# Cut-Based Photon Identification Studies at CMS

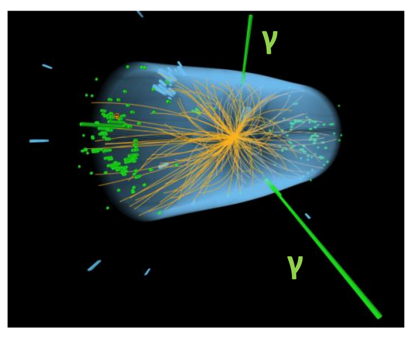

Shubham Dutta
On behalf of CMS Collaboration

Saha Institute of Nuclear Physics, A CI of HBNI, 1/AF Bidhannagar, Kolkata-700064, West Bengal, INDIA


## Introduction

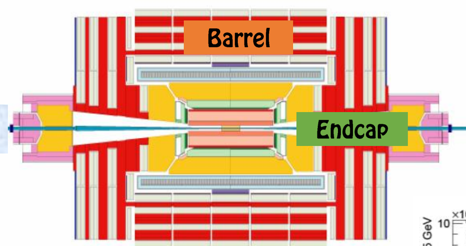
### Why are Photons important?

Photons provide a cleaner channel amidst the messy hadronic environment in proton-proton (p-p) collisions at the LHC. Nothing highlights its importance more than the discovery of **Higgs decaying into two photons** which has a **branching ratio of only 0.002 !!!**

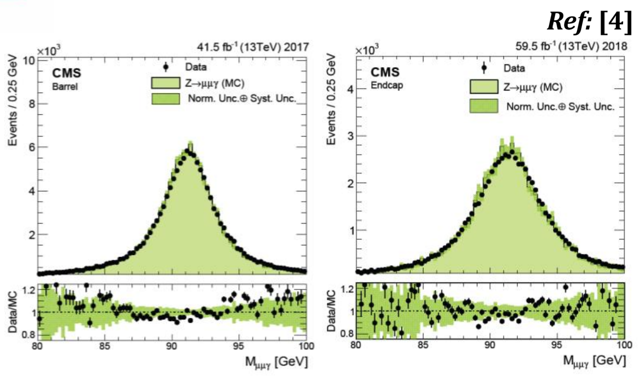



H → γγ event

Energy Resolution within 1% (3%) in Barrel (Endcap)



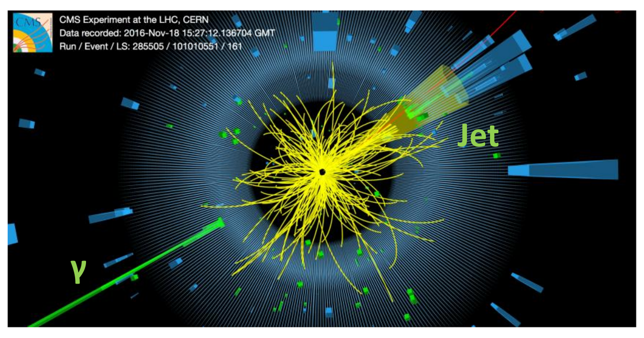

Ref: [4]



Ever since Higgs-discovery, it has been used as a probe in **Beyond Standard Model (BSM)** searches. One such analysis, for instance, looks for the decay of Higgs into a dark photon and an SM photon and uses the **cut-based photon identification (ID)** for selecting the SM photon[5]. Many other searches which have a photon as one of the final state particles use this ID.

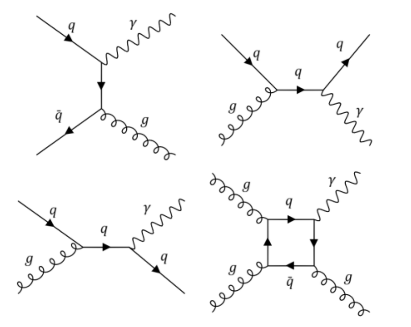
### Objective

- Photons come from two sources: i) **prompt photons** → produced in hard scattering processes **(signal)** ii) **non-prompt photons** → produced abundantly in hadron decays and fragmentation processes **(background)**.
- **ID variables** discriminate against these two "kinds" of photons. Combination of these variables are taken and the ID values - the value at which the discrimination threshold is kept, are tuned using **Genetic Algorithm (GA)** for **three working points (WP)** at **90% (loose), 80% (medium) and 70% (tight)** signal efficiency.



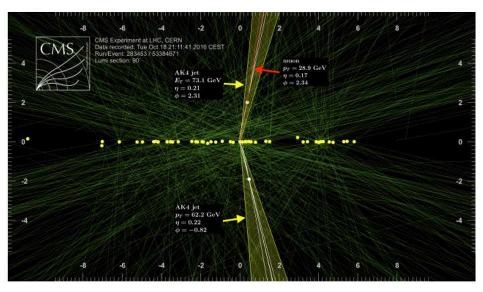

γ-jet event

Study done using **Monte Carlo (MC) samples** of **γ-jet** with latest pile-up (PU) conditions i.e. Run 3 conditions (CoM energy 13.6 TeV) for p-p collisions.



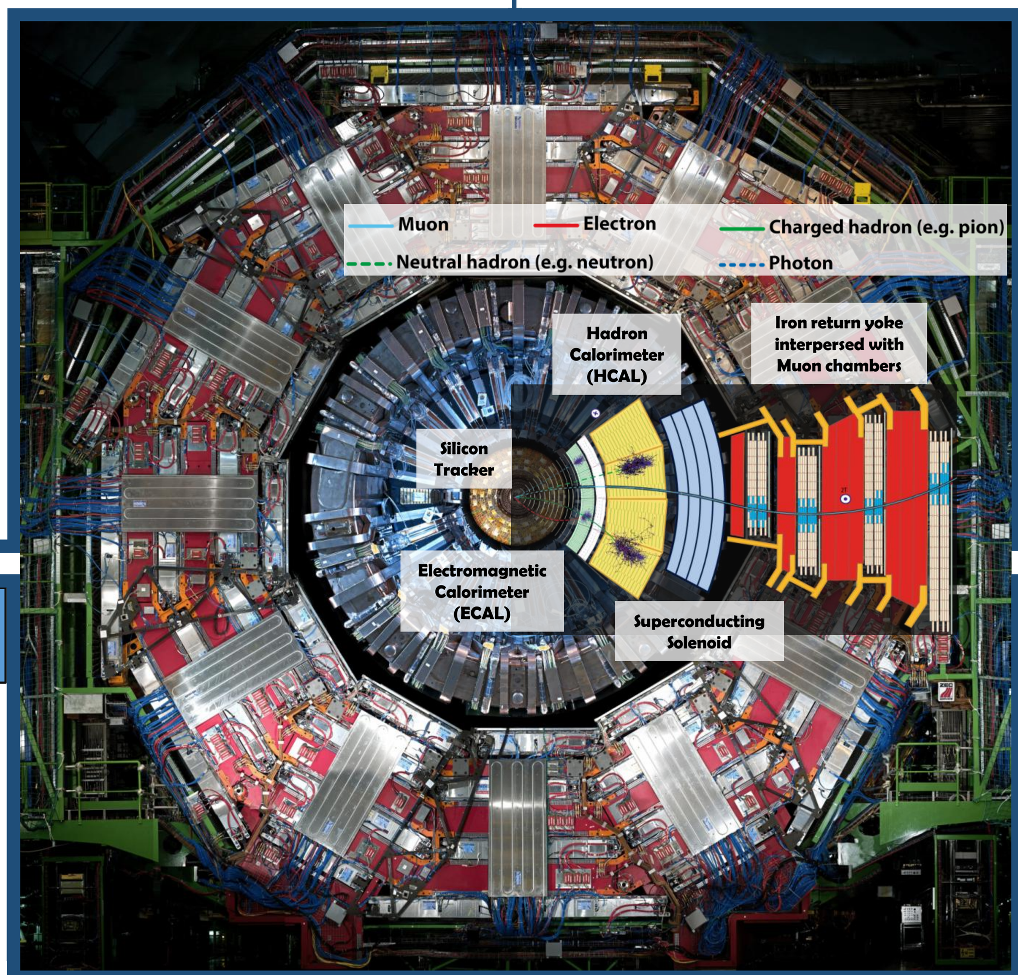
γ-jet MC sample physics process

### Pile-up



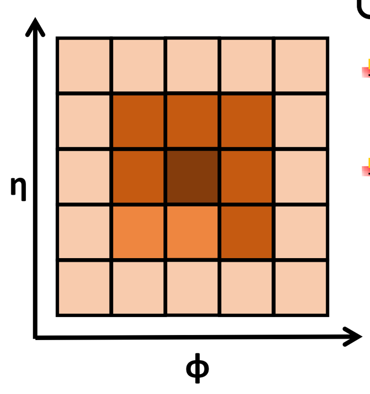
Pile-up events

- Multiple interactions within the same p-p bunch crossing.
- Increases with luminosity and energy
- Run 3 has an expected pile-up of about 52 interactions
- Quantified by **no. of vertices (nvtx)** and **ρ**, which is the **pile-up transverse momentum ($p_T$) density**.

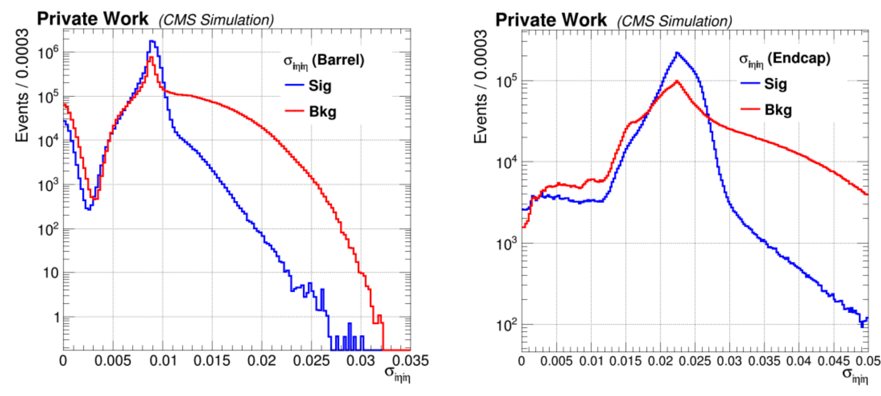



## ID Variables

### $\sigma_{i\eta i\eta}$

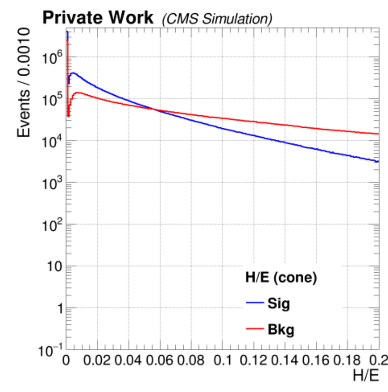


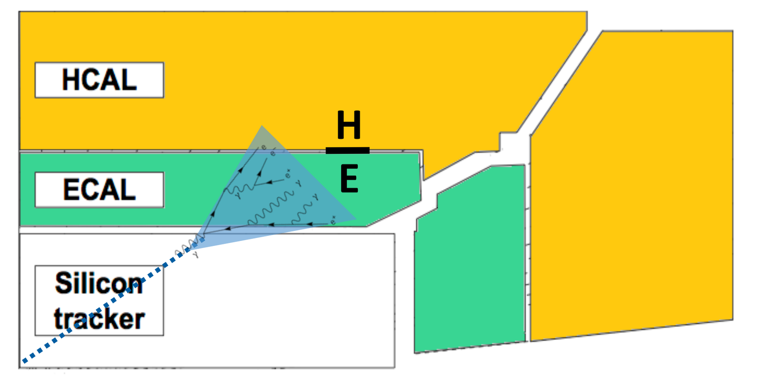

- Energy-spread along **η-direction** for a **5×5 cell array** in ECAL
- Prompt photons have lower values of $\sigma_{i\eta i\eta}$ since their shower is narrower than that of jets




### H/E

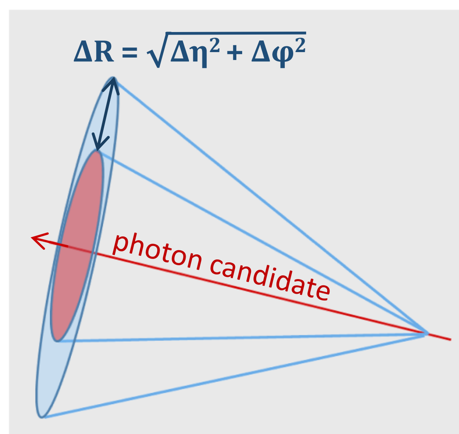
- Ratio of energy deposited in HCAL to that in ECAL
- Prompt photons have higher energy deposit in ECAL relative to HCAL, so low value of H/E

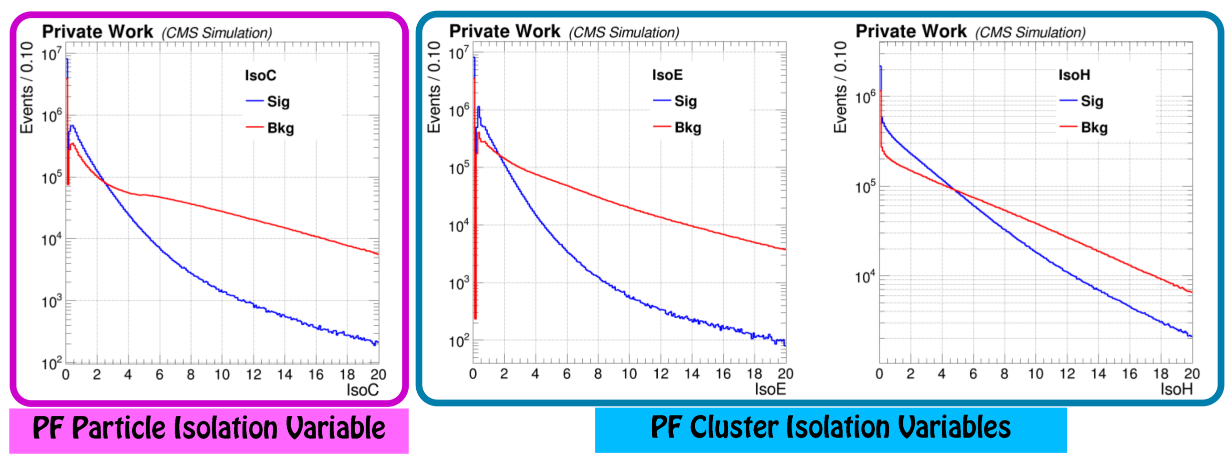



### Isolation

- **Σscalar transverse momentum ($p_T$)** of all detected particles within a hollow cone of **ΔR radius** around the photon candidate.
- Jets will typically have more particles surrounding it and thus, a higher value of $\Sigma p_T$
- **Two categories** of isolation variables:
  - **Particle Flow (PF) Cluster Isolation** → calculated based on the energy-deposit clusters in the sub-detectors, viz. **Tracker (IsoT), ECAL (IsoE)** and **HCAL (IsoH)**
  - **PF Particle Isolation** → calculated after forming particle candidates, **Charged hadrons (IsoC), Photons (IsoP)** and **Neutral hadrons (IsoN)**



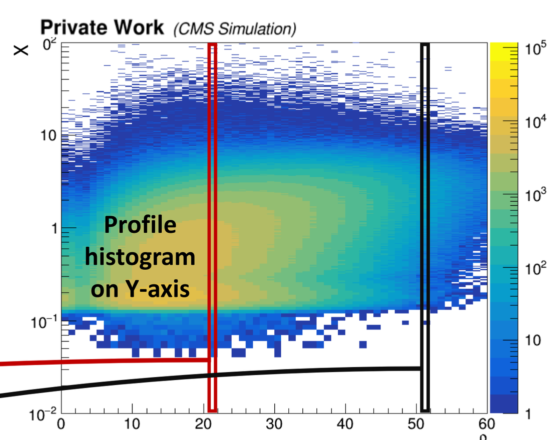



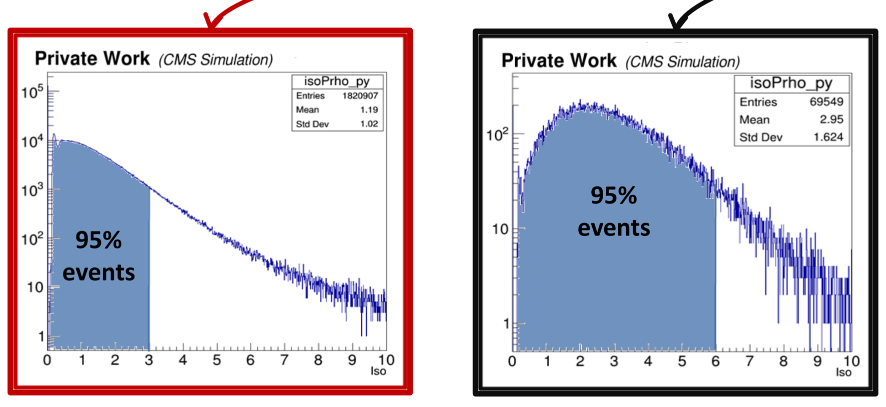

PF Particle Isolation Variable

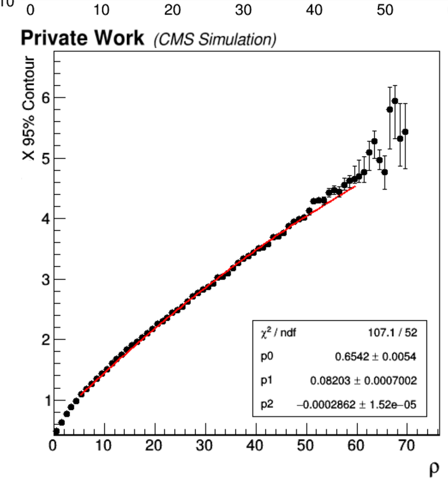
PF Cluster Isolation Variables

## Outline of Workflow

- **Isolation** and **H/E** are **PU-corrected**
  Effective areas (EA) are extracted with **95% contour profile** after separating the samples in **seven η-bins**
  $X_{p\text{-corrected}} = X - (\rho^2 \times EA2 + \rho \times EA1)$
  **EA2, EA1** → Coefficients of the fitted polynomial
- **$p_T$-scales** are calculated similarly for **IsoE, IsoH, IsoP** and **IsoN** for $p_T$-correction
- ID-training for **loose, medium and tight** is done. Five variables are used:
  **$\sigma_{i\eta i\eta}$, H/E,** and a combination of **three variables** from **PF Cluster** and **PF Particle isolation**
- Stability of the ID cut-values are checked by plotting the **efficiency** over **$p_T$, η** and **nvtx** range for the best performing combination

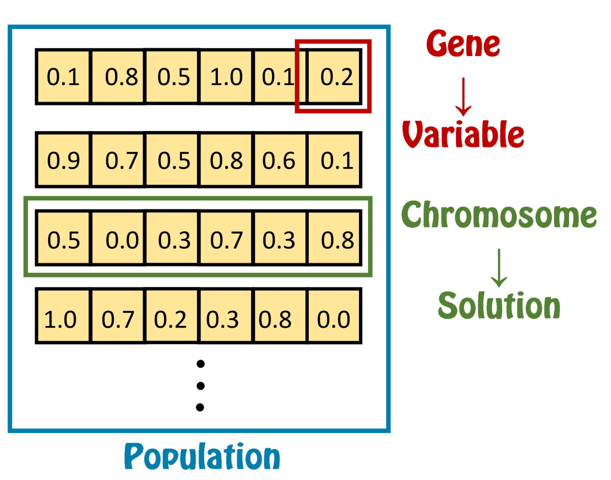


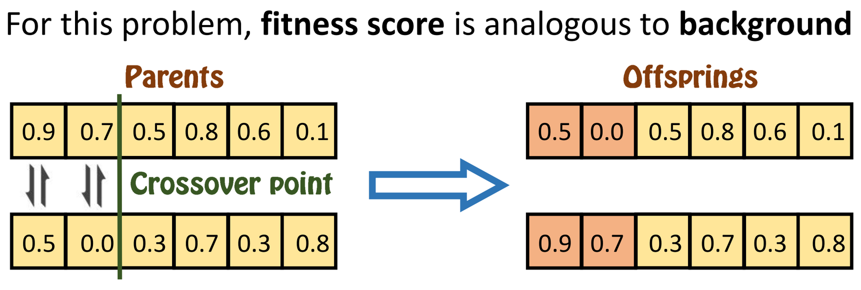



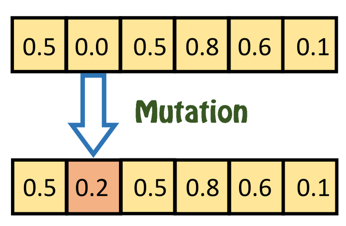




### Genetic Algorithm

- Optimization algorithm which is based on the idea of **'Survival of the Fittest'**
- Each **variable** is analogous to a **gene**, can take any value
- Set of **genes** form **chromosome** of an **individual**, represents **solution**
- Each chromosome has survivability or fitness score – a score associated with each solution. Individuals with high fitness score are more likely to be chosen as parents.
- For this problem, **fitness score** is analogous to **background rejection**

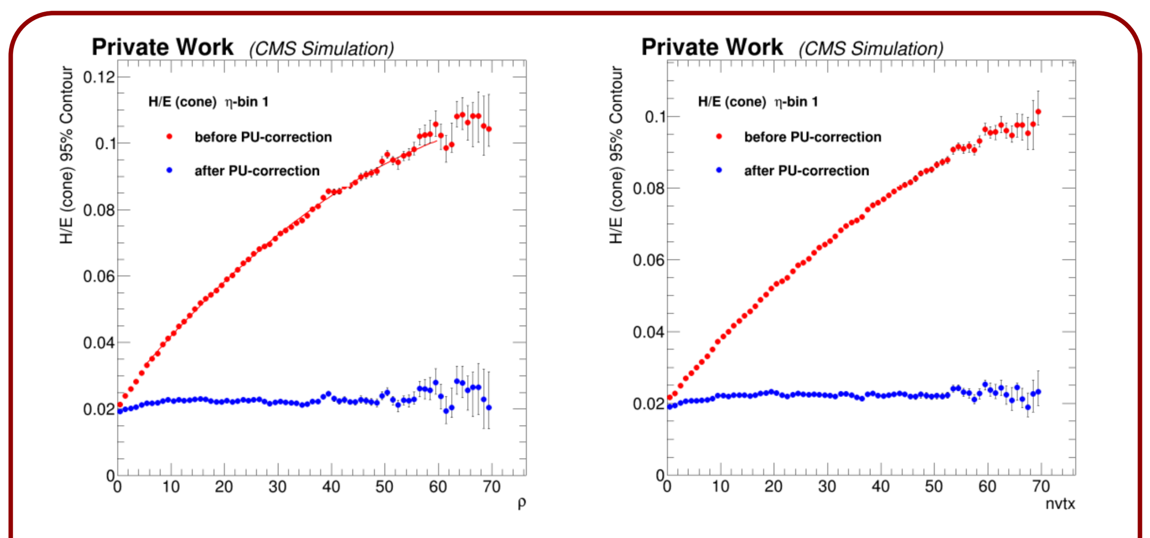


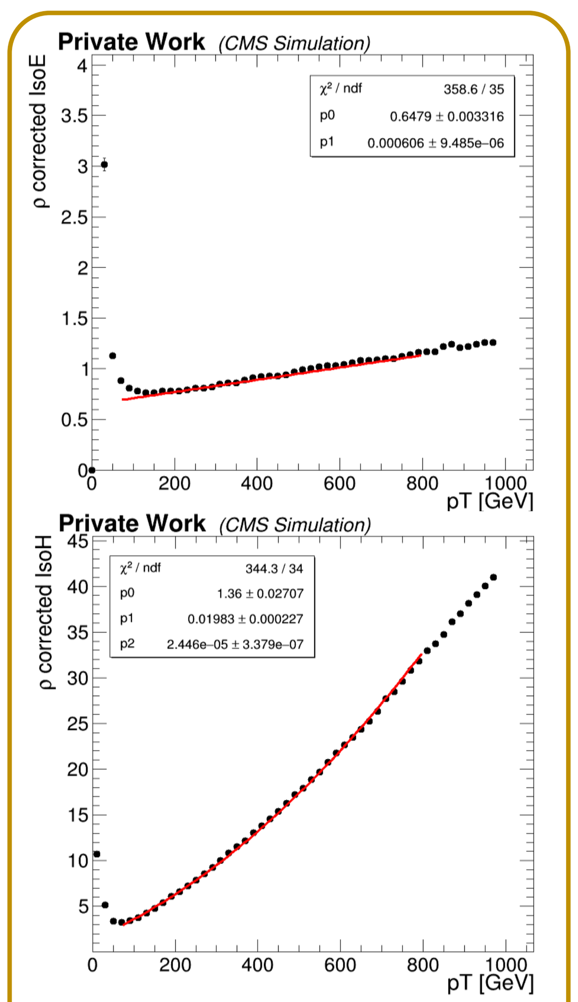




- The population in the next generation thus, has higher average fitness score.
- **Mutation** fosters exploring more optimum solutions.

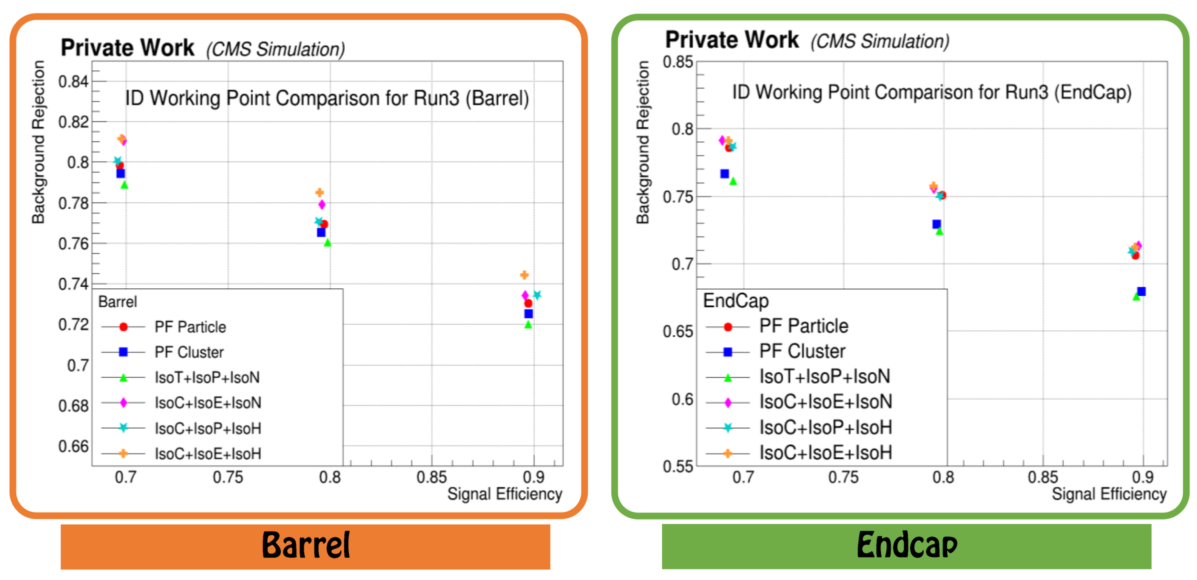
## Results




Flat profile of the variables after the PU-correction




IsoE and IsoP have linear dependence while IsoH and IsoN have quadratic dependence




Barrel | Endcap

The combination IsoC + IsoE + IsoH gives the best performance

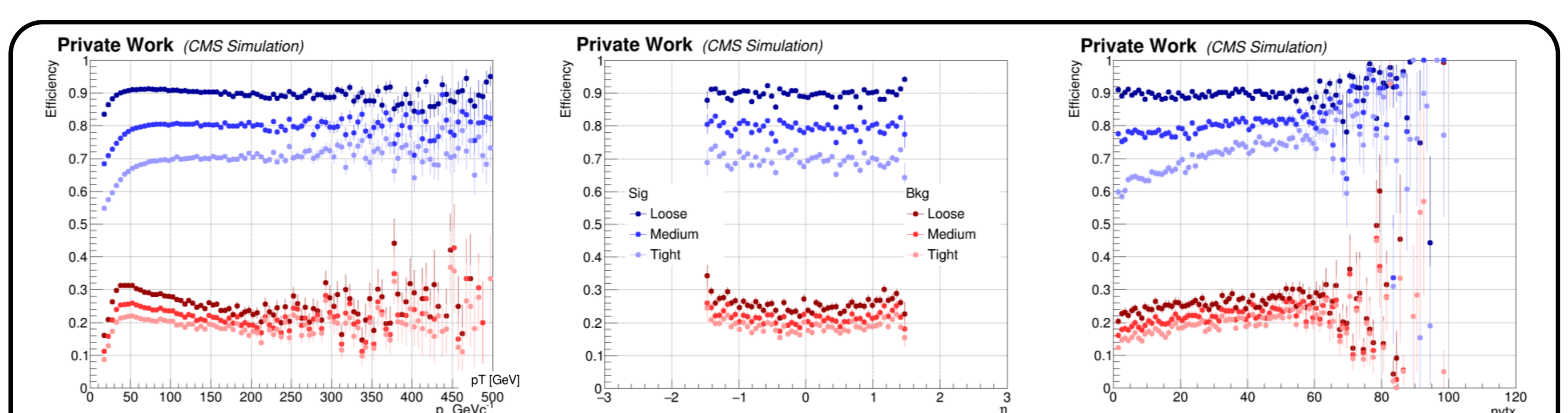
### Run 3 Photon ID for BARREL & ENDCAP

BARREL:

| WP | $\sigma_{i\eta i\eta}$ | H/E | IsoC | IsoE | IsoH |
|---|---|---|---|---|---|
| Loose | 0.0114 | 0.1299 | 1.8852 | 0.7038 | 6.3439 |
| Medium | 0.0100 | 0.0583 | 0.9393 | 0.2277 | 2.1890 |
| Tight | 0.0099 | 0.0418 | 0.3163 | 0.1419 | 0.3906 |

ENDCAP:

| WP | $\sigma_{i\eta i\eta}$ | H/E | IsoC | IsoE | IsoH |
|---|---|---|---|---|---|
| Loose | 0.0277 | 0.1534 | 1.6539 | 6.6158 | 1.8588 |
| Medium | 0.0269 | 0.0052 | 0.9703 | 1.1240 | 0.0337 |
| Tight | 0.0269 | 0.0025 | 0.2927 | 1.0427 | 0.0293 |




Stability of efficiency over pT , η and nvtx range


## References

1. CMS Physics, Technical Design Report (2006)
2. CMS Electromagnetic Calorimeter, Technical Design Report (1997)
3. 'Particle-flow reconstruction and global event description with the CMS detector ', JINST 12 (2017) P10003
4. 'Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC', JINST 16 (2021) P05014
5. Search for Dark Photons in ZH Decays at $\sqrt{s}$ = 13 TeV, JHEP 10 (2019) 139
6. 'An Introduction to Genetic Algorithms', M. Melanie, MIT Press (1996)