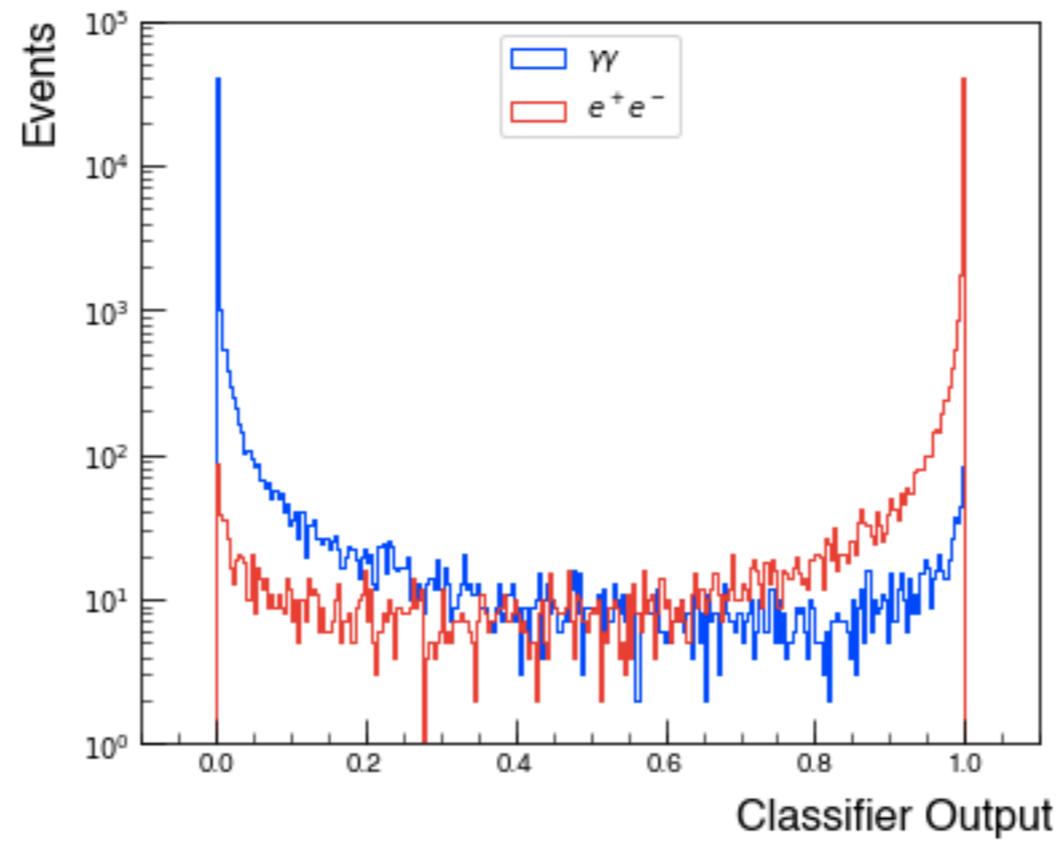|	ResNet-15
Test Set ROC AUC	0.788

Shower Correlations?

- **Simulated DoubleEG particle gun:** e^+e^- vs $\gamma\gamma$
- Flat $p_{T,gen}$ = [20,80] GeV, Flat $|\eta_{gen}|$ < 1.4
- Particles decay back-to-back
- With Poisson out-of-time pile-up: $\langle PU \rangle$ = 25, no underlying event
- Filter on interesting hits

e⁺e⁻ vs γγ

	ResNet-23
Test Set ROC AUC	0.997

Work In Progress

‣ **Realistic physics**

   ‣ How about for realistic *pp* physics processes?

   ‣ Can exploit *spatial* correlations?

   ‣ Transfer learning to different p_T regimes

   ‣ ECAL Barrel+Endcaps

   ‣ Data validation?

Conclusions

‣ Reconstructed energy hits seem to be most reasonable low-level data collection to use

‣ Convolutional neural nets on low-level detector data show good performance in extracting higher-order features in ECAL particle showers in simulated data

‣ End-to-end event classifiers are able to effectively exploit correlations between different objects in the event—beyond single-object performance—in simulated ECAL data

Plans

‣ **Multiple Sub-detectors?**

 ‣ Possible to seamlessly combine information across different images / sub-detectors?

 ‣ ECAL+HCAL? Full CMS detector?

BACKUP

Systematics

‣ **Due to finite convergence of NN**
Train to convergence by decaying learning rate to small epsilon then observe stability of metrics from last N epochs
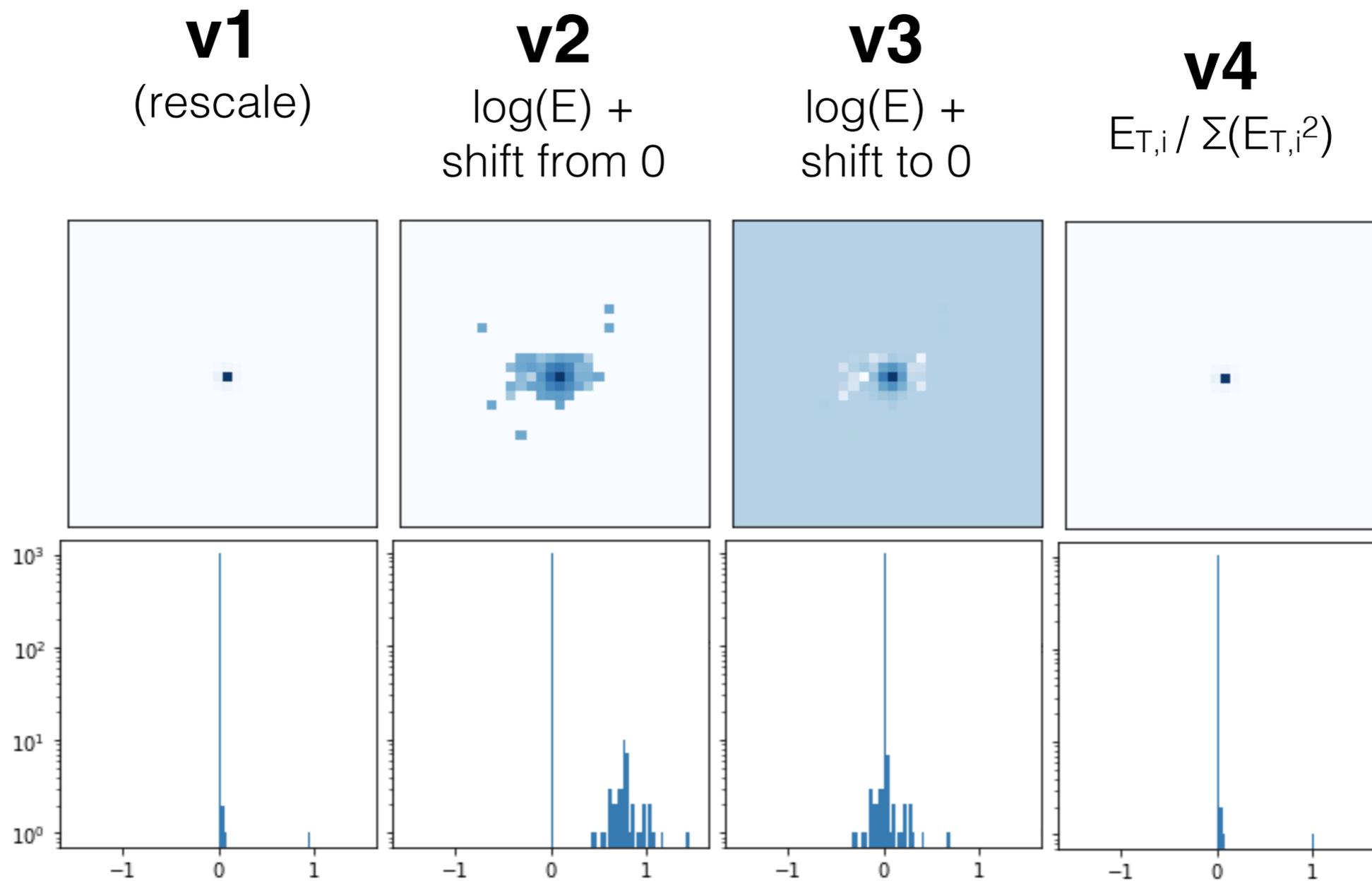 ‣ Validation and Test ROC AUC: ± 5.e-5

‣ **Due to non-deterministic GPU optimization**
Nature of parallelization means even if random seed is fixed, slightly different results appear each time you train
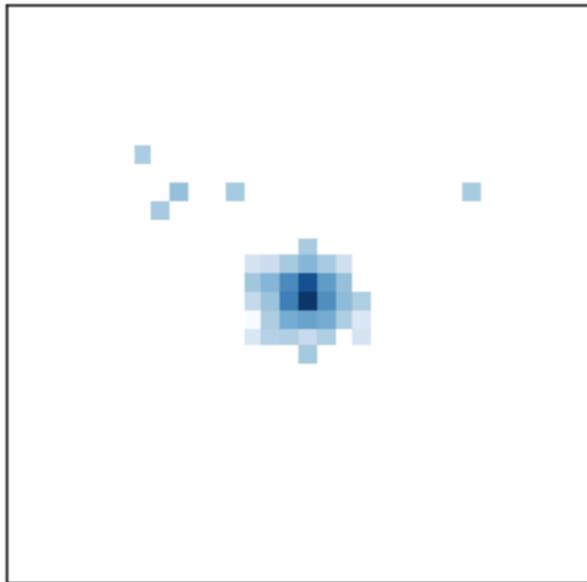 ‣ Validation and Test ROC AUC: ± 2.e-3

‣ **Due to representativeness in test set**

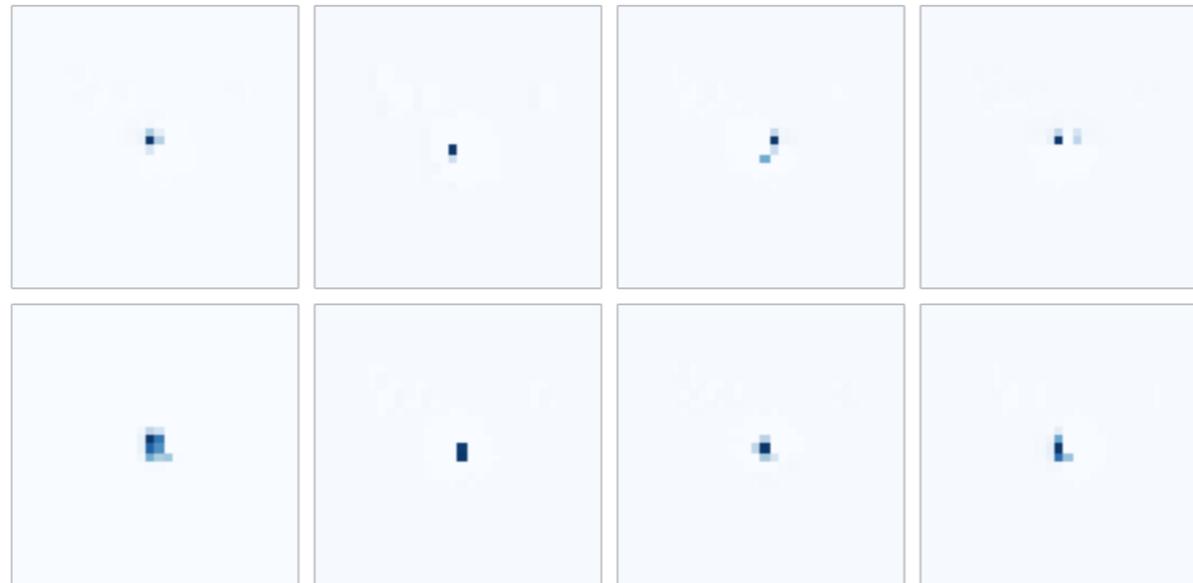# **Object Classification:** e⁻ vs γ
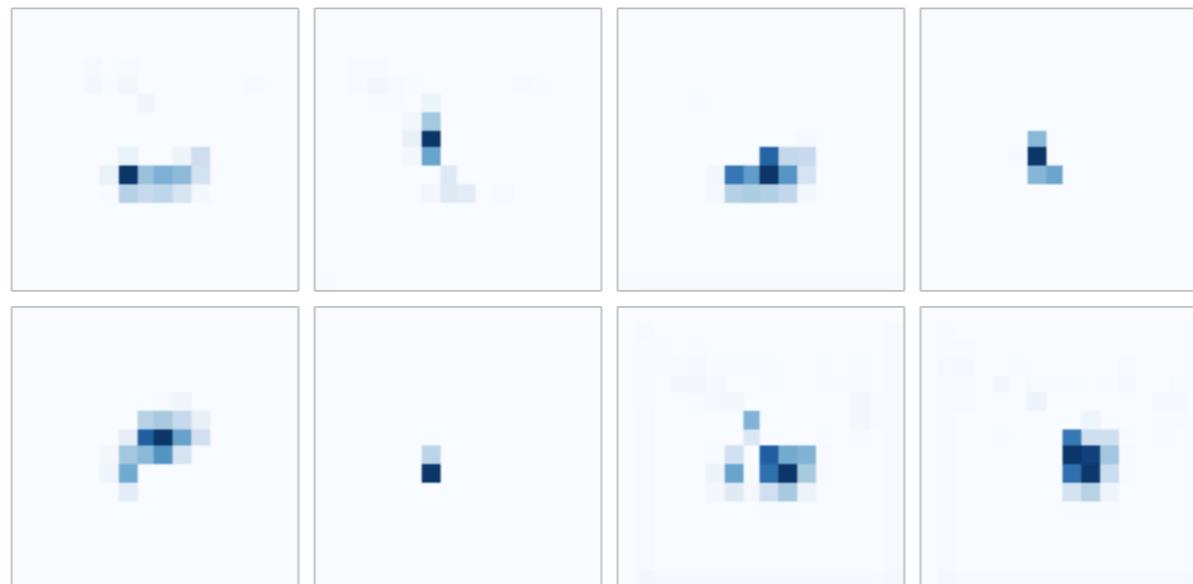
Visualizing CNN Layers

Input

Convolutional Layer 1

Convolutional Layer 3

Output

88%
Photon
(correct!)