https://indico.bnl.gov/event/30949/contributions/118693/attachments/67457/115897/JSTGslides_1_13_26_v1.pdf

γ-jet Calibration

Justin Bennett

01.14.26

Overview

Project code

Analysis Sample

- **MC** → Run-28 PYTHIA8 γ+jet (PhotonJet10 + PhotonJet20 – merging steps here), p+p 200 GeV, no pileup
  - **Photon 10 & Photon 20 GeV both have** ~ 3.56 million events **with z < 30 cm**

Discussion Points

- Overview of PPG12 cuts used to establish baseline γ+jet matching needed for in situ γ-jet calibration
- Truth/reco definitions used to calculated $x_{J\gamma}^{truth}$ & $x_{J\gamma}^{reco}$ from sim sample
- Discuss issue in reco-level output that needs further investigation
- Overview of next steps

Analysis Notes

- **Using local build** of PhotonClusterBuilder from version pulled on 12.20.25 to apply direct photon selection
- **JES calibration** from JetCalib (see Hanpu's JSTG slides)
- **Jet $p_T \geqq$ 5 GeV and $\Delta\varphi$(leading γ, leading jet) $\geqq \pi/2$**
- **E_T^{γ} = [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV** → match PPG12

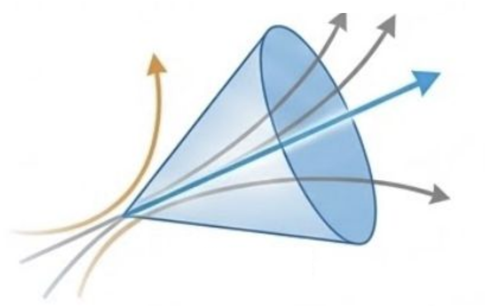
Isolated Photon Sample

sPH-CONF-JET-2025-02

Two ingredients of isolated photon selection:

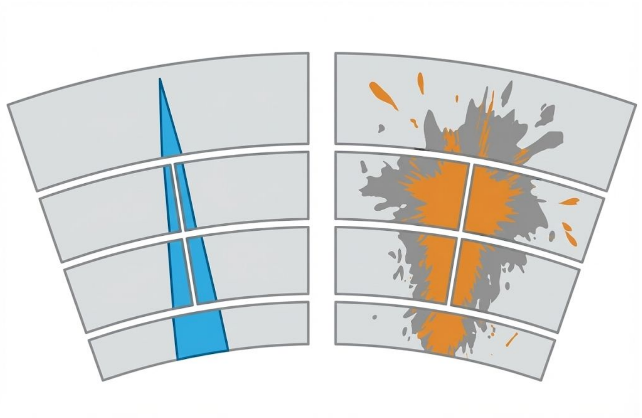
i) Cluster Isolation

$$\Sigma E_T^{\Delta R<0.3} = \sum_{i\in \mathrm{EMCal}\cup \mathrm{HCal}} E_{T,i}$$



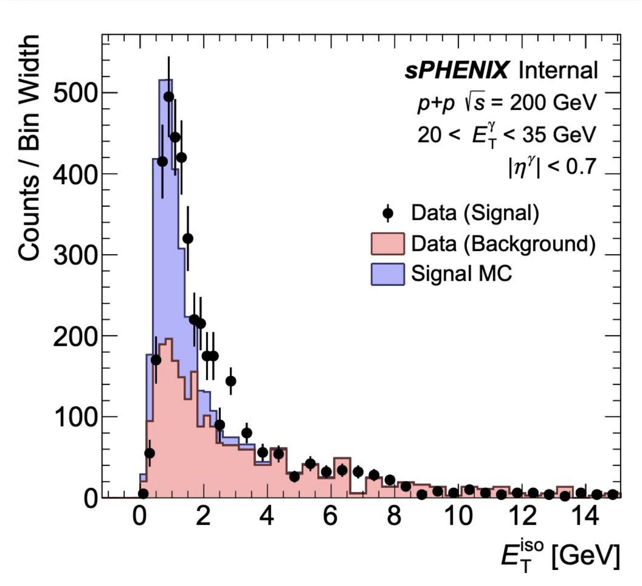
- **Sliding E_T^{iso} cut** tuned in signal MC for ~flat prompt-γ efficiency vs E_T^{γ}
- **Non-isolated sideband** background control region

ii) Shower-Shape (SS) ID



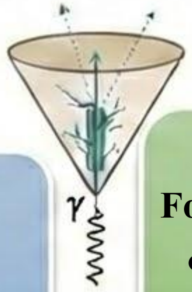
- **SS cuts** → compact EM γ vs broad hadronic/merged π^0
- **Loose** quality *preselection* → **tight** γ-ID working point

Validation (from PPG12)



Isolation E_T^{iso}: tight vs non-tight γ-ID (compared to signal MC)

Isolation Calculation – PhotonClusterBuilder



Seed (photon candidate) Kinematics

- η_{seed} = cluster_eta, φ_{seed} = cluster_phi
- **Seed contribution** → $E_T^{\gamma} = E_{cluster}/\cosh(\eta_{seed})$

Tower Selection (per calo layer)

For given calo layer with cone radius R = 0.3

- **Include** towers if "isGood" and is within cone:

  $\Delta R = [(\eta_{seed} - \eta_i)^2 + (\Delta\varphi)^2]^{1/2} < R(0.3)$

Isolation Energy Contribution

Each included tower contributes → $E_{T,i} = E_i/\cosh(\eta_i)$ if $E_i > E_{min} = 0.070$ GeV

$$E_T^{\text{layer}}(R) = \sum_{i \in \text{layer}} 1(\text{good}_i), 1(\Delta R_i < R), 1(E_i > E_{\min}), \frac{E_i}{\cosh(\eta_i)}$$

PhotonClusterBuilder stores cone sum for each layer (EMCal → subtract E_T^{γ}, HCal → raw sum)

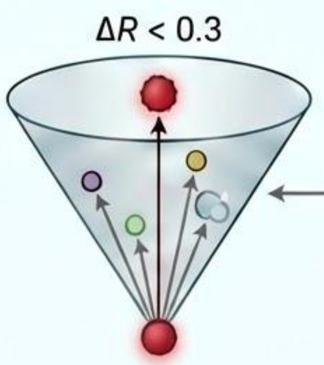
∑ **Analysis code** → sums 3 stored components with default R = 0.3

Truth vs Reconstructed Isolation Energies

Instead of using one fixed iso threshold → derived by PPG12 via fitted function of E_T^{γ}

- Allows prompt isolated photon efficiency to stay roughly *constant* vs E_T^{γ}

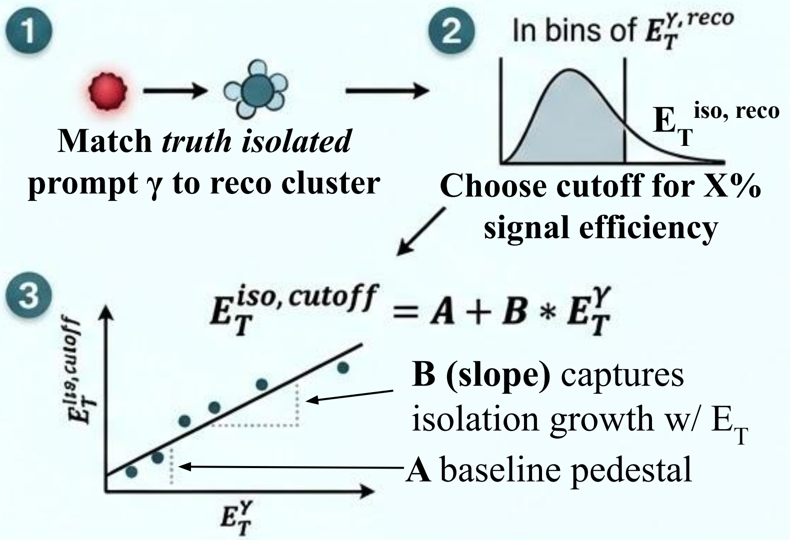
Truth-level Isolation (RecoilJets)



ΣE_T of all *final-state particles* inside **ΔR < 0.3 cone** & subtract photon contribution

Isolated if $E_T^{\text{iso, truth}} < 4.0$ GeV

Reco-level Sliding Isolation (PPG12 IAN)



Sliding Isolation Window

Current pipeline – uses PPG12 studies verbatim for reco isolation and data isolation definition at 90% efficiency cutoff

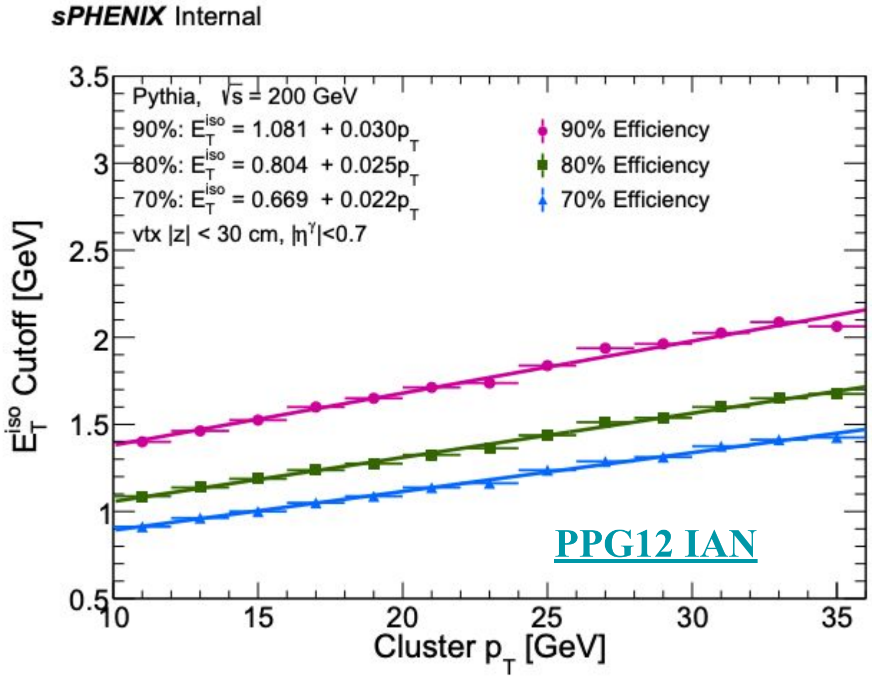
Is isolated:

$$E_T^{iso} < 1.08128 + 0.0299107*E_T^{\gamma,\,reco}$$

Non isolated:

$$E_T^{iso} > \mathbf{1} + 1.08128 + 0.0299107*E_T^{\gamma,\,reco}$$

Everything between the "+1" gap is excluded from ABCD regions



[PPG12 IAN]

γ Selection

Pre-selection (loose **γ-ID quality-gate w/ $|\eta^{\gamma}| < 0.7$):** $\boxed{\dfrac{E_{11}}{E_{33}} < 0.98\,, \quad 0.6 < \mathrm{et1} < 1.0\,, \quad 0.8 < \dfrac{E_{32}}{E_{35}} < 1.0\,, \quad w_{\eta}^{\mathrm{cogX}} < 0.6}$

Tight γ-ID (5 SS variables → compact, EM-like showers)

(1 & 2) Seed-excluded widths (around CoG, exclude the seed tower) → $\boxed{0 < w_a^{\mathrm{cogX}} < 0.15 + 0.006\,E_T^{\gamma} \quad \text{for } a \in \{\eta, \phi\}}$

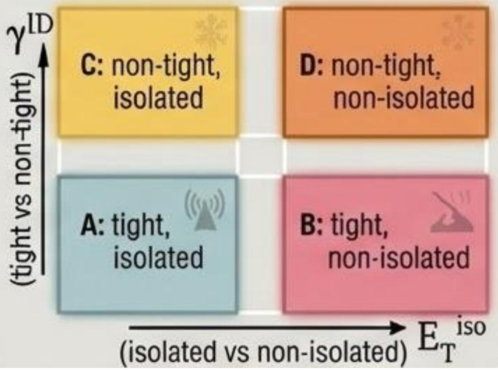
(3) Core-to-cluster compactness (odd×odd window sums) → $\boxed{0.4 < \dfrac{E_{11}}{E_{33}} < 0.98}$

(4) 2×2 core energy fraction (et1) → $\boxed{0.9 < \mathrm{et1} < 1.0}$ $\boxed{\mathrm{et1} = \dfrac{E_1 + E_2 + E_3 + E_4}{E_{\mathrm{tot}}}}$

(5) Elongation veto (strip ratio) → $\boxed{0.92 < \dfrac{E_{32}}{E_{35}} < 1.0}$

ABCD Regions

- **Purpose:** estimate the signal purity of your tight & isolated photon sample
- **Axes** → γ-ID (tight vs non-tight, w/ 5 SS variables), and isolated vs non-isolated
  - **Non-tight** → fails any two of five tight SS requirements
  - **Non-isolated** → $E_T^{\mathrm{iso}} > 1 + 1.08128 + 0.0299107 \cdot E_T^{\mathrm{reco}}$ (+1 gap > iso def)



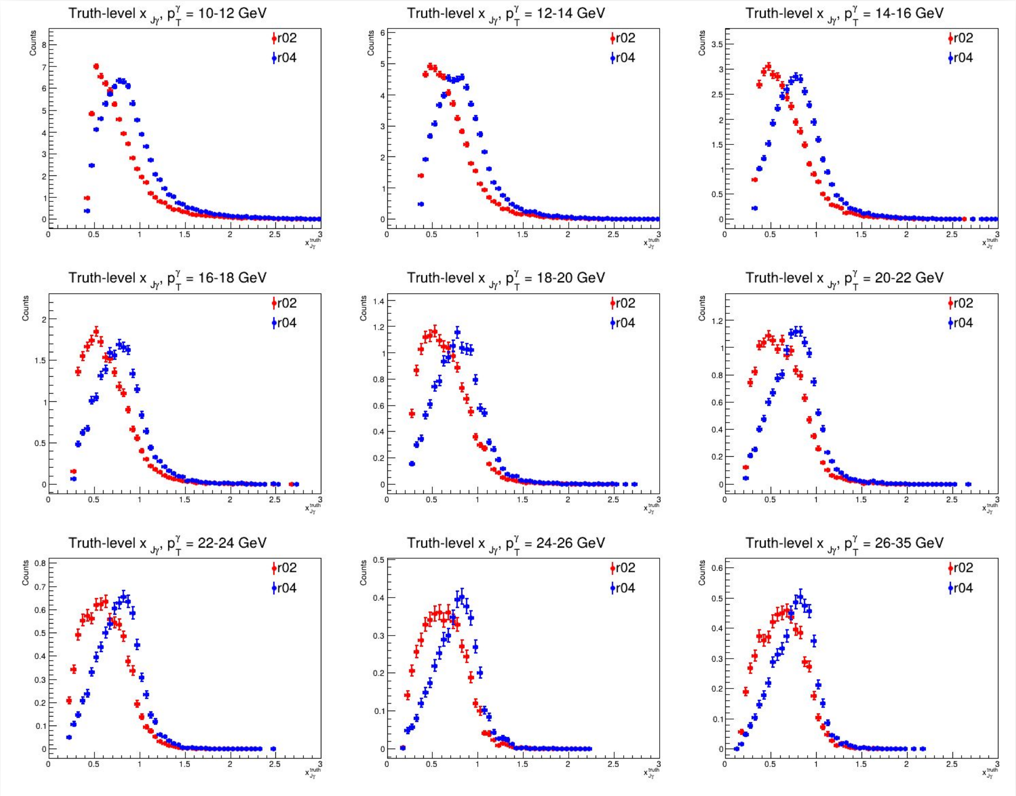
Truth-Matched Photon+Jet — $x_{J\gamma}^{truth} = p_T^{truth, jet}/p_T^{\gamma, truth}$

Truth photon — final state γ (PID = 22) with $|\eta_\gamma^{truth}| < 0.7$ that is prompt (direct or frag) & truth-isolated ($E_T^{iso, truth} \leq 4$ GeV)

Iso def — sum of E_T of all final-state particles $\Delta R < 0.3$ around the γ (excluding γ)

Truth $\leftrightarrow$ reco association — keep only truth γ's corresponding to the event's leading reco γ used for the recoil-jet selection

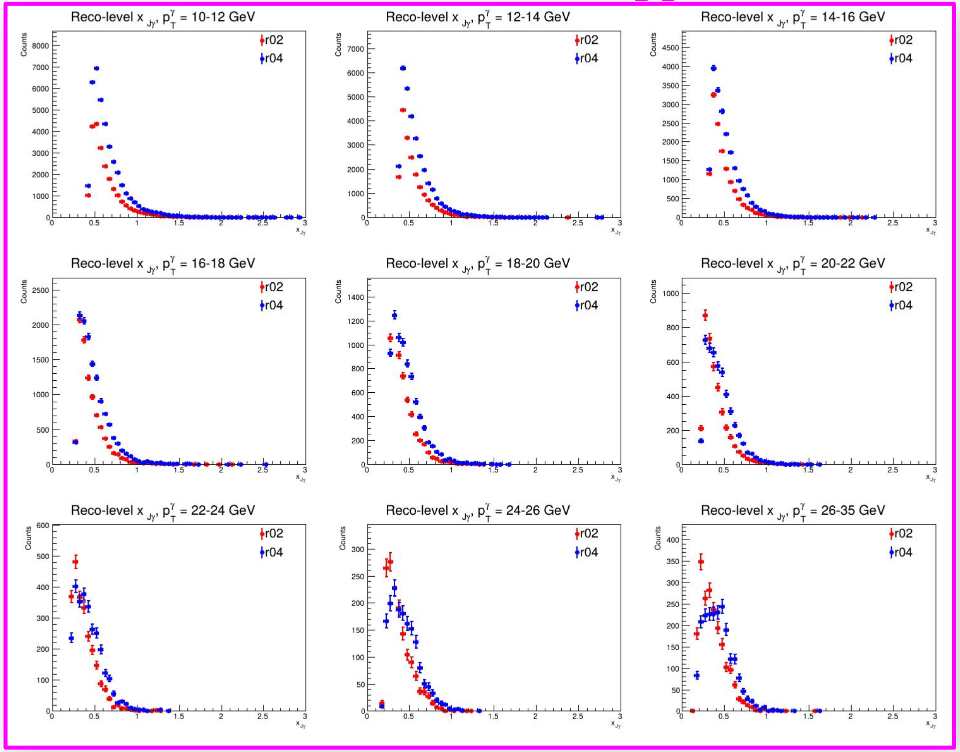
CaloRawClusterEval used to find best-matched truth photon candidate for the cluster *within $\Delta R<0.05$* of leading reco



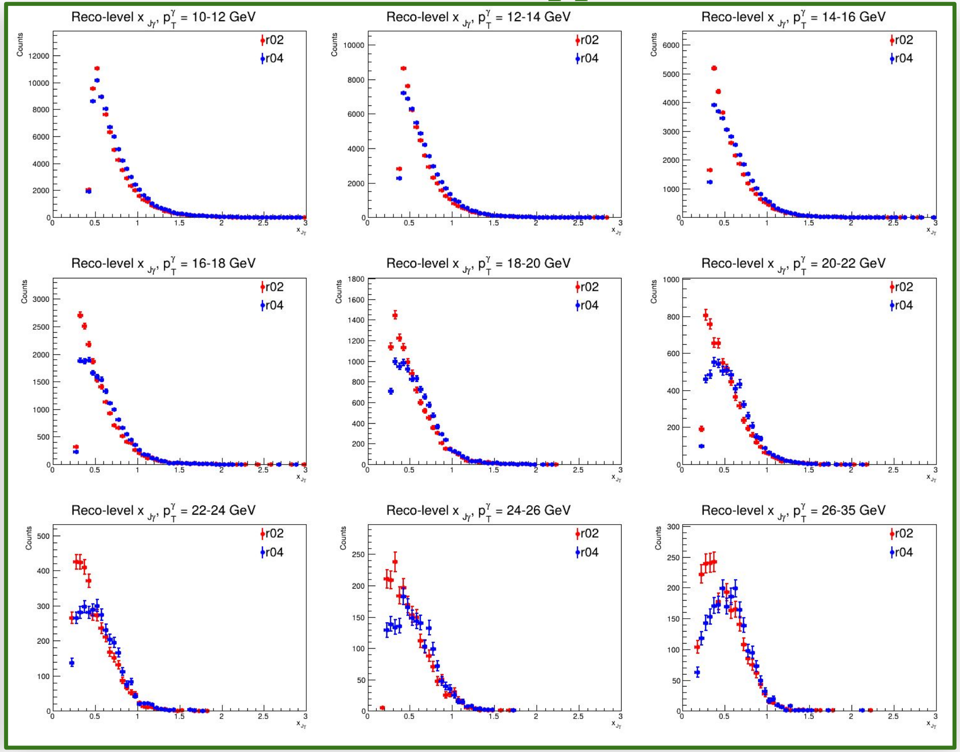
★ **Truth information filled iff** there is nearby reco photon + reco jet ($\Delta R < 0.2$) within matching criteria (more info)

$x_{J\gamma}^{reco}$ for Photon+Jet 10 GeV Only → before/after JES Calib

Before JES Calib Applied



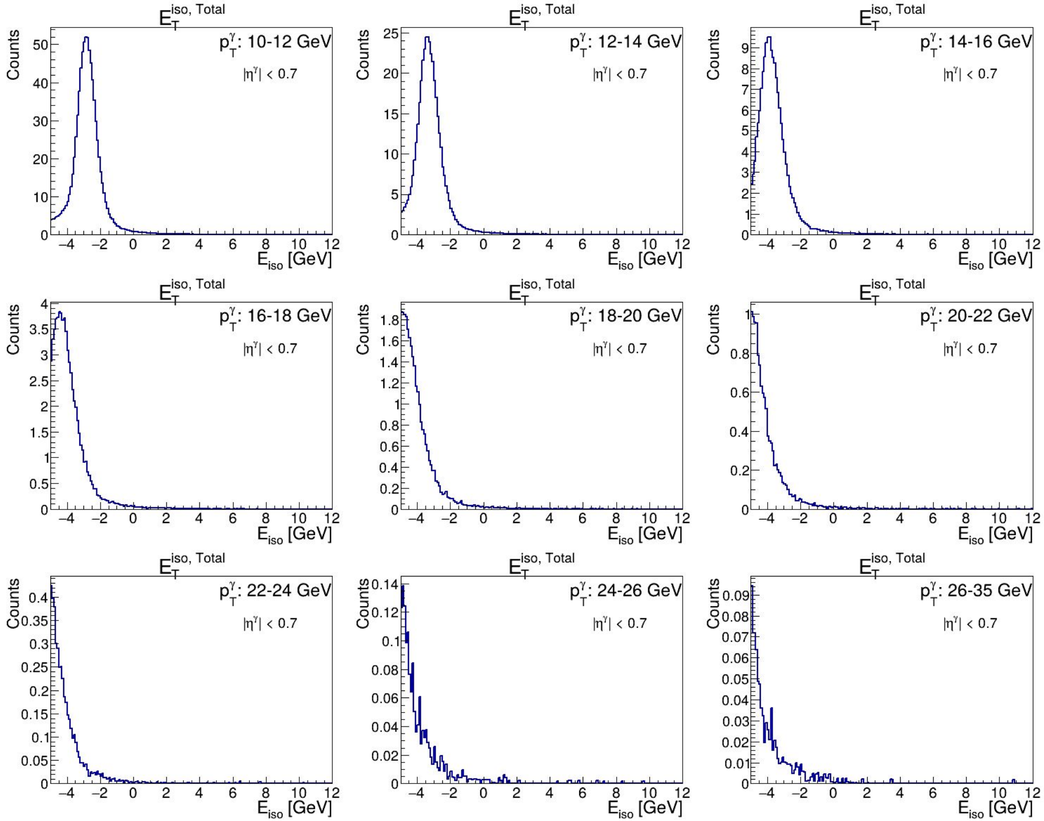
After JES Calib Applied



- Reco distributions from pure PPG12 tight+isolated cuts does not give the same distribution as truth
- Seems post JES calibration applied it scales up r02 jets to be closer to r04, but the width and peak position aren’t really affected much – differences require further investigation

Reco Isolation QA – Negative Peak in E_T^{iso}

Isolation energies peaking in negative GeV range – something is happening in calculation that causes cluster E_T subtraction to cause EMCal isolation go negative



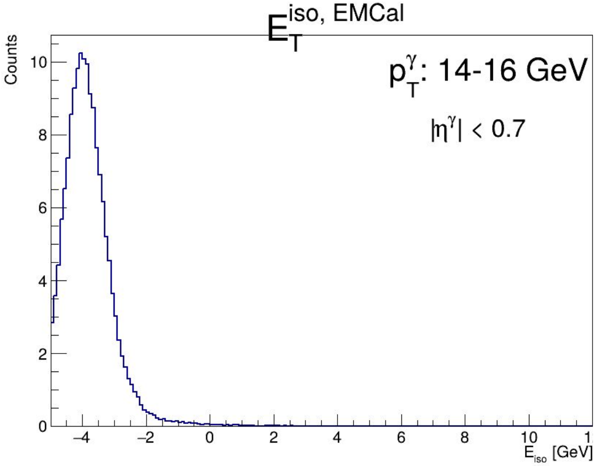
Reco Isolation QA – Negative Peak in E_T^{iso}

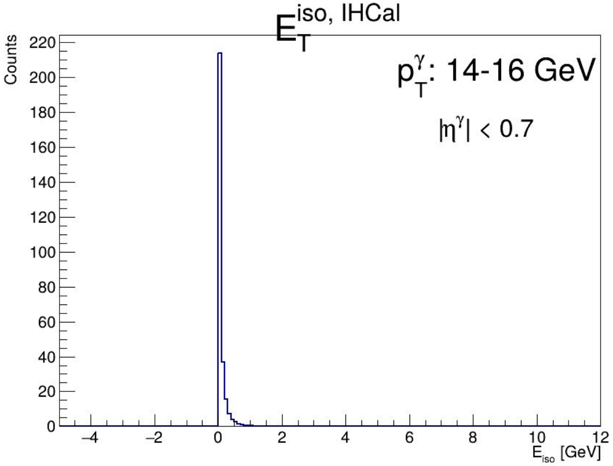
Isolation energies peaking in negative GeV range – something is happening in calculation that causes cluster E_T subtraction to make EMCal isolation go negative

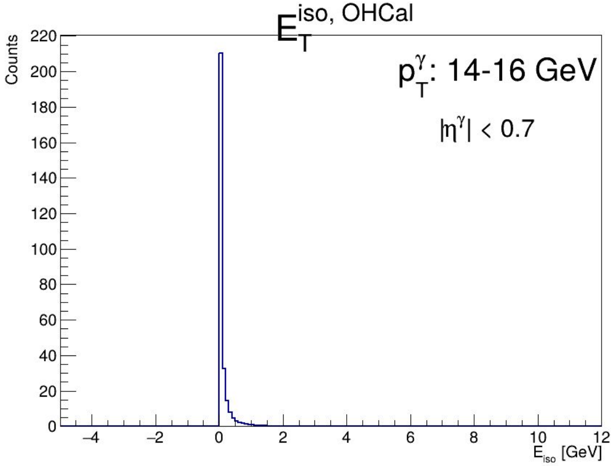
- By construction photon cluster builder feeds analysis macro:

$$E_T^{\text{iso, EMCal}} + E_T^{\text{iso, IHCal}} + E_T^{\text{iso, OHCal}} = (\Sigma E_T^{\text{CEMC}} - E_T^{\gamma}) + \Sigma E_T^{\text{IHCal}} + \Sigma E_T^{\text{OHCal}}$$

- Something is causing reco isolation for EMCal to accumulate counts at -4 GeV, so cut on E_T^{iso} is **not** going to reject fragmentation photons as intended







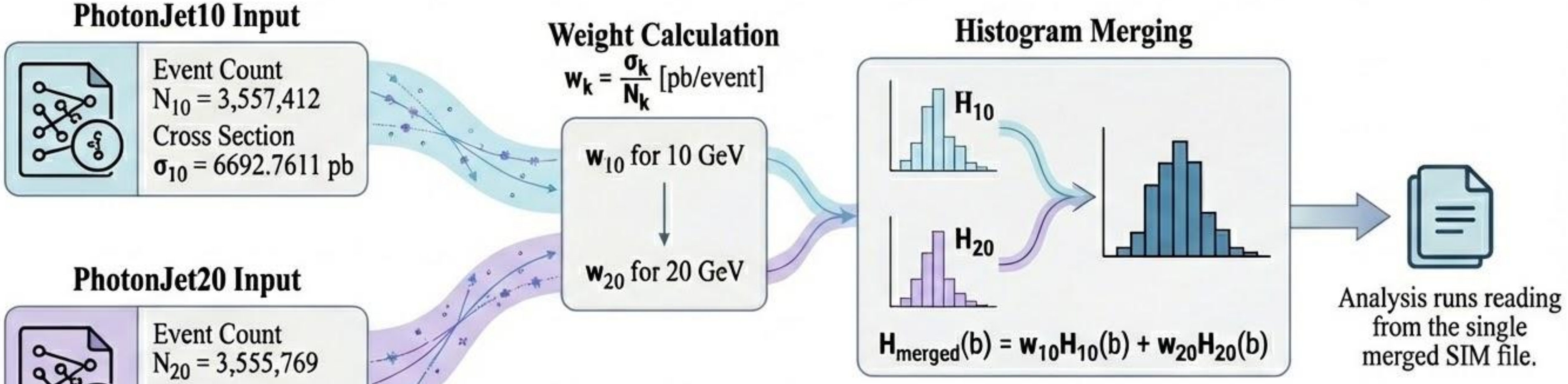
Next Steps

1. **Fix issue with isolation energies** → diagnose if this is cause of variation between truth and reco $x_{J\gamma}$ distributions
   a. Check if occurs when using ClusterIso package instead of PhotonClusterBuilder – from looking at this last year in pp it seemed to not have this issue for non UE-subtracted isolation
2. **With that fixed –** move onto pushing in-situ calibration through using PPG12 baseline cuts to calibrate reco → data $x_{J\gamma}$ distribution
   a. Fit distribution per p_T bin with iterative Gaussian approach to converge on mean x_J
   b. Take ratio of reco mean to data mean and plot/fit as function of p_T
   c. Apply calibration factors to data and funnel back through pipeline so ratio becomes ~1


BACKUP

Weighted Merging PhotonJet10 + PhotonJet20 GeV Sim Samples

Goal → Combine two MC slices into one physics-normalized SIM file.



Result → Merged histograms are in cross-section units (pb/bin), not raw counts.

$$L_k = \frac{N_k}{\sigma_k}\ (\mathrm{pb}^{-1})$$

Each slice **k** simulated an "integrated luminosity".

$$L_{10} = N_{10}/\sigma_{10}$$

$$L_{20} = N_{20}/\sigma_{20}$$

Total Effective Luminosity (z < 30 cm):

$$L_{eff} = L_{10} + L_{20} = 34140\ \mathrm{pb}^{-1}$$

Exact Cuts before Filling Truth x_J

Force (i) truth-isolated signal photon**, and (ii)** reco photon and reco jet to be consistent with that truth signal topology

1. **Truth photon (sets x-bin)**

   $\gamma^{\mathbf{truth}}$**:** $|\eta_\gamma^{truth}| < 0.7$, $p_T^{\gamma,\,truth} \in$ analysis bin, PID = 22, HepMC ancestry is direct/frag, $E_T^{iso,\,truth} < 4$ GeV

2. **Reco photon gate (defines event sample) –** choose event-leading γ^{reco} that passes:

   a. preselection shower-shape cuts

   b. tight $\wedge$ isolated (reco)

3. **Truth↔Reco photon association**

   $\gamma^{\mathbf{truth}} \leftrightarrow \gamma^{\mathbf{reco}}$**:** $\Delta R(\gamma^{truth}, \gamma^{reco}) < 0.05$ and best matched cluster via CaloRawClusterEval

4. **Reco recoil-jet1 (per radius, R) –** require a leading recoil jet, j^{reco}

   $p_T^{j,\,reco} \geq 5$ GeV, $|\eta^{j,\,reco}| < (1.1 - R)$, $|\Delta\phi(\gamma^{reco}, j^{reco})| \geq \pi/2$

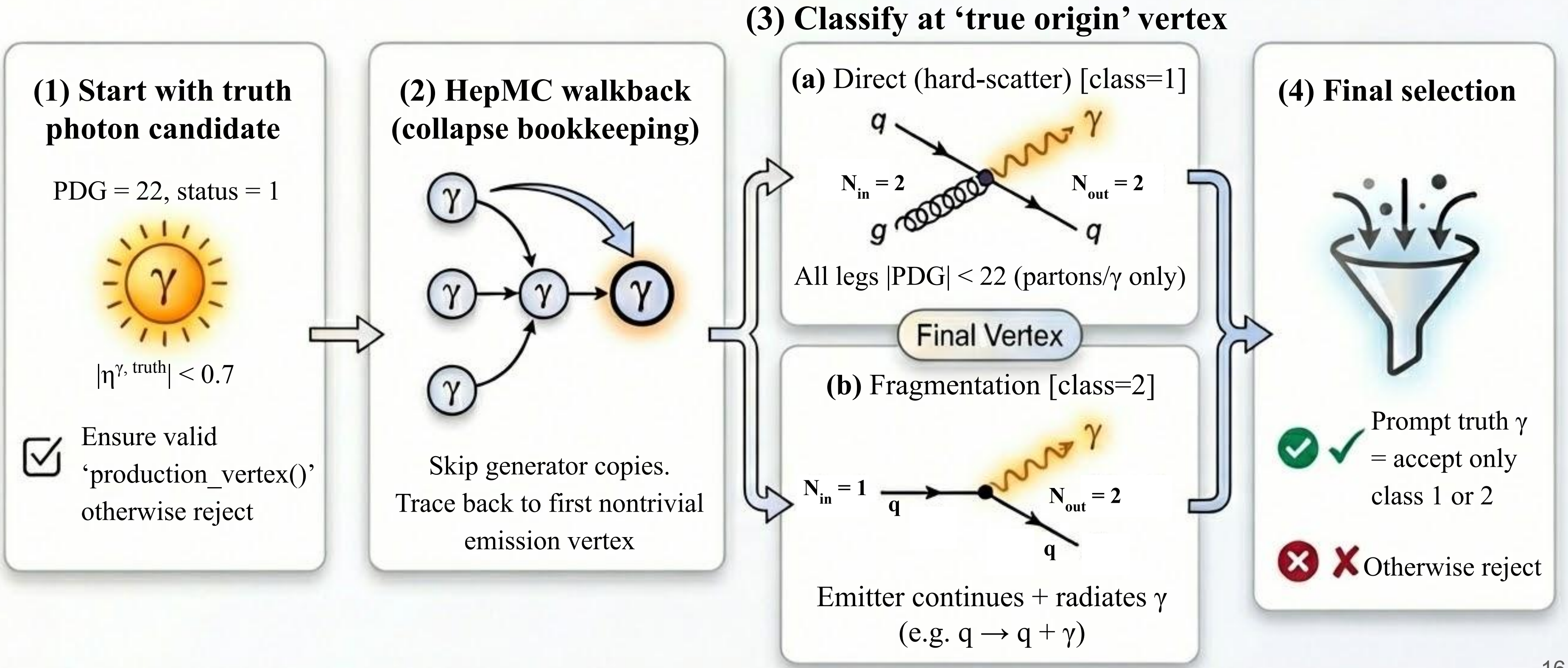
5. **Truth recoil-jet (same R) –** require leading recoil jet, j^{truth}

   $p_T^{j,\,truth} \geq 5$ GeV, $|\eta^{j,\,truth}| < (1.1 - R)$, $|\Delta\phi(\gamma^{truth}, j^{truth})| \geq \pi/2$

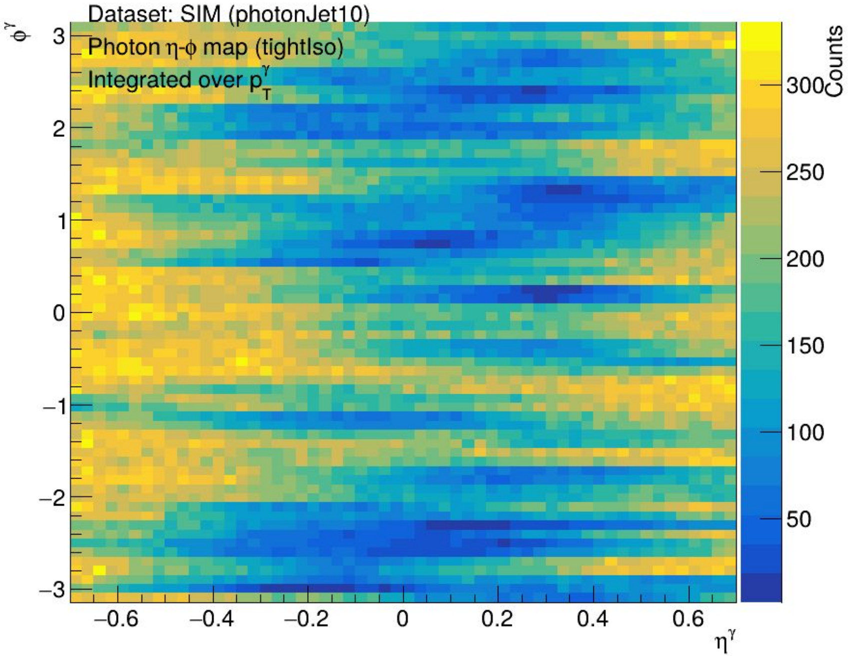
6. **Reco↔Truth jet match (final gate)** $\rightarrow \Delta R(j^{reco}, j^{truth}) \leq 0.2$

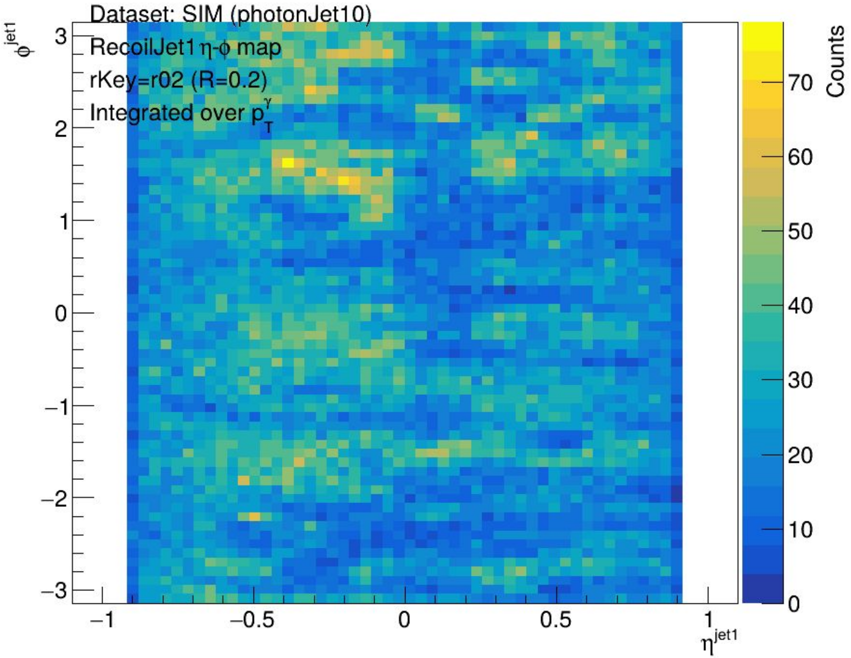
Then fill $x_J^{truth}(p_T^{\gamma,\,truth})$

Tracing HepMC Ancestry – Classifying Direct vs Fragmentation

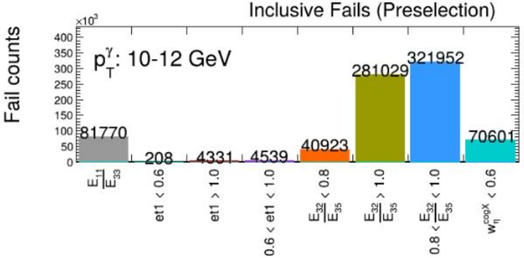


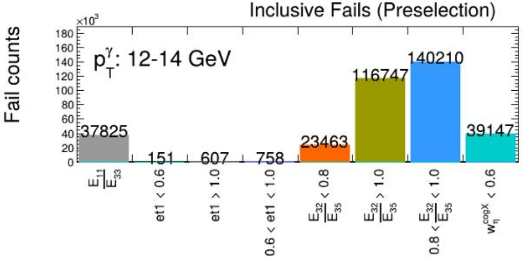
Leading Recoil QA – eta/phi maps for paired photon+jet

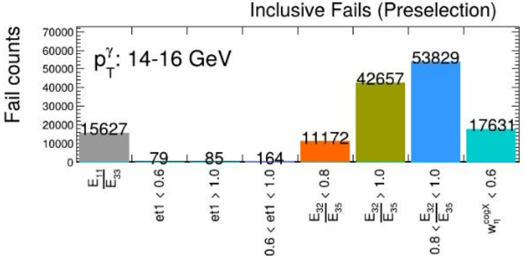


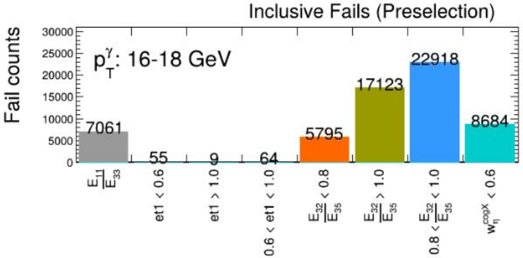


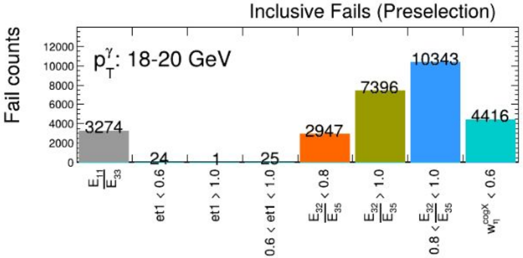
Counts of Preselection Failures by Cut – Photon Jet 10 Sim

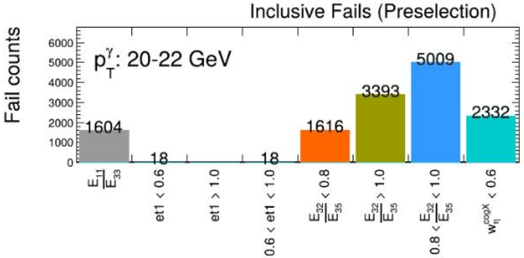


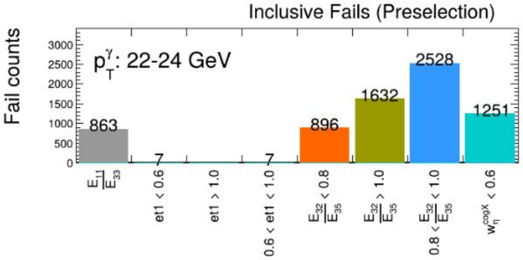


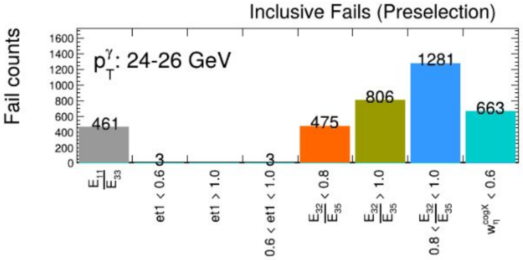


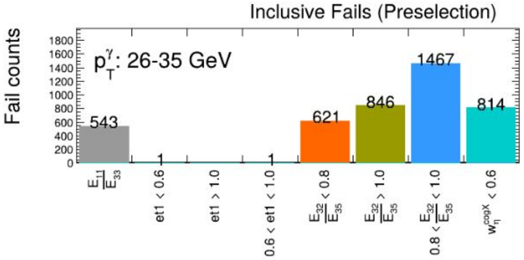




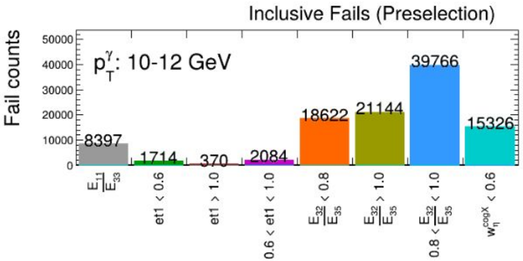


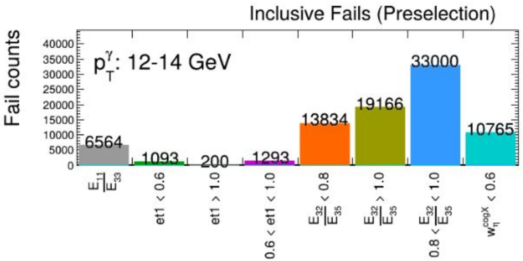


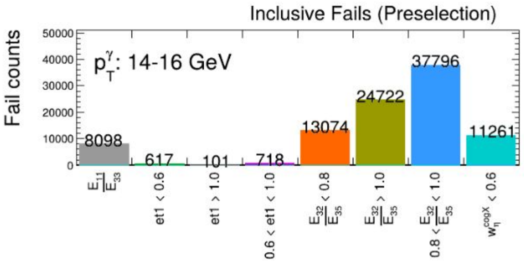


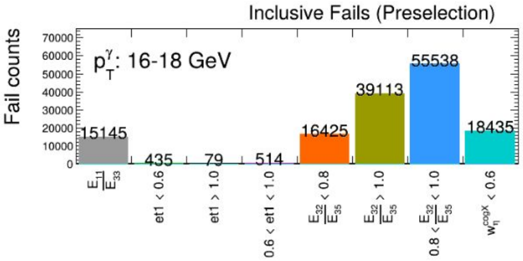


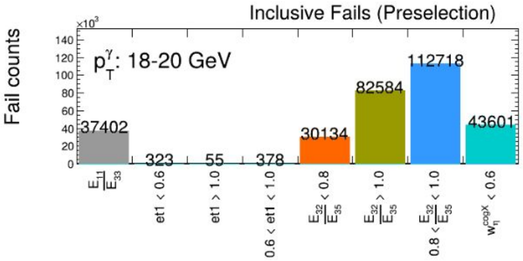
Counts of Preselection Failures by Cut – Photon Jet 20 Sim

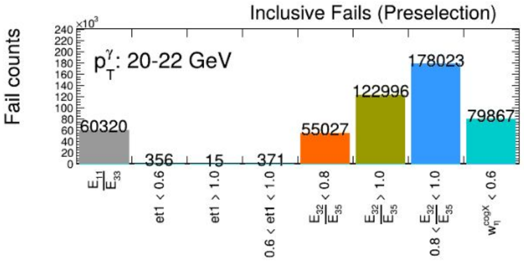


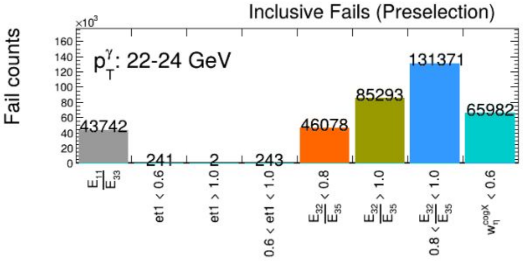


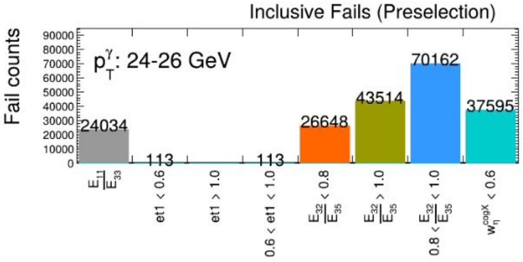


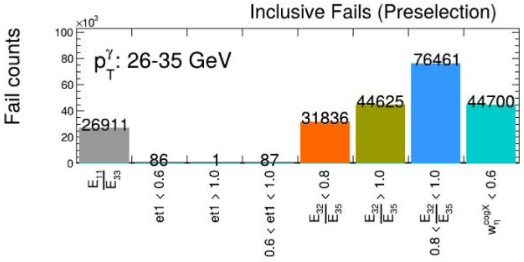




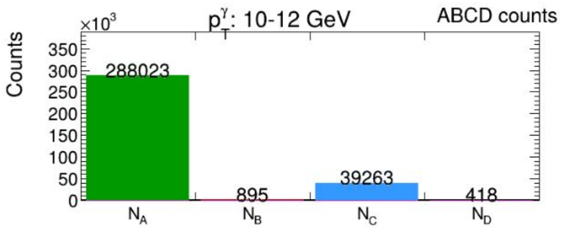


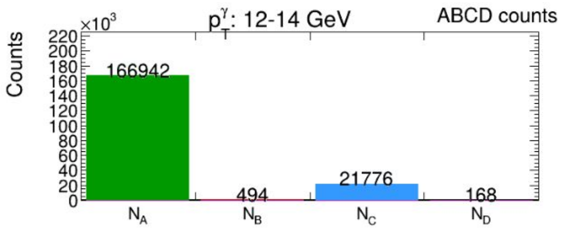


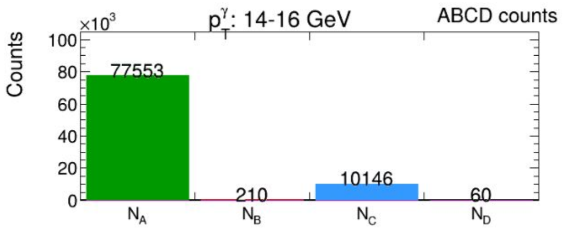


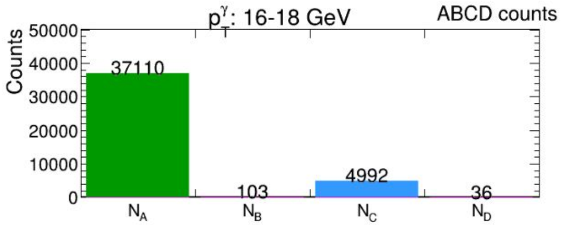


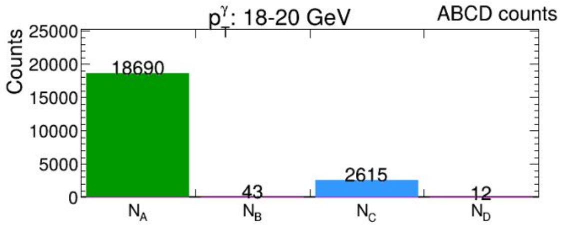
ABCD Bin Counts – Photon + Jet 10 Sim

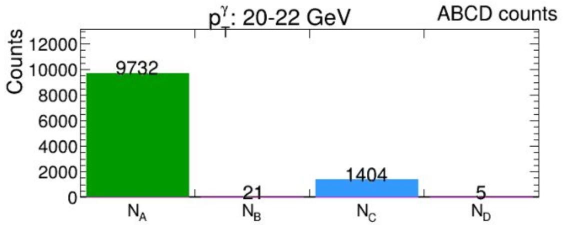


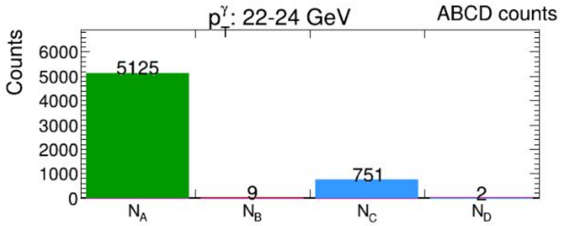


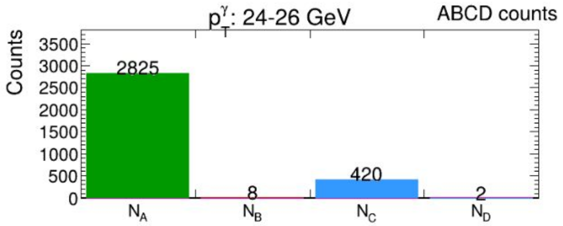


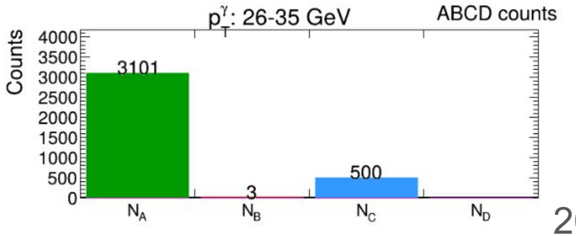




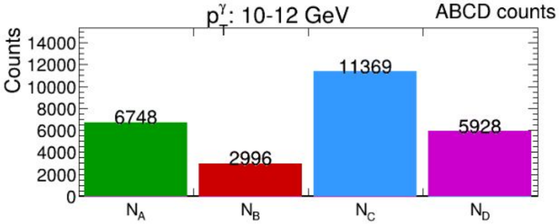


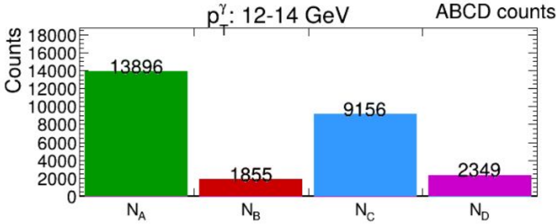


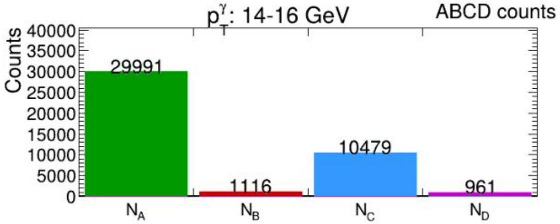


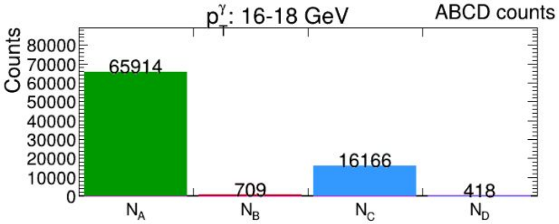


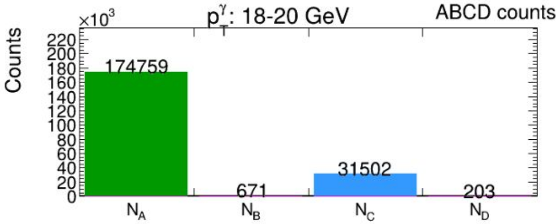
ABCD Bin Counts – Photon + Jet 20 Sim

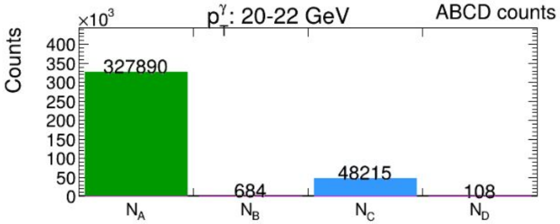


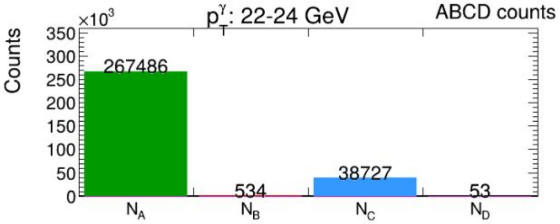


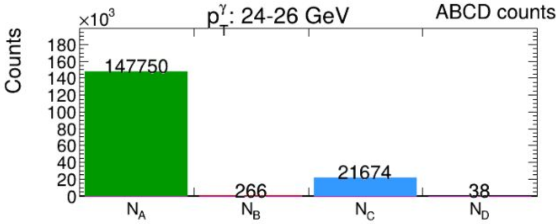


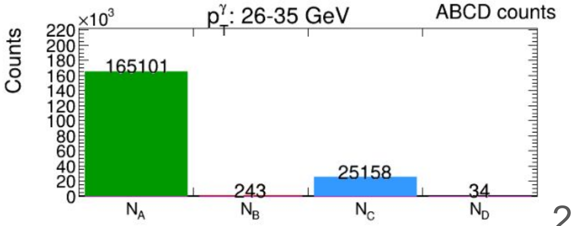












2D response matrix (1)

Why response matrix? → relevant truth isn't just $x_{J\gamma}$ but is build from two measured energies (p_T^{jet}/p_T^{γ}) and detector smearing acts on both

- Its like computing x axis ruler and y axis ruler before computing a derived ratio – if you only calibrate the ratio directly you can hide correlated distortions

EG → set bins E_T^{γ}: [15, 20], [20, 25], [25, 30] s.t. $N_\gamma = 3$ and p_T^{jet}: [10, 15], [15, 20], [20, 25] s.t. $N_J = 3$

($N_{2D} = N_\gamma \times N_J = 9$ total 2D-bins)

- Fill 3 histograms: $H_{truth}(E_T^{\gamma,\, true}, p_T^{jet,\, true})$, $H_{reco}(E_T^{\gamma,\, reco}, p_T^{jet,\, reco})$, & N(reco cell, truth cell) w 2D bin = cell
    - $H_{truth}(E_T^{\gamma,\, true}, p_T^{jet,\, true})$ → counts how many truth photon-truth jet pairs fall into each truth cell
    - $H_{reco}(E_T^{\gamma,\, reco}, p_T^{jet,\, reco})$ → same for reco pairs/cells
    - N(reco cell, truth cell) → response cross tab → response matrix in counts form
        - For the same event you record which truth cell it belonged to and which reco cell it landed in and increment that entry by 1

To plot a 2D matrix → flatten each 2D cell into a single integer label

- γ → {0, 1, 2} index E_T^{γ} bin (15-20, 20-25, 25-30) & j → {0, 1, 2} index p_T^{jet} bin (10-15, 15-20, 20-25)
    - Define flattened ID cell label → cellID = $j * N_\gamma + \gamma$ s.t. 9 truth cells are cellID 0…8 and 9 reco cells also 0…8
    - Plotted response matrix will be 9x9 heatmap of x-axis → cellID_{reco} and y-axis → cellID_{truth} with z(color) = counts

2D response matrix (2)

Say event 1 → loop over truth particles with PID = 22, in fiducial range and passes truth isolation

- And loop over jets in p_T and fiducial range and pair with $\Delta\varphi(\gamma, \text{jet})_{truth} > 7\pi/8$ and find for a pair:
  - $E_T^{\gamma, true}$ = 18 GeV → γ_{true} = 0 (15-20), $p_T^{jet, true}$ = 12 GeV → j_{true} = 0 (10-15)
    - $\text{cellID}_{truth} = 0*3 + 0 = 0$ → H_{truth}(truthID = 0) += 1
- Still in the same event → loop over reco clusters and find leading E_T cluster passing SS, iso, etc and matches to truth photon
  - Loop over reco jets passing p_T and fiducial cut and match to truth jet and is back to back and get:
    - $E_T^{\gamma, reco}$ = 28 GeV → γ_{reco} = 2 (25-30), $p_T^{jet, reco}$ = 22 GeV → j_{reco} = 2 (20-25)
    - $\text{cellID}_{reco} = 2*3 + 2 = 8$ → fill H_{reco}(recoID = 8) += 1
- Response cross-tab N(recoID, truthID) += 1

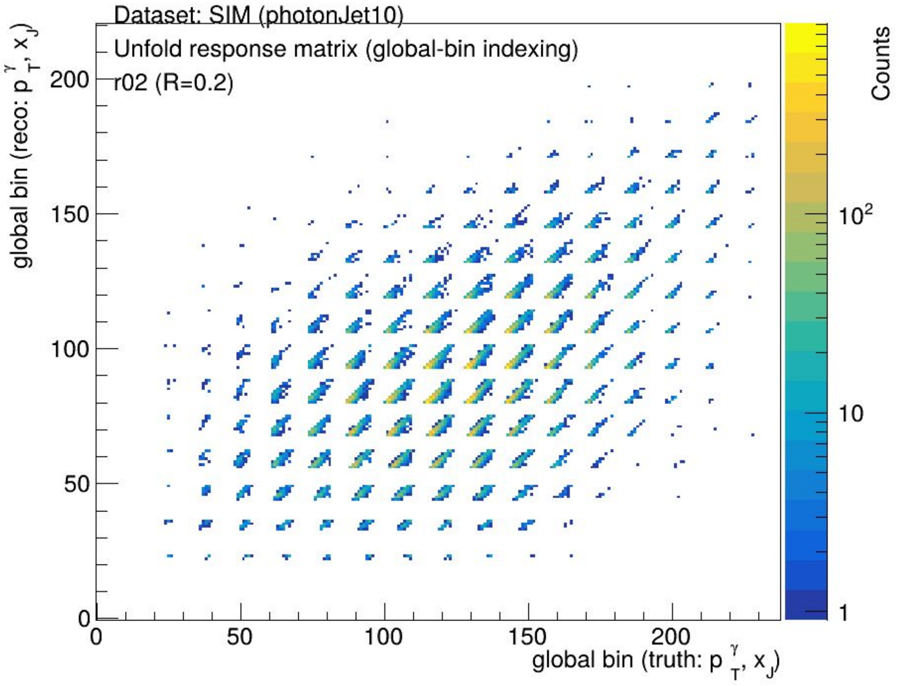
Event 2:

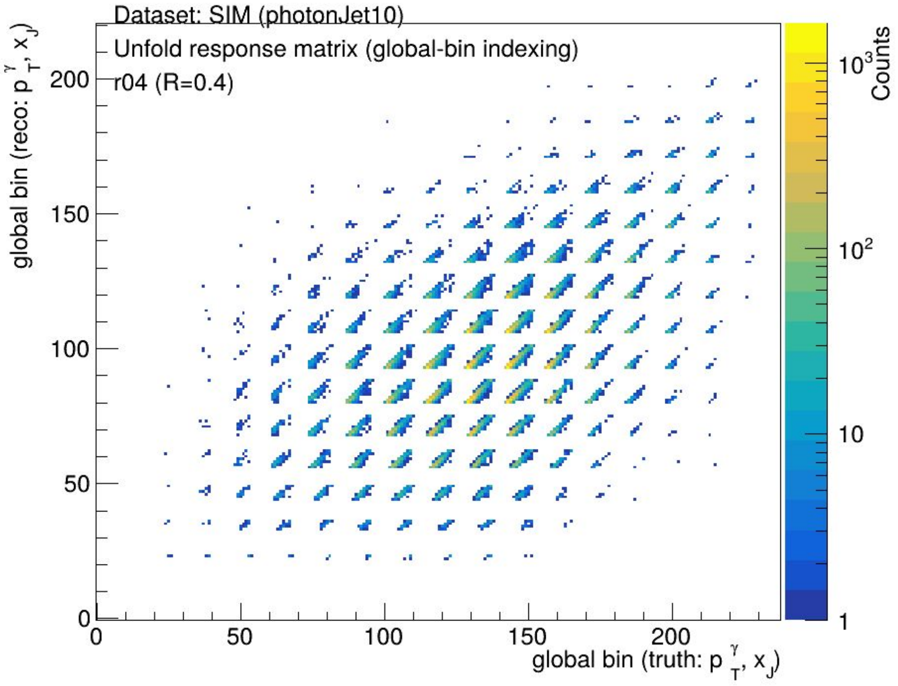
- truth pair → $E_T^{\gamma, true}$ = 23 GeV → γ_{true} = 1 (20-25), $p_T^{jet, true}$ = 17 GeV → j_{true} = 1 (15-20), $\text{cellID}_{true} = 1*3 + 1 = 4$
  - Fill $H_{truth}(4)$ += 1
- reco pair → $E_T^{\gamma, reco}$ = 21 GeV → γ_{reco} = 1, $p_T^{jet, reco}$ = 14 GeV → j_{reco} = 0, $\text{cellID}_{reco} = 0*3 + 1 = 1$
  - Fill $H_{reco}(1)$ +=1 and response N(1, 4) += 1

Event 3 → truth pair exists again in TruthID = 0 but suppose reco fails matching and nothing added to H_{reco} or N(reco, truth), incremen et miss counter for truthID = 0 and after the 3 events we have:

- $H_{truth}(0) = 2$, $H_{truth}(4) = 1$ & $H_{reco}(8) = 1$, $H_{reco}(1) = 1$
- N(8, 0) = 1, N(1, 4) = 1, and one miss for truthID = 0

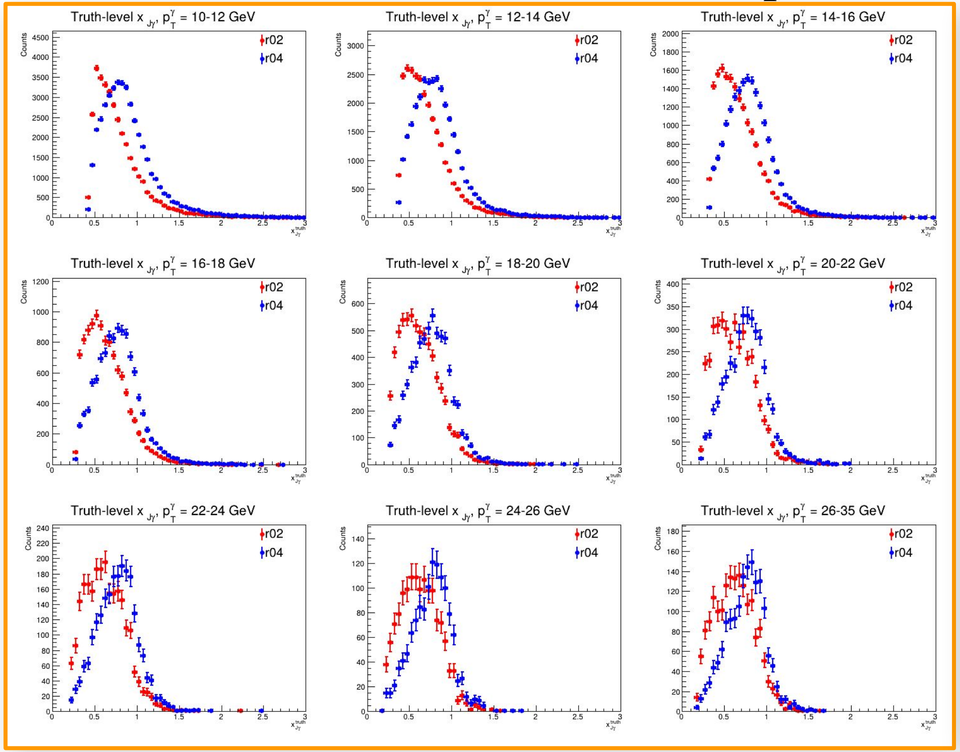
Unfolding


Dataset: SIM (photonJet10)
Unfold response matrix (global-bin indexing)
r02 (R=0.2)
global bin (reco: p_T^{γ}, x_J)
global bin (truth: p_T^{γ}, x_J)
Counts
0
50
100
150
200
1
10
10^2

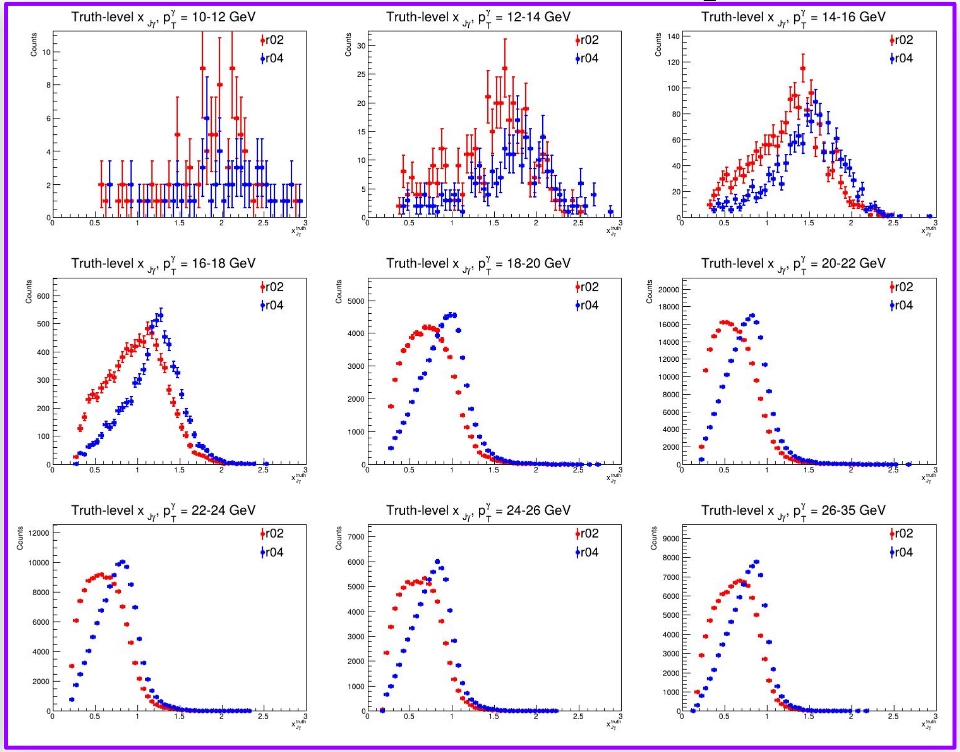

Dataset: SIM (photonJet10)
Unfold response matrix (global-bin indexing)
r04 (R=0.4)
global bin (reco: p_T^{γ}, x_J)
global bin (truth: p_T^{γ}, x_J)
Counts
0
50
100
150
200
1
10
10^2
10^3

Truth Matched Photon+Jet 10 GeV vs 20 GeV $x_{J\gamma}^{truth}$

Photon + Jet 10 GeV Sample



Photon + Jet 20 GeV Sample



- Previous slide was for combined weighted sample of photon+jet 10 & 20 GeV (weighting procedure in backup)

Shower-Shape (SS) Criteria – Window Sums

A) Window sums: let E_{ij} → energy in tower index (i, j) with (i ∈ [0,95] in η, j ∈ [0,255] in φ)

- **Choose a center (i_0, j_0)** (usually the cluster seed or CoG) – for **odd** *m, n* define half-sizes: $h_\eta = \frac{m-1}{2}, \quad h_\phi = \frac{n-1}{2}$
  - **Then the window sum** (symmetric patch) is: $\boxed{E_{m\times n}(i_0, j_0) = \sum_{i=i_0-h_\eta}^{i_0+h_\eta} \sum_{j=j_0-h_\phi}^{j_0+h_\phi} E_{ij}}$
    - **h_η= (m−1)/2** → how many tower steps you extend up and down in η from the center index i_0
    - **h_ϕ = (n-1)/2** → how many steps you extend left and right in ϕ from the center index j_0
  - **ϕ-wrapping:** if *j* steps outside [0, 255], wrap modulo 256 so sums cross the 0/2π boundary cleanly
- **EGS → 1x1:** $h_\eta = h_\phi = 0$ → just the center tower, $E_{11} = E_{i0,\, j0}$, **3x3:** $h_\eta = h_\phi = 1$ → indices $[i_0 - 1, i_0 + 1] \times [j_0 - 1, j_0 + 1]$

Strip windows (when not odd): variables that *probe elo* $E_{32} = \sum_{|i|\le 1} \sum_{j\in\{0,s\}} E_{ij}$, $E_{35} = \sum_{|i|\le 1,\, |j|\le 2} E_{ij}$, $\frac{E_{32}}{E_{35}} = \frac{\sum_{|i|\le 1} \sum_{j\in\{0,s\}} E_{ij}}{\sum_{|i|\le 1,\, |j|\le 2} E_{ij}}$

- **Rectangular strips** instead of odd x odd squares:
  - where s = ±1 picks the CoG side in ϕ

Shower-Shape (SS) Criteria – et1 and 2nd Moments

B) 2x2 core energy fraction (et1) – let a clusters *total* energy be $\mathbf{E_{tot}} = \sum_{\mathbf{i} \in \mathbf{cluster}} \mathbf{E_i}$

- **Find the 4 EMCal towers** whose centers are *closest* to the cluster CoG – these **form a 2x2 core with energies $\mathbf{E_{1\rightarrow4}}$**
  - **et1** → fraction of the cluster's E_{total} carried by the four towers immediately surrounding the cluster CoG
  - $\boxed{\mathrm{et1} = \dfrac{E_1 + E_2 + E_3 + E_4}{E_{\mathrm{tot}}}}$ → tight photons are expected to be compact, so et1 **should be ~1 for signal**

C) Seed-excluded, energy-weighted second moments (cluster "widths") – let each EMCal tower in a cluster have coordinates (η_i, ϕ_i) and energy E_i

- **Let (η_{COG}, ϕ_{COG})** be a cluster center-of-gravity used by the photon ID
- **For either axis a ∈ {η, ϕ}**, define the seed-excluded second moment →

$$w_a^{\mathrm{cogX}} \equiv \frac{\sum_{i \neq \mathrm{CoG}} E_i \left(a_i - a_{\mathrm{COG}}\right)^2}{\sum_{i \neq \mathrm{CoG}} E_i}$$

- **Measures energy-weighted spread** of the cluster around the CoG
  - *Narrow*, EM-like clusters give small $w_{\eta,\phi}$, *broader*, multi-tower (hadronic/merged π^0) clusters push them up
- **Seed tower dominates the energy sum** → removing it makes width sensitive to the shape of the surrounding shower

Combined Photon+Jet 10/20 GeV MC