End-to-end Jet ID for quark/gluon discrimination using CMS Open Data

M. Andrews[1], J. Alison[1], S. An[1], P. Bryant[1], M. Paulini[1], B. Poczos[1]
S. Gleyzer[2]
B. Burkle[3], M. Narain[3], E. Usai[3]
[1]Carnegie Mellon University, [2]University of Florida, [3]Brown University

ML4Jets Workshop 2018, 14-16 NOV 2018
Fermilab

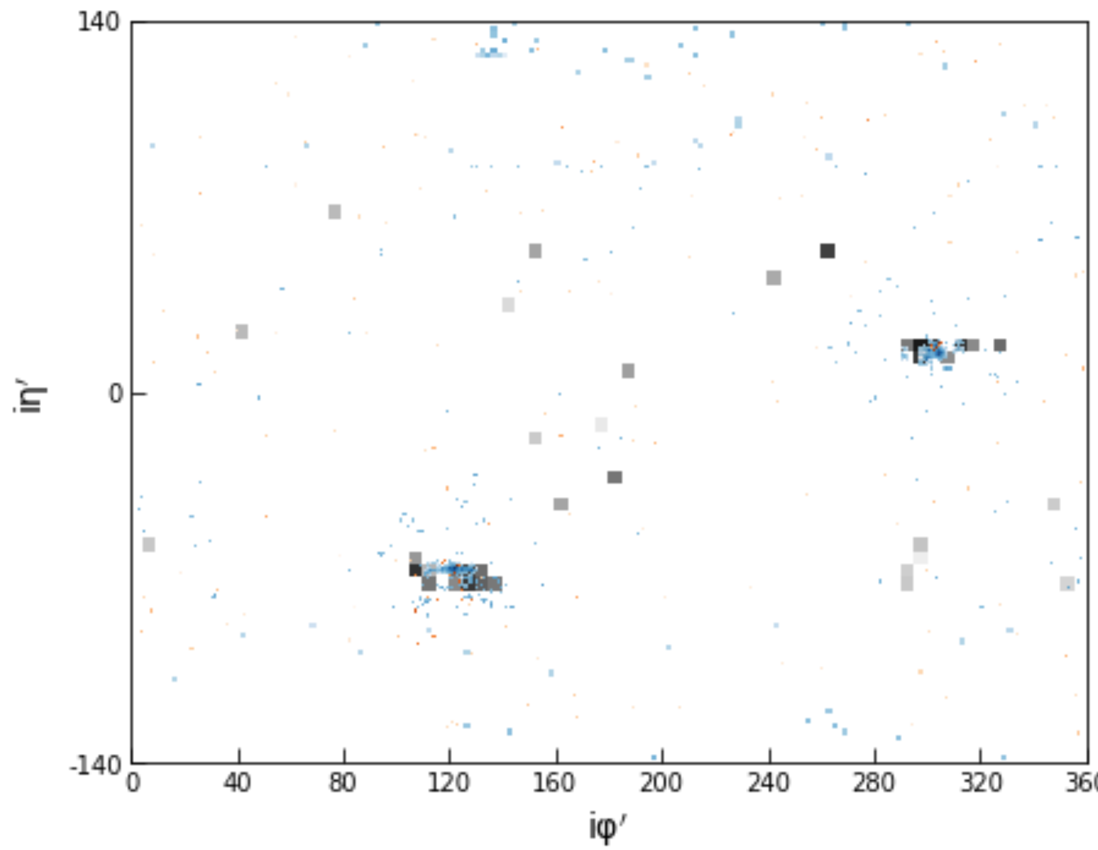
E2E | Outline

- Motivation
- The CMS Detector
- The End-to-end Approach
- Quark vs. Gluon Jet Identification
- Di-quark vs. Di-gluon Event Identification
- Conclusions

Motivation

Typical
Jet ID

Break down classification into different sub-steps which are optimized separately



Detector Data

↓ Physics Reconstruction

	Energy	**p_T**	**η**	**ϕ**	…
1	74.0	70.6	0.44	1.92	
2	47.2	45.3	0.69	0.39	
..					

Particle Data

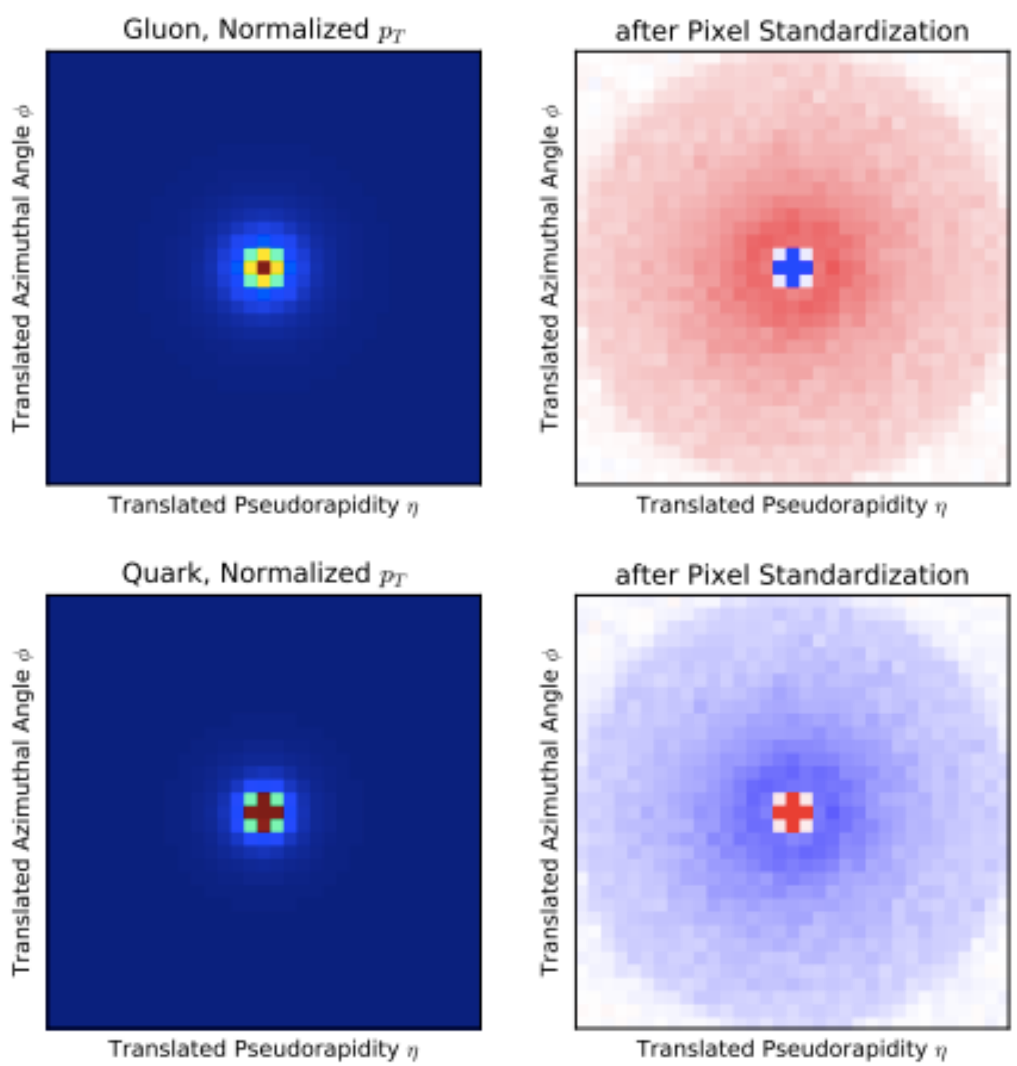
↓ Classification

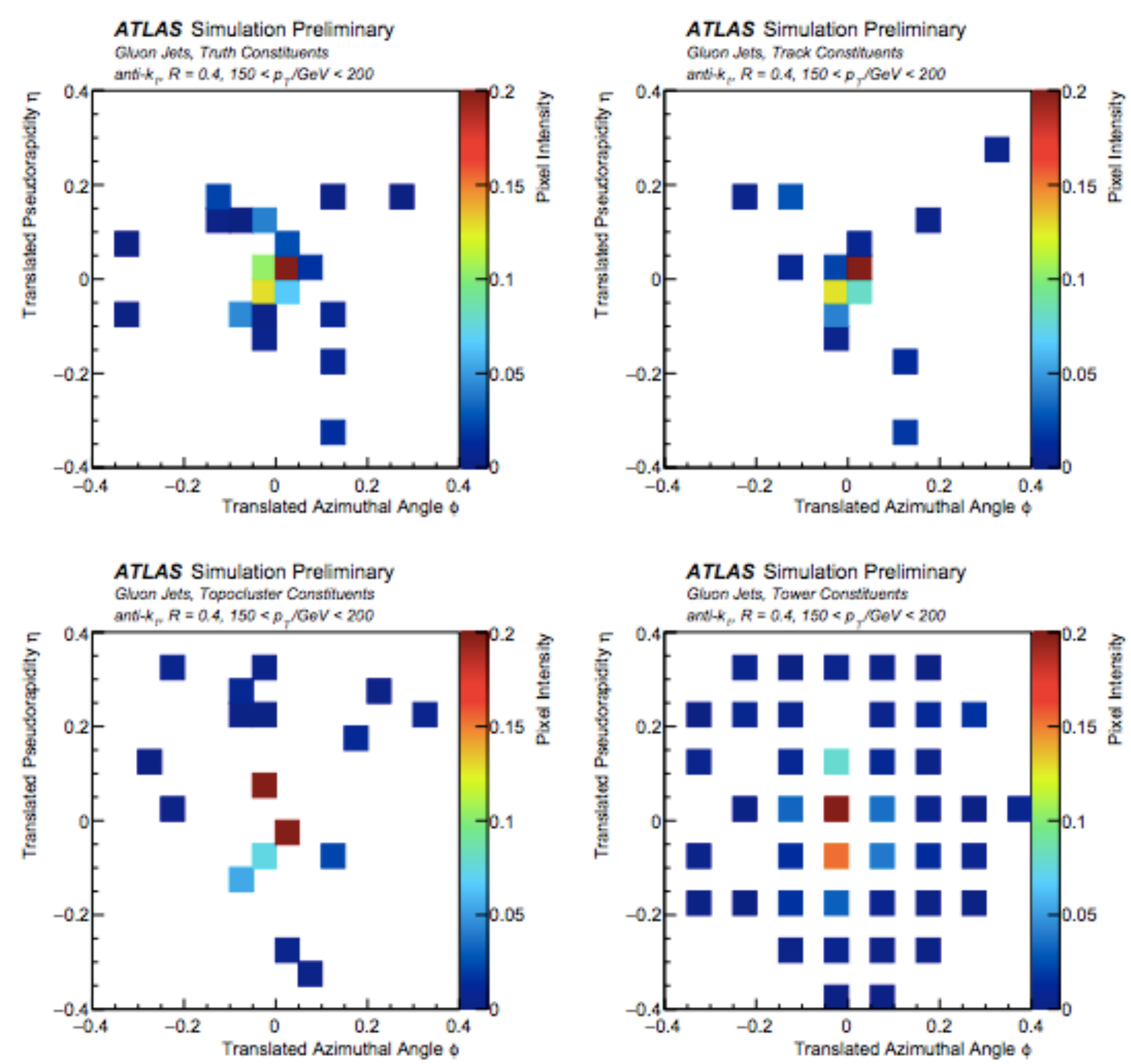
Jet Class

(e.g. gluon vs. light quark)

Jet ID | Images

- **Jet images for quark vs gluon discrimination not new:**
  - See *P. Komiske et al.:* https://arxiv.org/abs/1612.01551
  - See *ATLAS:* http://cds.cern.ch/record/2275641





Jet ID | Images vs High-level features

- **RecNN, Jet ID for QCD vs boosted *W* jet**
  - *K. Cranmer et al.:* https://arxiv.org/pdf/1702.00748.pdf
  - DELPHES detector simulation
  - Applied to quark vs gluon by *T. Cheng:* https://arxiv.org/pdf/1711.02633.pdf
  - **Traditional jet images perform less well than 4-momenta**

Traditional Jet images

Projected into images

towers	MaxOut	**0.8418**	–
towers	k_t	0.8321 ± 0.0025	**12.7 ± 0.4**
towers	k_t (gated)	0.8277 ± 0.0028	12.4 ± 0.3

RecNN

With gating (see Appendix A)

towers	k_t	0.8822 ± 0.0006	25.4 ± 0.4
towers	C/A	0.8861 ± 0.0014	26.2 ± 0.8
towers	anti-k_t	0.8804 ± 0.0010	24.4 ± 0.4
towers	asc-p_T	0.8849 ± 0.0012	27.2 ± 0.8
towers	desc-p_T	**0.8864 ± 0.0007**	**27.5 ± 0.6**
towers	random	0.8751 ± 0.0029	22.8 ± 1.2

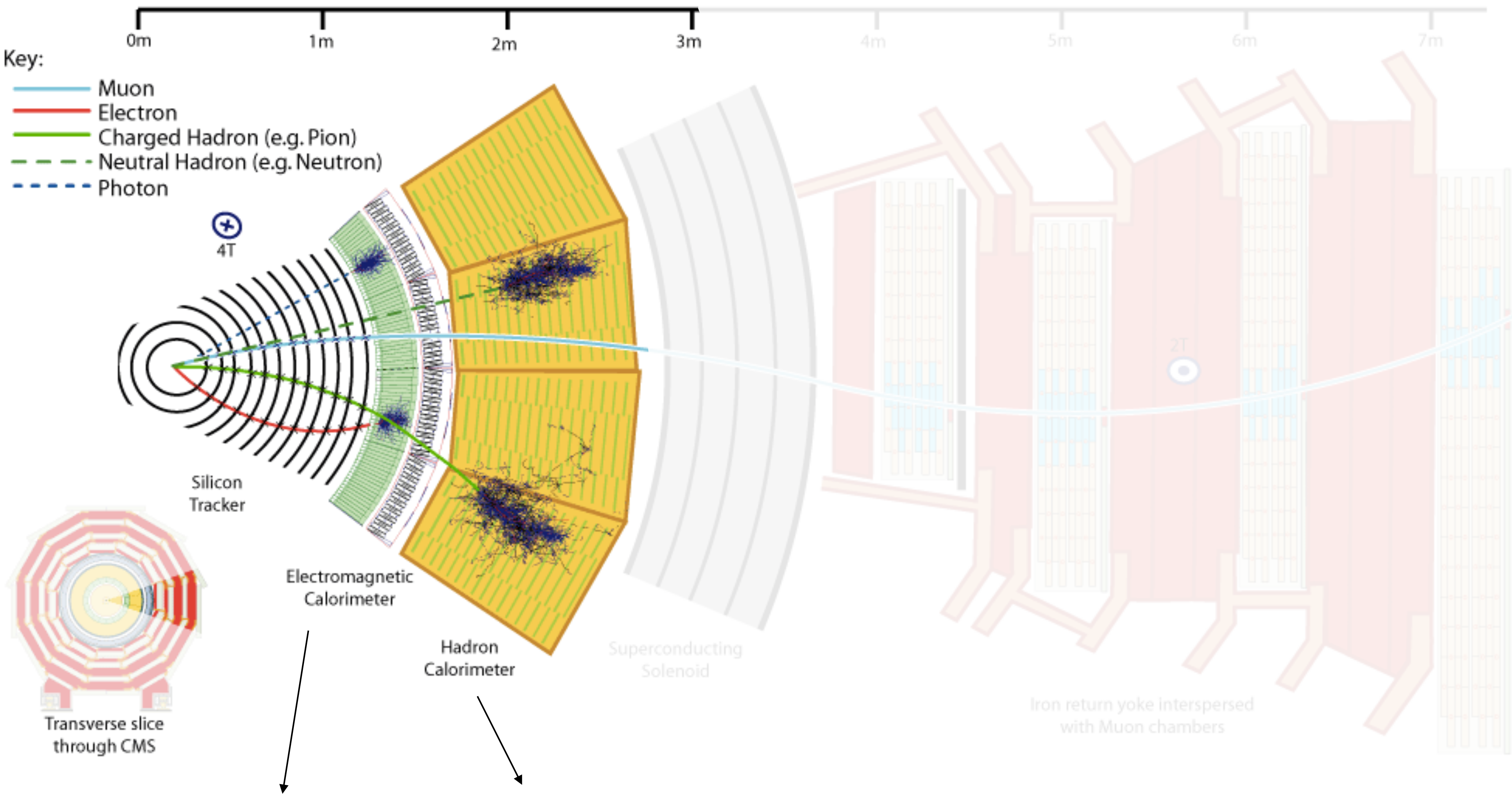
Jet ID | Images vs High-level features

- **CMS DeepJet, Jet ID for quark vs gluon jet**
  - *CMS:* https://cds.cern.ch/record/2275226
  - Jets from QCD dijet, with PU, $|\eta| < 1.3$ or $1.3 < |\eta| < 2.4$
  - CMS GEANT4 full detector simulation
  - **DeepJet comparable to RecNN***

ROC AUC*		Area under ROC	ϵ(tight)	ϵ(medium)	ϵ(loose)
	QCD $\hat{p}_T = 80 - 120\,\mathrm{GeV}$, jet $p_T > 70\,\mathrm{GeV}$				
0.796	DeepJet central	0.204	0.17	0.51	0.65
0.797	DeepJet forward	0.203	0.15	0.50	0.65
0.789	Convolution central	0.211	0.15	0.49	0.64
0.785	Convolution forward	0.215	0.13	0.47	0.63
0.795	Recurrent central	0.205	0.16	0.51	0.65
0.795	Recurrent forward	0.205	0.14	0.49	0.65

***NOTE:** *"In addition, the p_T and η of the jet, the number of charged and neutral candidates, and the number of secondary vertices within the jet are given to the following dense layer with 128 nodes."*

CMS | Geometry & Particle ID

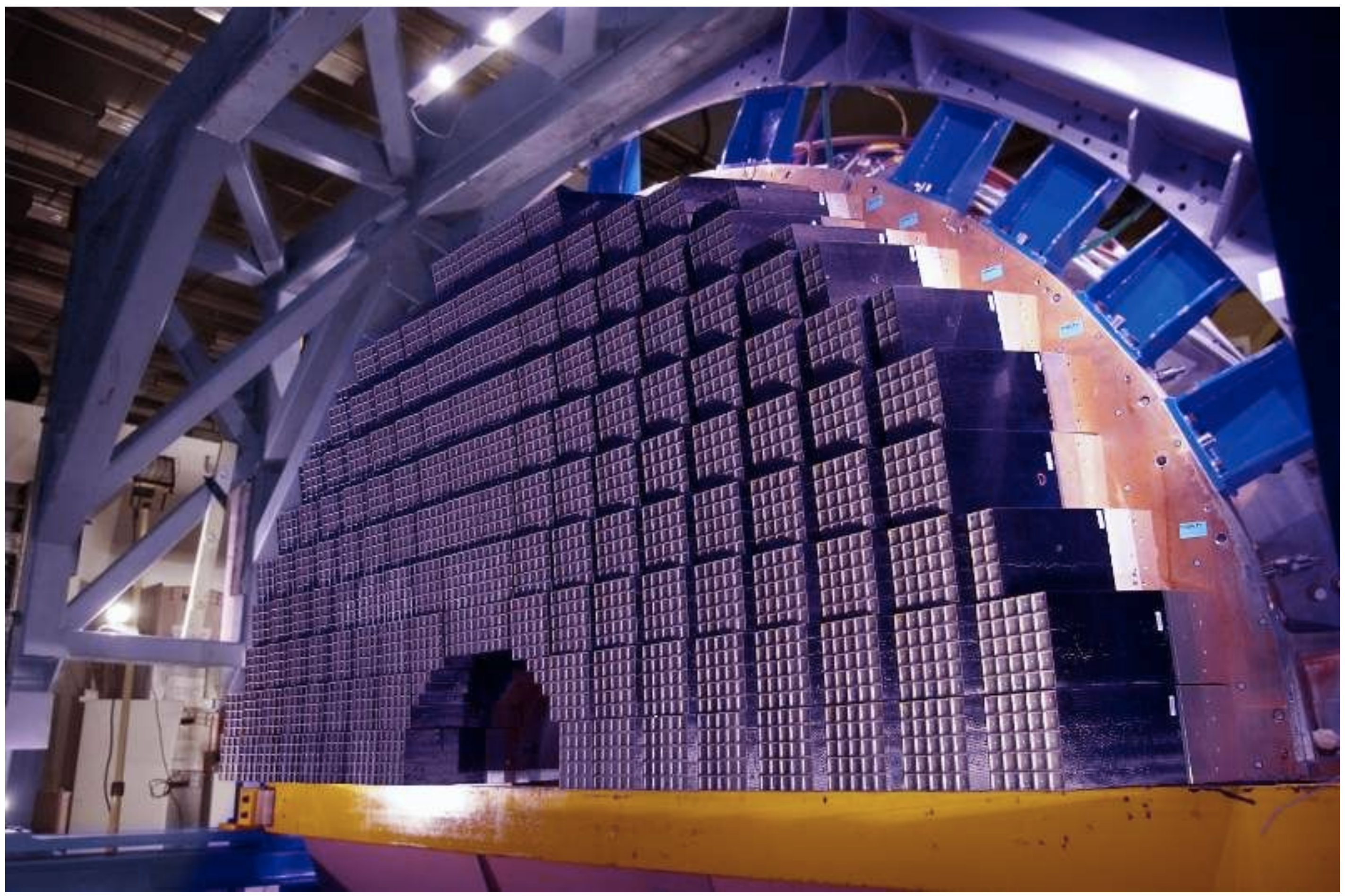


crystal-based
$\Delta\eta \times \Delta\phi \sim 0.0174 \times 0.0174$

tower-based
$\Delta\eta \times \Delta\phi \sim 0.087 \times 0.087$

Image Credit: CERN

CMS | Detector Segmentation

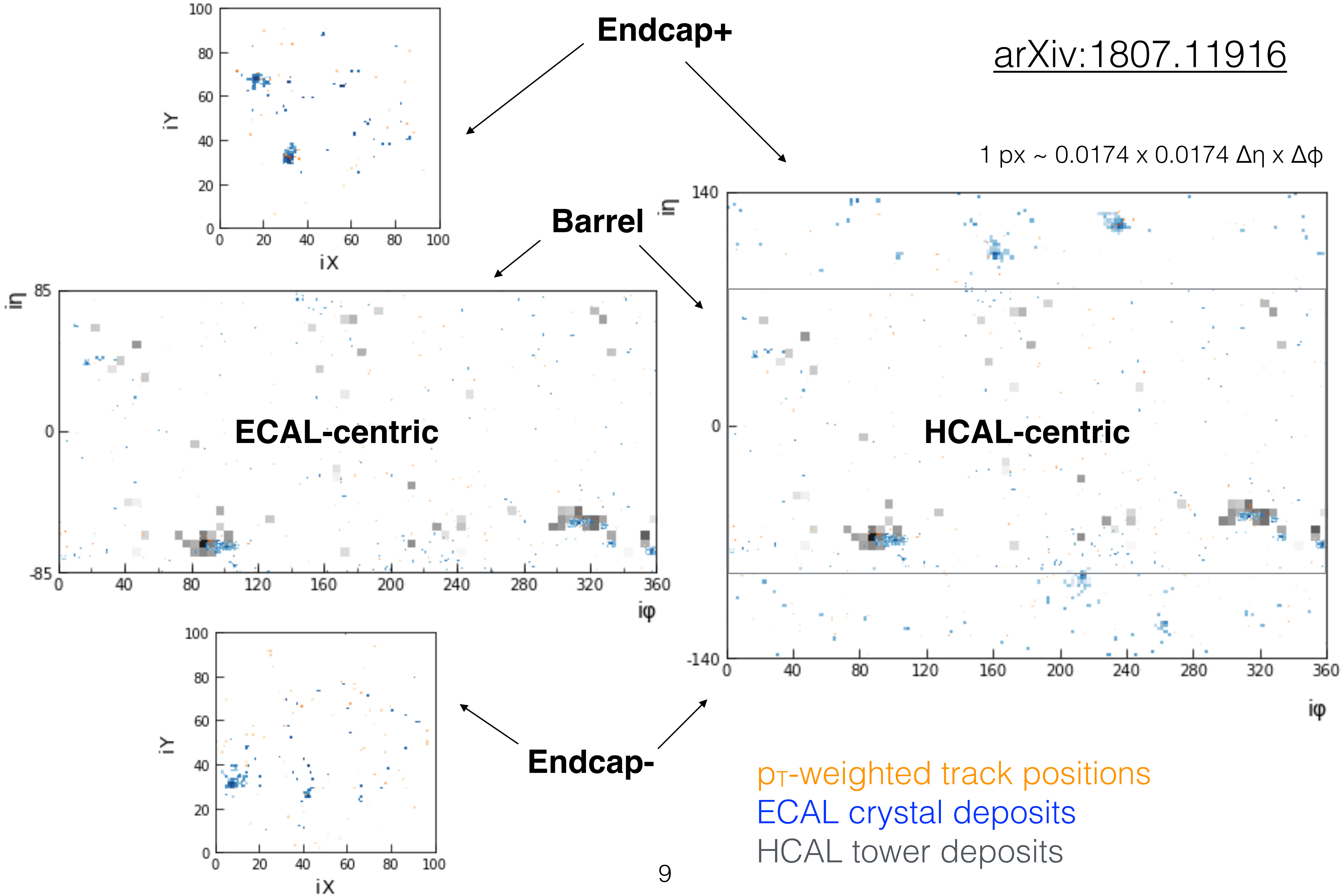


ECAL Endcap
(iX, iY)

HCAL Endcap
(iϕ, iη)

Image Credit: CERN

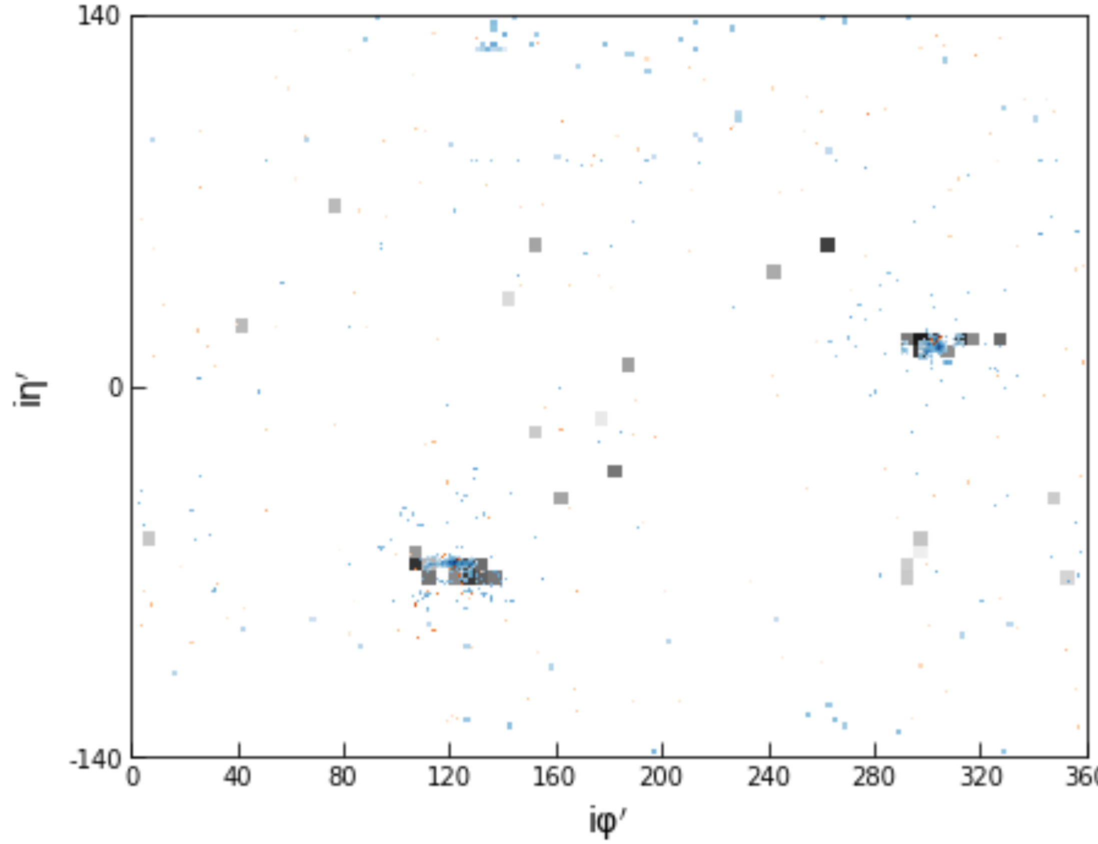
CMS | Detector Geometry



End-to-end Event ID

Optimize for the final classification objective

Detector data as fundamental (maximum?) measured information



Detector Data

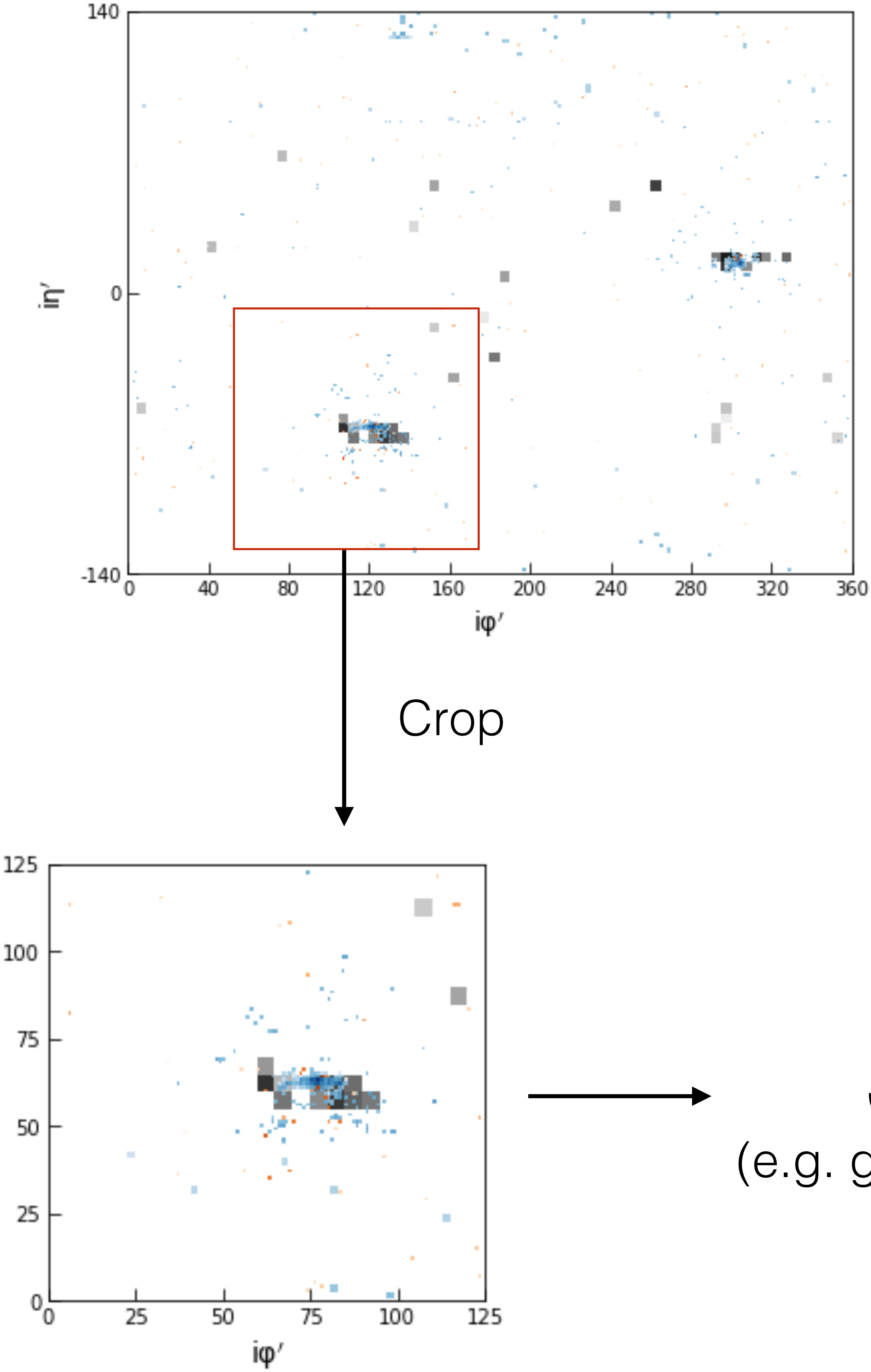
Event Class

(e.g. digluon vs. diquark)

End-to-end Jet ID

Optimize for the final classification objective

Detector data as fundamental (maximum?) measured information



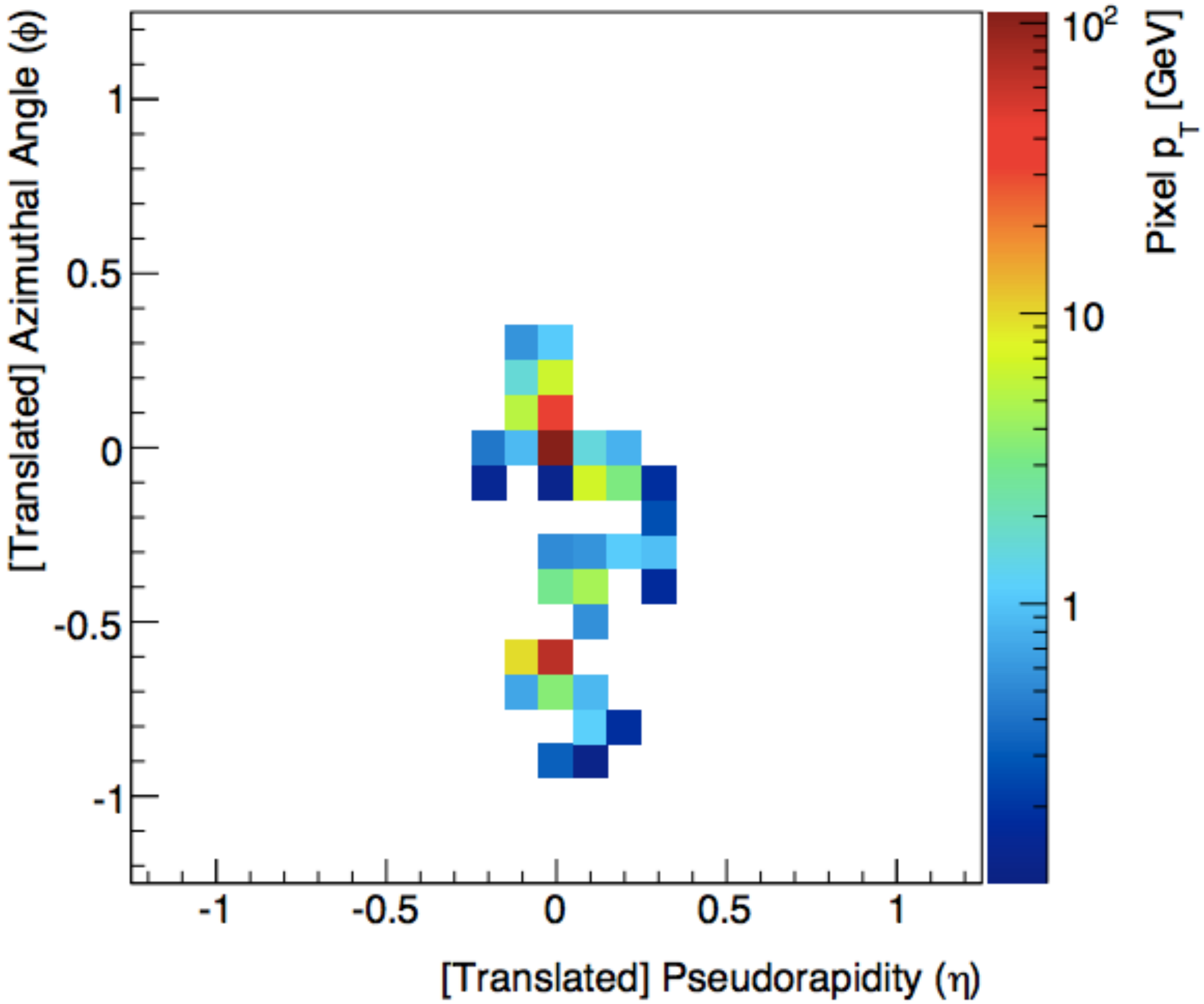
Detector Data

Jet Class

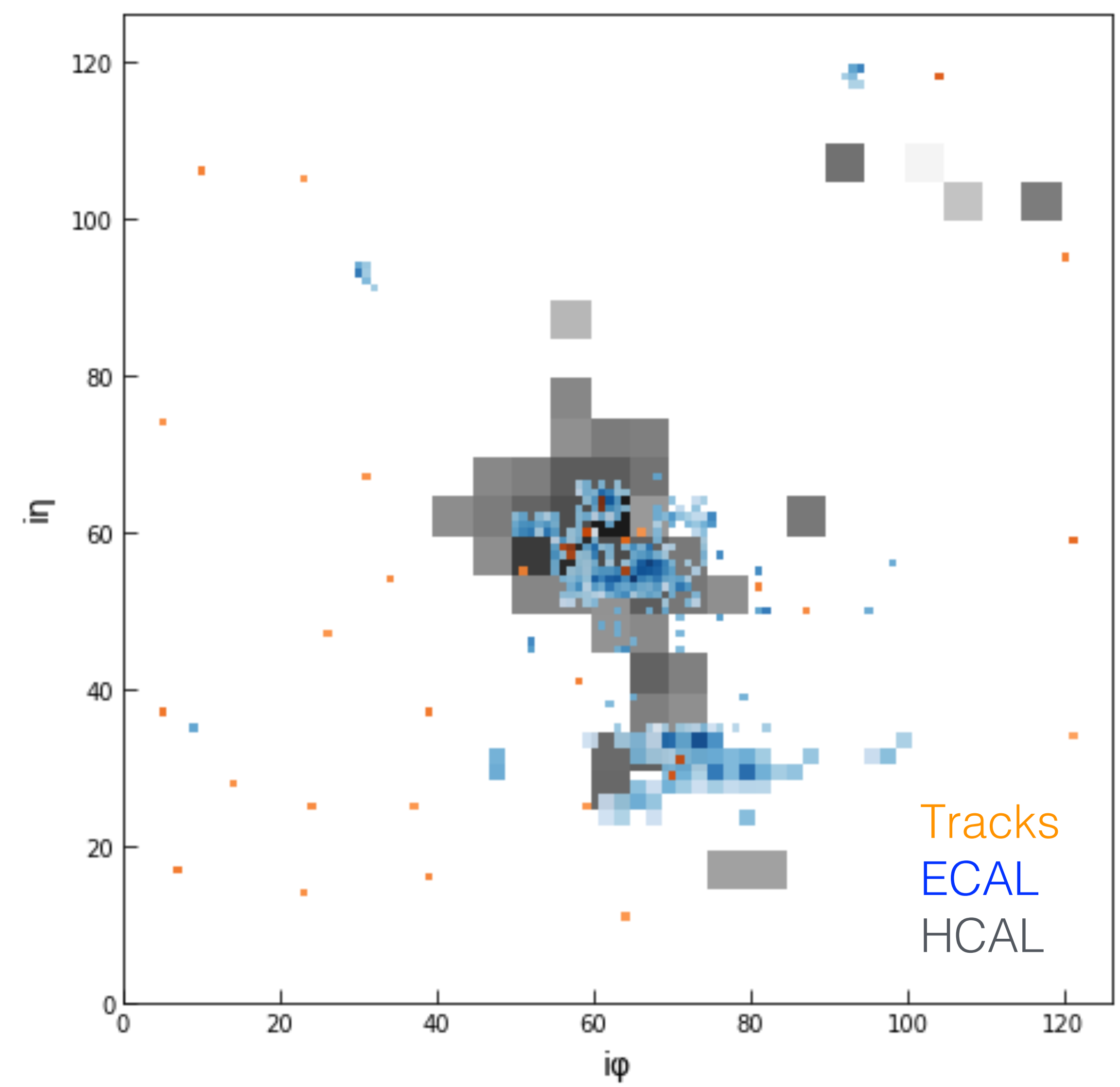
(e.g. gluon vs. light quark)

Jet ID | Traditional vs E2E Image

Traditional
jet image



E2E
jet image



Note: Not the same jet.

Jet ID | quark vs gluon

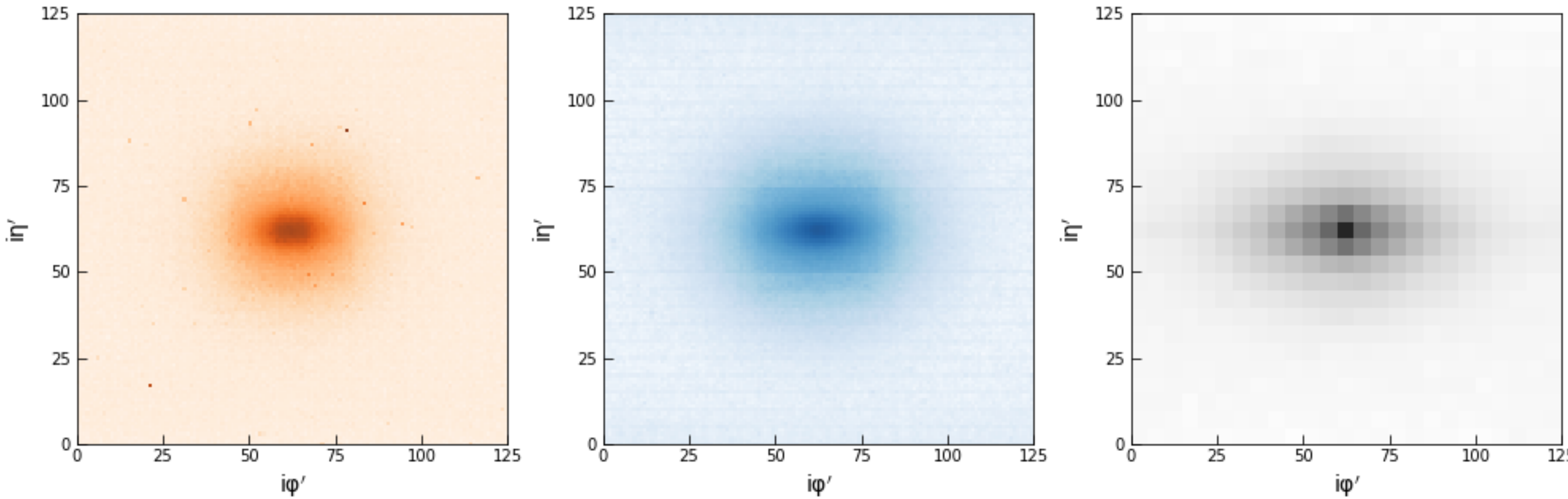
- **CMS OpenData QCD Samples**
  - Leading jet from QCD dijet qq’ (*uds*) or gg, EMenriched @ 8 TeV
  - CMS GEANT4 full detector simulation, PTYHIA 6
  - $\hat{p}_T$: 80-170 GeV, reco $p_T > 70$ GeV, $|\eta| < 1.8$
  - Run-dependent $\langle PU \rangle$: 18-21
  - Produced and ntuplized with CMSSW 5_3_32
  - **Sample split:**
    - Training set: 576k jets (of which, 26k jets for validation)
    - Test set: 139k jets
    - Balanced samples per class
    - Balanced PU representation per class
  - **Architecture:** ResNet-15 trained from scratch on an NVIDIA Titan X/p using Pytorch 0.4

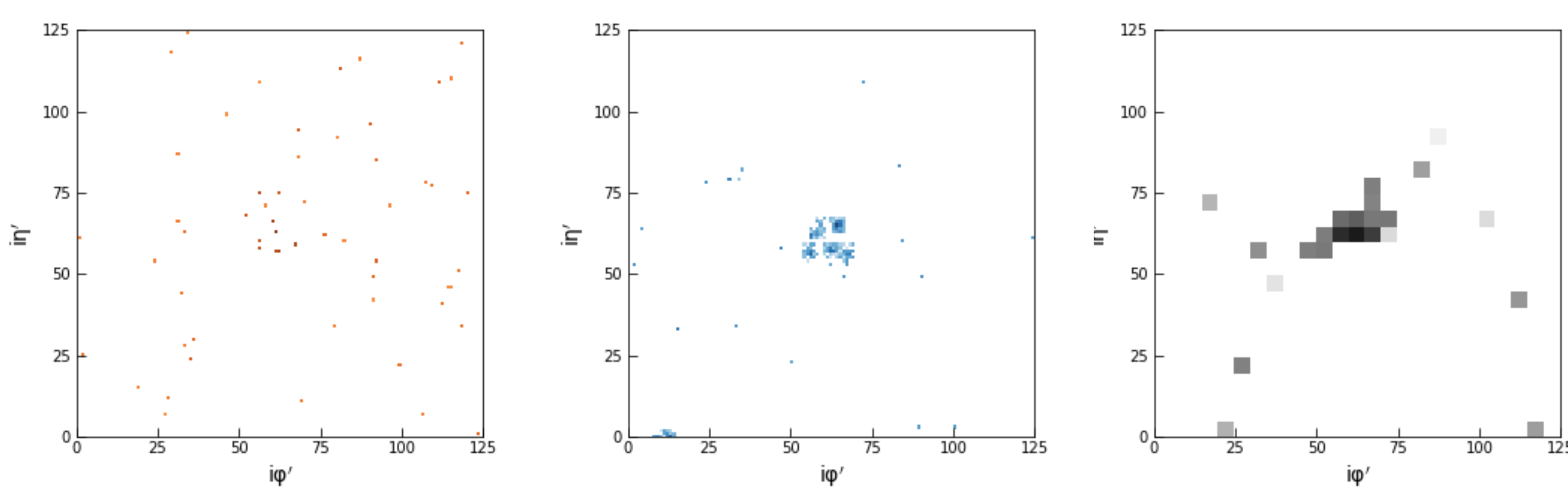
1 px = 0.0174 x 0.0174 Δη x Δϕ

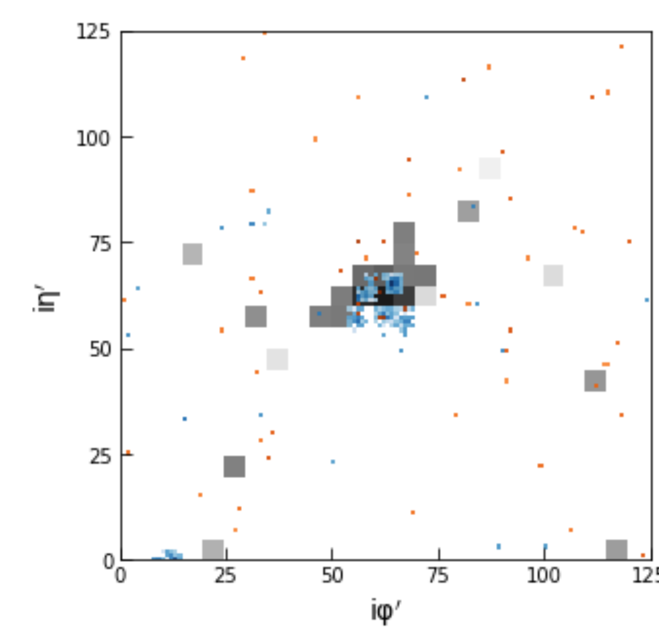
E2E Image | gluon

Radiation pattern more dispersed (top: overlays, bottom: single jet)

Tracks
ECAL
HCAL



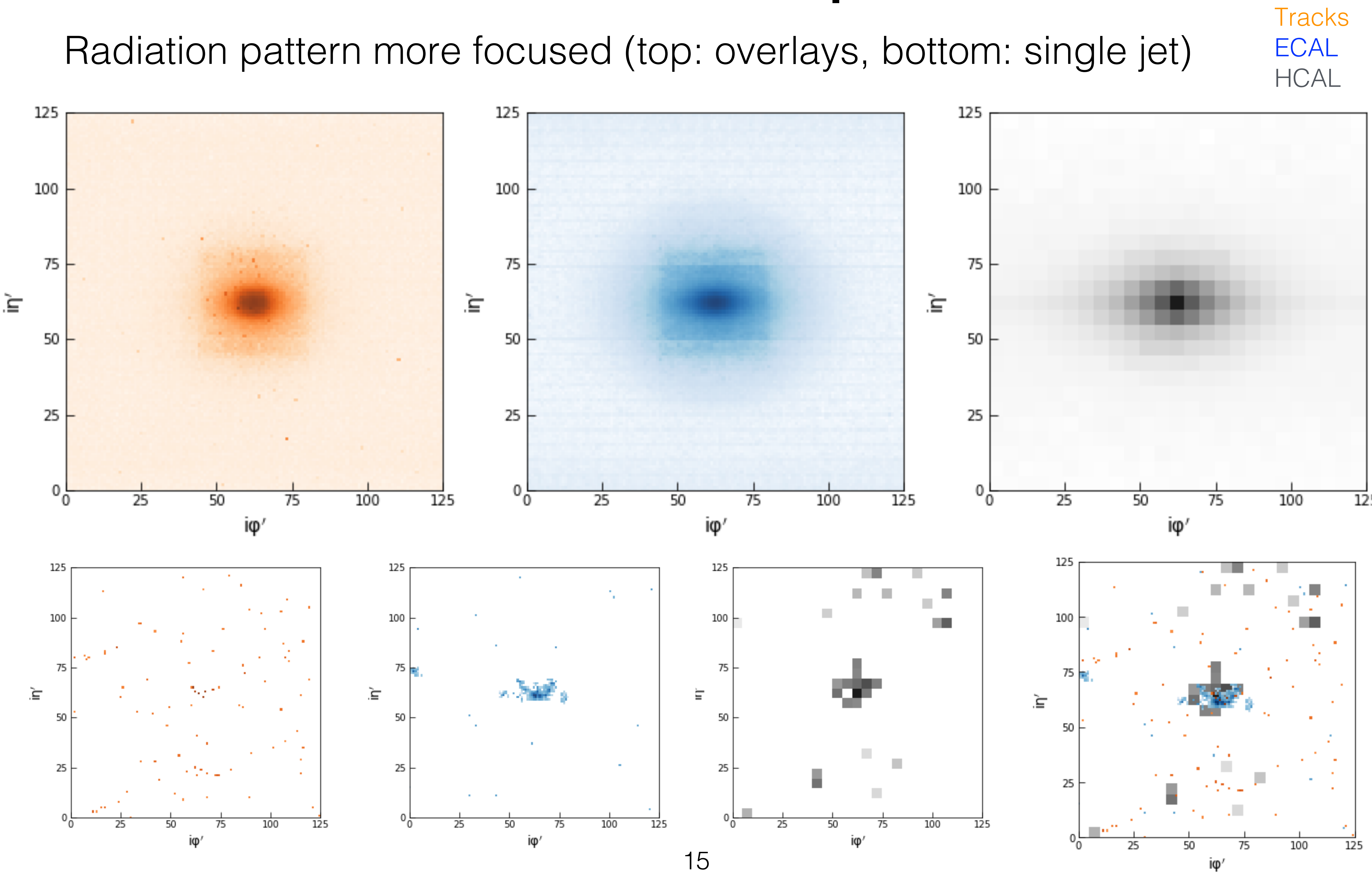




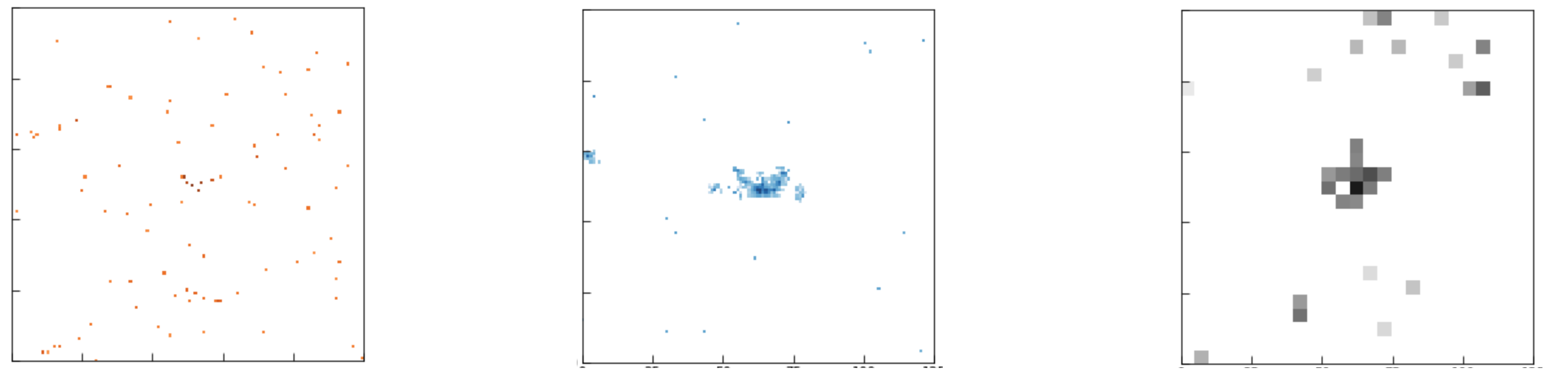
1 px = 0.0174 x 0.0174 Δη x Δϕ

E2E Image | quark

Radiation pattern more focused (top: overlays, bottom: single jet)



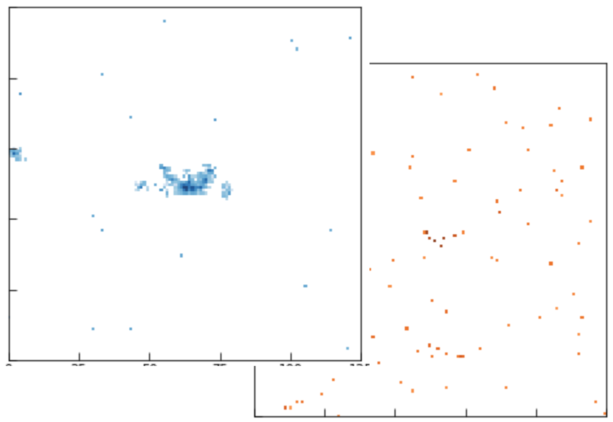
Jet ID | quark vs gluon

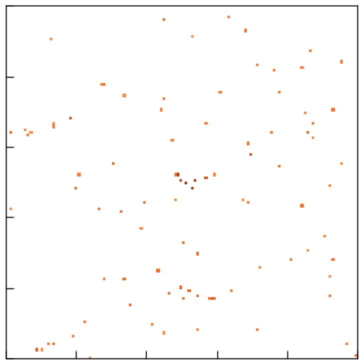


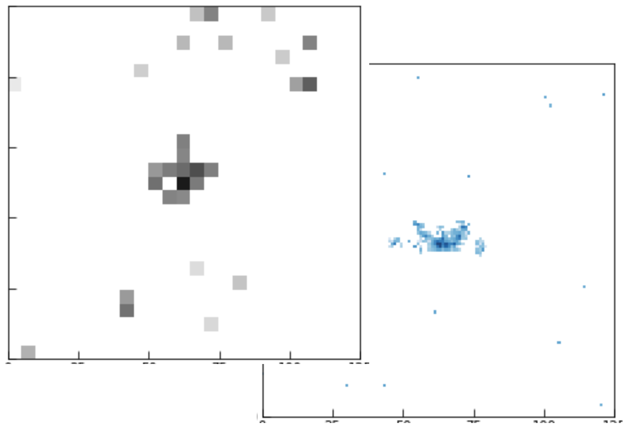
	ROC AUC
E2E jet image, Tracks	0.782
E2E jet image, ECAL	0.760
E2E jet image, HCAL	0.682

- **E2E Results, Jet ID**
  - Provides insight into detector performance / particle ID
  - **Spatial resolution important:** track info more valuable than shower/energy information from any one calorimeter
  - **Handles sparsity well**

Jet ID | quark vs gluon



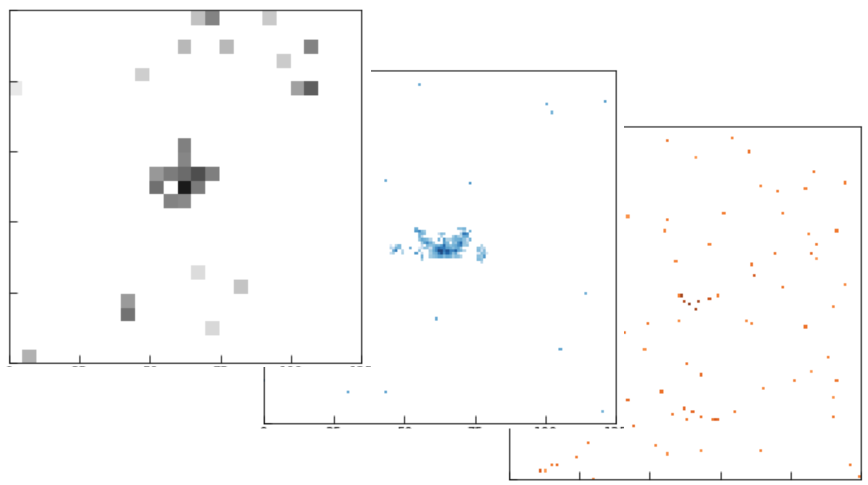


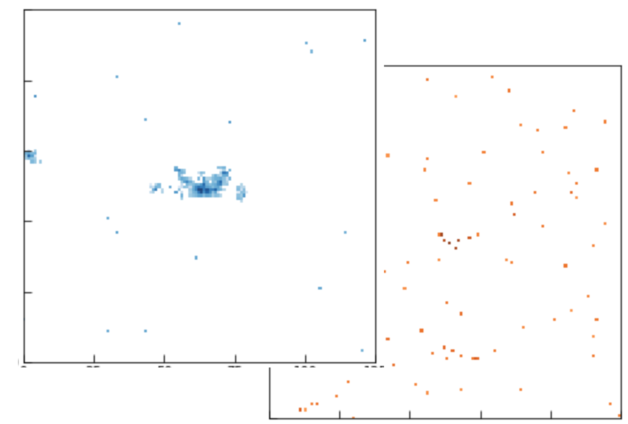


	ROC AUC
E2E jet image, ECAL+Tracks	0.804
E2E jet image, Tracks	0.782
E2E jet image, ECAL+HCAL	0.781

- **E2E Results, Jet ID**
  - Combine two subdetector images
  - **Spatial resolution important:** charged hadron info from Tracks more valuable than from HCAL
  - **Track info alone as valuable as combined calo info**

Jet ID | quark vs gluon





	ROC AUC
E2E jet image, ECAL+HCAL+Tracks	0.808
E2E jet image, ECAL+Tracks	0.804

- **E2E Results, Jet ID**
  - Combine ECAL+HCAL+Tracks images
  - **ECAL+Tracks sufficient for strong discrimination:** HCAL info not so important
  - **Track info supplemented with calo info works best.**

Jet ID | quark vs gluon

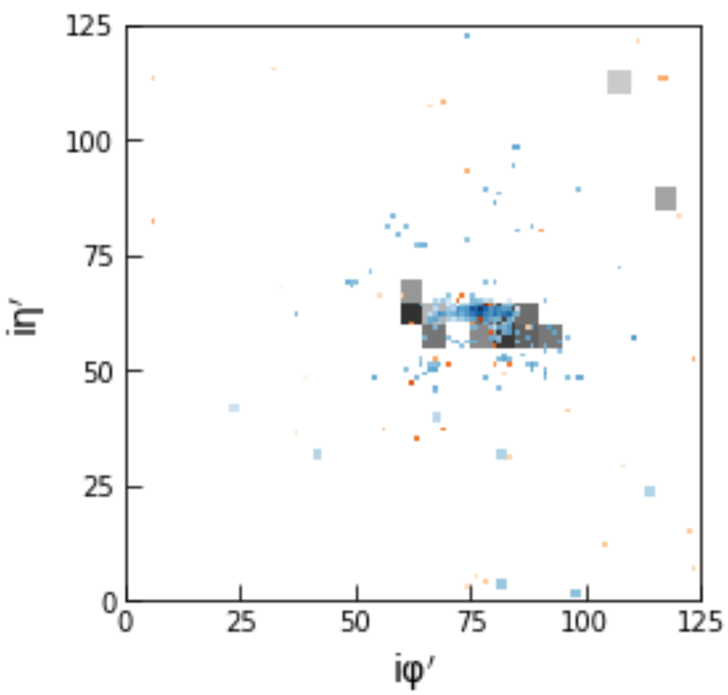
	ROC AUC
E2E image, ECAL+HCAL+Tracks	**0.8077 ± 0.0003***
RecNN, ascending-p_T	0.8017 ± 0.0003*
RecNN, descending-p_T	0.802
RecNN, anti-k_T	0.801
RecNN, Cambridge/Aachen	0.801
RecNN, no rotation/reclustering	0.800
RecNN, k_T	0.800
RecNN, k_T-colinear10-max	0.799
RecNN, random	0.797

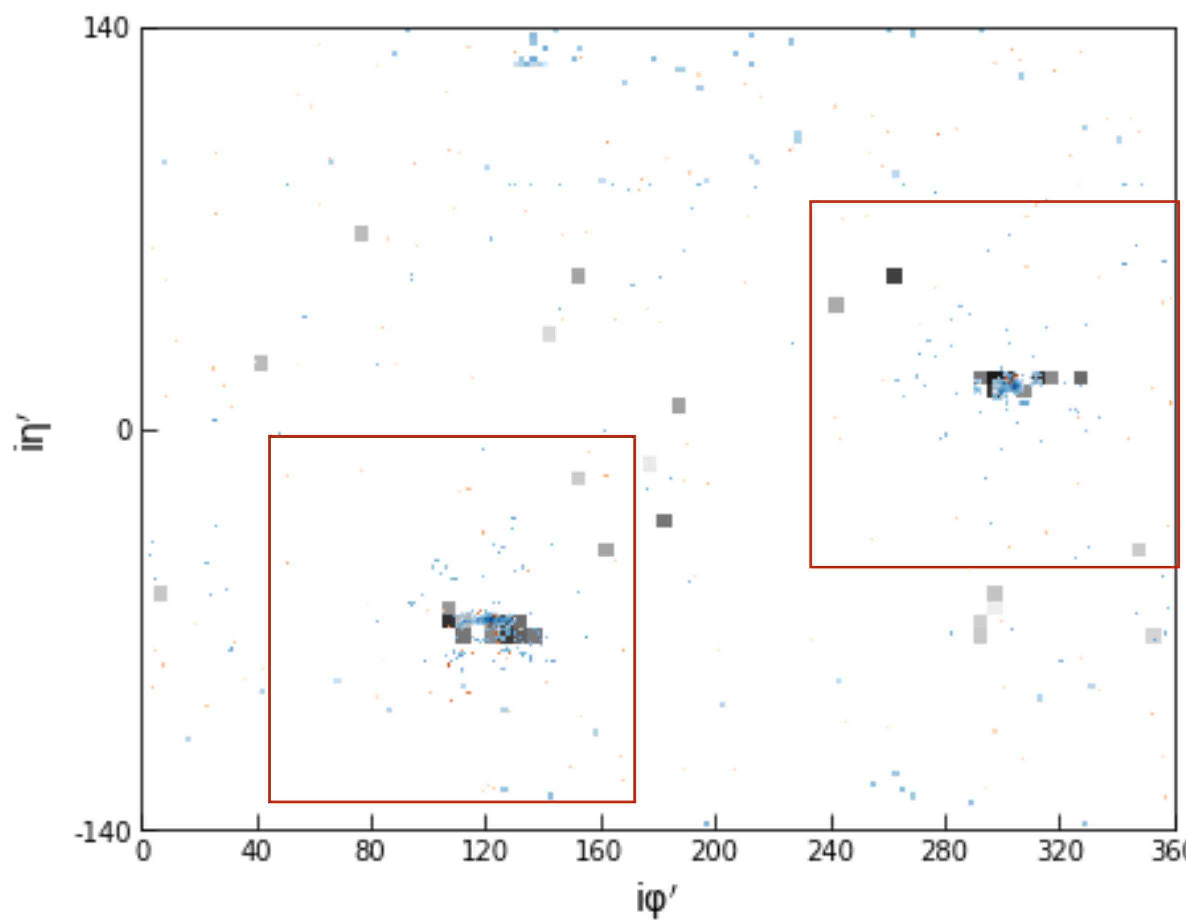
- **RecNN Results, Jet ID**
  - Use 4-momenta derived from CMS Particle Flow
  - Perform boost/rotation, then reclustering with different algos
  - **E2E jet images perform well**

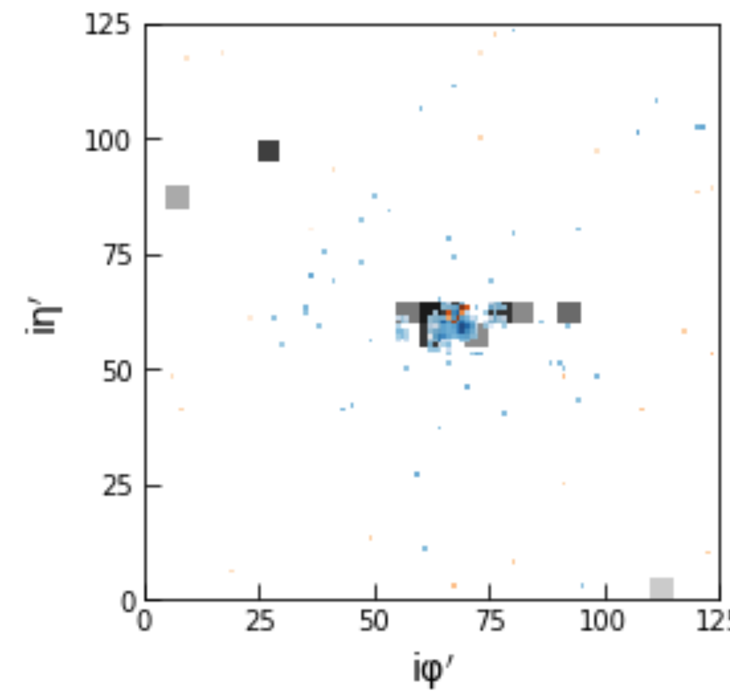
*ROC AUC mean ± 1σ over 5 trials

Event ID | qq vs gg

- Classify the full event as either QCD di-quark or di-gluon
- In addition to local jet physics, **global event-level physics factors in:** jet 4-momenta, *qq* spin-correlations and polarization
- Problem becomes much richer!

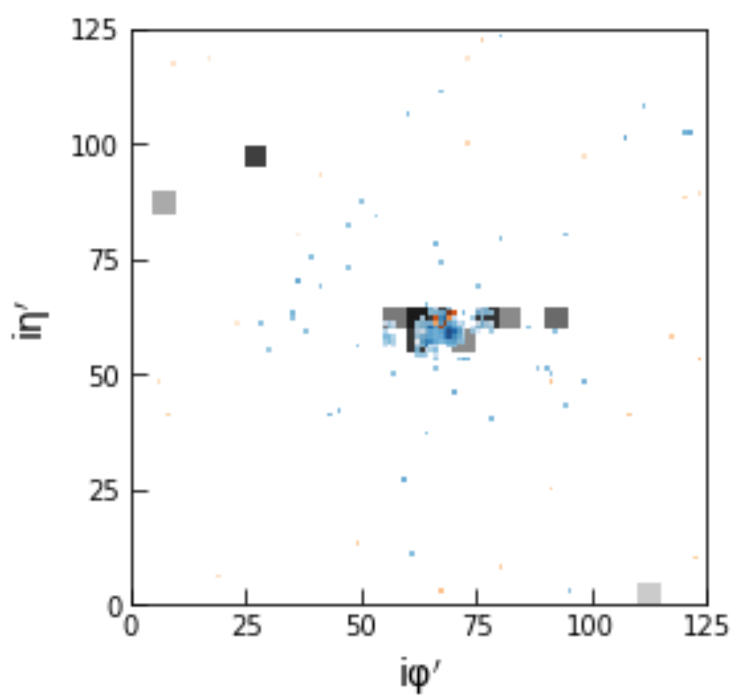




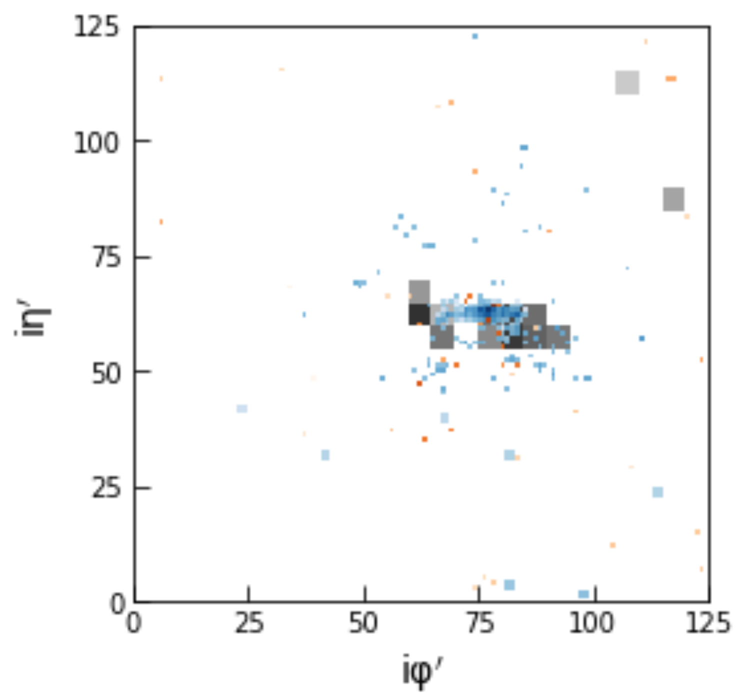


Event ID | qq vs gg

- **Scenario A:** 2 x jet images



ResNet-15, convolutional output

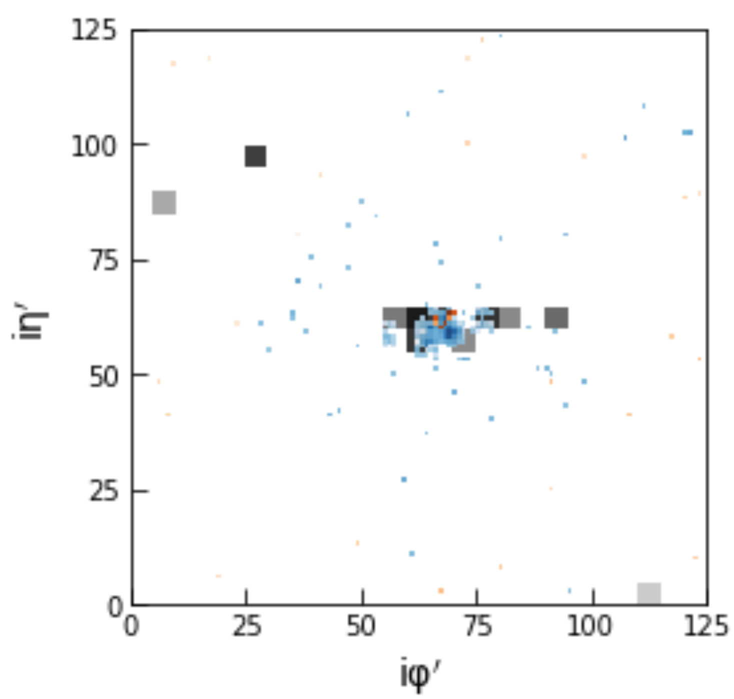


ResNet-15, convolutional output

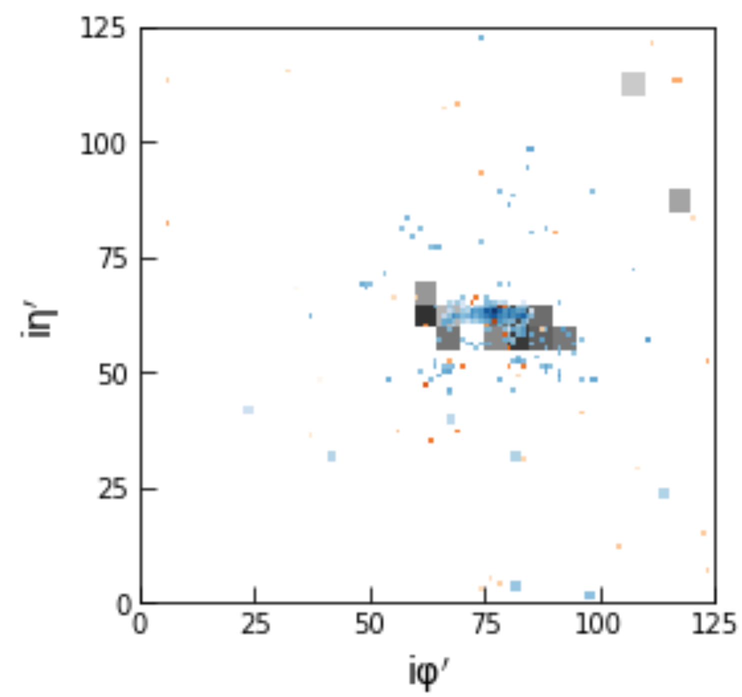
Fully-connected, 128 x 2

Event ID | qq vs gg

- **Scenario B:** 2 x jet images + jet 4-momenta



ResNet-15,
convolutional output



ResNet-15,
convolutional output

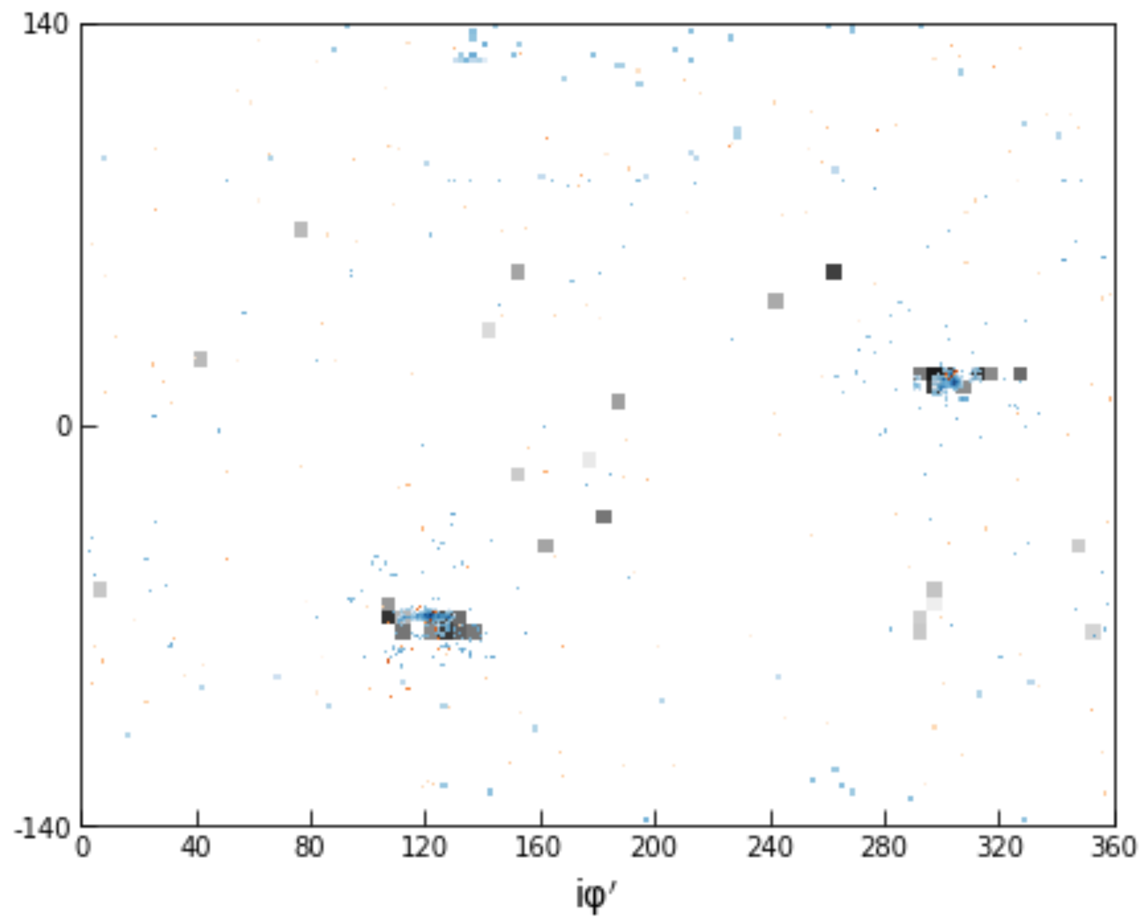
$p_{T,i}$
η_i
$\Delta\phi_{ij}$

dijet 4-momenta

Fully-connected, 128 x 2

Event ID | qq vs gg

- **Scenario C:** Fully end-to-end detector image



ResNet-15

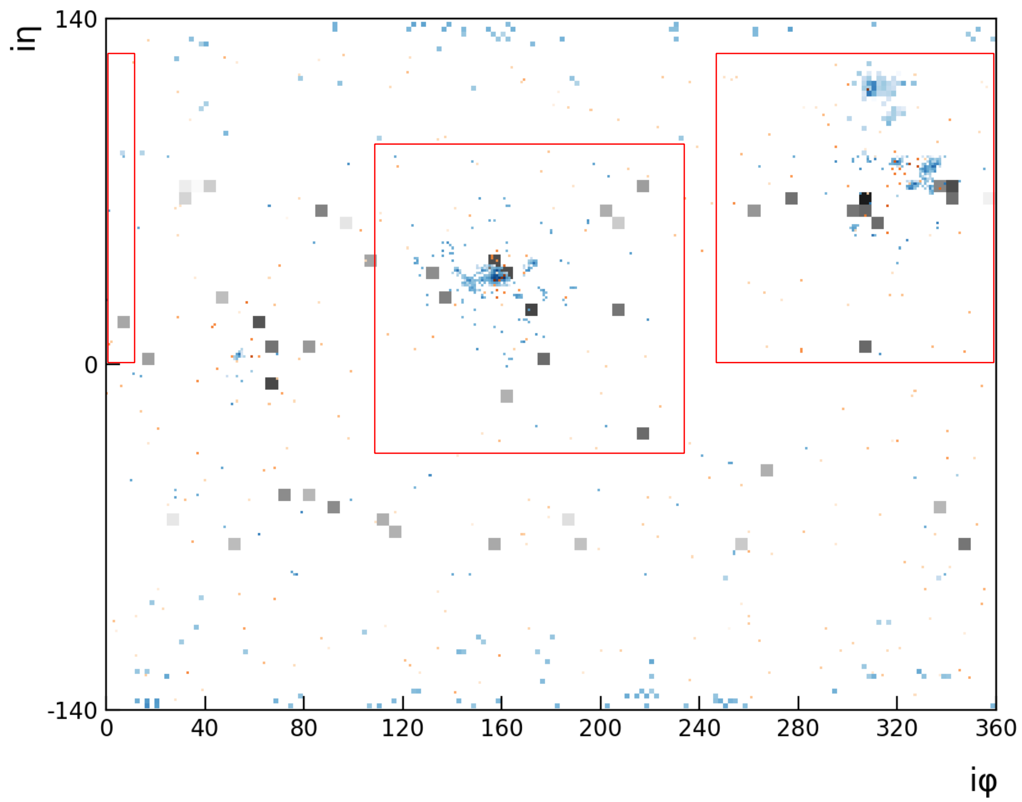
Event ID | qq vs gg

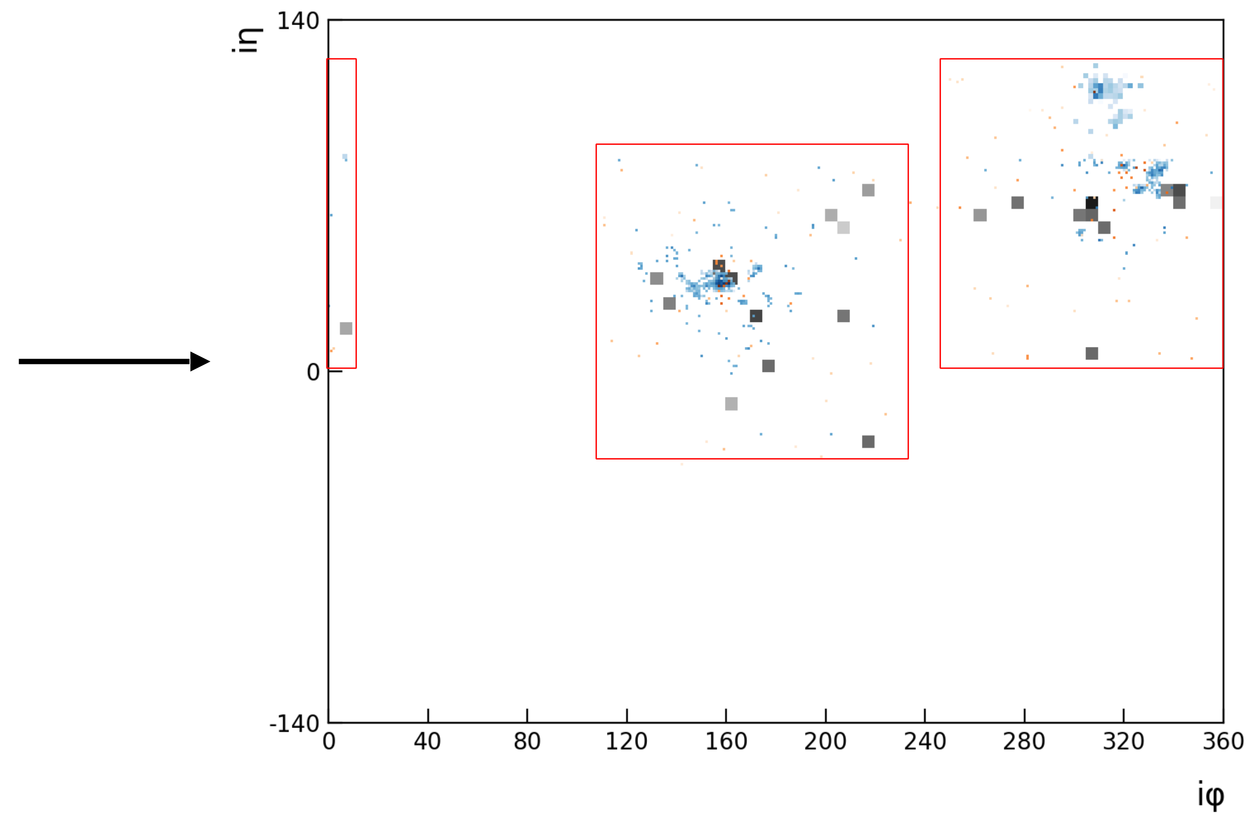
	ROC AUC
Scenario A	0.876
Scenario B	0.878
Scenario C	0.889

- **Local or global physics? Part I.**
  - Performance dominated by jet-level differences (Scenario A vs. B or C)
  - Both dijets are non-resonant decays, so jet 4-momenta doesn't hold much discrimination power (Scenario B vs. A)
  - **Fully E2E approach (Scenario C) picking up on subtle, event-level effects not captured by either B or A?**

Event ID | qq vs gg

- **Is the E2E relying on the underlying event/PU?**
  - Try **Scenario C-Zero:** zero out all pixels outside of the two jet windows





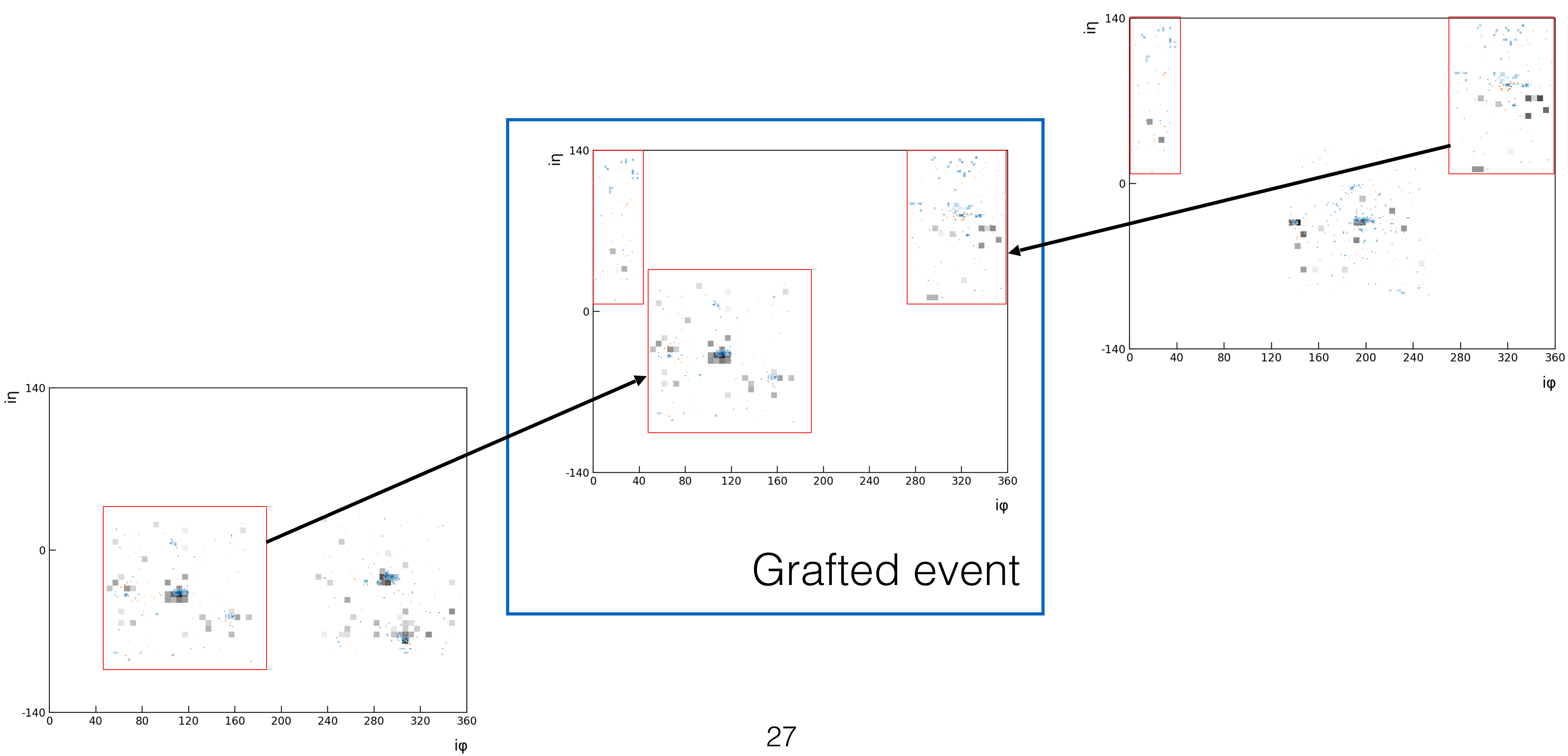
Event ID | qq vs gg

	ROC AUC
Scenario C	0.889
Scenario C-Zero	0.887
Scenario C, evaluated on C-Zero	0.883
Scenario C-Zero, evaluated on C	0.884

- **Is the E2E relying on the underlying event/PU?**
  - E2E event classifier not sensitive to underlying event and PU outside of jet region of interest

Event ID | qq vs gg

- **Local or global physics? Part II.**
    - **Scenario C-Zero-Graft:** Graft jets from different events onto a new image with fake event-level info but otherwise real jets



Event ID | qq vs gg

- **Local or global physics? Part II.**
  - Use model trained on Scenario C-Zero and evaluate on grafted events, Scenario C-Zero-Graft

	ROC AUC
Scenario C-Zero	0.887
Scenario C-Zero, evaluated on C-Zero-Graft	0.877
Scenario A	0.876

- **Consistent with findings from Part I:**
  - Performance from jet-level differences preserved
  - The subtle event-level info is lost in Scenario C-Zero-Graft—score now similar to 2 x jet images (Scenario A)
  - **E2E learns event-level correlations**

E2E | Conclusions

E2E Jet ID:

- Achieves quark vs. gluon discrimination competitive with existing state-of-the-art jet ID classifiers
- **Not all jet images are created equally:** E2E techniques help to optimize full detector performance

E2E Event ID:

- Able to capturesubtle, event-level correlations not present at jet-level that may otherwise be difficult to model by hand
  - **Capable of learning particle phenomenology**
  - Can be be “reversed-engineered” to understand what deep physics is being learned
- Smart enough to know what is noise/irrelevant in the image without any human intervention

E2E | Outlook

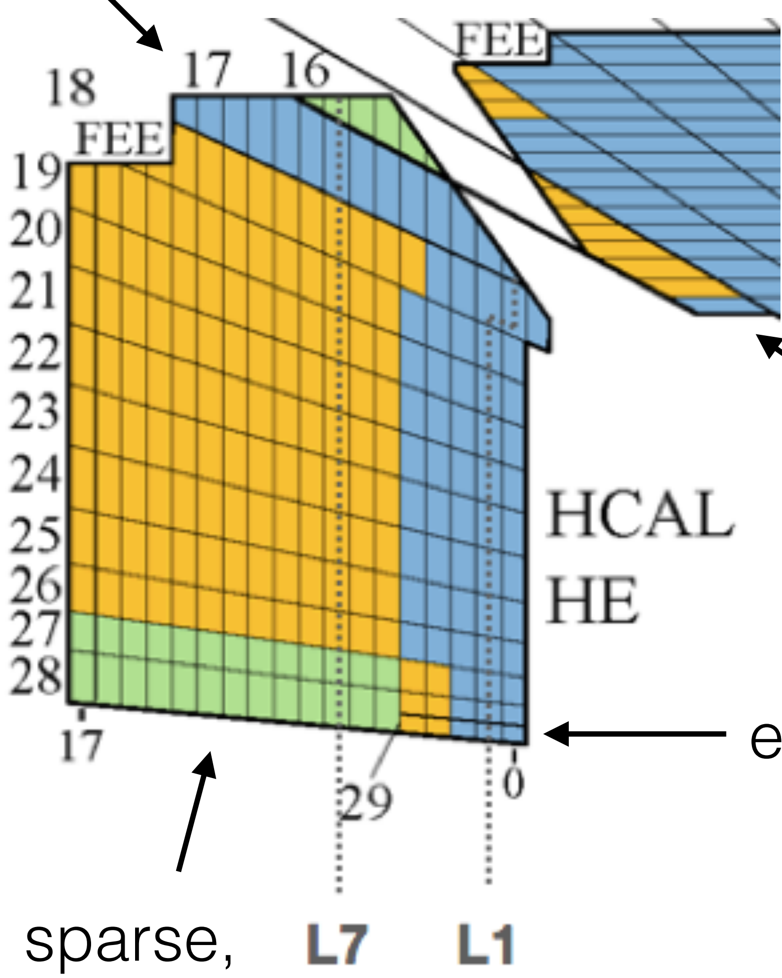
- **How far can we take E2E approach?**
  - Use the full Tracker information?
  - Add Muon Trackers
  - Effects of higher pile-up?
  - Apply to boosted topologies

BACKUP

HCAL | Segmentation

iphi planes of ECAL crystals vs HCAL towers not aligned!

EB/EE boundary



HB/HE overlap

embedding

sparse, non-trivial segmentation of readout depths

Granularity in phi halved

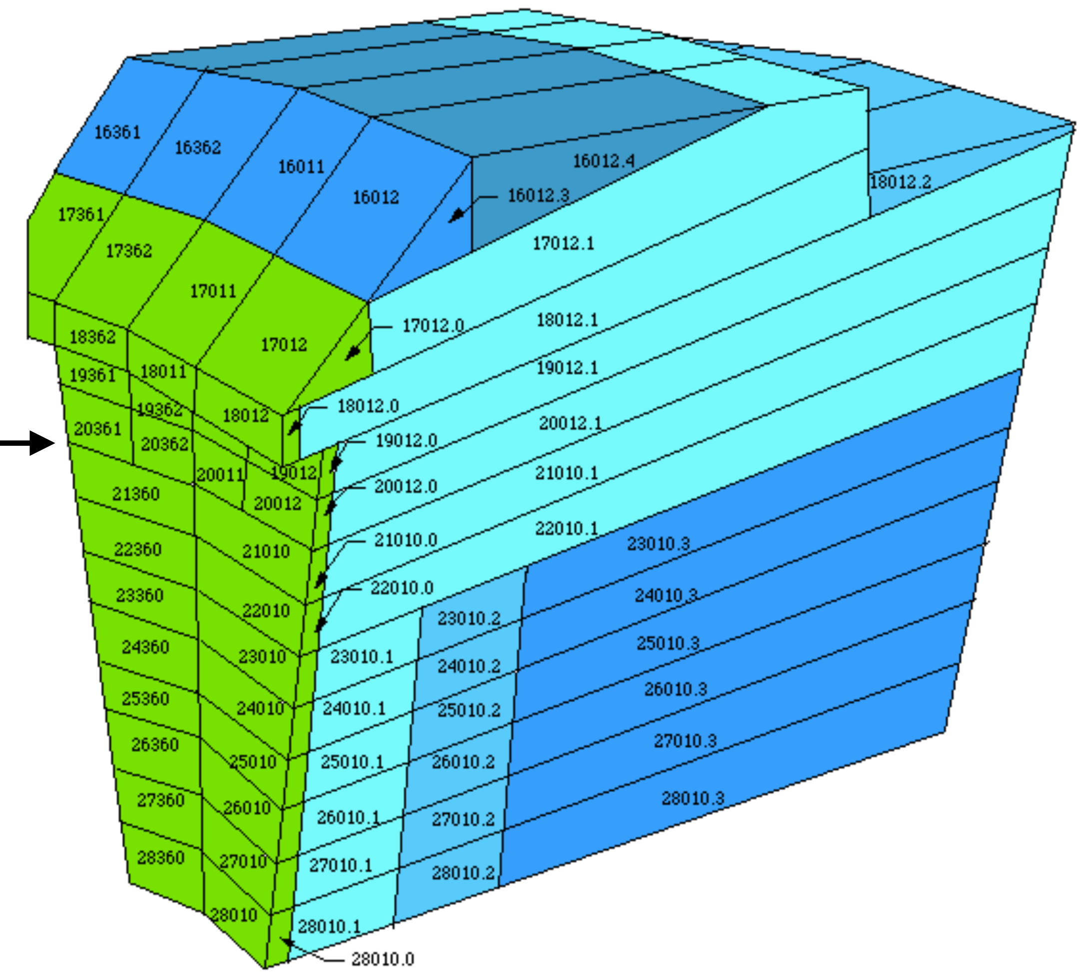
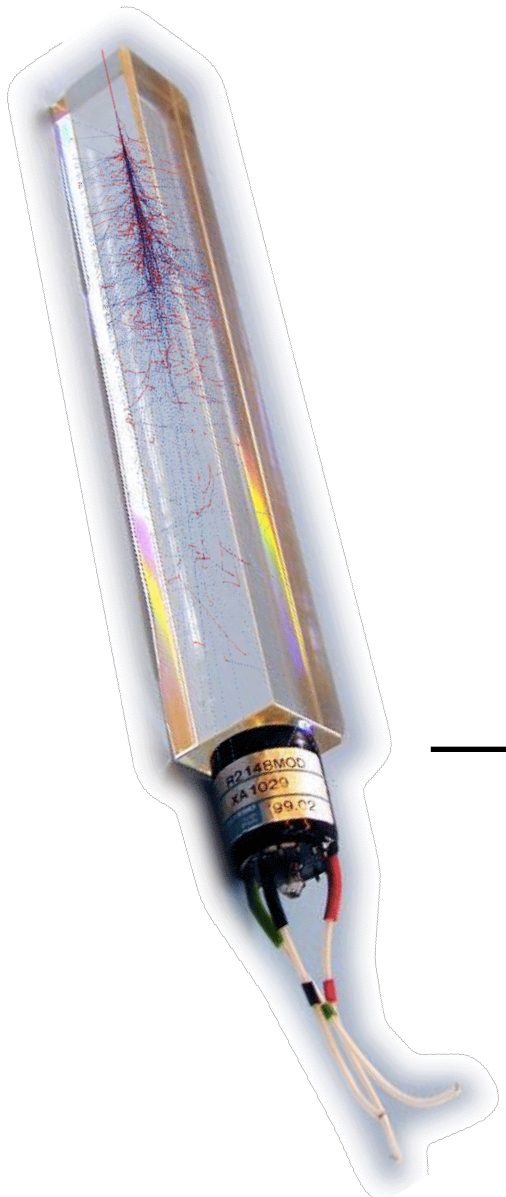


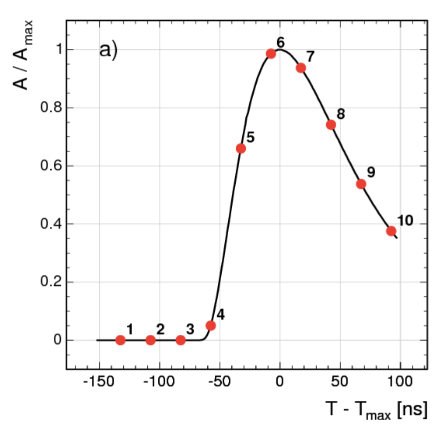
Image Credit: CERN, UMD

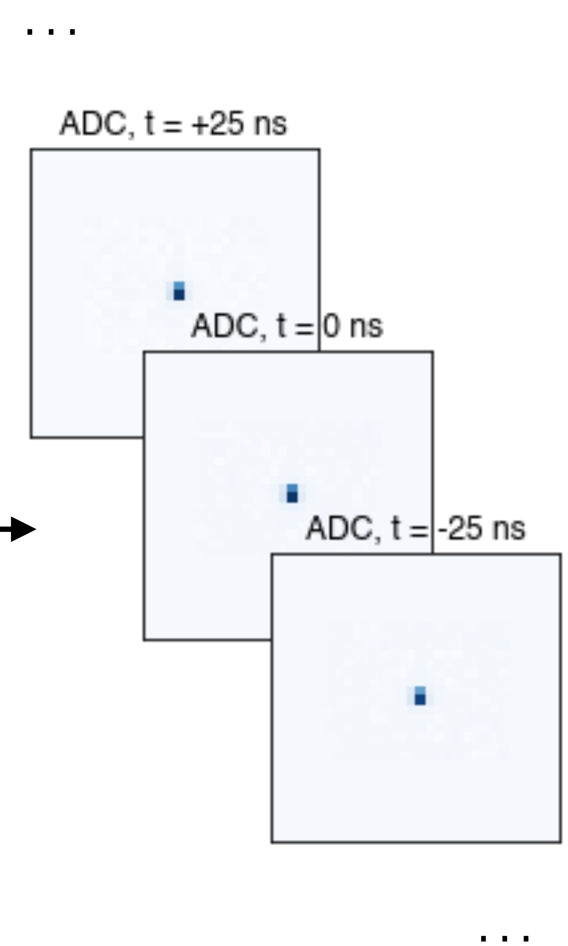
ECAL | Hit Reconstruction

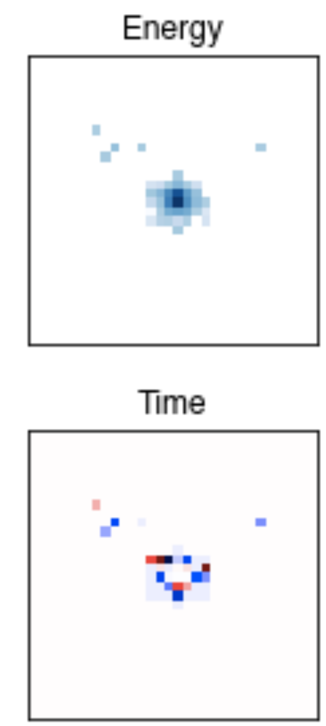
http://iopscience.iop.org/article/10.1088/1742-6596/1085/4/042022

Scale: 1 pixel = 1 crystal









Scintillating Crystal

Signal Pulse

Digitized Hit ("DIGI")

Reconstructed Hit ("RecHit")

Jet ID | q vs g

- **RecNN, Jet ID for quark vs gluon jet**
  - *T. Cheng:* https://arxiv.org/pdf/1711.02633.pdf
  - Jets from QCD dijet gg or qq events, no PU, |η| < 2.5
  - DELPHES detector simulation

ROC AUC | $R_{\epsilon=80\%}$ | $R_{\epsilon=50\%}$	200 GeV	300 GeV	500 GeV	1000 GeV
BDT	0.8164 | 3.1 | 10.5	0.8443 | 3.8 | 16.5	0.8385 | 3.5 | 14.1	0.8421 | 3.6 | 16.1
RecNN without pflow identification	0.8344 | 3.4 | 12.9	0.8390 | 3.6 | 14.4	0.8505 | 3.9 | 16.9	0.8623 | 4.2 | 21.9
RecNN with categorical pflow	**0.8392** | 3.6 | **14.0**	0.8443 | 3.8 | **16.5**	0.8517 | 4.0 | 17.8	**0.8637** | 4.4 | **22.0**
RecNN with pt-weighted charge	0.8340 | 3.5 | 12.8	**0.8453** | 3.9 | 14.5	**0.8525** | 4.0 | **18.6**	0.8616 | 4.3 | 20.4

Table 2 AUCs and background rejection rates for different jet $p_T s$. The baseline BDT and three scenarios concerning particle flow identification are considered. The largest AUCs and $R_{\epsilon=50\%}$s are highlighted in bold.