Preliminary Studies for Higgs to $Z\gamma$ with $Z \to c\bar{c}$

Presented by Daniel Li, Northwestern University

Introduction and Motivation

$H \to Z\gamma$ is a rare SM process with branching ratio of 1.53×10^{-3}

- Published searches only include $Z \to ee$ and $Z \to \mu\mu$

$Z \to q\bar{q}$ plagued by large backgrounds and poor $m_{q\bar{q}}$ resolution

- $\sigma(H \to Z\gamma \to q\bar{q}) \approx \mathcal{O}(10)$ fb and $\sigma(H \to q\bar{q}) \approx \mathcal{O}(10^3)$ fb
- Low efficiency charm-tagging algorithms (DeepAK8)

$H \to Z\gamma$ with $Z \to \ell^-\ell^+$ is limited to $\mathrm{BR}(Z \to \ell^-\ell^+) \approx$ 7% (no τ)

- $\mathrm{BR}(Z \to c\bar{c}) \approx 12\%$ and $\mathrm{BR}(Z \to b\bar{b}) \approx 15\%$

Question: How much sensitivity can be gained by including $H \to Z_{c\bar{c}}\gamma$ (and possibly $H \to Z_{b\bar{b}\gamma}$) along with improved flavor-tagging algorithms?

- ...and is it worth the effort?

Relevant Processes and Backgrounds

$HZ\gamma$ backgrounds in $\ell\ell$

- Resonant Higgs decays
- Drell-Yan
- SM $Z\gamma$

Differences with $c\bar{c}$ final state

- Reduce $H \to c\bar{c}$ backgrounds with γ requirement, fewer FSR γ with $c\bar{c}$
- DY and SM $Z\gamma$ reducible and more easily filtered
- Small contribution $< \mathcal{O}(1)$ fb from SM $ZW\gamma$

Yields further adjusted by available triggers

- VH and ttH are primary $c\bar{c}$ targets with available lepton triggers
- Targeting a 4/5-body final state $c\bar{c}\ell\ell(\nu)\gamma$

$VBF \to Z\gamma$ could factor in with multi-jet trigger

ggH will likely "contaminate" other channels

BR	ee	$\mu\mu$	$c\bar{c}$
H	$< 10^{-8}$	2.2×10^{-4}	2.9×10^{-2}
Z	3.4×10^{-2}	3.4×10^{-2}	1.2×10^{-1}
$H \to Z\gamma$	5×10^{-5}	5×10^{-5}	1.8×10^{-4}

$\sigma \times 138\ \mathrm{fb}^{-1}$	ee	$\mu\mu$	$c\bar{c}$
H	< 1	1.7×10^{3}	2.2×10^{6}
$VH \to Z\gamma$	16	16	57
$ttH \to Z\gamma$	4	4	13
$VBF \to Z\gamma$	26	26	94
$ggH \to Z\gamma$	335	335	1207
$ZW^{\pm}\gamma$	80	80	137

Trigger Considerations and $c\bar{c}$ Strategy

Fully hadronic final state $HZ\gamma$ not favored due to un-prescaled mono-photon trigger p_T requirements

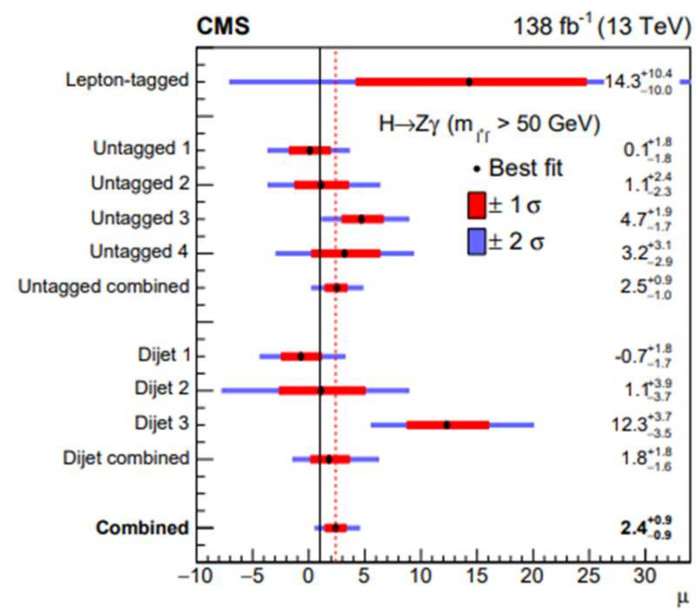
- $HZ\gamma$ peaks at $p_T^\gamma \approx 20$ GeV while triggers are $p_T^\gamma \geq 60$ GeV
- Potential VBF strategy with trijet L1 trigger seed ($VBF \to b\bar{b}$)

Considering dilepton triggers *and* single lepton triggers

- In Run 2 HIG-014-19, only dilepton triggers and $ZH \to \ell\ell c\bar{c}$ filtered by $M(\ell\ell\gamma)$ cut or consolidated into "VBF" category
- Inclusion of single lepton allows access to $W^\pm H/t\bar{t}H \to \ell\nu c\bar{c}$

70 additional $c\bar{c}$ events to the 40 "additional lepton", 52 dijet and 670 ggH categories (assuming full efficiency)

- Exploit fewer FSR photons in $c\bar{c}$ and improved $Z \to c\bar{c}$ and $H \to c\bar{c}$ mass reconstruction for $M(c\bar{c})$ and $M(c\bar{c}\gamma)$ selection
- Avoids the substantial SM $Z\gamma$ background by requiring four final state objects (SM $Z\gamma \to c\bar{c}\gamma$ filtered by lepton trigger)



$VHc\bar{c}$ Strategy with $Z \to c\bar{c}$ at 5.7σ

Defines two search regions: merged (AK15) and resolved (AK4)

- Custom regression for merged AK15 jets with ~50% improvement in mass resolution over traditional algorithms
- Three final states: $Z \to \nu\nu$ (large E_T^{miss} trigger), $W \to \ell\nu$ (single lepton trigger), $Z \to \ell\ell$ (dilepton trigger)
- Split regions by $p_T(H) > 300$ GeV (only 5%, but better sensitivity)

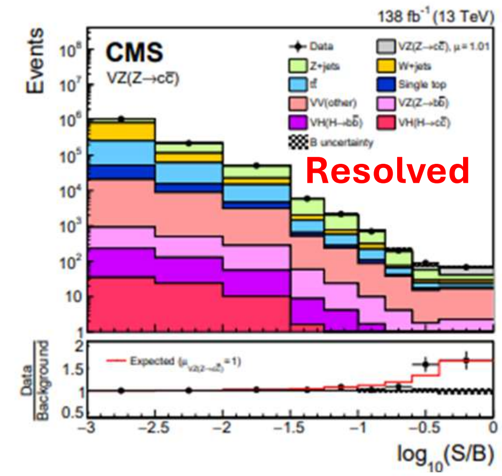
Merged c-tagging using ParticleNet, resolved c-tagging using DeepJet

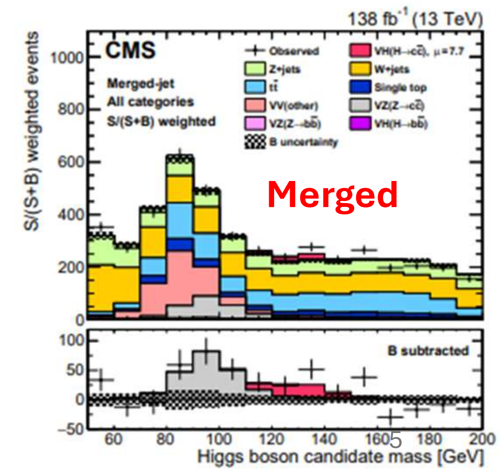
- 58% PN $c\bar{c}$ efficiency with 2% LF fake rate
- 40% DJ $c\bar{c}$ efficiency with 4% LF fake rate

Reduce $H \to b\bar{b}$ contamination from $\times\, 20 \to \times\, 2$ $H \to c\bar{c}$ yield

- In $Z \to c\bar{c}$ validation, $\times\, 2 \to \times\, 1.10$ reduction from $Z \to b\bar{b}$

Train BDT (resolved) for fit using event-level kinematics, c-jet discriminants, and Higgs candidate variables





Flavor-Tagging Options

DeepJet and ParticleNet are readily available for Run 2, Run 3

- Improvement upon DeepCSV as used for most pre-UL Run 2 analyses
- Only DeepJet resolved for an "easier" add-on to existing $HZ\gamma$

Particle Transformer (ParT) "incorporates pairwise particle interactions in attention mechanism"

- Supported by BTV, but is not available out-of-the-box

PAIReD resonant jet tagger for both resolved, boosted topologies on AK4 jets

- Developed by Spandan Mondal (Postdoc) at Brown
- Avoids using AK15 jets, targets $H \rightarrow c\bar{c}/b\bar{b}$ and $Z \rightarrow c\bar{c}/b\bar{b}$

Updated Categorization Strategy

1. Loosen the lepton multiplicity requirement from $N_\ell \geq 2 \rightarrow N_\ell \geq 1$
   - Targets Higgs production with associated $W^{\pm}H$ and single leptonic top-pair
2. Additional lepton category targeting leptonic $VH/t\bar{t}H$ with at least 3 leptons
3. New $HZ_{q\bar{q}}\gamma$ category targeting hadronic $VH/t\bar{t}H$ requiring two HF jets and 1 or 2 leptons
   - $\Delta m(q\bar{q}, Z)$ and $\Delta m(q\bar{q}\gamma, H)$ requirements
   - Can further define a ZH where the associated $Z \rightarrow c\bar{c}$
   - Separates "fake" VBF events which are hadronic $VH/t\bar{t}H$ events
4. VBF category requiring at least two jets with inverted $\Delta m(q\bar{q}, Z)$ and $\Delta m(q\bar{q}\gamma, H)$ requirements
   - Include MVA model for optimized categorization boundary
5. Un-tagged category (mostly ggH)
   - Include MVA model for optimized template likelihood fit

Pre-Selection Requirements

Apply ΔR matching between GEN-level quarks and jets

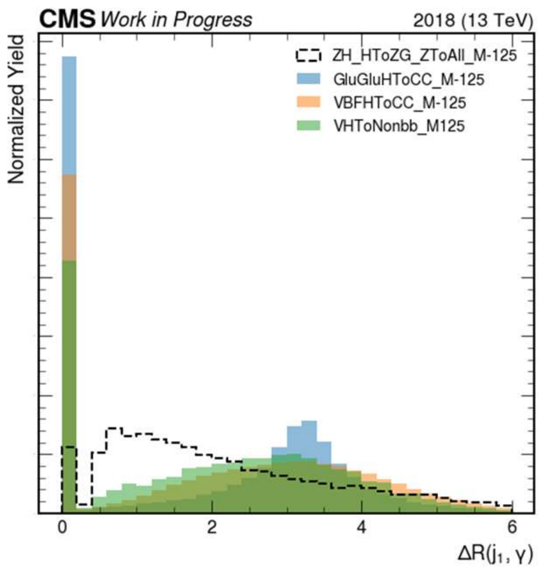
- Requiring $c\bar{c}$ to come from $H \to Z\gamma$ (70 events), but RECO will also include 5 (old VBF) $H \to Z\gamma \to \ell\ell\gamma$ events with the associated $Z \to c\bar{c}$

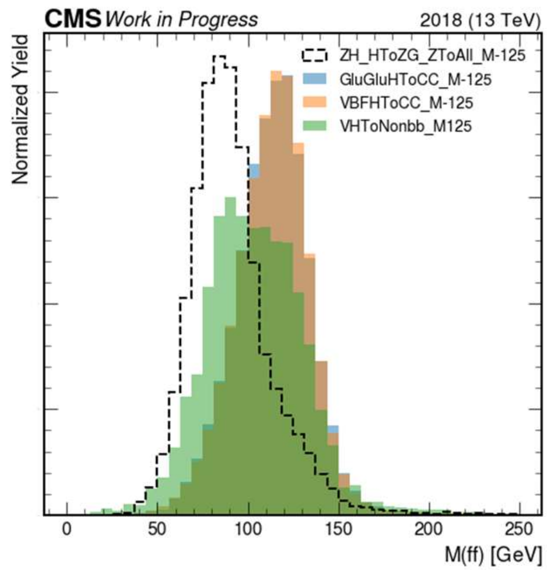
Using RECO-level objects with GEN-matching:

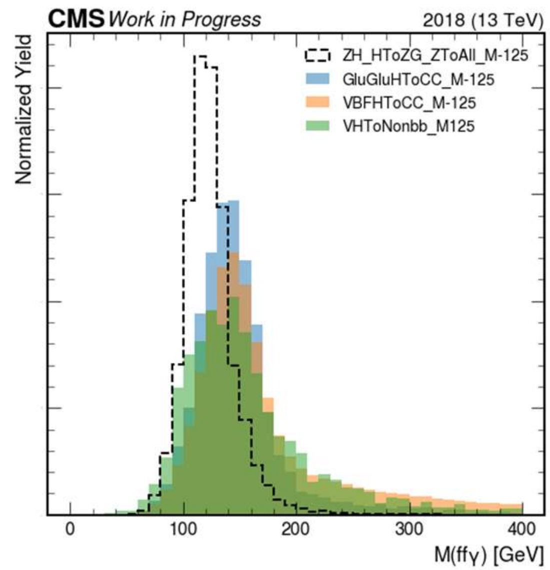
- ($N_\ell \geq 1$ and $N_{c/b} \geq 2$) or $N_\ell \geq 2$ with $|\eta| < 2.5$ (2.4) for electrons (muons, quarks)
  - Lead lepton $p_T \geq 25$ (20) GeV for electrons (muons)
  - Sub-leading lepton $p_T \geq 15$ (10) GeV for electrons (muons) when applicable
  - c-Jet $p_T \geq 25$ GeV and c-Jet $|\eta| < 4.0$
- $N_\gamma \geq 1$ with $|\eta| < 2.5$
- $HZ\gamma$ photon $p_T \geq 10$ GeV* and $\Delta R(\ell, \gamma) > 0.4$ or $\Delta R(j, \gamma) > 0.4$

Pre-Selection	Signal Yield	Efficiency
RECO/GEN-Matching Objects	52.8	100%
Photon $p_T > 10$ GeV and $\|\eta\| < 2.4$	48.7	92.2%
$N_\ell \geq 1$ (Trigger Requirement)	10.9	20.7%
Jet $p_T > 30$ GeV and $\|\eta\| < 2.4$	7.9	14.9%

Basic Cut Selection for $c\bar{c}$







Selection	Signal Yield (52.9)	Efficiency
$\Delta R(\gamma, j_1) > 0.4$ and $\Delta R(\gamma, j_2) > 0.4$	7.6	14.3%
$M(c\bar{c}) > 50$ GeV and $M(c\bar{c}) < 120$ GeV	6.4	12.0%
$M(c\bar{c}\gamma) > 80$ GeV and $M(c\bar{c}\gamma) < 150$ GeV	6.2	11.8%

Categorization Yields: $VH/t\bar{t}H \rightarrow HZ(c\bar{c})\gamma$

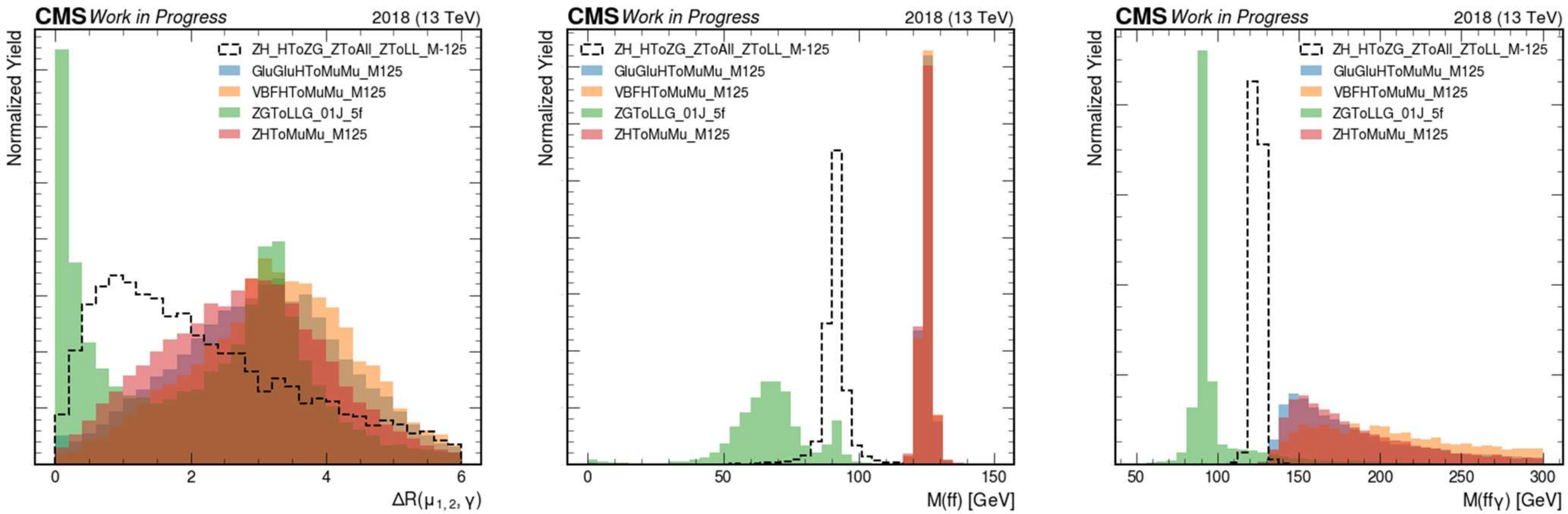
Category	N_ℓ	N_{HF}	N_{LF}	Sig.	VH	VBF	ggH	$S/\sqrt{B}$
Leptonic $VH/t\bar{t}H$	≥ 3			0.3	6.6	3.1	7.7	**0.07**
Hadronic HF $ZH/t\bar{t}H$	$= 2$	≥ 2 $m_{low}^{Z} < m(q\bar{q}) < m_{high}^{Z}$ $m_{low}^{H} < m(q\bar{q}\gamma) < m_{high}^{H}$		1.5	53.0	30.0	146.4	**0.10**
Hadronic HF $W^{\pm}H/t\bar{t}H$	$= 1$	≥ 2 $m_{low}^{Z} < m(q\bar{q}) < m_{high}^{Z}$ $m_{low} < m(q\bar{q}\gamma) < m_{high}$		3.9	218.5	183.1	1733.4	**0.08**
VBF	$= 2$	$m(q\bar{q}) \geq m_{high}^{Z}$ $m(q\bar{q}\gamma) > m_{high}^{H}$	≥ 2	0.0	0.0	0.0	0.0	**0.0**
ggH	$= 2$		≤ 1	0.0	1.9	1.0	7.7	**0.0**

- Signal is $VH/t\bar{t}H \rightarrow Z\gamma \rightarrow c\bar{c}\gamma$ and background is all $H \rightarrow c\bar{c}$, SM $Z\gamma \rightarrow c\bar{c}\gamma$ is negligible due to lepton trigger + $c\bar{c}$ dijet selection
- Using loose WP for DeepJet to define “HF” and “LF” categories
- The Δm_Z and Δm_H needs to be optimized, using $\Delta m_Z = 15$ GeV and $90. < m(q\bar{q}\gamma) < 140$ GeV → an ML method would perform better
- The Hadronic HF $ZH/t\bar{t}H$ category should improve sensitivity of VBF category as well
- The Hadronic HF $W^{\pm}H/t\bar{t}H$ category introduces events not included in the Run 2 selection
- The Hadronic HF $W^{\pm}H/t\bar{t}H$ category can include a E_T^{miss} and/or $M_T(\ell, E_T^{miss})$ cut to optimize non-prompt background filtering

Comparison to HIG-19-14

$138\ \mathrm{fb}^{-1}$	Lepton		Dijet 1	Dijet 2	Dijet 3	Untagged 1	Untagged 2	Untagged 3	Untagged 4
SM signal yield									
ggH	0.51	e^+e^-	1.10	1.62	9.44	6.89	7.35	29.8	22.5
		$\mu^+\mu^-$	1.41	2.05	12.1	8.52	9.17	38.0	29.0
VBF	0.09	e^+e^-	1.94	0.76	1.13	0.71	0.35	0.92	0.51
		$\mu^+\mu^-$	2.40	0.97	1.43	0.89	0.43	1.18	0.65
$VH + t\bar{t}H$	1.84	e^+e^-	0.04	0.13	1.89	0.31	0.17	0.45	0.27
		$\mu^+\mu^-$	0.05	0.16	2.36	0.39	0.21	0.57	0.33
SM resonant background									
$H \to \mu^+\mu^-$	0.14	$\mu^+\mu^-$	0.27	0.27	0.43	0.62	0.49	2.02	1.78
Mass resolution (GeV)	2.12	e^+e^-	1.91	2.06	2.15	1.80	1.97	2.12	2.33
		$\mu^+\mu^-$	1.52	1.61	1.72	1.37	1.42	1.62	1.83
Data yield	1485		168	589	11596	1485	1541	2559	17608
$S/\sqrt{B}$	0.06		0.54	0.24	0.26	0.45	0.35	0.53	0.30

Basic Cut Selection for $\mu\mu$



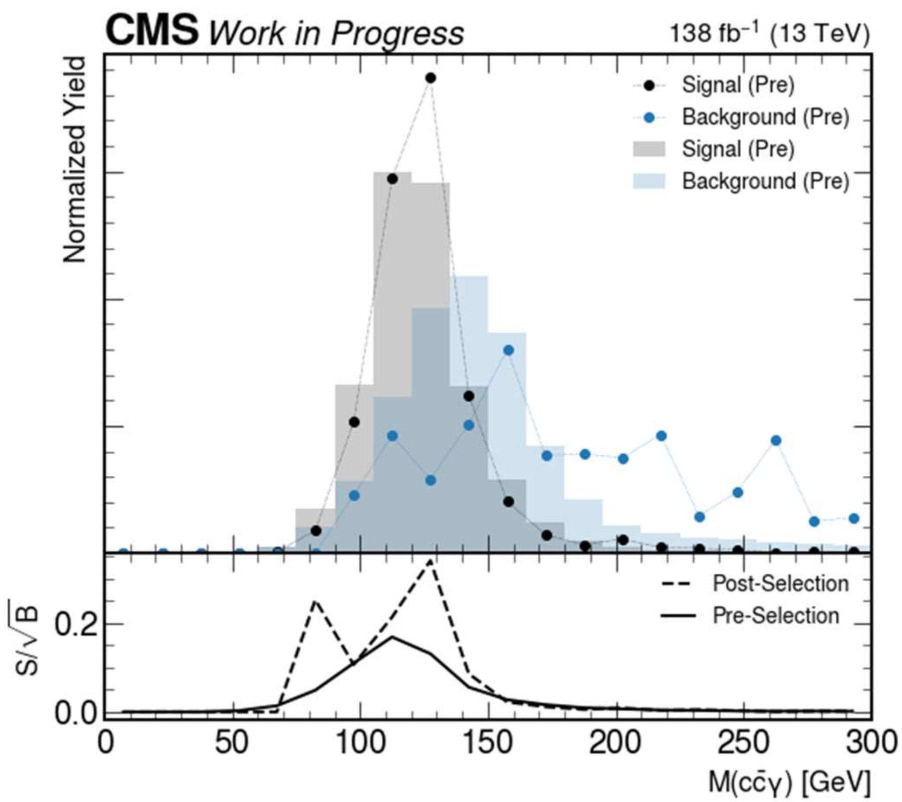
Selection	Signal Yield (41.4)	Efficiency
$\Delta R(\gamma, \mu_1) > 0.4$ and $\Delta R(\gamma, \mu_2) > 0.4$	36.7	88.6%
$M(\mu\mu) > 70$ GeV and $M(\mu\mu) < 110$ GeV	36.6	88.4%
$M(c\bar{c}\gamma) > 105$ GeV and $M(c\bar{c}\gamma) < 140$ GeV	36.6	88.4%

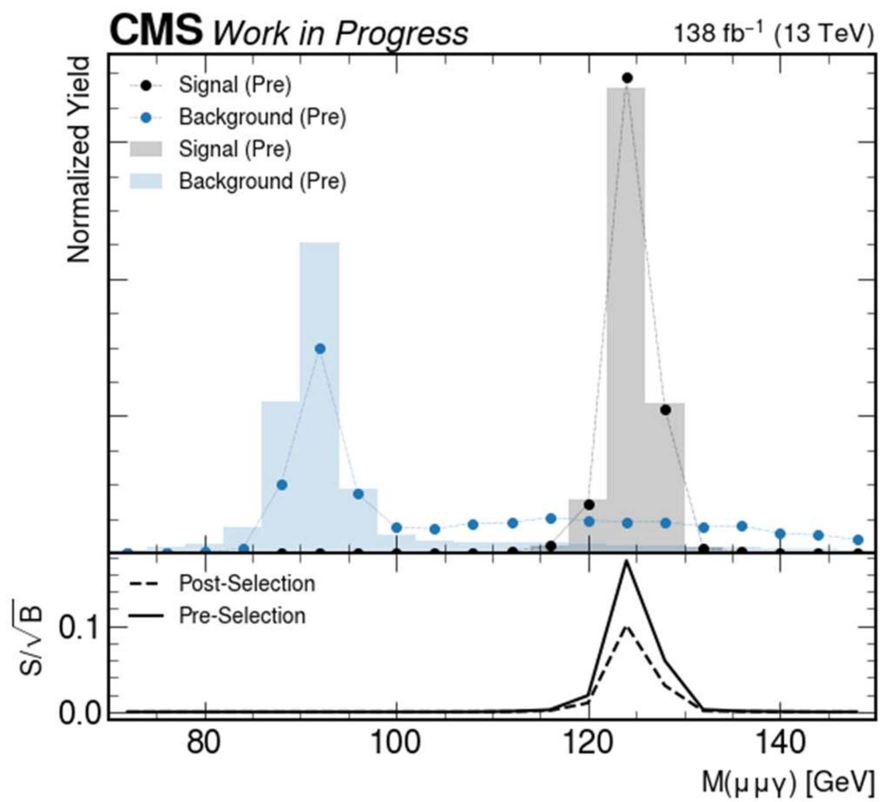
Categorization Yields: $VH/t\bar{t}H \rightarrow HZ(\mu\mu)\gamma$

Category	N_ℓ	N_{HF}	N_{LF}	Sig.	SM $Z\gamma$	VH	VBF	ggH	$S/\sqrt{B}$
Leptonic $VH/t\bar{t}H$	≥ 3			14.8	458.7	0.0	0.0	0.0	**0.69**
Hadronic HF $ZH/t\bar{t}H$	$= 2$	≥ 2 $m_{low}^{Z} < m(\mu\mu) < m_{high}^{Z}$ $m_{low}^{H} < m(\mu\mu\gamma) < m_{high}^{H}$		18.8	521.3	0.0	0.0	0.0	**0.82**
Hadronic HF $W^{\pm}H/t\bar{t}H$	$= 1$	≥ 2 $m_{low}^{Z} < m(\mu\mu) < m_{high}^{Z}$ $m_{low} < m(\mu\mu\gamma) < m_{high}$		0.0	0.0	0.0	0.0	0.0	**0.00**
VBF	$= 2$	$m(\mu\mu) \geq m_{high}^{Z}$ $m(\mu\mu\gamma) > m_{high}^{H}$	≥ 2	52.2	1328.4	0.0	0.0	0.0	**1.43**
ggH	$= 2$		≤ 1	17.9	779.5	0.0	0.0	0.0	**0.64**

- Signal is all $H \rightarrow Z\gamma \rightarrow \mu\mu\gamma$ and background is all $H \rightarrow \mu\mu$, SM $Z\gamma \rightarrow \mu\mu\gamma$
- Using loose WP for DeepJet to define "HF" and "LF" categories
- The Δm_Z and Δm_H needs to be optimized, using $\Delta m_Z = 20$ GeV and $105 < m(\mu\mu\gamma) < 140$ GeV → an ML method would perform better
- The Hadronic HF $ZH/t\bar{t}H$ category should improve sensitivity of VBF category as well
- The Hadronic HF $W^{\pm}H/t\bar{t}H$ category introduces events not included in the Run 2 selection
- The Hadronic HF $W^{\pm}H/t\bar{t}H$ category can include a E_T^{miss} and/or $M_T(\ell, E_T^{miss})$ cut to optimize non-prompt background filtering

$M(ff\gamma)$ Resolution after Selection $c\bar{c}$ vs $\mu\mu$





Possible Scenarios by Required Effort/Payoff

	Scenario	Effort Required	Payoff	Tasks
1	Baseline (Repeat Run 2)	🔥	$	New NanoAOD UL framework
2	Include Single Lepton triggers	🔥	$	Derive some corrections
3	Include Custom Photon ID	🔥 🔥	$ $	Dedicated ML model training Additional processing step
4	Include $c\bar{c}$ on AK4 Jets using DeepJet with SL/DL triggers	🔥 🔥	$ $ $	Additional event selection, category Kinematic fit to improve resolution?
5	Include $b\bar{b}$ on AK4 Jets using DeepJet with SL/DL triggers	🔥 🔥	$ $ $	Similar to $c\bar{c}$ but contributes less sensitivity due to $H \to b\bar{b}$
6	Include PAIRed tagger for HF AK4 Jets with a resolved/merged categorization	🔥 🔥 🔥	$ $ $	Implement custom dijet tagger Derive some corrections, studies
7	Include ParticleTransformer/Net for AK15 jets with a merged categorization	🔥 🔥 🔥 🔥	$ $ $	Implement custom jet tagger Inclusion of AK15 jets complicated
8	Include trijet 2016/2018 trigger for fully hadronic final states	🔥 🔥 🔥 🔥	$ $ $ $	Fully new analysis, requires new strategy

Considerations for Discussion

In addition to the gain in sensitivity, the appeal is:

- Successful precedent shown in the $VH \to c\bar{c}$ analysis
- Novel final state in $H \to Z\gamma$, not just repeating Run 2 strategy
  - Interesting physics differences with backgrounds (i.e. less SM $Z\gamma$), FSR radiation, dijet kinematics
- Incorporating the “latest and greatest” CMS machine learning algorithms
  - Improve beyond simple kinematic cuts for dedicated $HZ\gamma \to c\bar{c}$ vs $H \to c\bar{c}$ ML MVA ID
- Justification for a separate analysis from the “$t\bar{t}H$-motivated” approach
- Achieve CMS-only observation using Run 2 + partial Run 3 (2022-2024)
  - Not sufficient to combine $\ell^+\ell^-$ Run 2 + *full* Run 3 ($\sim$350 fb^{-1}) to make CMS-only observation
  - Run 2 ATLAS (2.2σ) and CMS (2.6σ) achieved 3.4σ on 280 fb^{-1}
- Possibility of targeting the $\text{H} \to Z_{\nu\bar{\nu}}\gamma$ final state for VH and $t\bar{t}$ associated production
- Possibility of fully hadronic VBF strategy using trijet trigger

Challenges in including the $H \to Z_{c\bar{c}}\gamma$ final state:

- Limited trigger options with focus on $V_{\ell^\pm}H$ and $t\bar{t}H$ production modes
- Non-trivial additional strategy that requires redefinition of Run 2 analysis categories
  - Also requires a broader analysis framework to handle jet flavor identification

Backup Slides

Trigger Selection

VH where $V = Z, W^{\pm}$ allows for using the leptonic decays of the Z and $W^{\pm}$ as the trigger signature

- Single lepton and opposite-sign, same-flavor dilepton

$t\bar{t}H$ allows for using the single lepton and dileptonic decays of the top-pair as the trigger signature

- Single lepton and opposite-sign dileptons (ee, $e\mu$ and $\mu\mu$)

Monophoton triggers have too large of a photon p_T requirement $\mathcal{O}(100)$ GeV while the γ from $H \to Z\gamma$ peaks around $10 - 20$ GeV

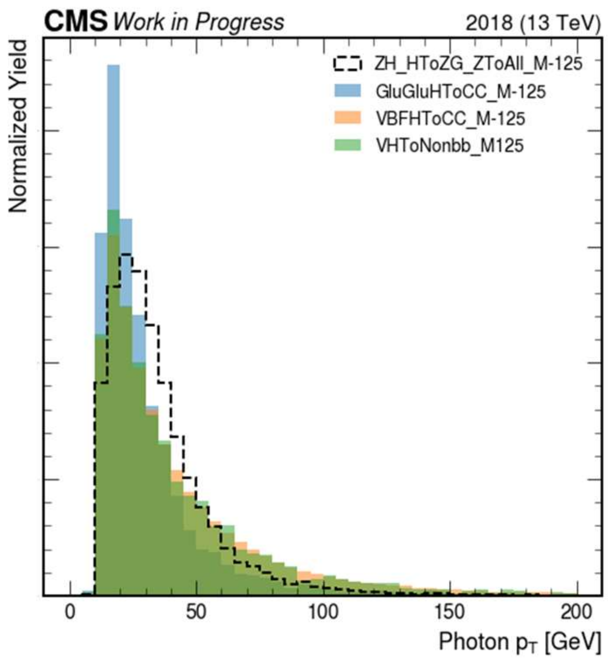
- Crossing photon trigger has $p_T > 60$ GeV with $H_T > 350$ GeV → possible non-negligible contribution from VBF production

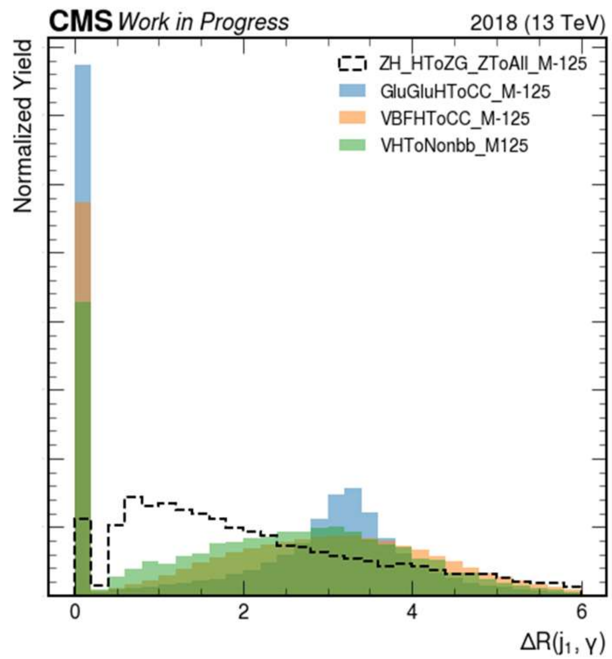
Monte Carlo Sample Details

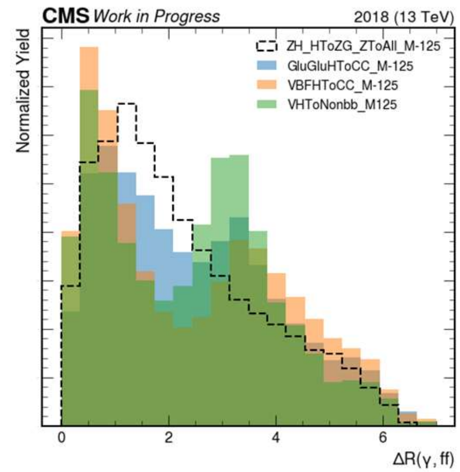
Using 2018 UL NanoAODv9 samples $\mathcal{O}(10^3 - 10^4)$ events from each:

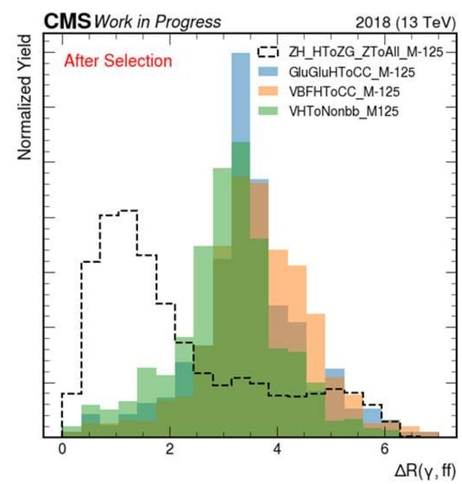
- ZH_HToZG_ZToAll_M-125_TuneCP5_13TeV-powheg-pythia8
- ZH_HToZG_ZToAll_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
- VBFHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
- GluGluHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
- ZH_HToCC_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
- ZHToMuMu_M125_CP5_13TeV-powheg-pythia8
- VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8
- VBFHToCC_M-125_TuneCP5_13TeV-powheg-pythia8
- VBFHToMuMu_M125_TuneCP5_withDipoleRecoil_13TeV-powheg-pythia8
- GluGluHToCC_M-125_TuneCP5_MINLO_NNLOPS_13TeV-powheg-pythia8
- GluGluHToMuMu_M125_TuneCP5_13TeV-powheg-pythia8
- ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8

Photon Kinematics: $c\bar{c}$



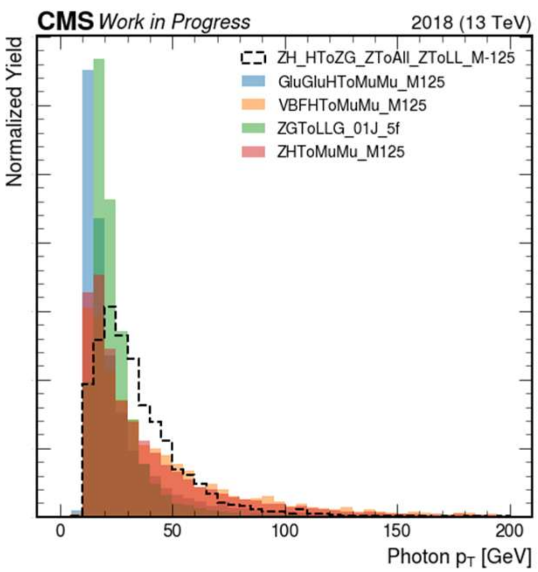


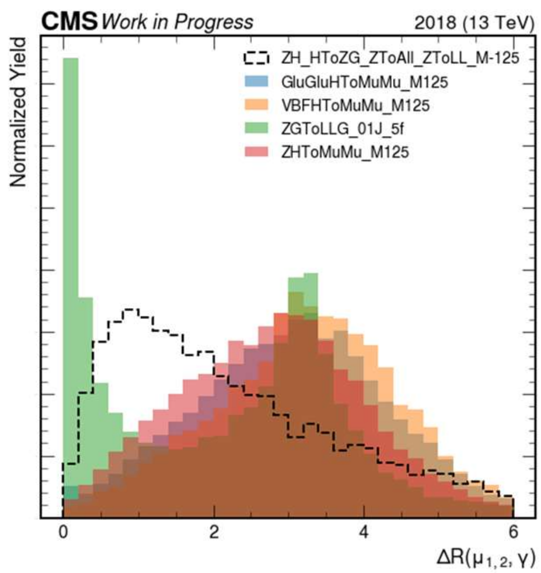


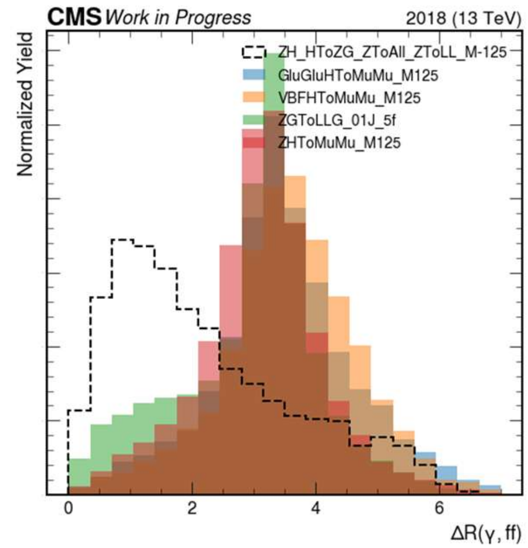


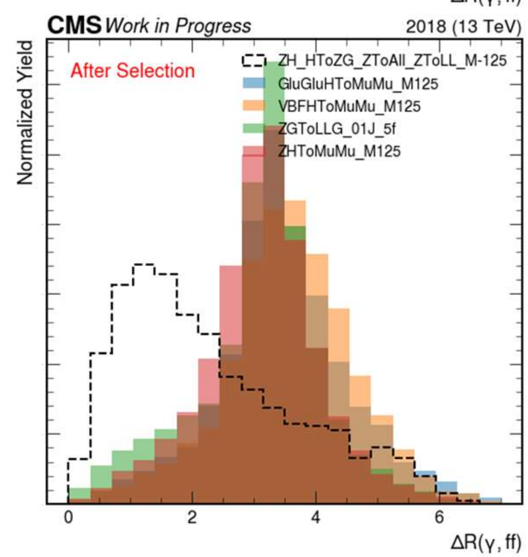
*Selection with detector constraints + $\Delta R(j, \gamma)$, no mass cuts

Photon Kinematics: $\mu\mu$



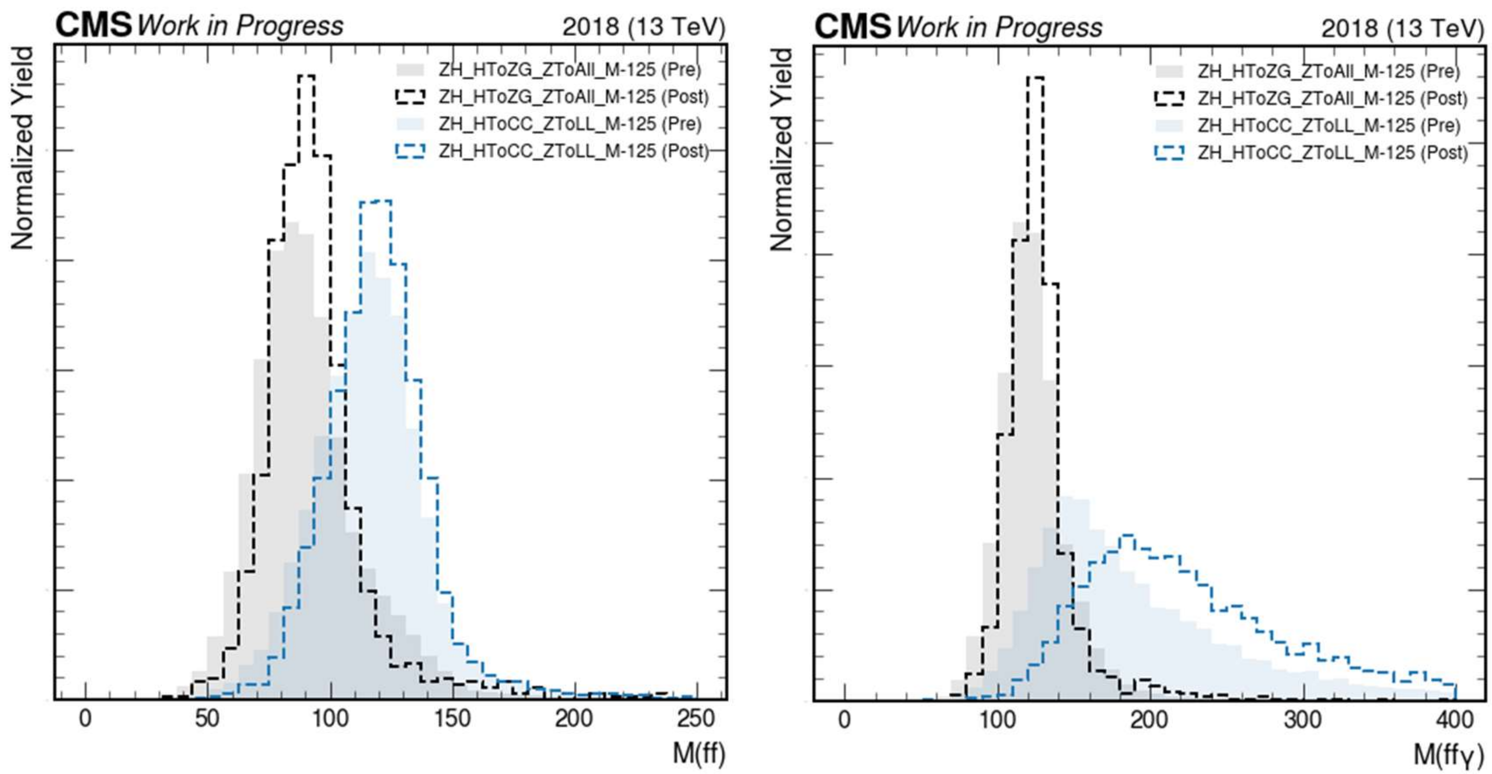




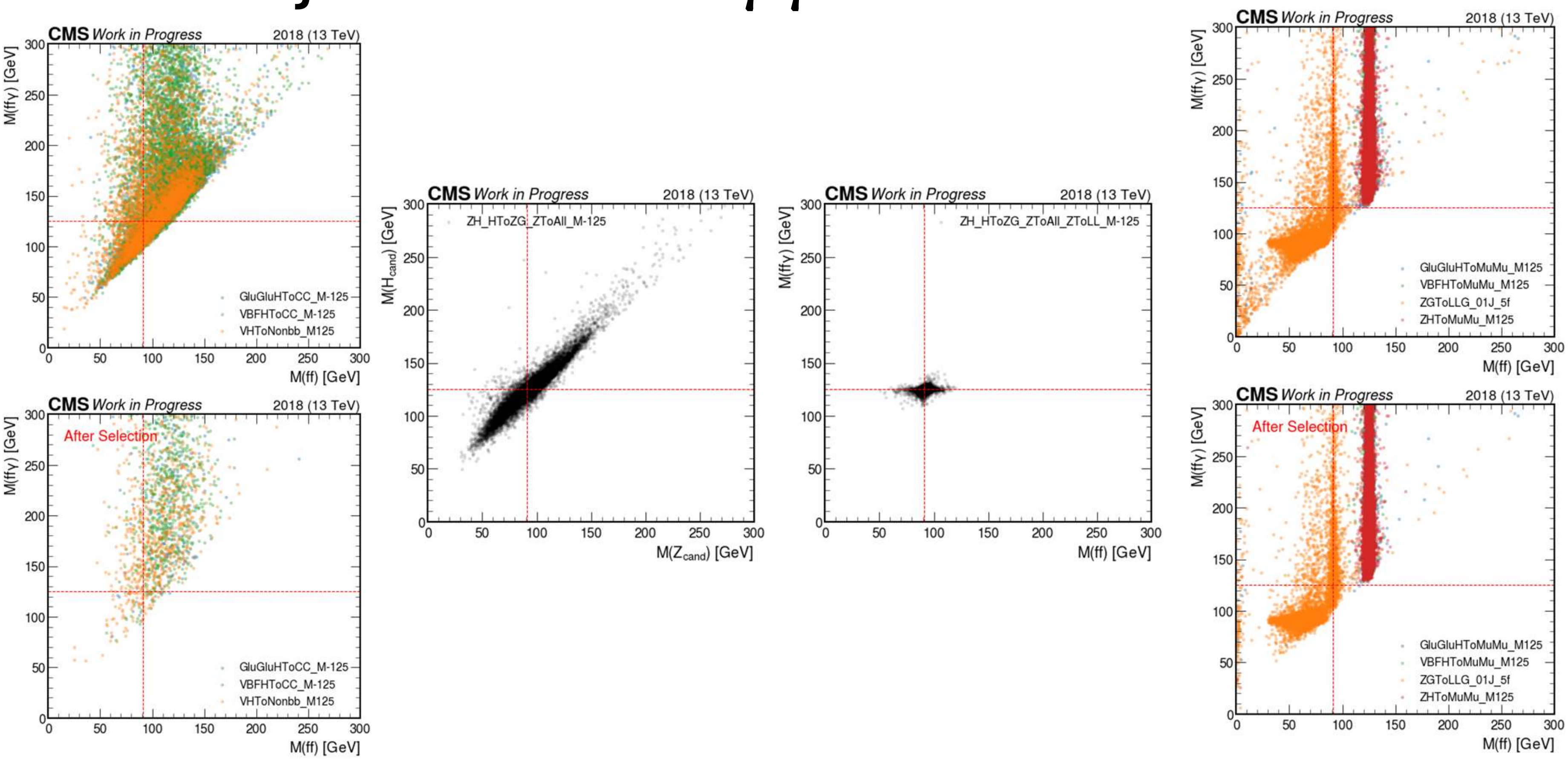


*Selection with detector constraints + $\Delta R(j,\gamma)$, no mass cuts

Z and Higgs Reconstruction: $c\bar{c}$



FSR Rejection: $c\bar{c}$ vs $\mu\mu$



Efficiency vs Photon p_T Cut