Methods for off-shell Higgs boson production simulation used in CMS analyses

Ulascan Sarica
for
the CMS Collaboration

LHC Higgs Off-shell subgroup meeting
Oct. 17, 2022

Introduction

We describe methods for off-shell Higgs simulation used in past CMS analyses:

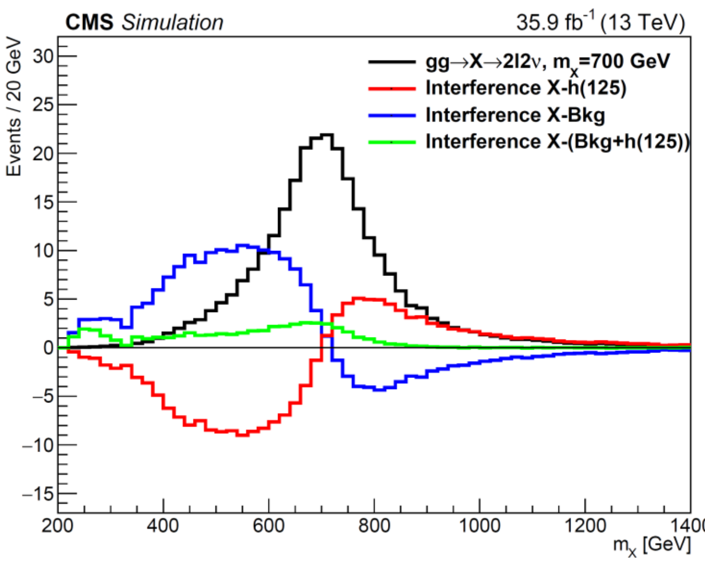
→ Need to include interference effects, consistent perturbative order for all components

→ Analyses use event categorization for gluon fusion and EW (mostly VBF in the SM)

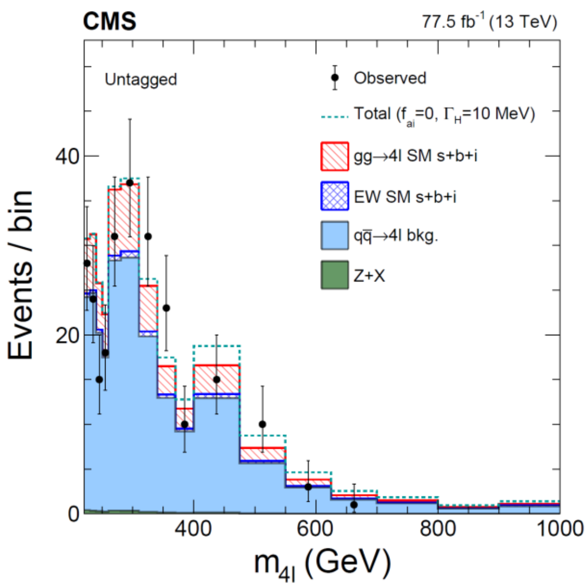
  - Need reliable info. on associated jets

→ Focus on ZZ→ $2\ell 2\nu$ or 4ℓ, but method also usable in WW analyses

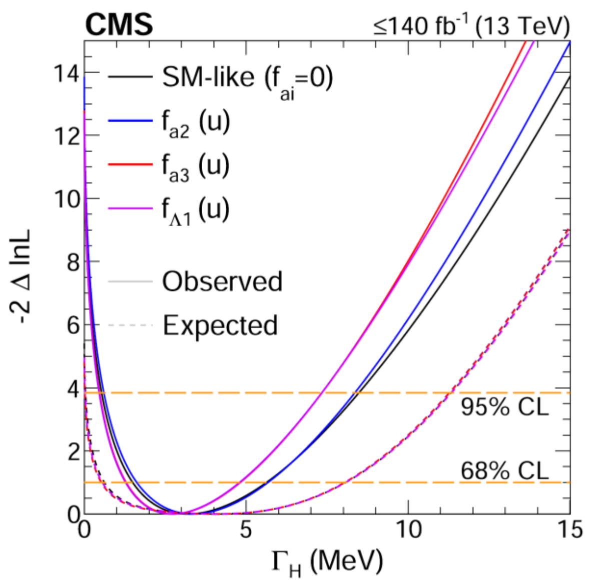
→ Recent CMS Note 22-010 with updated comparisons released last month



JHEP 03 (2020) 034



PRD 99, 112003 (2019)

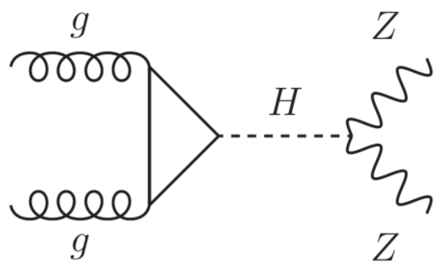


Acc. to Nature Phys.

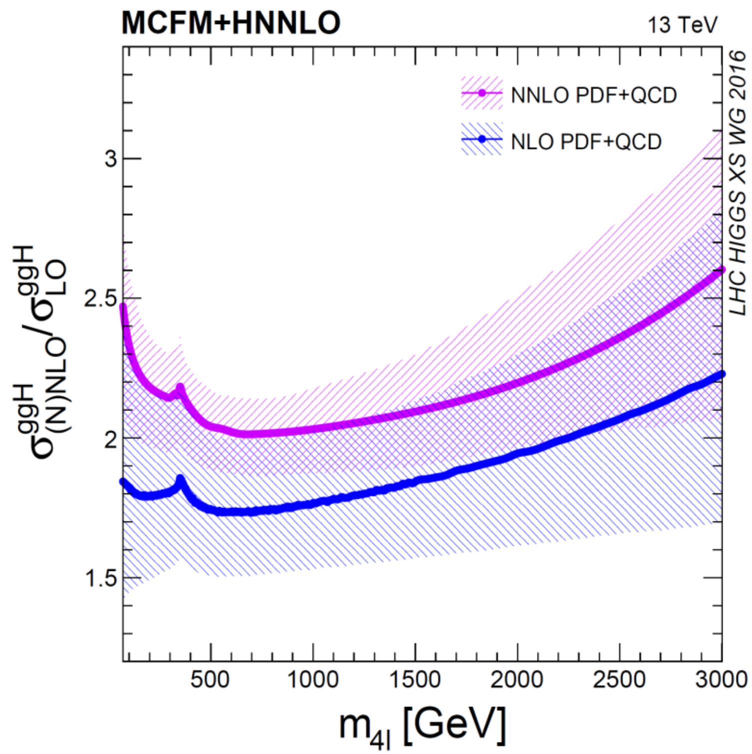
Gluon fusion process

Gluon fusion: Higgs amplitude

$gg \to H \to ZZ$:

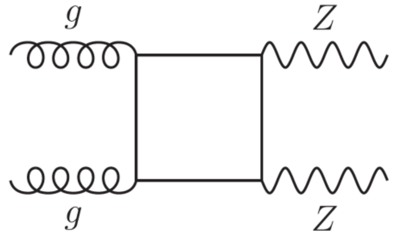


Full cross section calculation is available at different orders for the different components:
→ $gg \to H \to ZZ$: $\mathrm{N^3LO}$ in QCD around $m_H = 125$ GeV, NNLO for the full m_{ZZ} dependence, NLO or LO for event simulation
→ K-factors are large for NLO/LO (~1.7-1.8), smaller and flatter for NNLO/NLO (~1.2-1.3), and the $\mathrm{N^3LO}$/NNLO K-factor is 1.10.



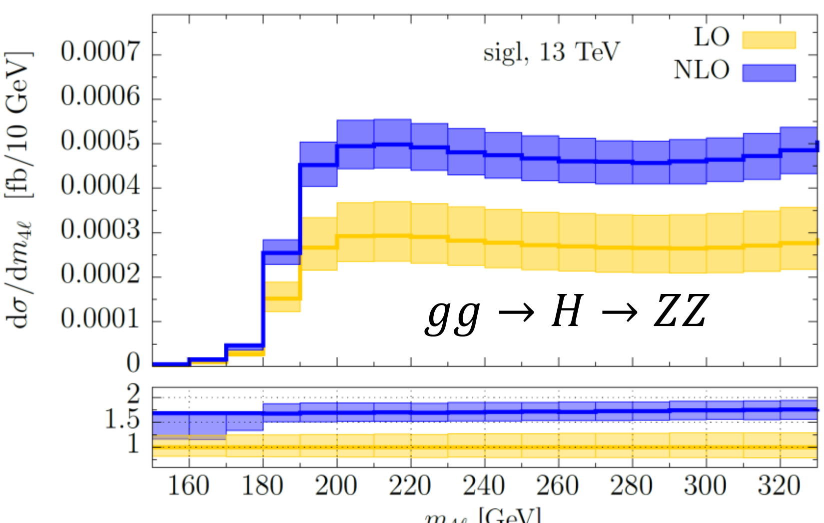
Gluon fusion: Continuum amplitude

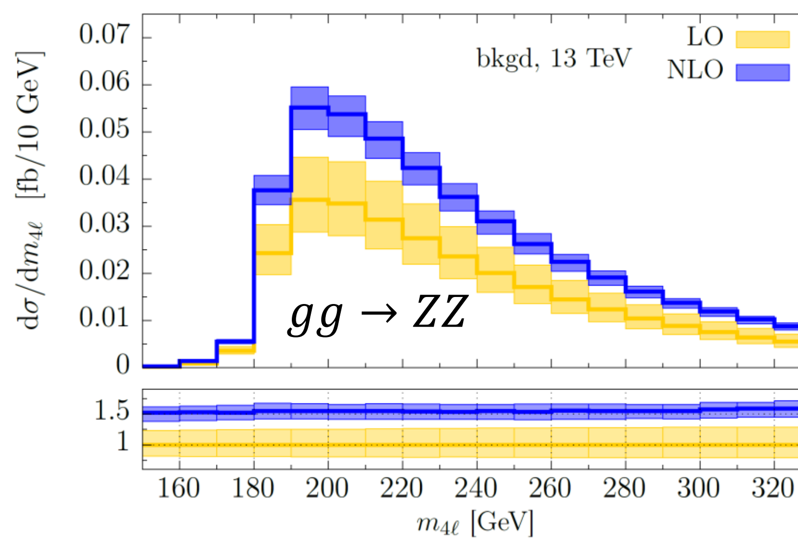
$gg \to ZZ$:



Full cross section calculation is available at different orders for the different components:
→ $gg \to ZZ$ continuum (and interference): Only full calculation and simulation with loop effects available at LO in QCD

- → Approximate NLO calculations (with Padé approximation with orders of m_t) show K-factors for $gg \to ZZ$ continuum, $gg \to H \to ZZ$, and their interference within ~10% suggesting corrections are mostly of soft/collinear nature
- → Current procedure is to use K-factors for $gg \to H \to ZZ$ on all components, and unc. $\kappa_{ggZZ} = 1 \pm 0.1$ on continuum with related scale $\sqrt{\kappa_{ggZZ}}$ on interference.





Gluon fusion: Event generation

For the Higgs amplitude contribution, continuum ZZ, or interference, MC event generation can be done in two ways:

→ Use JHUGen/MCFM to produce events at LO in QCD, apply NNLO K-factors and $\mathrm{N}^3\mathrm{LO}$ flat normalization

→ Relies on Pythia for jet multiplicity and kinematics

→ Use POWHEG to produce $gg \to H$, JHUGen for $H \to ZZ$, and the MELA matrix elements from JHUGen/MCFM (instead of event generation) to obtain continuum ZZ and interference.

→ POWHEG cannot produce off-shell line shape. Instead, produce samples for Higgs samples at m_H =125, 160 ... 200 ... 3000 GeV, which have increasingly larger widths.

→ $\mathrm{hfact} = m_H/10 + 37.5$ GeV to match p_T^H to NNLO+NNLL HRES predictions.

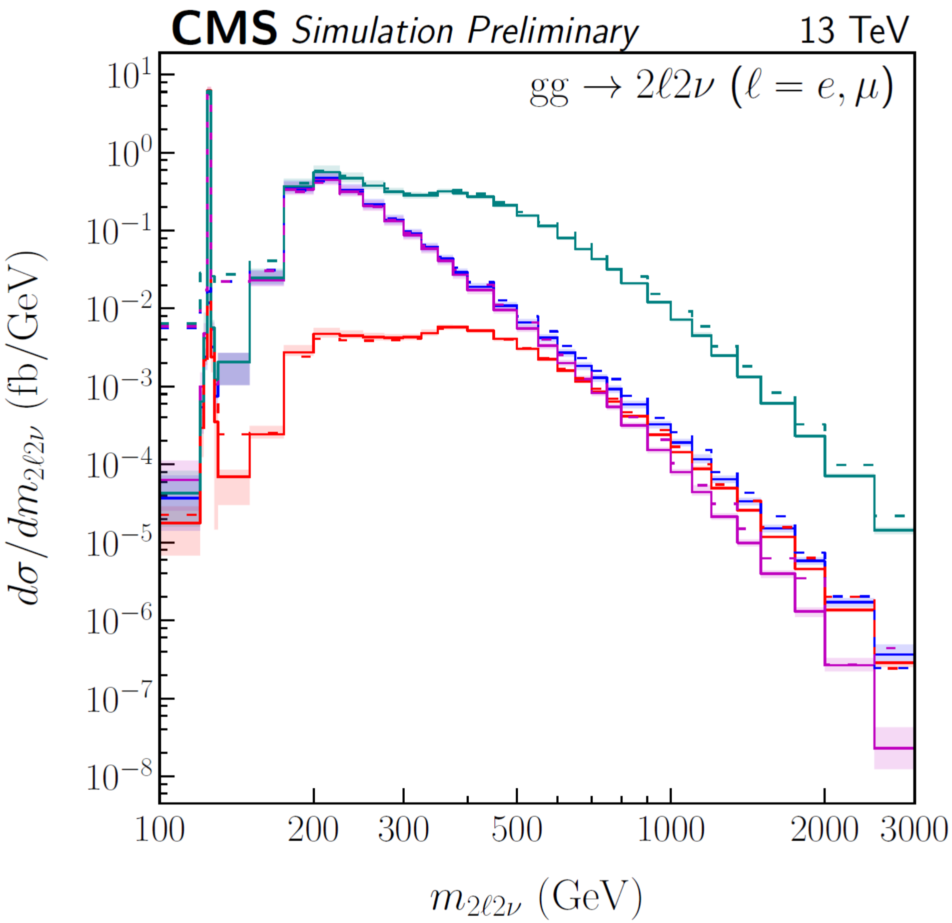
→ For the $gg \to H(125) \to ZZ$ amplitude, the only differences in these samples are the propagator and the correction of the m_{ZZ} line shape for the evolution of BR($H \to ZZ$).

The former is just reweighting the propagator to a BW($m_H = 125$ GeV, $\Gamma_H = 4.1$ MeV), so it is basically part of the MELA reweighting procedure, and the latter is added as a modification of event weights when running the JHUGen decay step.

→ The samples are glued together in the end to produce the full spectrum. The mathematical formulation is provided extensively in the note.

→ We observe this approach produces stable results in jet multiplicity and other kinematics after Pythia parton shower.

Gluon fusion: Distributions



→ Jet-inclusive m_{ZZ} distribution comparison exemplified with the $2\ell 2\nu$ final state

→ Almost perfect overlap between POWHEG+JHUGen and JHUGen/MCFM predictions

→ PS refers to the purely-pseudoscalar $H \to ZZ$ coupling with the same on-shell gluon fusion signal strength as SM since we also show an illustration of off-shell effects for a BSM scenario.

SM H signal ($|\mathrm{H}|^2$)　Total SM ($|\mathrm{H}+\mathrm{C}|^2$)

SM contin. ($|\mathrm{C}|^2$)　Total PS ($|\mathrm{PS}+\mathrm{C}|^2$)

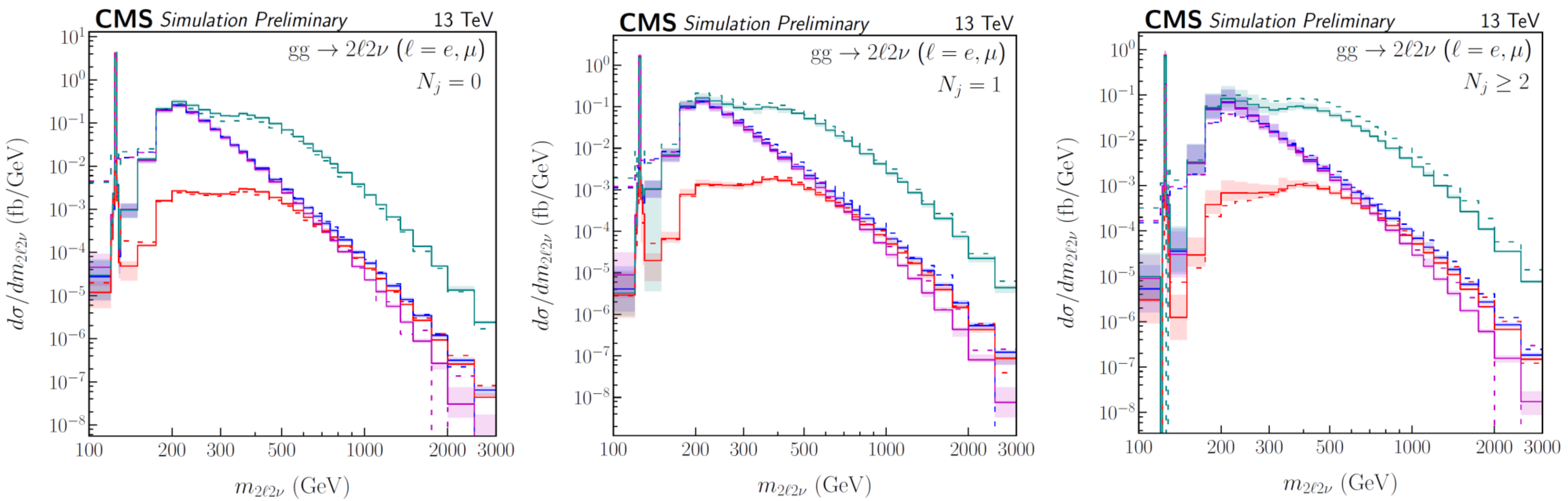
POWHEG+JHUGen

JHUGen/MCFM (LO QCD)

Legend is the same in other plots, so I won't repeat it on the next slides.

Constraints: $\mu_{\mathrm{sig}}^{\mathrm{on-shell}}(\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ} \to 2f2f') = 1$

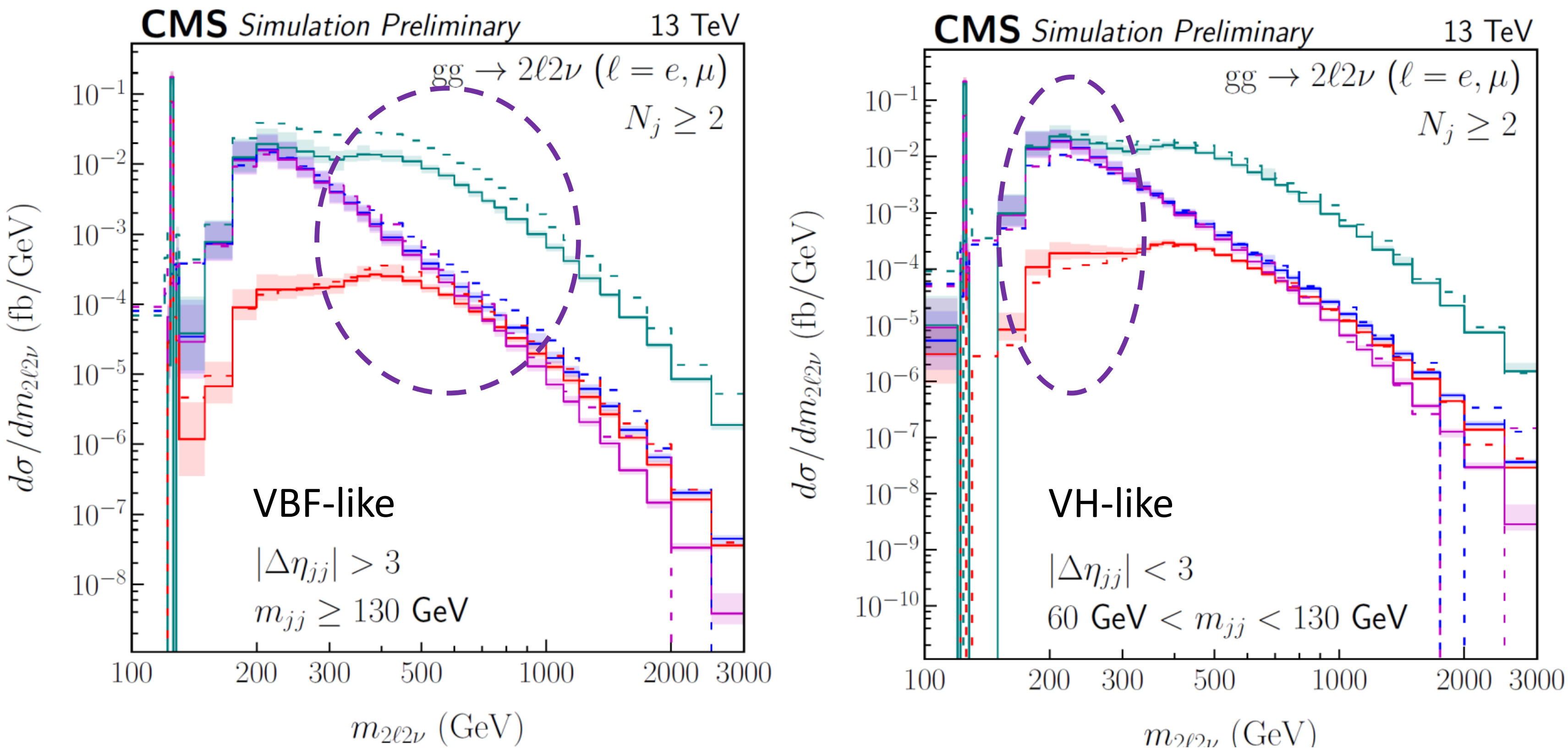
Gluon fusion: Jet-exclusive



When split by jet (*) multiplicity, $N_j = 0{,}1$ have similar levels of agreement
→ LO m_{ZZ} distribution distorted at $N_j \geq 2$

(*) Gen.-level anti-k$_T$ $\Delta R = 0.4$ jets with $p_T > 30$ GeV, $|\eta| < 4.7$

Gluon fusion: Focus on $N_j \geq 2$



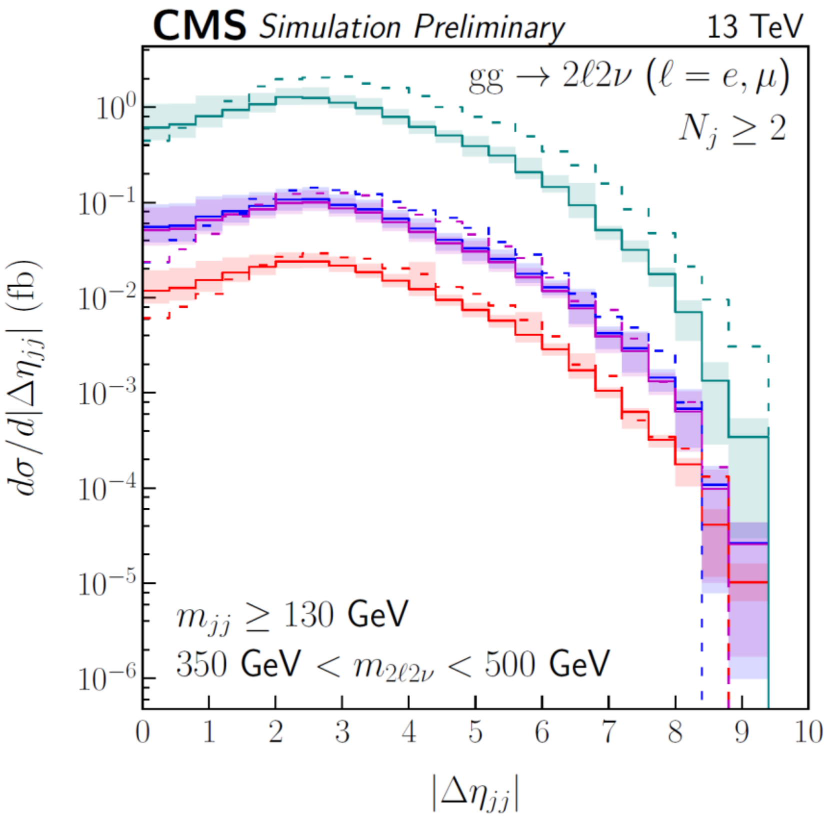
Since event cat. in CMS targets Higgs production mechanism, apply simple categorization on $N_j \geq 2$ events further into regions with a

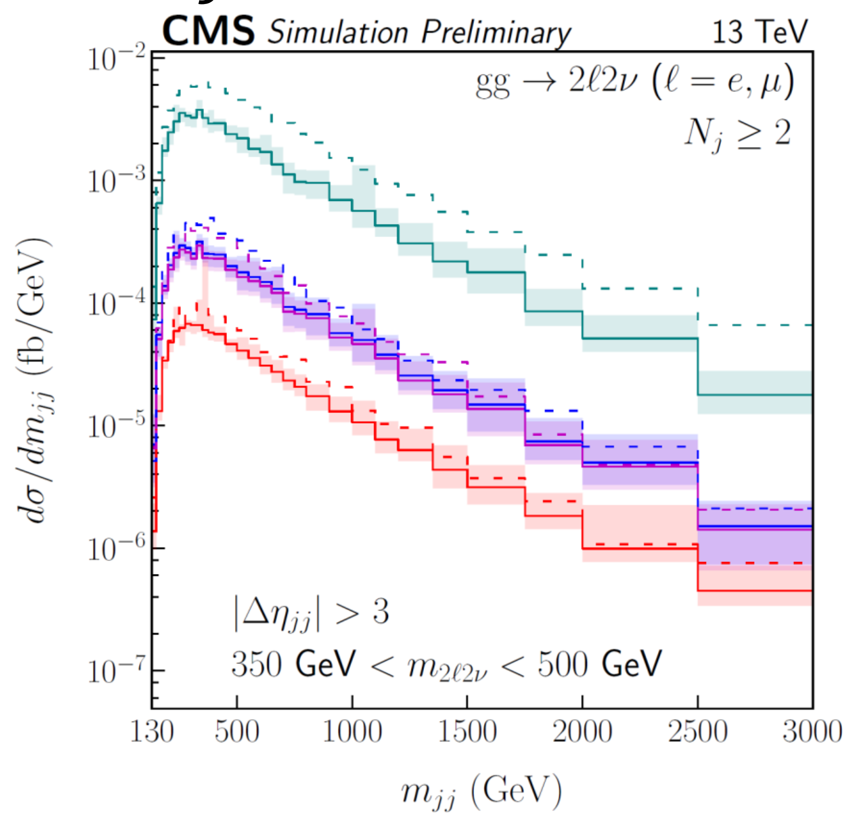
→ **VBF-like** topology: $m_{jj} > 130$ GeV, $|\Delta\eta_{jj}| > 3$

→ **VH-like** topology: $m_{jj} = 60 - 130$ GeV, $|\Delta\eta_{jj}| < 3$

Shape differences in seemingly opposite directions between the two regions

Gluon fusion: Focus on $N_j \geq 2$, VBF-like





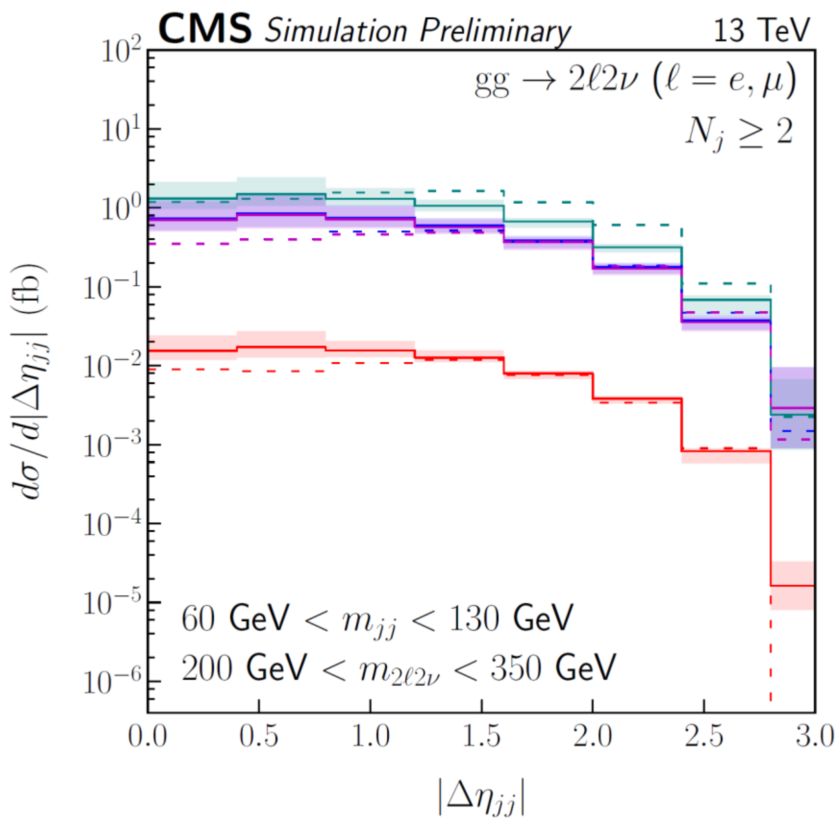
Zooming into $m_{ZZ} = 350 - 500$ GeV for the VBF-like topology,

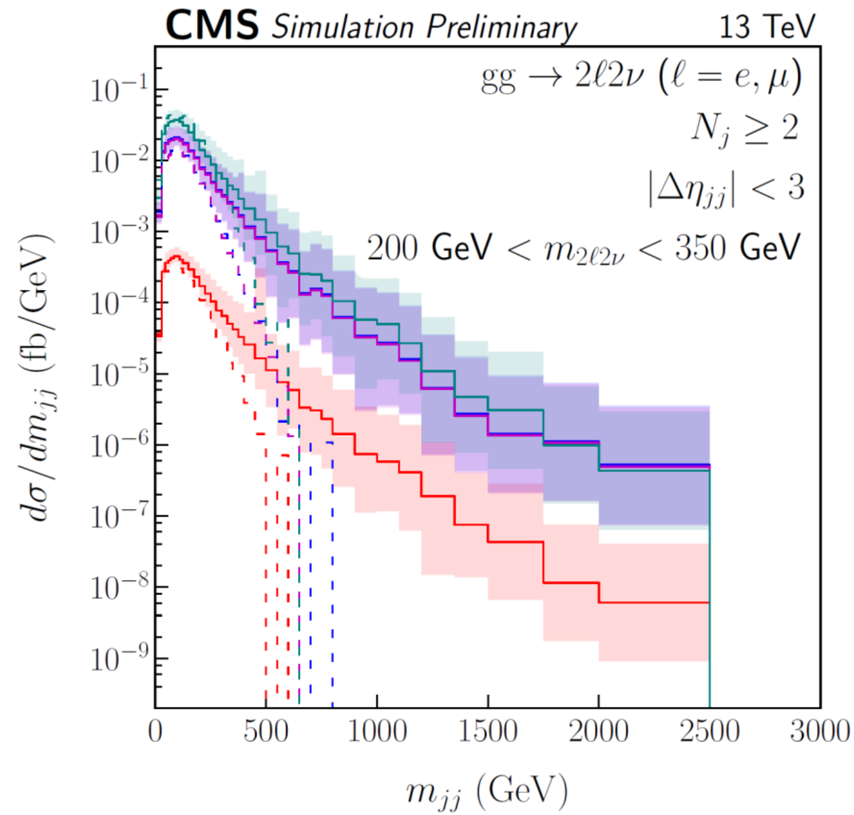
→ $|\Delta\eta_{jj}|$ shapes differ between LO and NLO samples at high m_{jj}

→ At $|\Delta\eta_{jj}| > 3$, m_{jj} distributions are peaked more at ~200 GeV in the LO sample

Distributions are also more stable in the NLO sample across m_{ZZ}.

Gluon fusion: Focus on $N_j \geq 2$, VH-like





Zooming into $m_{ZZ} = 200 - 350$ GeV for the VH-like topology,

→ Disagreement observed in $|\Delta\eta_{jj}|$ near 0 when $m_{jj} \sim m_V$.

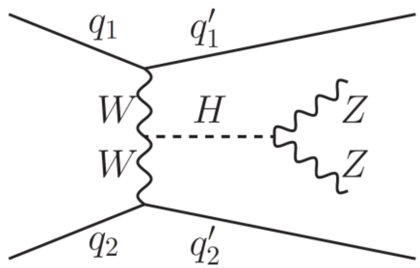
→ Faster decline in m_{jj} in the LO sample after parton shower at $|\Delta\eta_{jj}| < 3$

Distributions are also more stable in the NLO sample across m_{ZZ}.

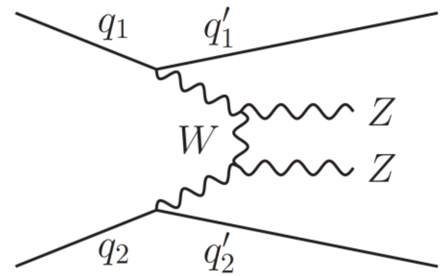
EW (VBF+VH/VBS+VZZ) process

EW process: Available methods

$q_1 q_2 \to q_1' q_2' H \to q_1' q_2' ZZ$:



$q_1 q_2 \to q_1' q_2' ZZ$:



Matrix element (MELA) and event simulation (MCFM/JHUGen) available for SM or BSM Higgs hypotheses, and continuum at LO in QCD consistently.

→ Improve event simulation technique for jet kinematics by

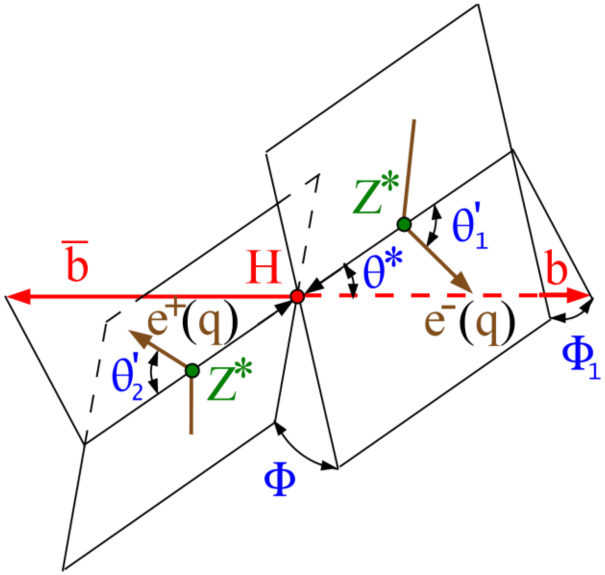
- starting with POWHEG+JHUGen samples for NLO VBF, and ZH and WH NLO + MiNLO HVJ) - apply MELA ME reweighting

→ Account for the extra partons from POWHEG by merging four-momenta of gluons (or $g \to q\bar{q}$ decays) to the closest quark

→ We demonstrate that the LO topology is approximated decently

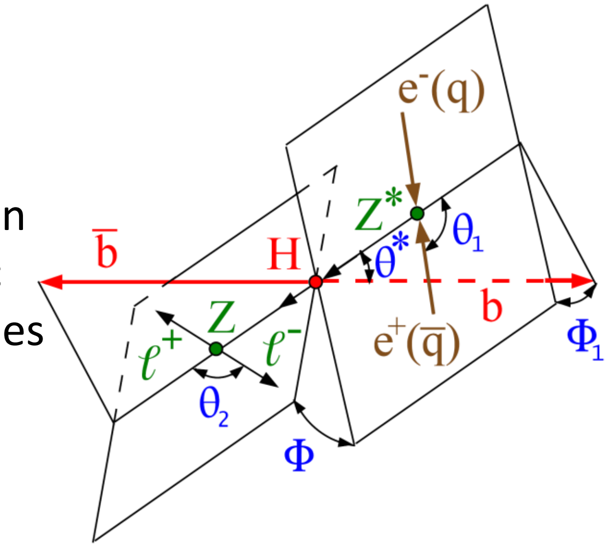
→ See backup for the merging details from the note

EW process: LO topology kinematics



VBF production kinematic observables

VH production kinematic observables

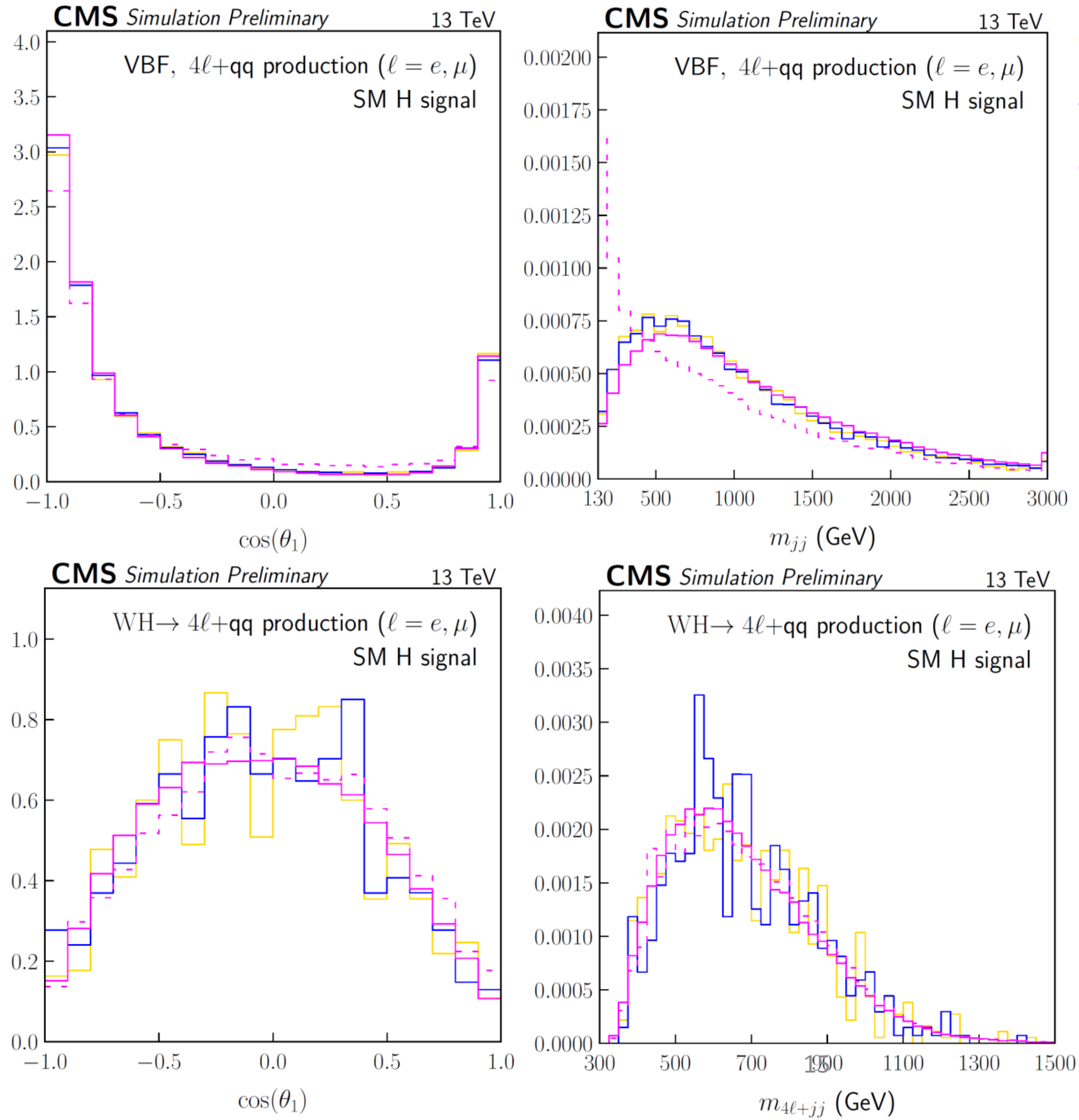


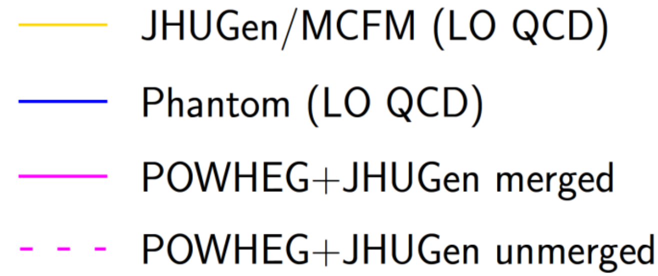
Can compare independent kinematic d.o.f.s in the VVH production vertex between NLO-merged vs actual LO distributions from the hard process:

→ This tests the validity of using LO MEs in reweighting.

→ Next couple of slides show a few examples.

EW process: LO topology comparisons



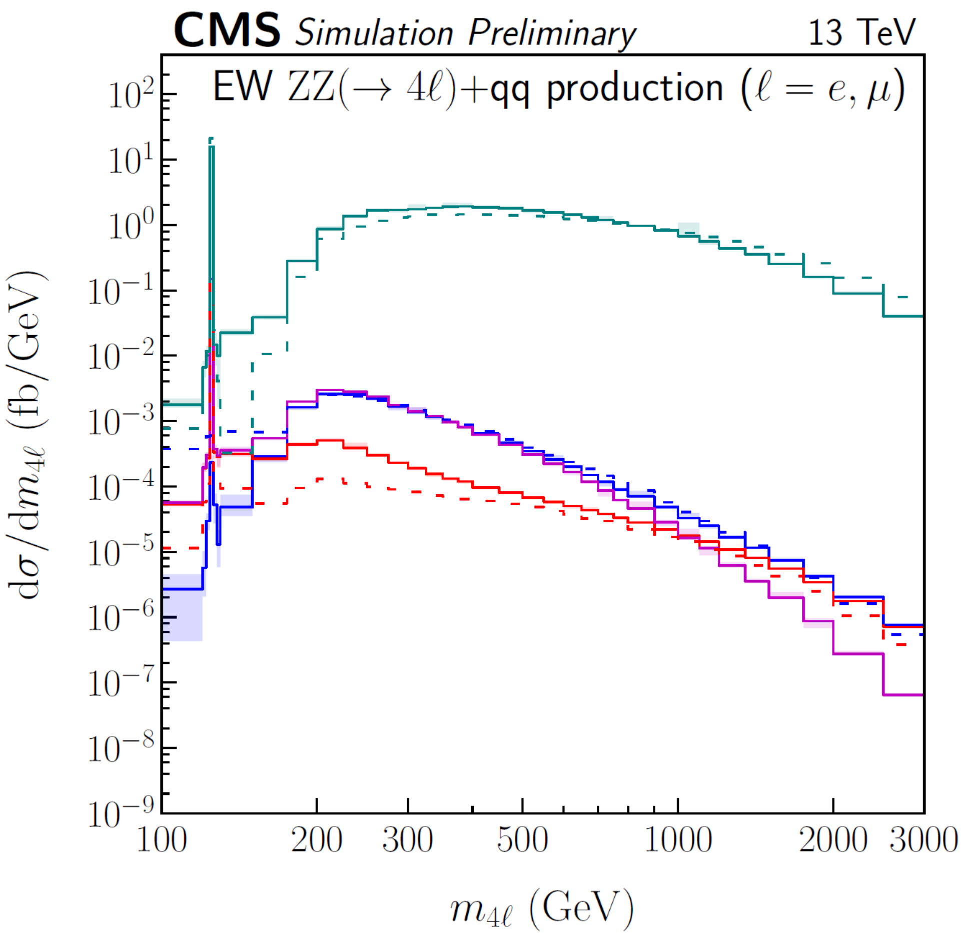


→ POWHEG unmerged uses only the two leading-p_T partons instead of deducing the LO topology, shown for illustration

→ Top plots for VBF-like topology ($m_{qq} > 130$ GeV), whereas bottom are for WH ($m_{qq} \sim 80$ GeV)

→ Decent matches in many kinematic variables, also in continuum hypotheses

EW process: Distributions after Pythia

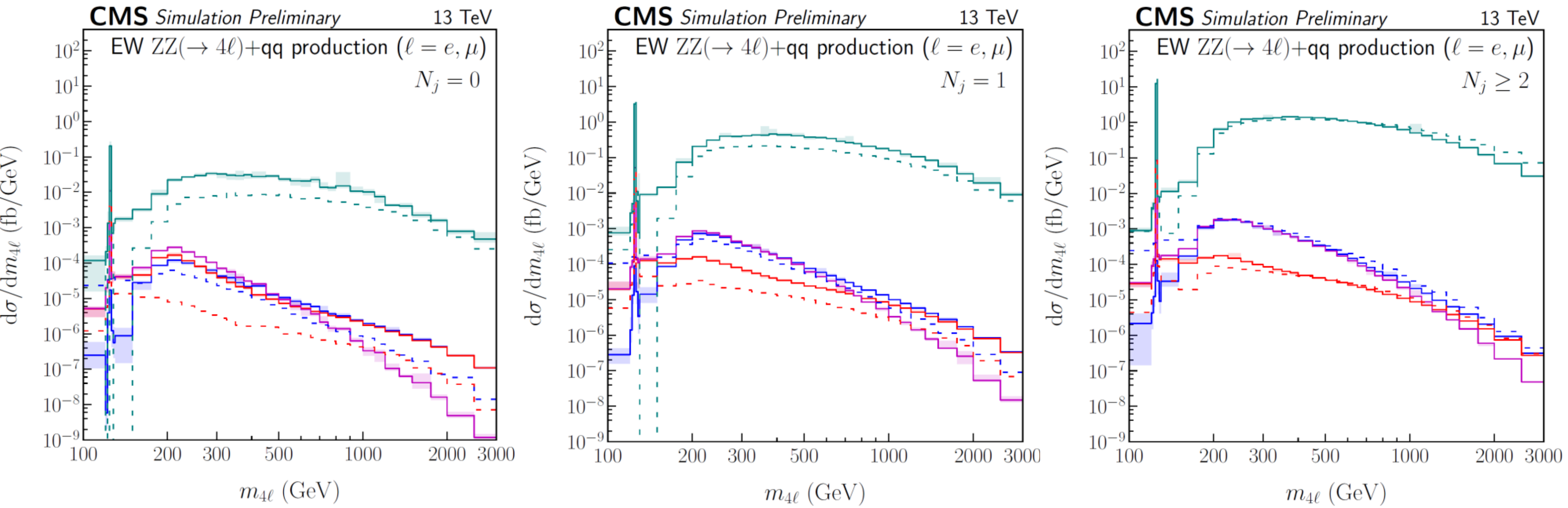


→ Jet-inclusive m_{ZZ} distribution comparison exemplified with the $4\ell + qq$ final state

→ Differences at $m_{ZZ} < 500$ GeV investigated further in jet-exclusive distributions

Constraints: $\mu_{\mathrm{sig}}^{\mathrm{on-shell}}(\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ} \to 2f2f') = 1$

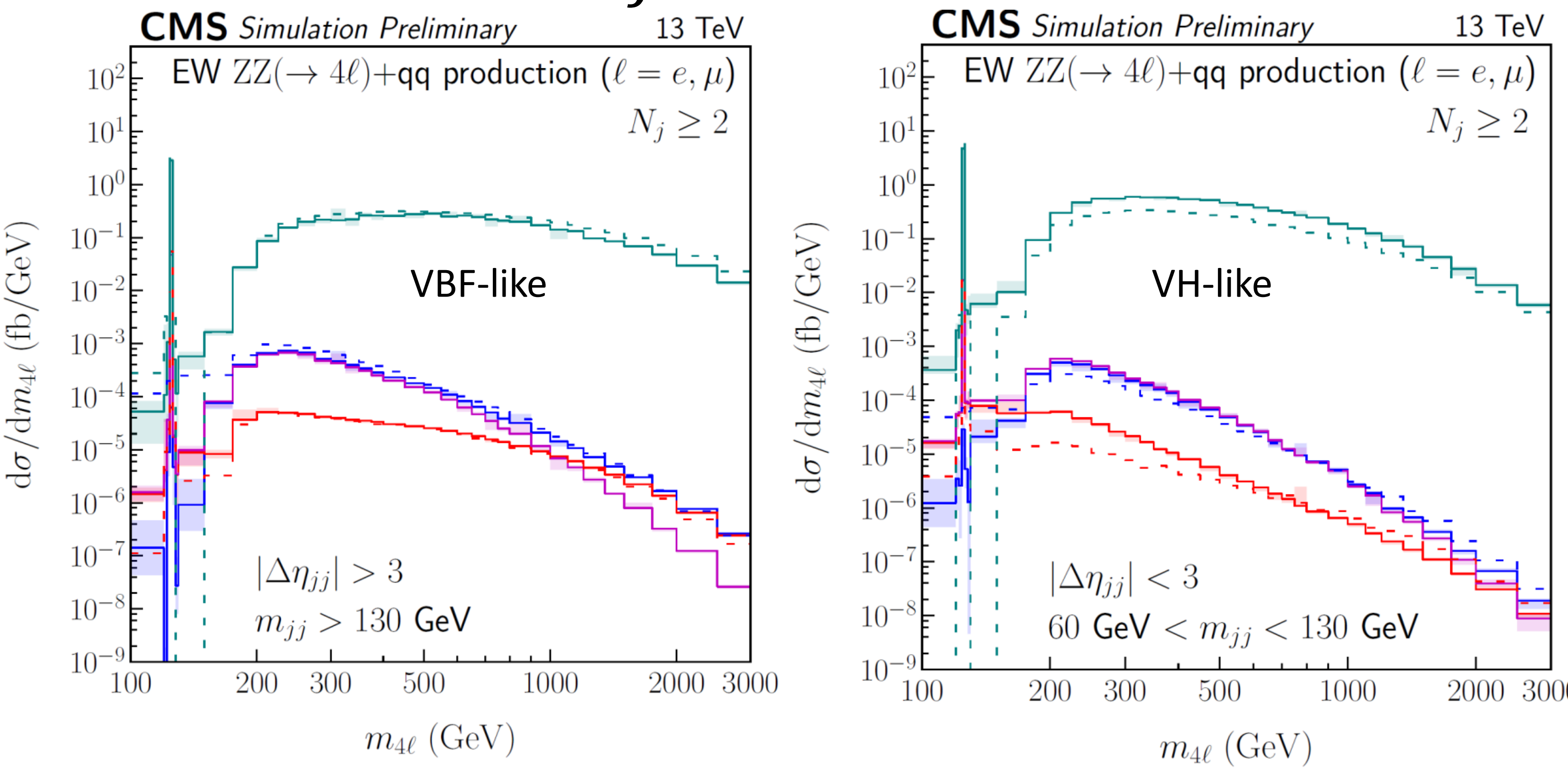
EW process: Jet-exclusive distributions



When events are split by jet multiplicity, we find differences in $N_j = 0$ and 1 larger.

→ Differences in H-only and continuum seem to go in similar directions, so the features are not attributable to the LO topology recasting.

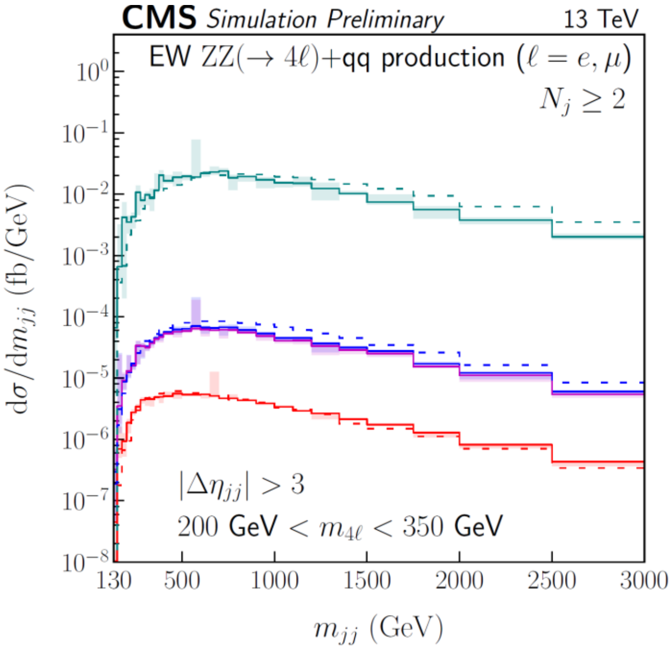
EW process: $N_j \geq 2$, VBF- and VH-like

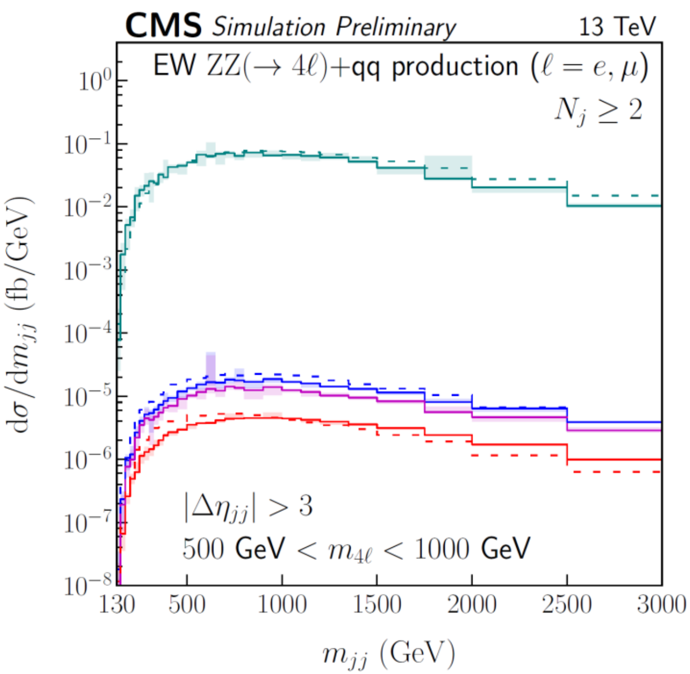


→ Decent match between LO samples and POWHEG prediction for the VBF-like topology

→ VH-like topology disagreement at $m_{ZZ} < 500$ GeV drives what is observed in the inclusive distribution shown before.

EW process: m_{jj}

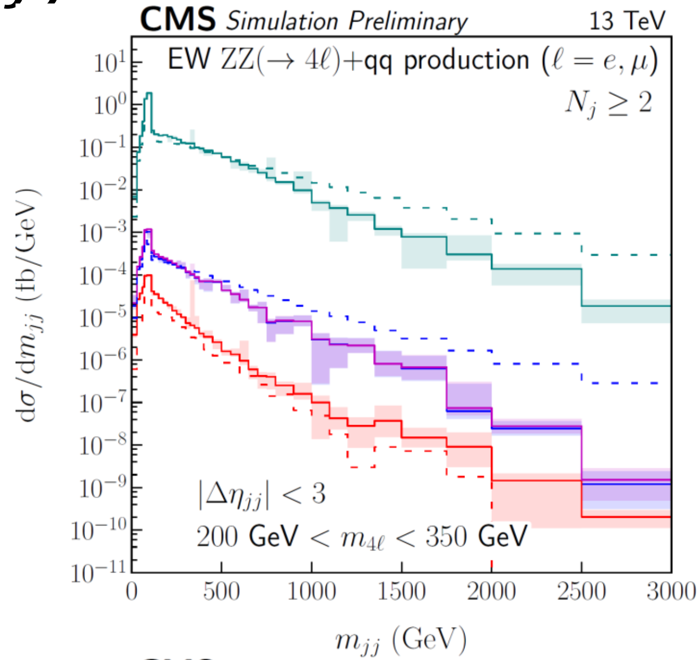


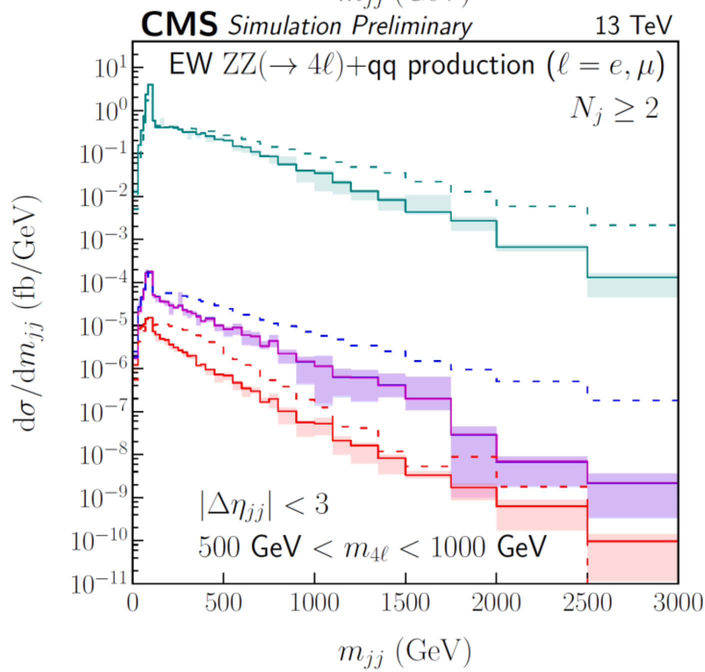


→ Decent agreement in at $|\Delta\eta| > 3$, but usually within the edge of the small uncertainties in POWHEG

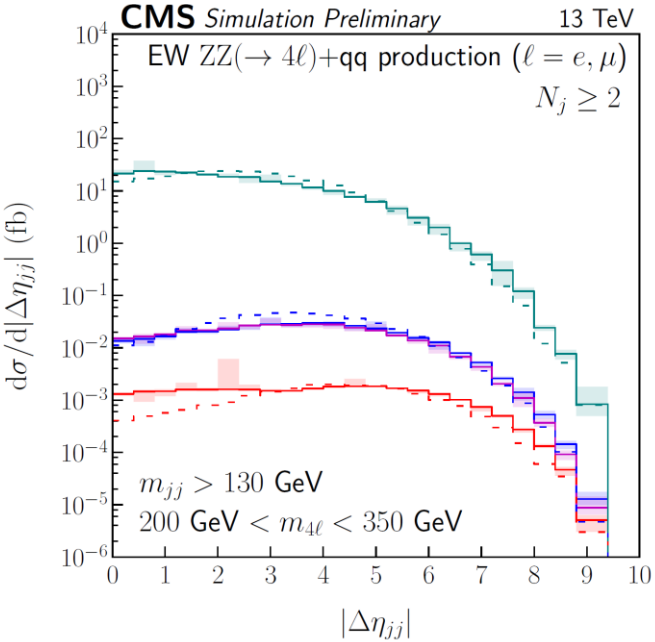
→ Continuum m_{jj} underneath $m_{jj} \sim m_V$ differs at $|\Delta\eta| < 3$

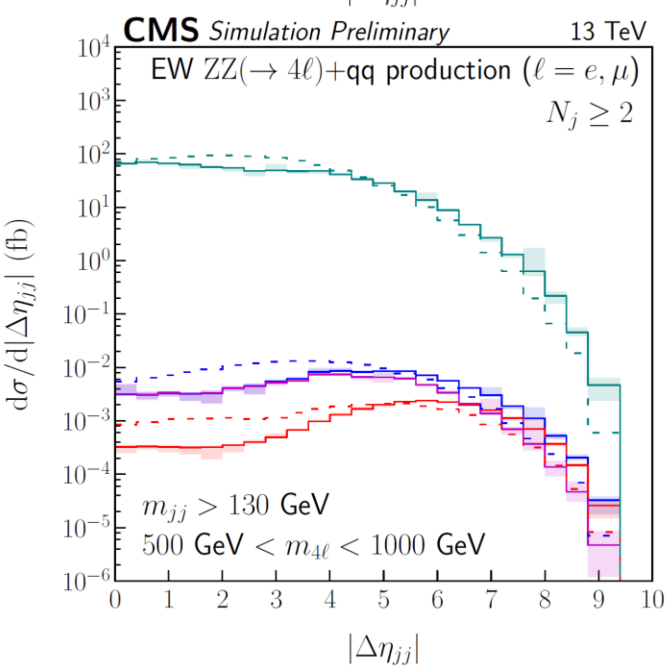
→ m_{ZZ} evolution of m_{jj} spectrum from POWHEG more stable at $|\Delta\eta| < 3$





EW process: $|\Delta\eta_{jj}|$

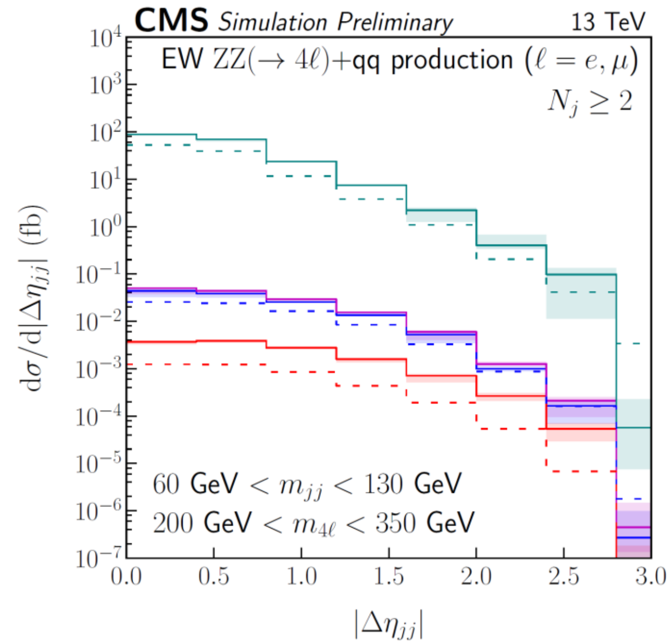


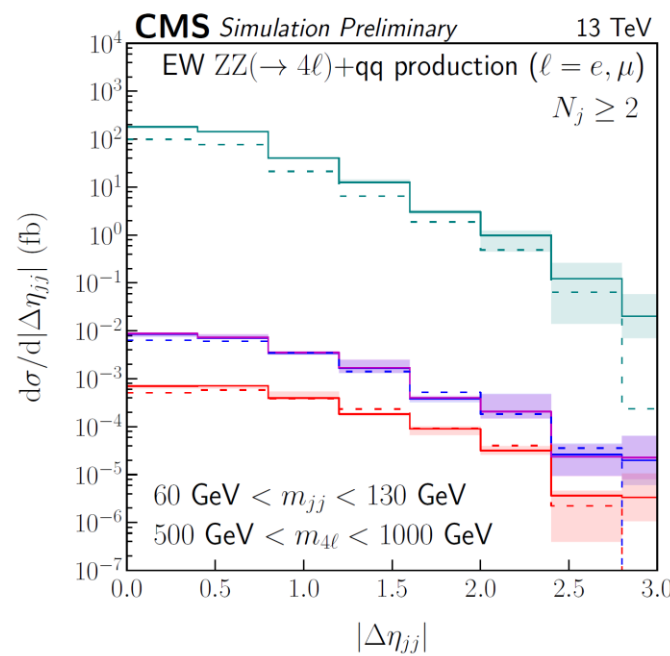


Disagreements on the previous slide can also be seen in these distributions.

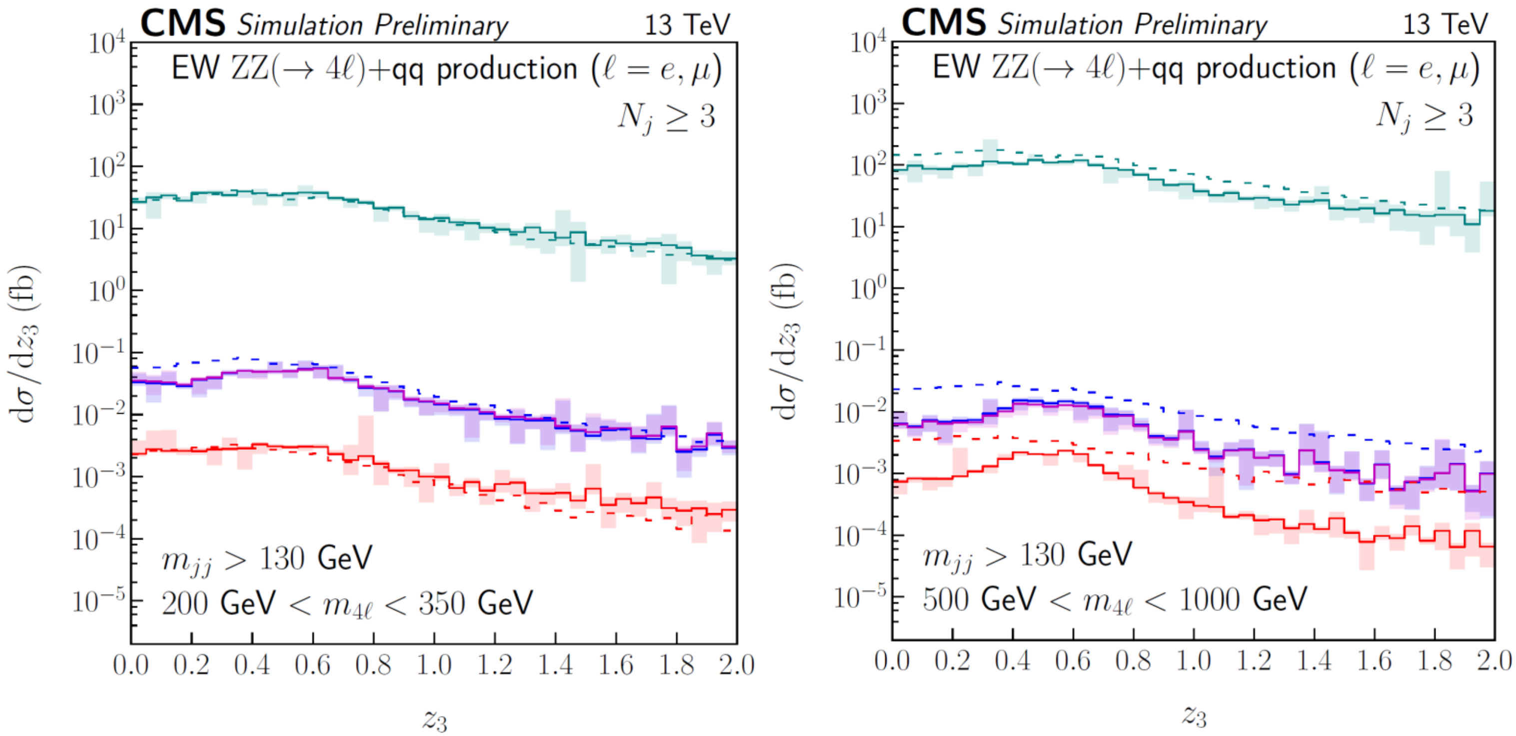
→ In opposite direction at low vs high m_{ZZ} in the $m_{jj} > 130$ GeV, $|\Delta\eta| < 3$ region.

→ Also at low m_{ZZ} when $m_{jj} \sim m_V$





EW process: Centrality of third jet (z_3)



Third jet more sensitive to parton showering schemes, but can still look at its centrality to understand jet radiation kinematics from Pythia better

→ The Zeppenfeld $z_3 = \frac{y_3 - (y_1+y_2)/2}{|\Delta y_{12}|}$ variable → 0 for central third jets

→ Distributions from POWHEG are close to best predictions at high m_{jj}, except for small known discrepancies near $z_3 = 0$.

→ While LO prediction agrees at lower m_{ZZ}, too many jets near $z_3 = 0$ at higher regions.

Summary

Presented simulation methods used in the off-shell/high mass Higgs analyses
→ Main Higgs production modes are gluon fusion, and the EW processes VBF, VH
→ Need to include interference effects with continuum ZZ, or BSM Higgs, at consistent orders in perturbative QCD

Analyses use jet kinematics to distinguish the Higgs production mechanism
→ Using matrix element reweighting methods on samples NLO in QCD are found to produce more stable distributions in jet multiplicity and kinematics across m_{ZZ}
→ Merging of the extra gluons at NLO needed in the EW process for reweighting to work, validated the recasting of event topology to LO equivalent

The CMS Note 22-010 would be a reference point for the CMS contributions to the LHC Higgs Off-shell subgroup studies in their future notes.
→ Can include these comparisons in a future write-up from this subgroup.

Backup

EW processes: Recasting NLO topology to LO

The MELA and MELAANALYTICS packages impose several rules on the merging procedure for the EW processes in order to make sensible predictions:

- An incoming gluon is never merged into an incoming quark. This rule is invoked implicitly as the q^2 of the incoming partons is always the largest compared to that of any other pair of partons in the event.
- Gluons are never merged into the decay products of the H boson from the JHUGen step as they are produced during the production of the H boson with no prior knowledge about the boson's decay.
- Gluons are also never merged into the decay products of the associated W or Z boson in the VH samples. Doing so distorts the BW nature of these resonances significantly.
- All merging is done in the convention of outgoing particles. This means the four-momenta and charge of incoming particles are reversed in the intermediate steps when those of the two merged particles are summed.
- When an incoming gluon is merged into an outgoing quark, the charge (i.e., PDG id) and the four-momentum of the quark are reversed in the final step of the LO topology construction. This reversion is done so that the event topology ensures having exactly two incoming quarks as expected in the LO matrix elements.
- In the VH samples, when extra gluons are encountered, the merging of individual gluons and that of a combined gluon (i.e., from a $g \to gg$ process) are all considered separately.
- In the VH samples, it is also possible to encounter two extra quarks instead of gluons. These extra quarks are merged into a gluon substitute first, as they are from a $g \to q\overline{q}$ branching process, before the merging of this gluon substitute is considered.
- Every merging permutation is considered, rated with the product of the dot-products between the merged quarks and gluons, and those that do not produce an incoming-outgoing parton composition that is compatible with the main physics process of the sample (i.e., VBF, ZH, or WH) are skipped.
- A momentum redistribution procedure is applied on the incoming and outgoing particles associated with H boson production so that the resultant topology features massless particles, which is what is required from the use of massless spinors in matrix elements. Denoting the momenta of the two final incoming or outgoing partons as p_1 and p_2, an intermediate four-momentum k is added to p_1 and subtracted from p_2 such that $|p_1 + k|^2 = |p_2 - k|^2 = 0$. This step is common to any matrix element computed using the MELA package. Because event-by-event reweighting is done through a ratio of matrix elements, which are invariant under any arbitrary boost of the event topology, and because factors coming from PDFs cancel in the ratio, the common boost of all particles does not affect reweighting as long as momentum conservation is maintained strictly, and is therefore adjusted arbitrarily.

Systematic uncertainties considered

- Renormalization scale: The renormalization scale is varied by a factor of 0.5 or 2, and the effect is included through per-event weights. In the case of the gg process, the weights are applied such that the relative variations resulting from these weights are scaled as a function of true m_{VV} to those for the inclusive k-factor variation predictions [20].
- Factorization scale: This source is considered to be uncorrelated from other uncertainty sources, and the scale is varied in the same way as the renormalization scale. The gg process is adjusted to match the inclusive k-factor variations in the same way as above.
- $\alpha_s(m_Z)$: A variation of $\alpha_s(m_Z) = 0.118 \pm 0.0015$ is considered. The variations for the gg process are adjusted to match those for the inclusive k-factor in the same way as above.
- PDF variations: This variation is taken as a conservative, envelope-type variation evaluated on a per-event basis over 100 variations from NNPDF [56] using the Hessian method. The gg process is adjusted to match the inclusive k-factor variations in the same way as above.
- Scale variations in parton shower: The scale variations are taken from additional per-event weights calculated from PYTHIA. The variation multiplicative factors considered are 0.25 and 4. The weights for the variations of initial- (ISR) and final-state radiation (FSR) are calculated separately, and the "down" and "up" variations are obtained by multiplying the ISR and FSR weights computed with the factors of 0.25 and 4, respectively.
- Simulation of the second jet in the gg samples: The uncertainty is evaluated as the difference of the nominal POWHEG samples generated at the pole masses $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ and $m_{\mathrm{H}} = 300\,\mathrm{GeV}$ from the simulation with the MINLO HJJ program [57], generated at the same pole masses. The reweighting factors are parametrized in three dimensions, in bins of m_{VV} below or above 150 GeV, in bins of $p_{\mathrm{T}}^{\mathrm{VV}}/m_{\mathrm{VV}}$, evaluated for the hard process, and in bins of 0, 1, and ≥ 2 in the number of jets (N_{j}), with the jet definition as aforementioned.
- Approximation of the NNLO QCD corrections to the continuum gg $\to$ ZZ contribution: Because the signal k-factor is applied on all contributions in the gg process, the continuum ZZ contribution is multiplied by a factor $k_{\mathrm{gg}} = 1 \pm 0.1$. The interference between the continuum amplitude, and the SM or BSM signal amplitudes is multiplied by $\sqrt{k_{\mathrm{gg}}}$ accordingly. The uncertainty of 0.1 is determined based on the level of disagreement observed between the amounts of corrections the SM signal and continuum ZZ contributions receive at approximately NLO in QCD [58].