γ-jet Calibration

Justin Bennett

01.15.26

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

- Cut consolidation to put on mattermost for others to use
- Updated isolation distributions
- Updated x_J distributions

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

γ+jet definition – Reco, Truth, and Reco-Truth Matching Criteria

Reco-level (data/sim)

→ **Pick event-leading photon that is Tight $\wedge$ Isolated** – (details on iso criteria *here* & on SS criteria *here*)

- Reco-jet matching – per radius R $\in$ {0.2, 0.4} → $p_T^{jet, reco} \geq 5$ GeV, $|\eta_{jet}^{reco}| < 1.1$ - R, $|\Delta\varphi(\gamma_{reco}, jet_{reco})| \geq \pi/2$
- E_T^{γ} (PPG12) = [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV

Truth-level (sim)

→ **Truth signal photon definition (γ_{truth})** – for **all PDG=22** truth particles

- **Fiducial cut** → $|\eta_{\gamma}^{truth}| < 0.7$ and and fill in same p_T^{γ} bins as reco
- **Prompt signal type requirement** → direct|fragmentation: *trace HepMC ancestry* to parent vertices corresponding to direct/fragmentation production (details)
- **Truth isolation** → $E_T^{iso, truth} < 4$ **GeV** (details)

→ **Truth recoil jet combination: $p_T^{jet, truth} \geq 5$ GeV, $|\eta_{jet}^{truth}| < 1.1$ - R, $|\Delta\varphi(\gamma_{truth}, jet_{truth}) \geq \pi/2$**

Truth-Reco matching (sim)

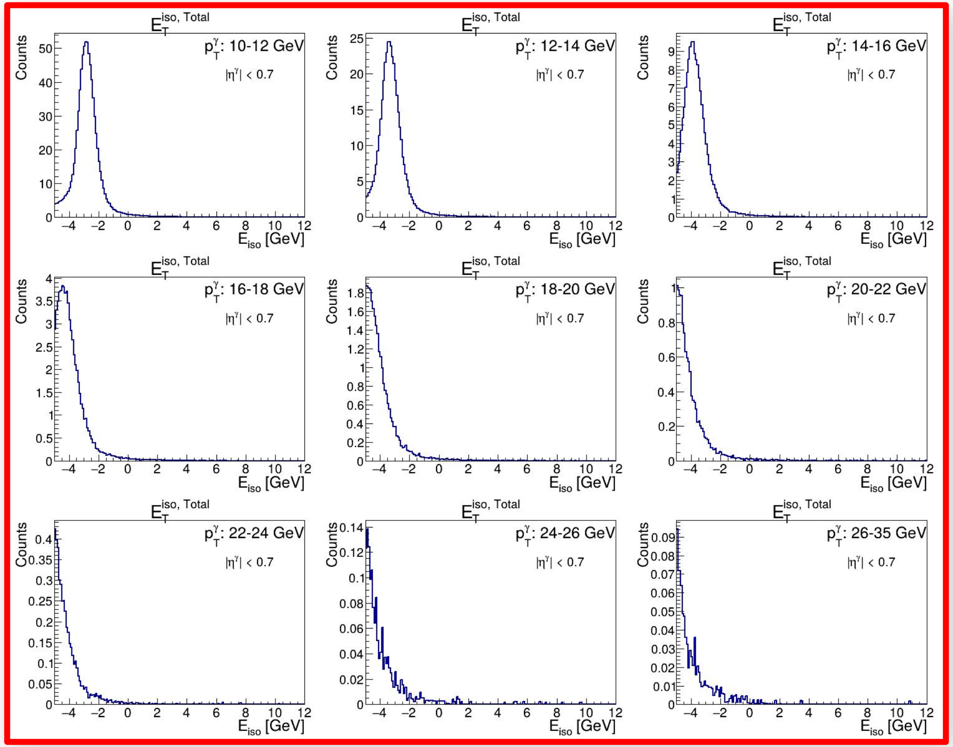
→ γ_{truth} **matching to γ_{reco} = $\Delta R(\gamma_{truth}, \gamma_{reco}) \leq 0.05$** AND CaloRawClusterEval "best match" condition

→ **For a jets passing truth/reco level away-side topology:** Accept matched pair iff $\Delta R(jet_{reco}, jet_{truth}) < 0.3$

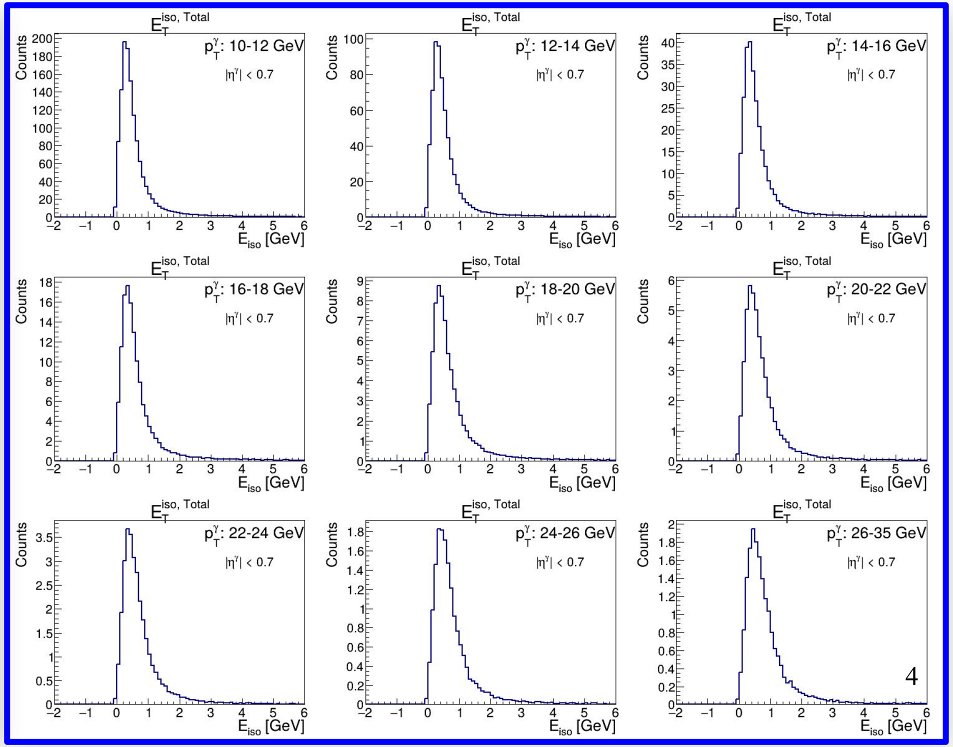
Reco Isolation QA – Fix Implemented for Negative Iso

Fixed after I stopped using calibration steps in sim since already had calibrated nodes – and now conditioning on MBD vertex – before it was getting overridden with truth vertex

Before Fix



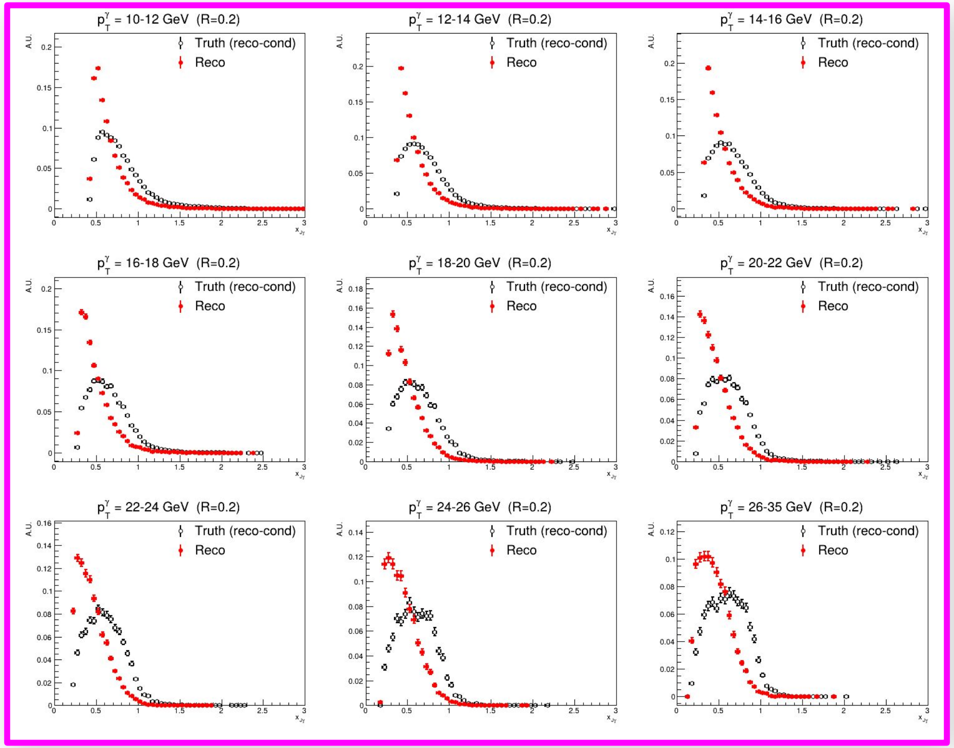
After Fix



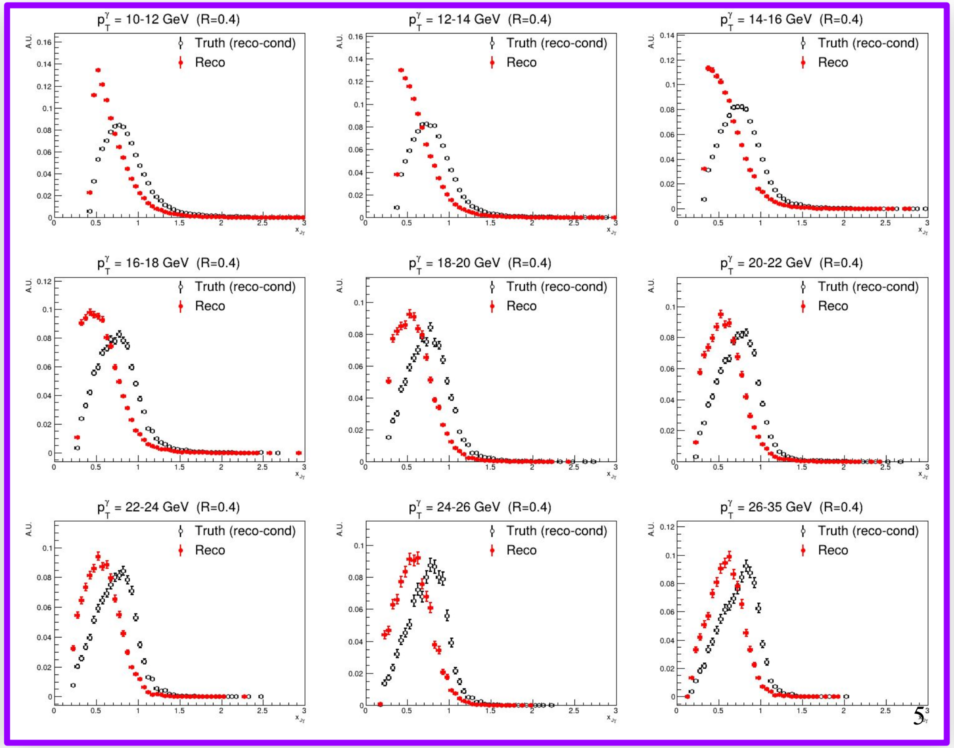
Truth vs Reco $x_{J\gamma}$ Overlays – R = 0.2 & R = 0.4 Jets

Before I think issue was a mixture of the problem with isolation energies, and [TowerJetInput] source code hard setting z = 0 without reco MBD vertex to use – now reco and truth are aligning more at high energies

R = 0.2



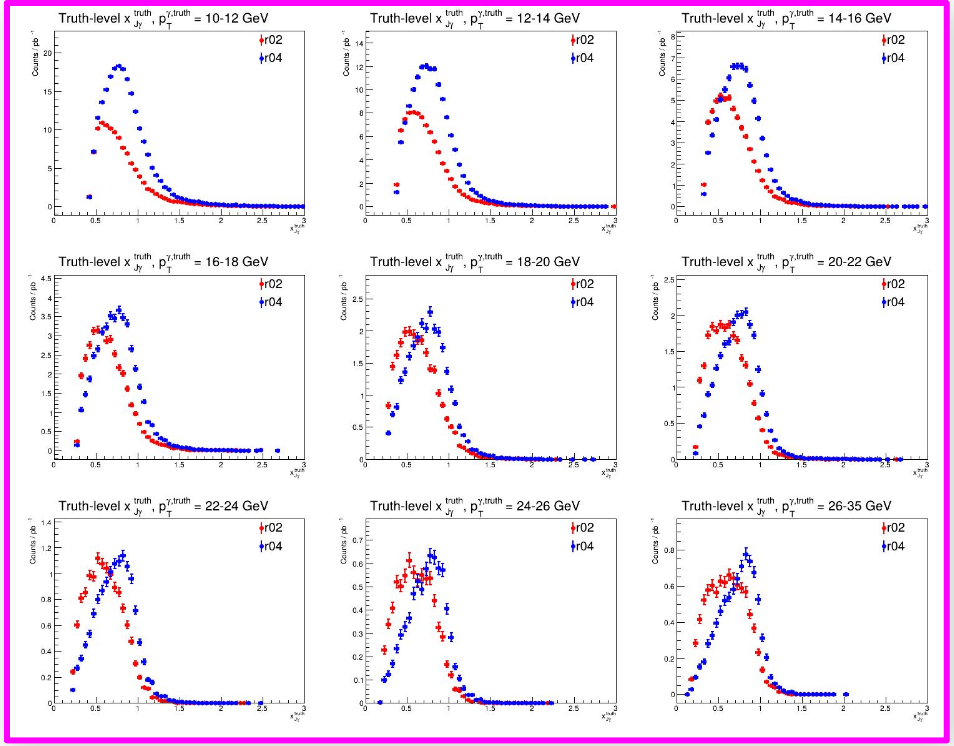
R = 0.4



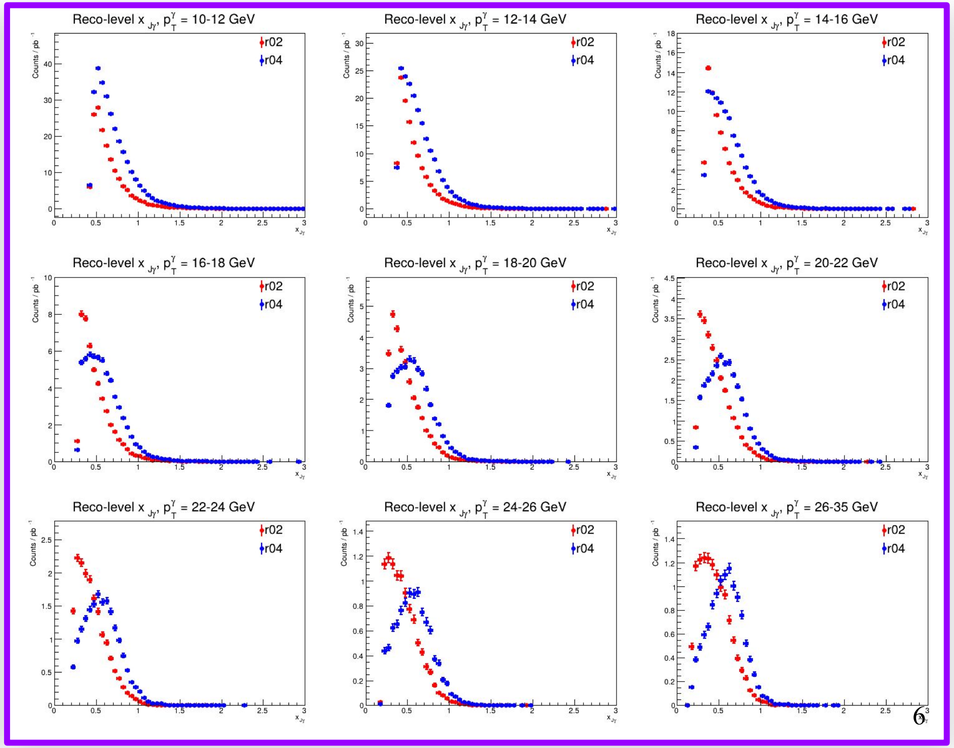
Truth vs Reco $x_{J\gamma}$ – R = 0.2 & R = 0.4 Overlays

At least for R = 0.4 jets starting around 16 GeV – seems x_J distribution has a 'Gaussian-enough' shape to go through first pass of in situ calibration – will run over data and go through the steps next

Truth (Reco Conditioned)



Reco



Next Steps

1. **With isolation 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

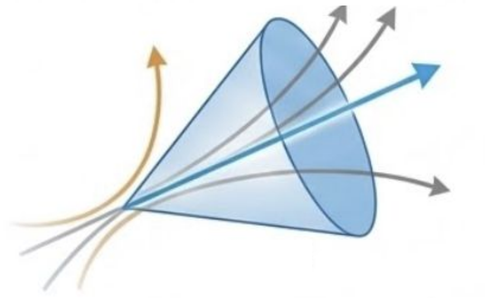

BACKUP

Isolated Photon Sample

Two ingredients of isolated photon selection:

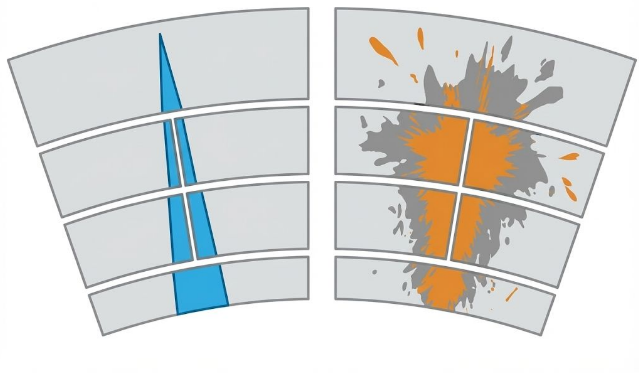
i) Cluster Isolation

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



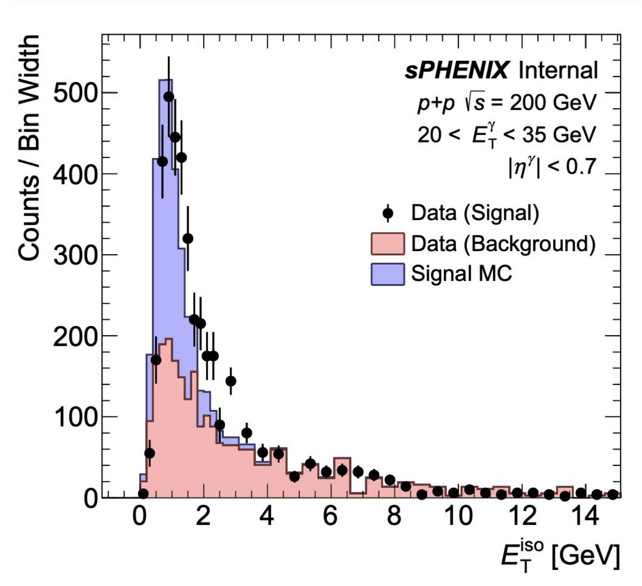
- **Sliding E_T^{iso} cut** tuned in signal MC for ~flat prompt-γ efficiency vs E_T^{γ}
- **Non-isolated sideband** background control region

ii) Shower-Shape (SS) ID



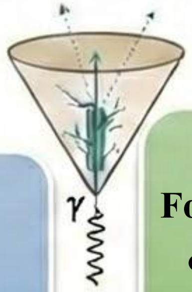
- **SS cuts** → compact EM γ vs broad hadronic/merged π^0
- **Loose** quality *preselection* → **tight** γ-ID working point

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)

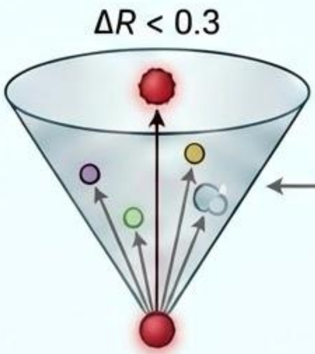
∑ **Analysis code** → sums 3 stored components with default R = 0.3

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}

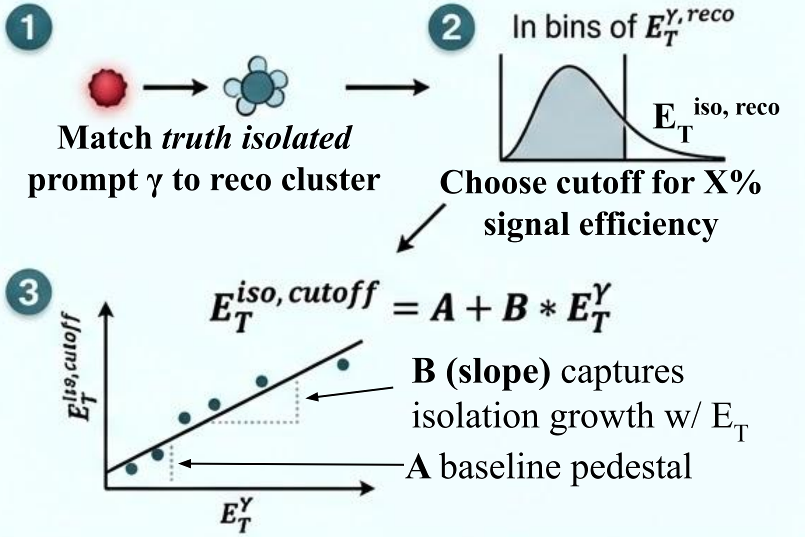
Truth-level Isolation (RecoilJets)



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

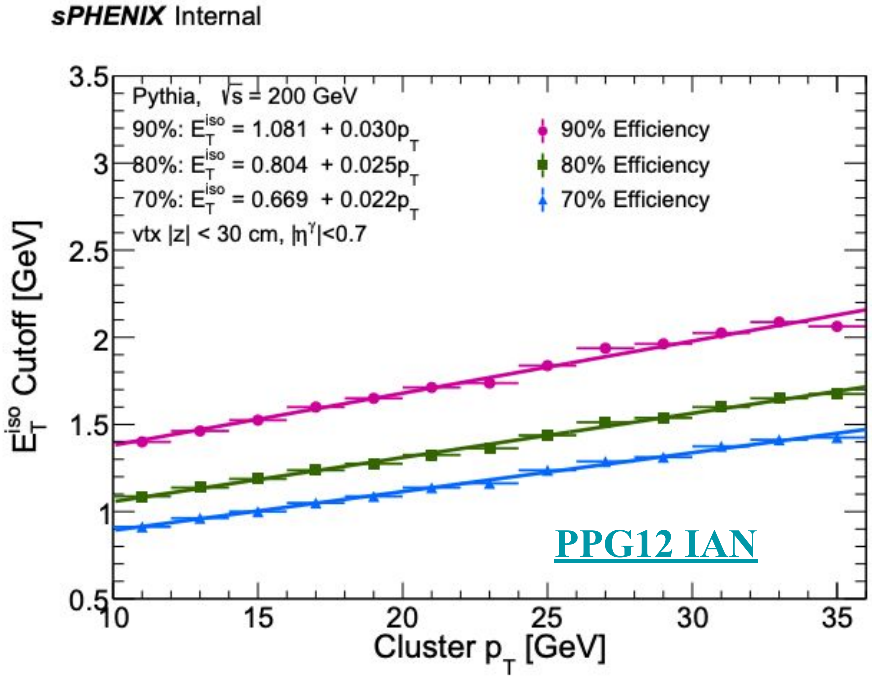
Is isolated:

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

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



γ Selection

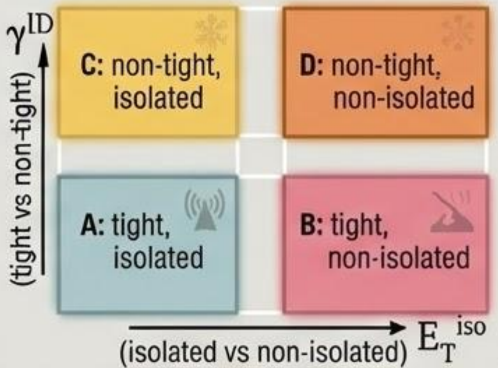
Pre-selection (loose **γ-ID quality-gate w/ $|\eta^\gamma| < 0.7$**): $\frac{E_{11}}{E_{33}} < 0.98\,, \quad 0.6 < \mathrm{et1} < 1.0\,, \quad 0.8 < \frac{E_{32}}{E_{35}} < 1.0\,, \quad w_\eta^{\mathrm{cogX}} < 0.6$

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

- **(1 & 2) Seed-excluded widths** (around CoG, exclude the seed tower) → $0 < w_a^{\mathrm{cogX}} < 0.15 + 0.006\,E_T^\gamma \quad \text{for } a \in \{\eta, \phi\}$
- **(3) Core-to-cluster compactness** (odd×odd window sums) → $0.4 < \frac{E_{11}}{E_{33}} < 0.98$
- **(4) 2×2 core energy fraction (et1)** → $0.9 < \mathrm{et1} < 1.0$, $\mathrm{et1} = \frac{E_1 + E_2 + E_3 + E_4}{E_{\mathrm{tot}}}$
- **(5) Elongation veto (strip ratio)** → $0.92 < \frac{E_{32}}{E_{35}} < 1.0$

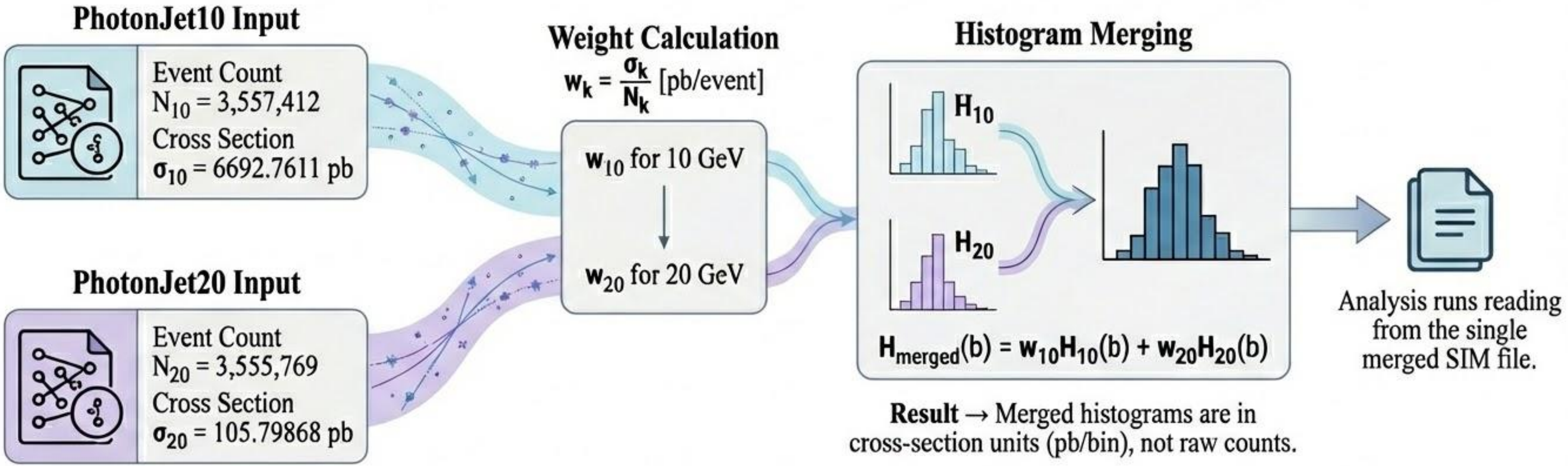
ABCD Regions

- **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)



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)**

   $\boldsymbol{\gamma^{truth}}$**:** $|\eta_{\gamma}^{truth}| < 0.7$, $p_T^{\gamma,\,truth} \in$ 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**

   $\boldsymbol{\gamma^{truth} \leftrightarrow \gamma^{reco}}$**:** $\Delta R(\gamma^{truth}, \gamma^{reco}) < 0.05$ 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, $|\eta^{j,\,reco}| < (1.1 - R)$, $|\Delta\phi(\gamma^{reco}, j^{reco})| \geq \pi/2$

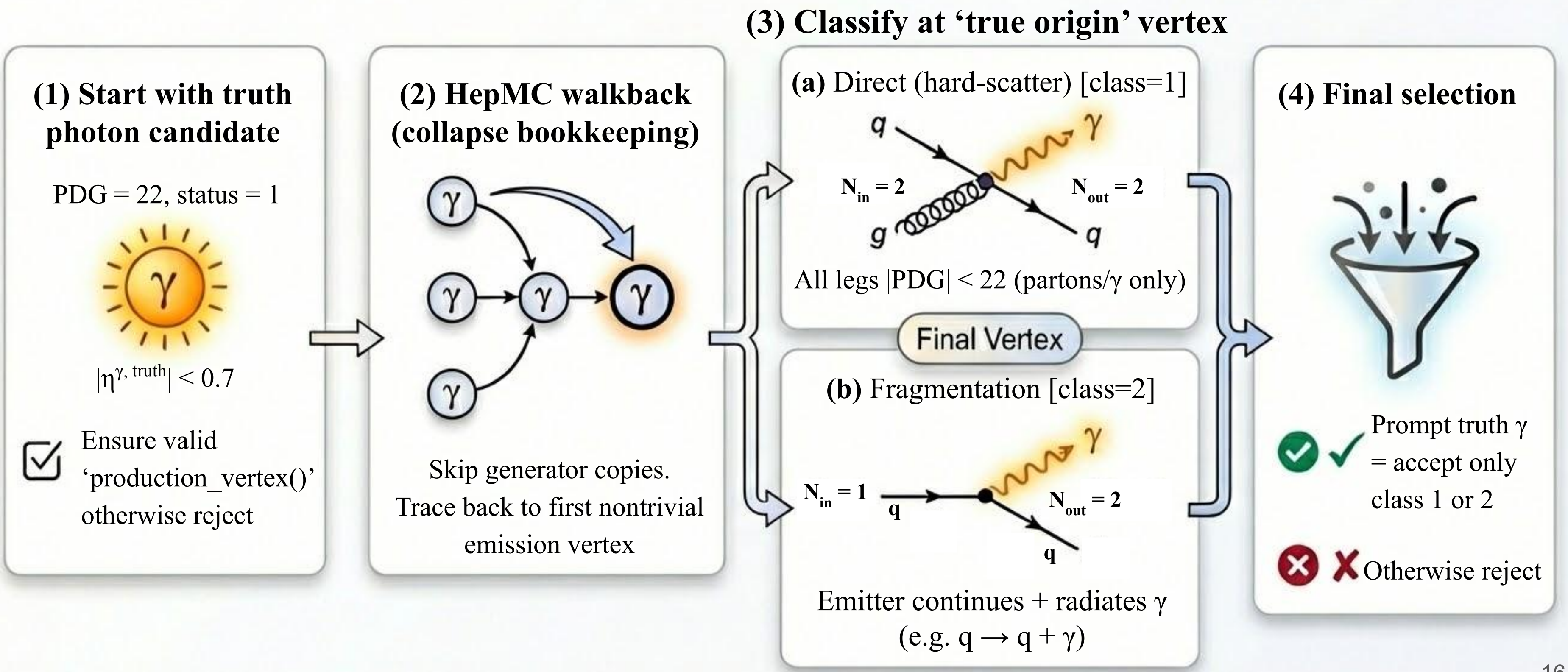
5. **Truth recoil-jet (same R) –** require leading recoil jet, j^{truth}

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

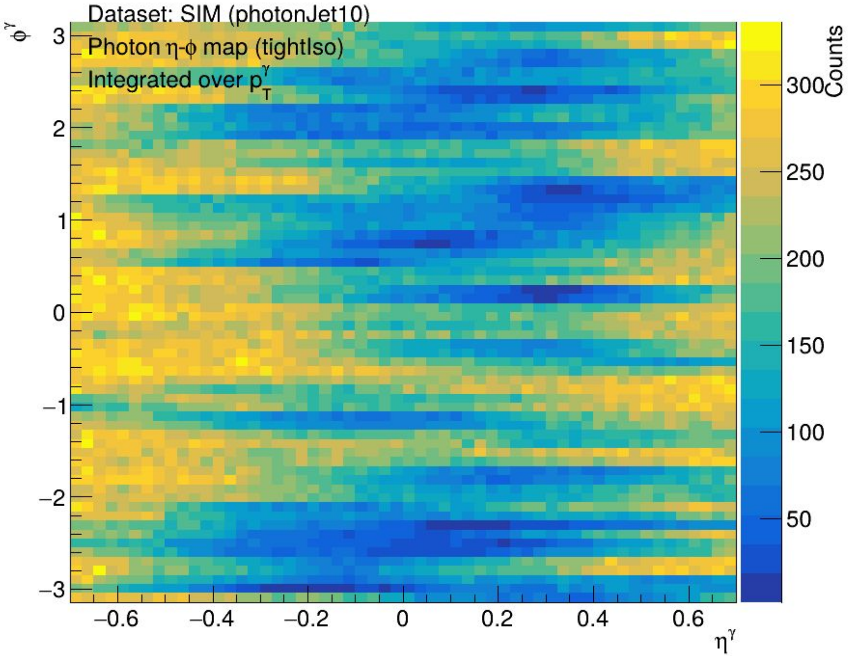
6. **Reco↔Truth jet match (final gate)** $\rightarrow \Delta R(j^{reco}, j^{truth}) \leq 0.2$

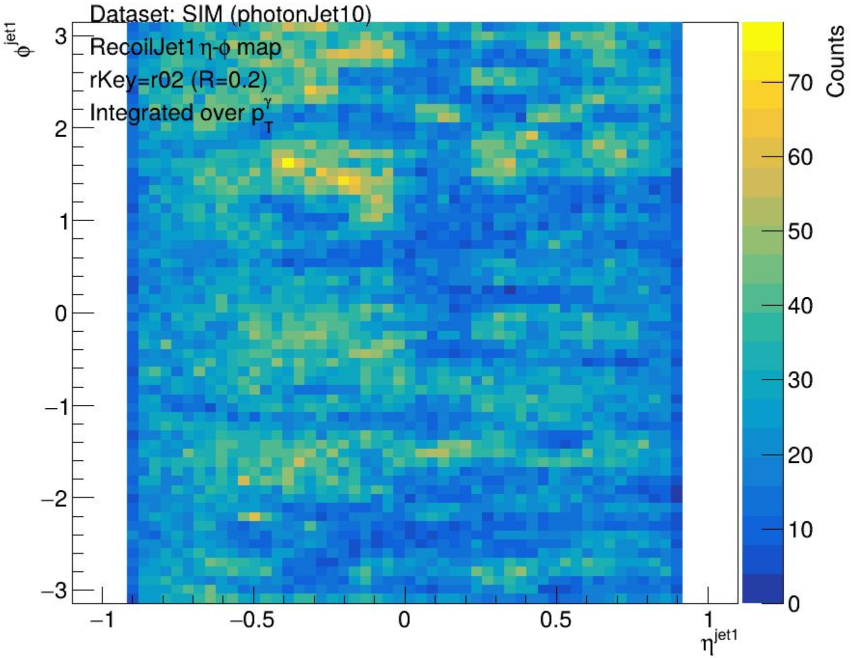
Then fill $x_J^{truth}(p_T^{\gamma,\,truth})$

Tracing HepMC Ancestry – Classifying Direct vs Fragmentation



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





Shower-Shape (SS) Criteria – Window Sums

A) Window sums: let E_{ij} → energy in tower index (i, j) with ($i \in [0,95]$ in η, $j \in [0,255]$ 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 $\mathbf{E_{tot} = \sum_{i \in cluster} E_i}$

- **Find the 4 EMCal towers** whose centers are *closest* to the cluster CoG – these **form a 2x2 core with energies $\mathbf{E_{1\to4}}$**
  - **et1** → fraction of the cluster's E_{total} carried by the four towers immediately surrounding the cluster CoG
  - $\boxed{\mathrm{et1} = \frac{E_1 + E_2 + E_3 + E_4}{E_{\mathrm{tot}}}}$ → 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

Project code

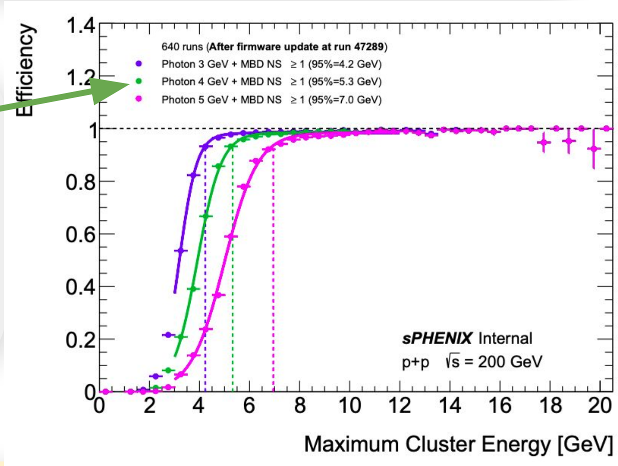
Overview

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**

- **pp** → *DST_CALOFITTING_run2pp_ana509_2024p022_v001*
  - Golden run list v3 from coresoftware
  - Using **Photon 4 GeV + MBD NS** trigger for data
  - Events analyzed = X (placeholder for % of GRL I have for tomm)

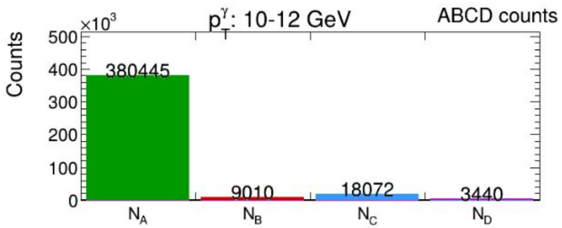
95% efficient @ 5.3 GeV

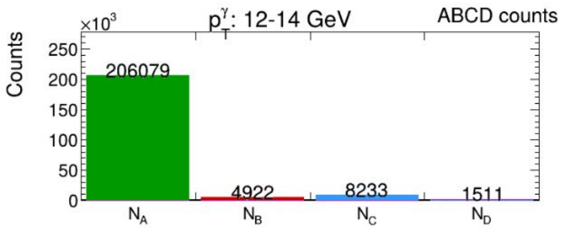


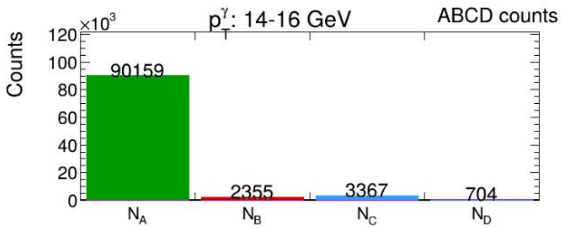
Analysis Notes

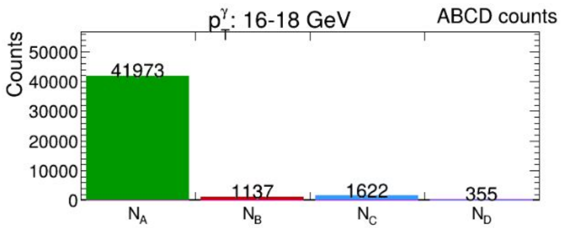
- **Using local build** of PhotonClusterBuilder from version pulled on 01.14.26 (updated to allow sim info)
- **JES calibration** from JetCalib (see Hanpu's JSTG slides)
- **Jet p_T ≧ 5 GeV and Δφ(leading γ, leading jet) ≧ π/2**
- **E_T^{γ} = [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV** → match PPG12

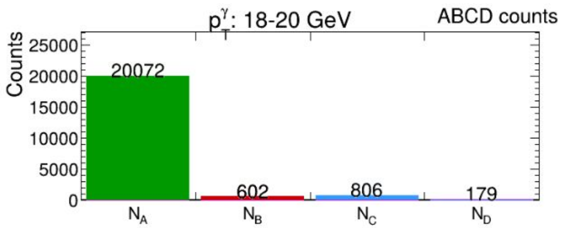
ABCD Bin Counts – Photon + Jet 10 Sim

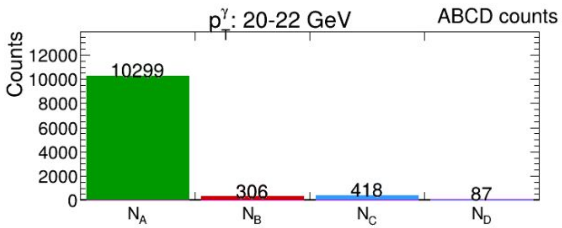


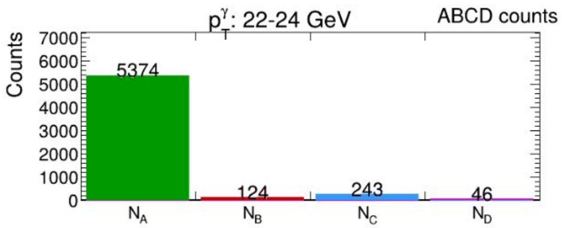


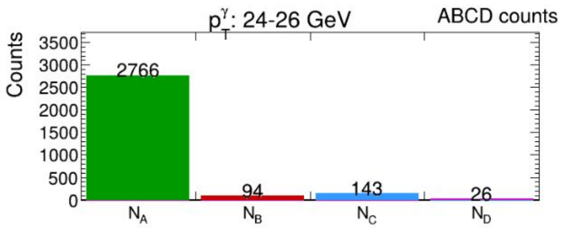


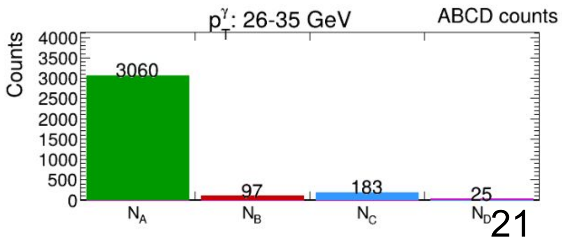




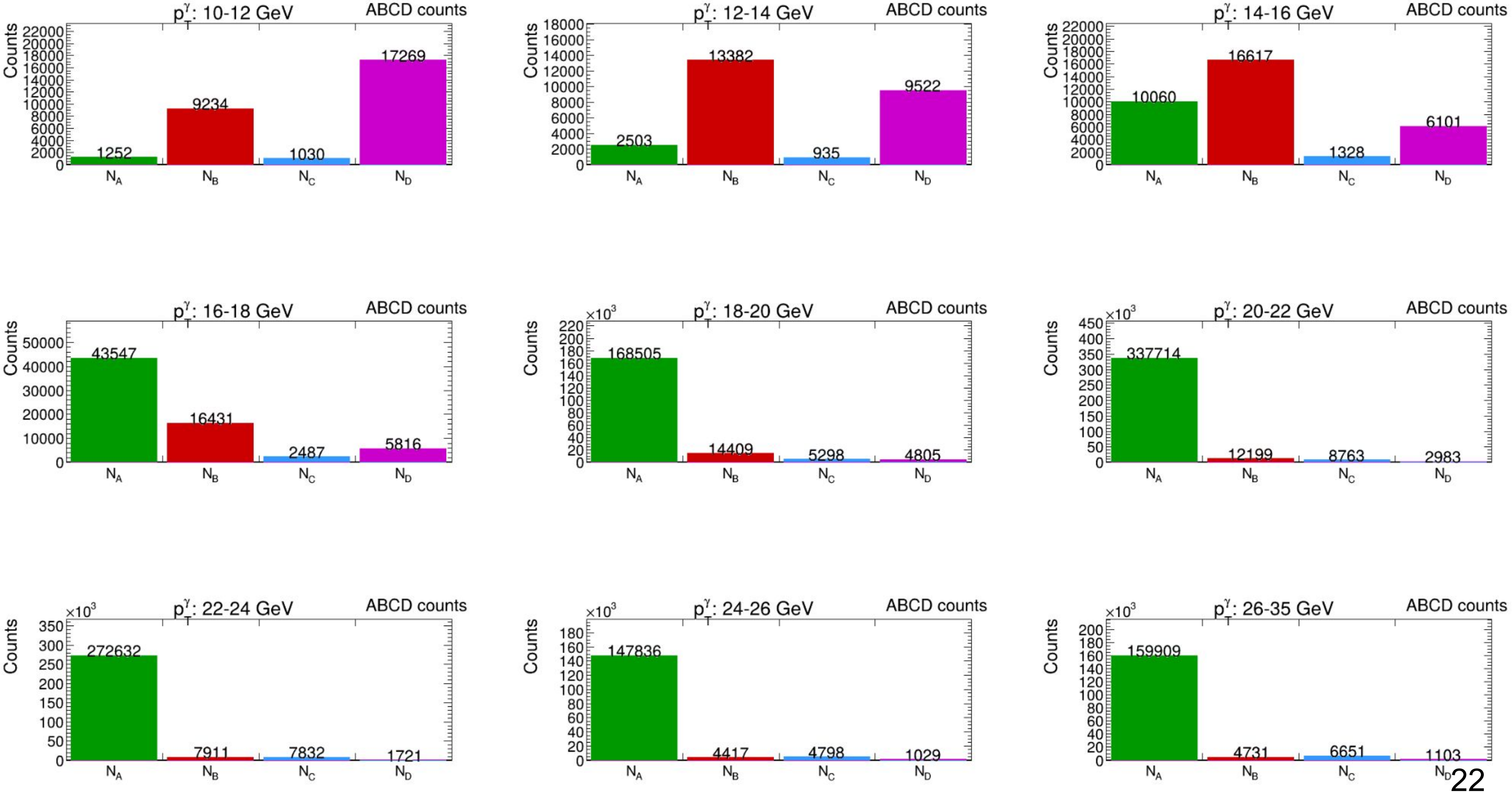




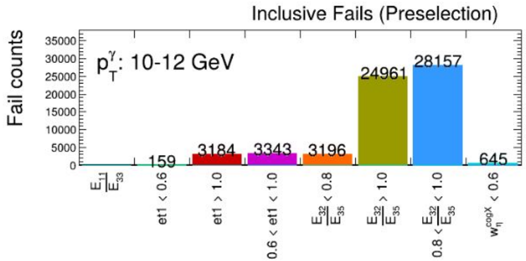


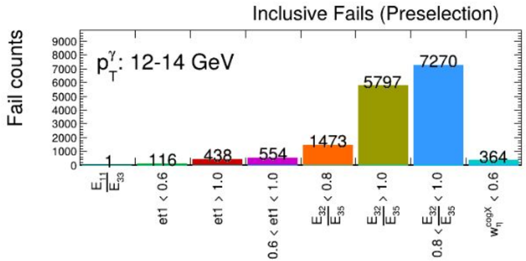


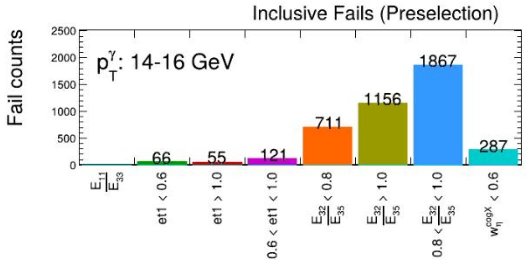
ABCD Bin Counts – Photon + Jet 20 Sim

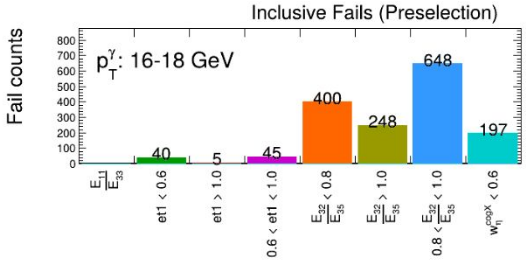


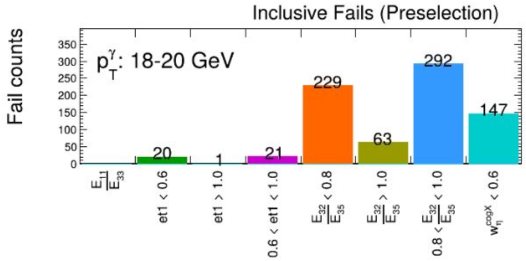
Counts of Preselection Failures by Cut – Photon Jet 10 Sim

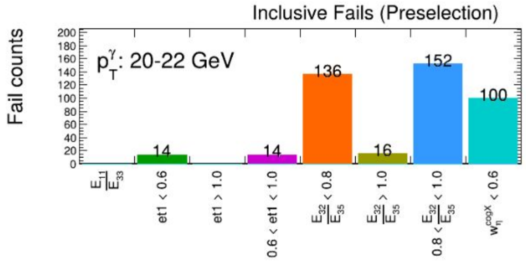


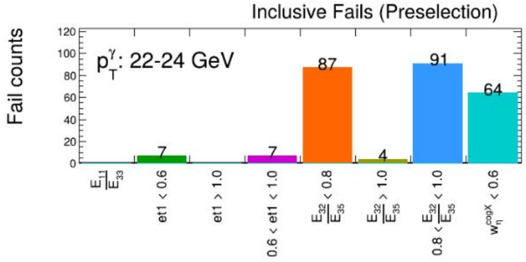


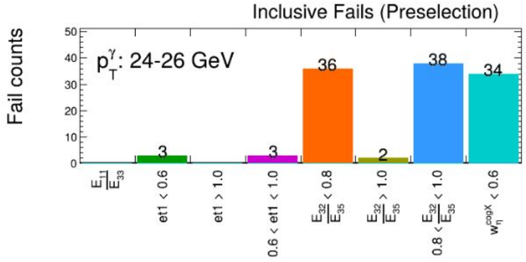


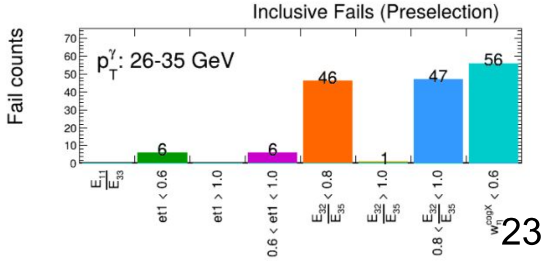




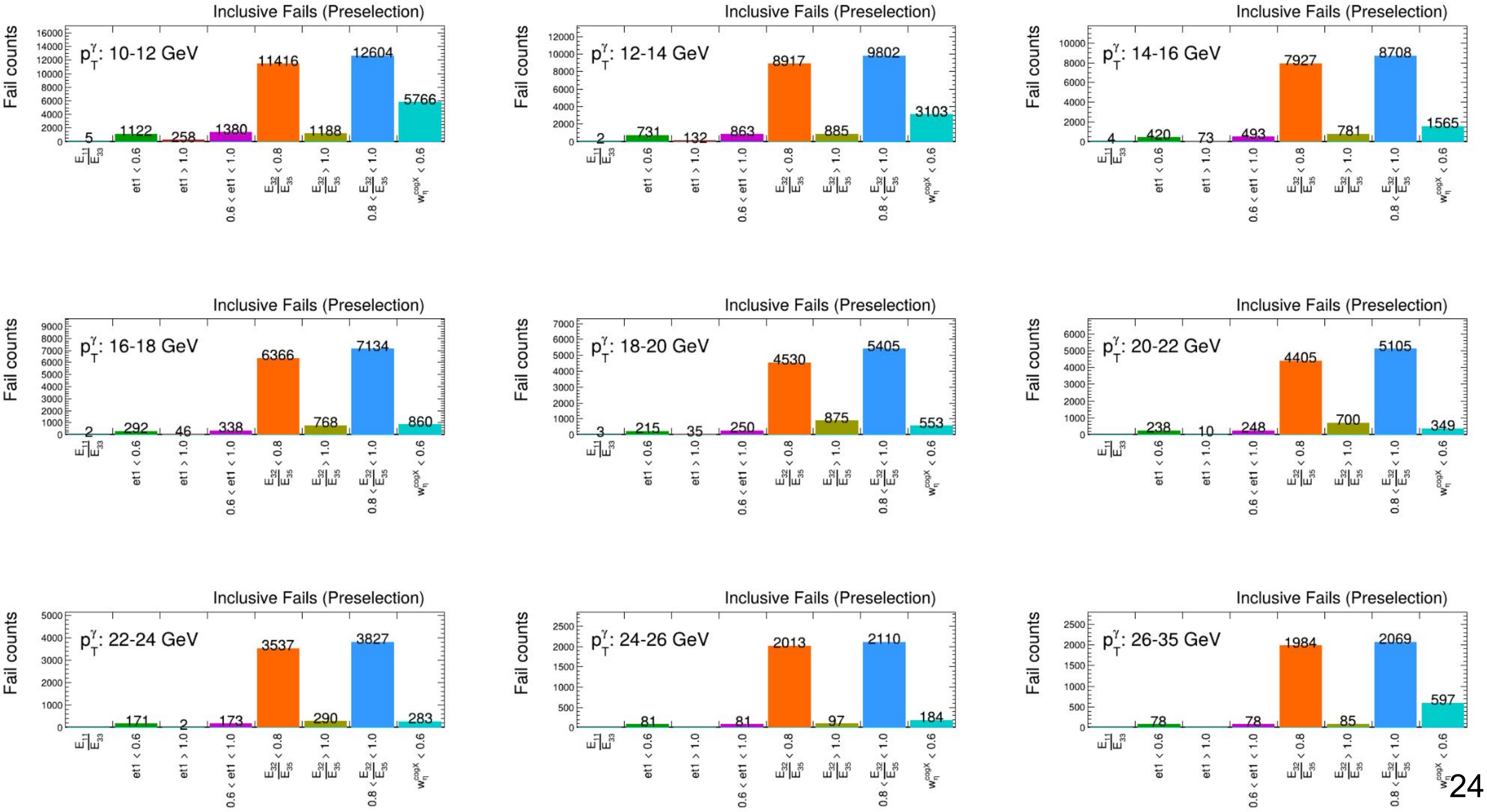








Counts of Preselection Failures by Cut – Photon Jet 20 Sim



Prepared Response Matrix for R = 0.2 & R = 0.4 Jets

- 2D response matrix for unfolding (p_T^γ, $x_{J\gamma}$) shown for photonJet10 and 20 for R = 0.4 jets
- Each pixel is the number of events that migrate from a truth global-bin (x-axis) to a reco global-bin (y-axis), with log-Z counts (yellow = high occupancy).
- xJγ is a ratio, but smearing acts on the two measured energies (pTγ and pTjet) in a correlated way. Building the response in the joint observable space preserves these correlations