Simulation-based inference of Higgs trilinear self-coupling via off-shell production

Aishik Ghosh, Maximilian Griese, Ulrich Haisch, Tae Hyoun Park

EPS-HEP 2025

THE UNIVERSITY OF CALIFORNIA IRVINE
MAX-PLANCK-INSTITUT FÜR PHYSIK

Motivation

- Rich phenomenology in off-shell Higgs boson production at the LHC.
  - Measurement of signal strength → Higgs width under SM-like assumptions [1]. ✅
  - **Breaks degeneracy in modifications to Higgs couplings under SMEFT treatment [2].** 🤔
  - **Probes the trilinear Higgs self-coupling through loop corrections [3].** 🤔

- Neural simulation-based inference, a new way to perform parameter estimation.
  - Neural networks can be trained to predict otherwise intractable likelihood ratios [4]. ✅
  - **"Gold can be mined" from simulation to expedite NN training & accuracy [5].** 🤔

What is the future potential of off-shell Higgs physics?

1. ATLAS Collaboration (and previous theory work)
2. ATLAS Collaboration (and previous theory work)
3. Ulrich Haisch, Gabriël Koole
4. Kyle Cranmer, Juan Pavez, Gilles Louppe
5. Johann Brehmer, Gilles Louppe, Juan Pavez, Kyle Cranmer

Outline

1
- Off-shell Higgs production at the LHC
  - SM processes
  - SMEFT modifications

2
- Neural simulation-based inference
  - Classification vs. Regression-based methods
  - Probability ratio mixture model

3
- Results: SM-expected sensitivity at HL-LHC
  - C_H from $gg \to (h^* \to)ZZ$
  - C_H, C_{tH}, C_{HG} from $pp \to ZZ$

Pre-print:
arXiv:2507.02032

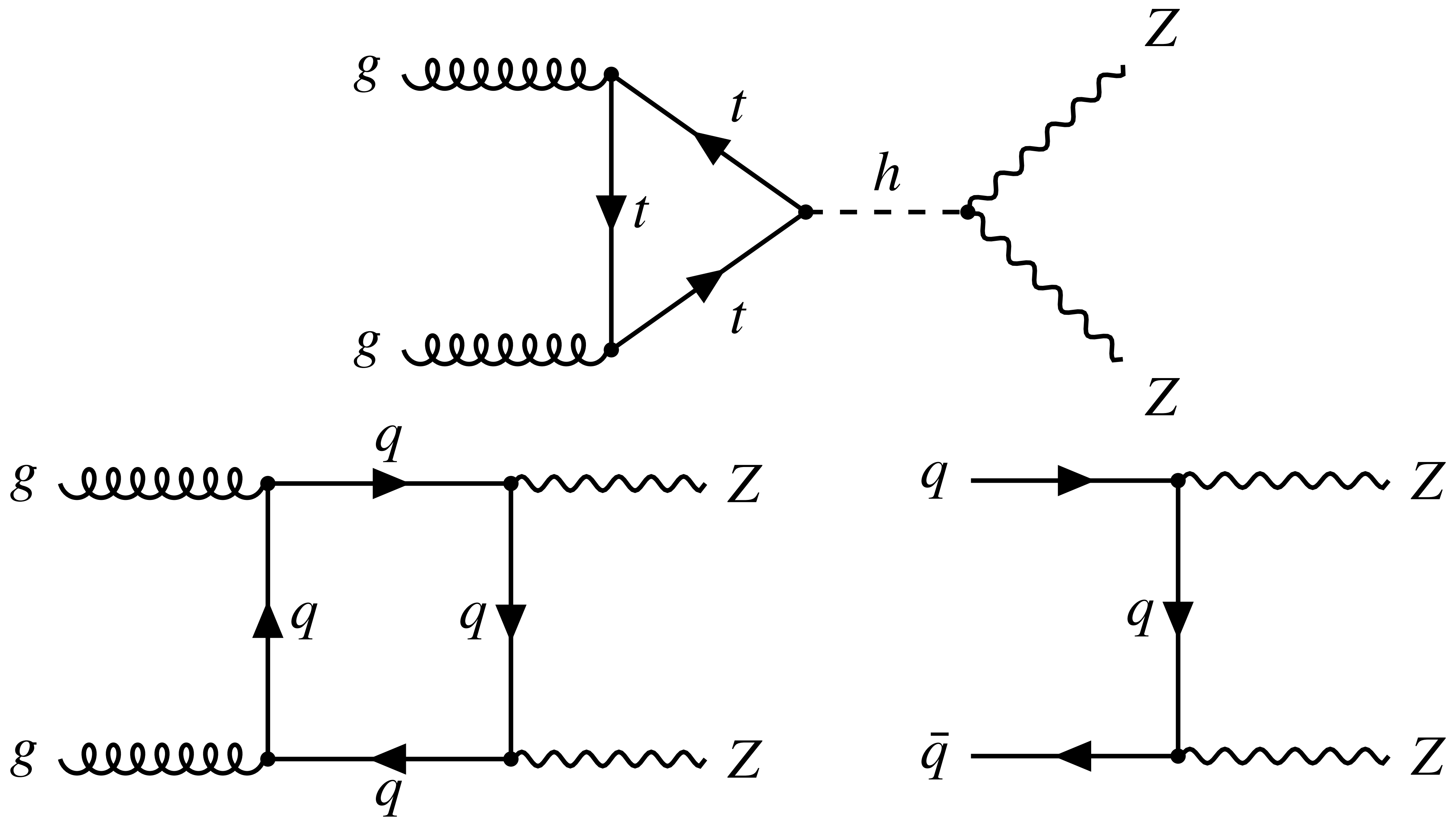
1

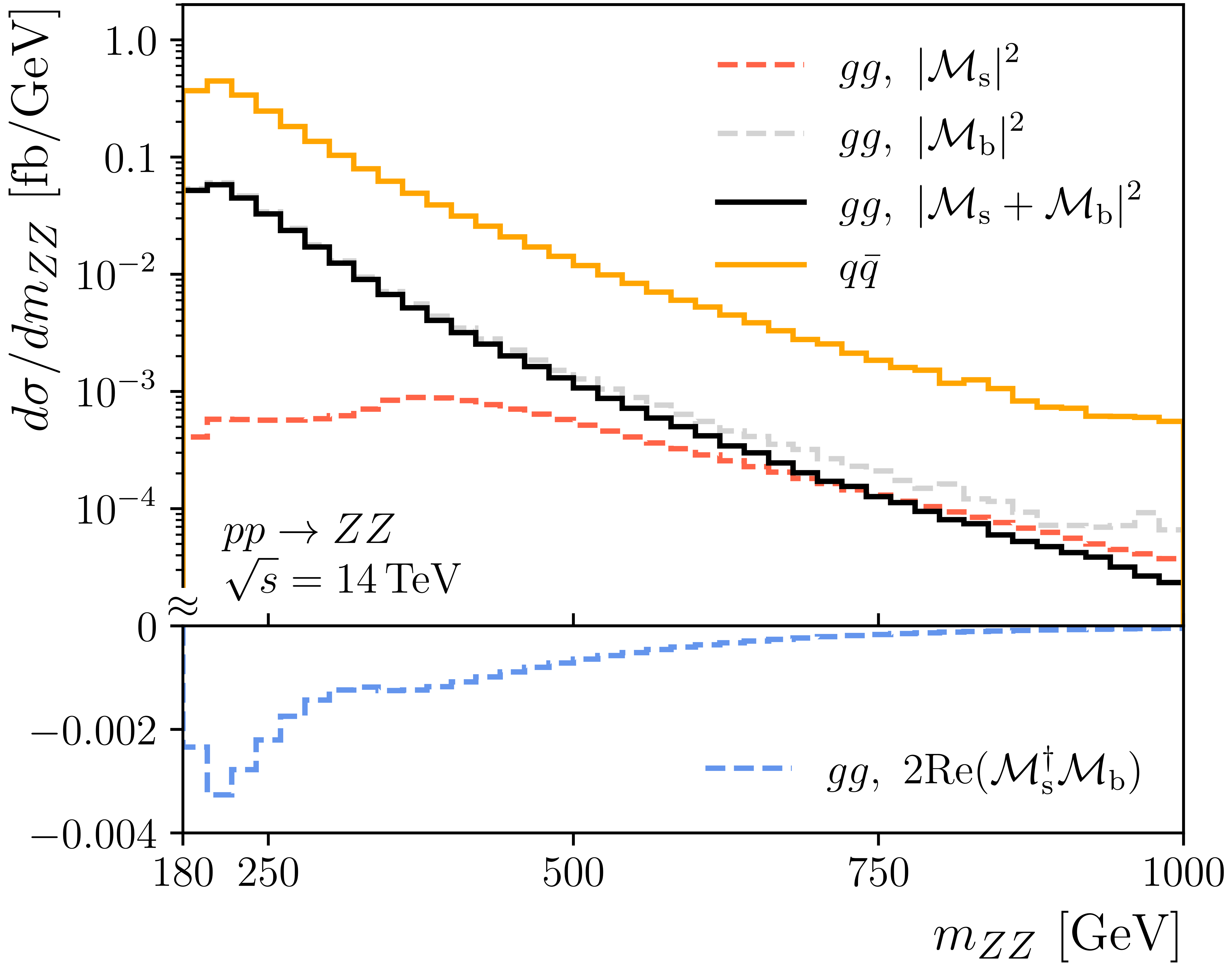
Off-shell Higgs production at the LHC

What happens in SM? What can happen in BSM?

Off-shell Higgs production at the LHC

- Higgs boson produced with high virtuality, $E^2 - |\vec{p}\,|^2 \gg m_h^2$.
- Both Z bosons can go on-shell above $m_{ZZ} \geq 2m_Z$.
- Negative interference with continuum $gg \to ZZ$.
- Large, non-interfering $qq \to ZZ$ background.





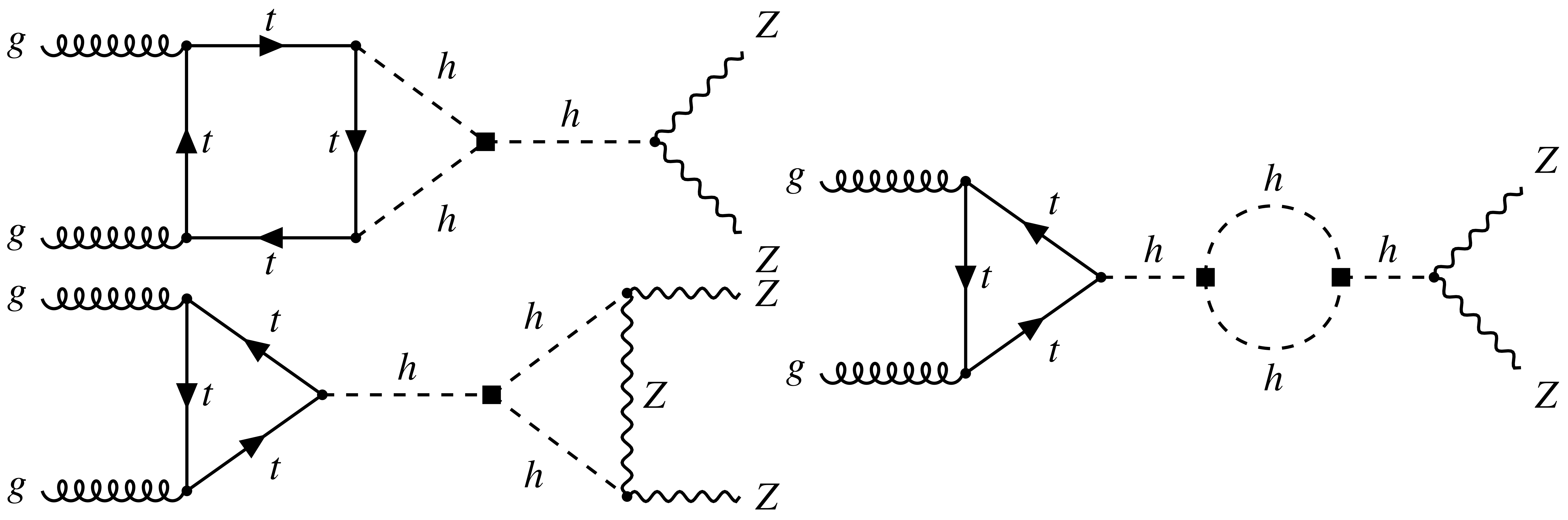
Predictions obtained from MCFM 10.3

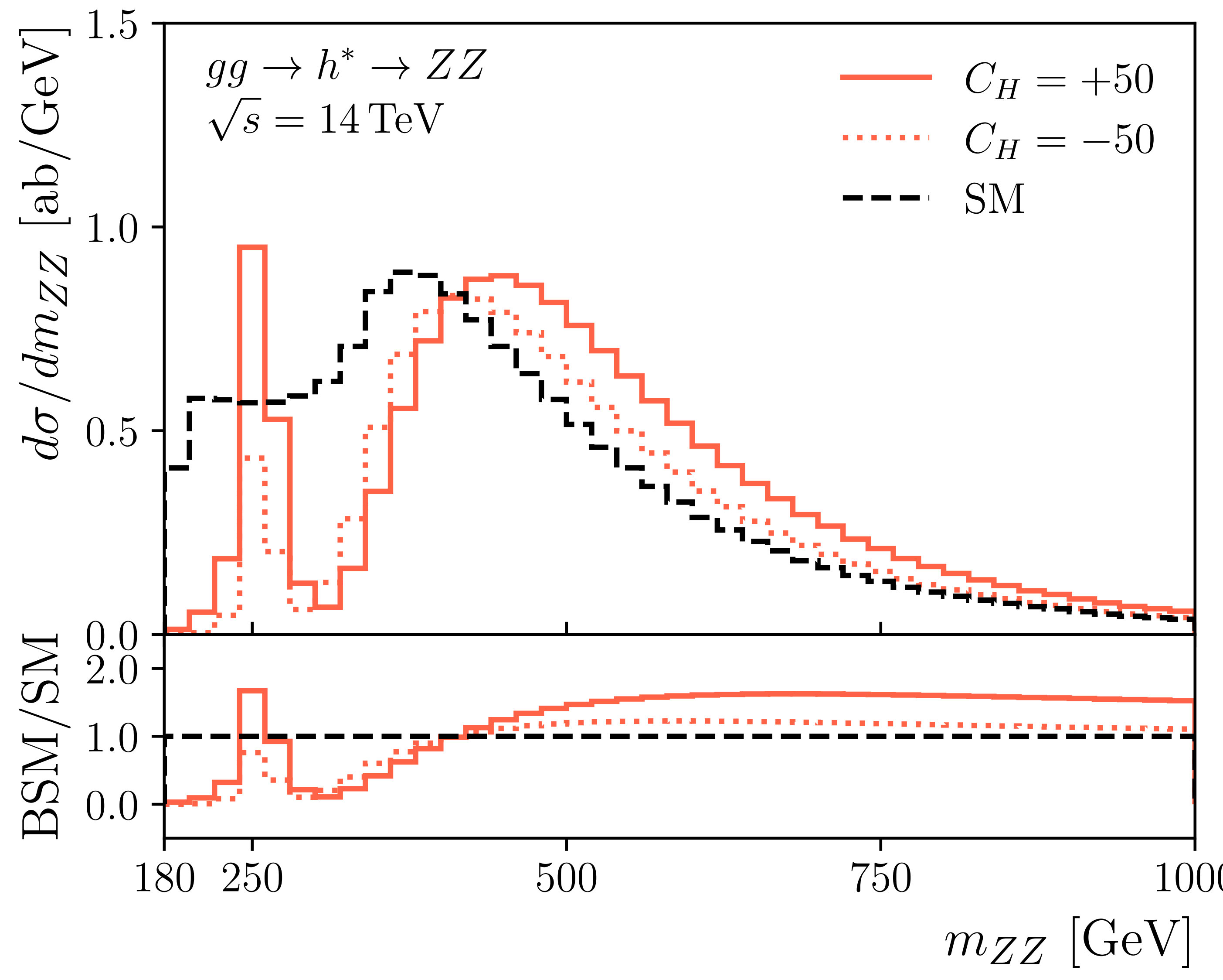
BSM modifications to $gg \to (h^* \to)ZZ$

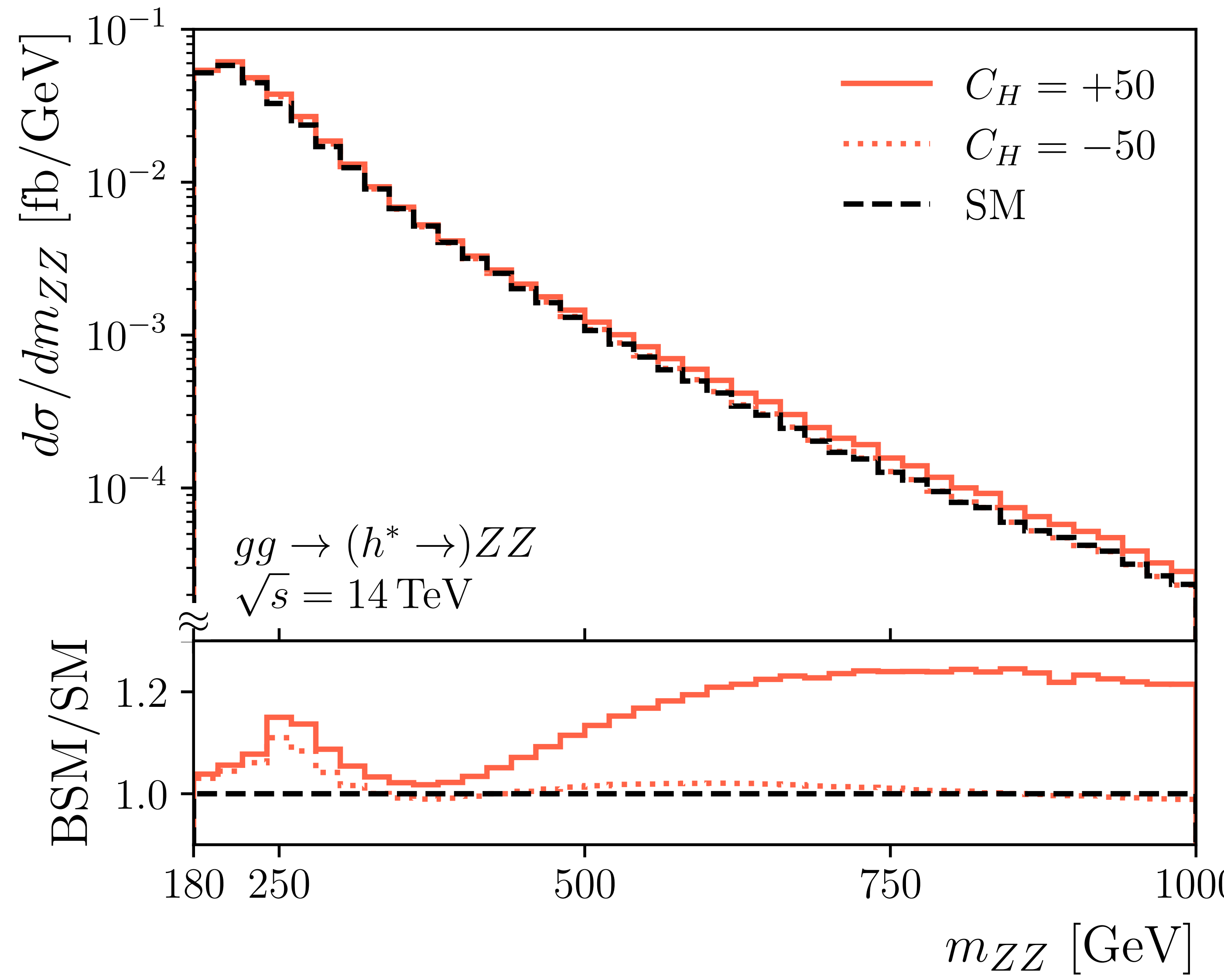
- SMEFT $d = 6$ operator that modifies trilinear self-coupling:

$$Q_H = \left(H^\dagger H\right)^3 \Rightarrow \kappa_\lambda = 1 - \frac{2v^2}{m_h^2}\frac{v^2}{\Lambda^2}C_H$$

  - Appears as two-loop contributions.
  - Resonance-like structure around $m_{ZZ} \approx 2m_h$.







Predictions obtained from (modified) MCFM 10.3 (to be made public)

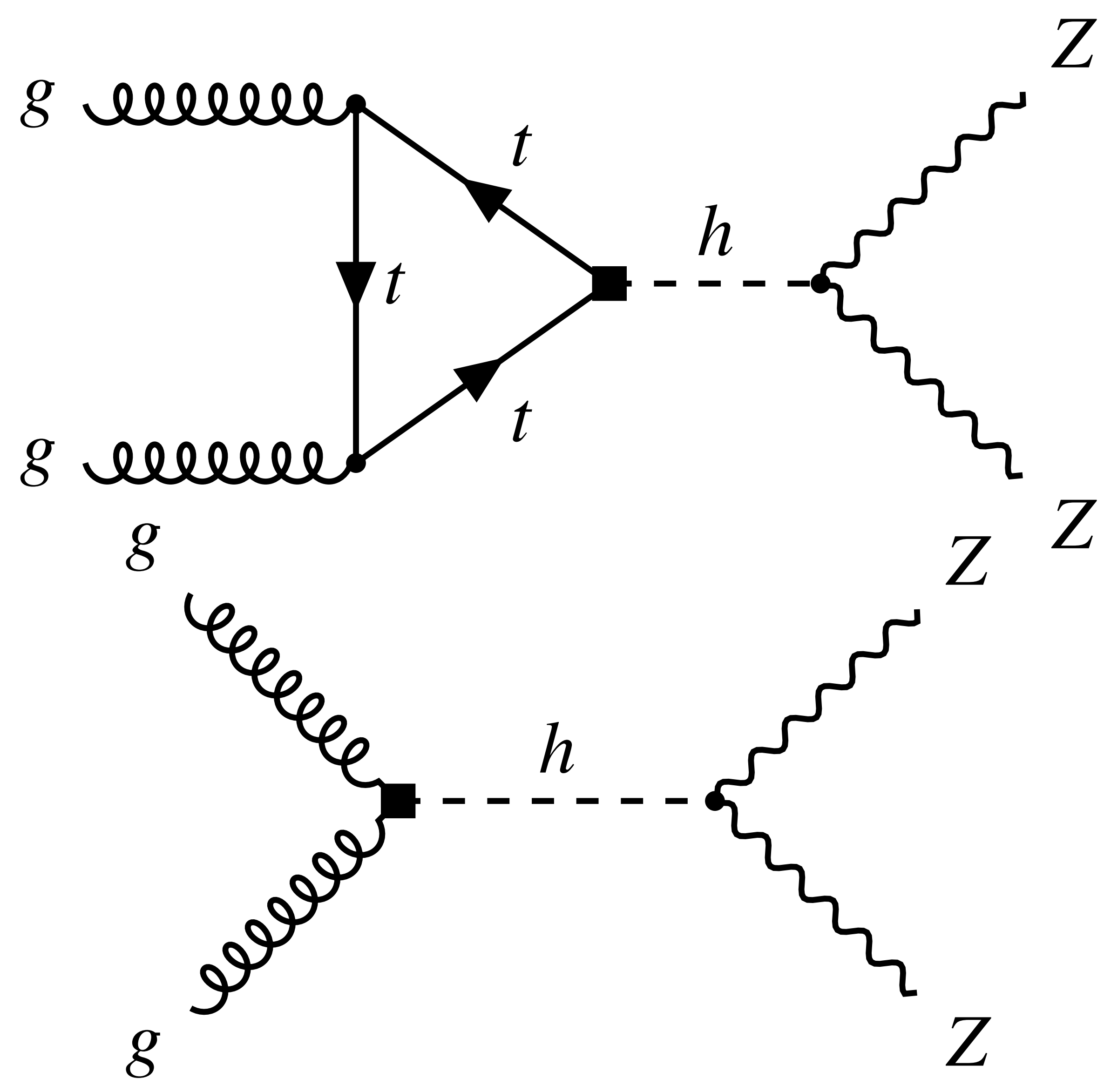
BSM modifications to $gg \to (h^* \to)ZZ$

- Additional production mode associated SMEFT operators:

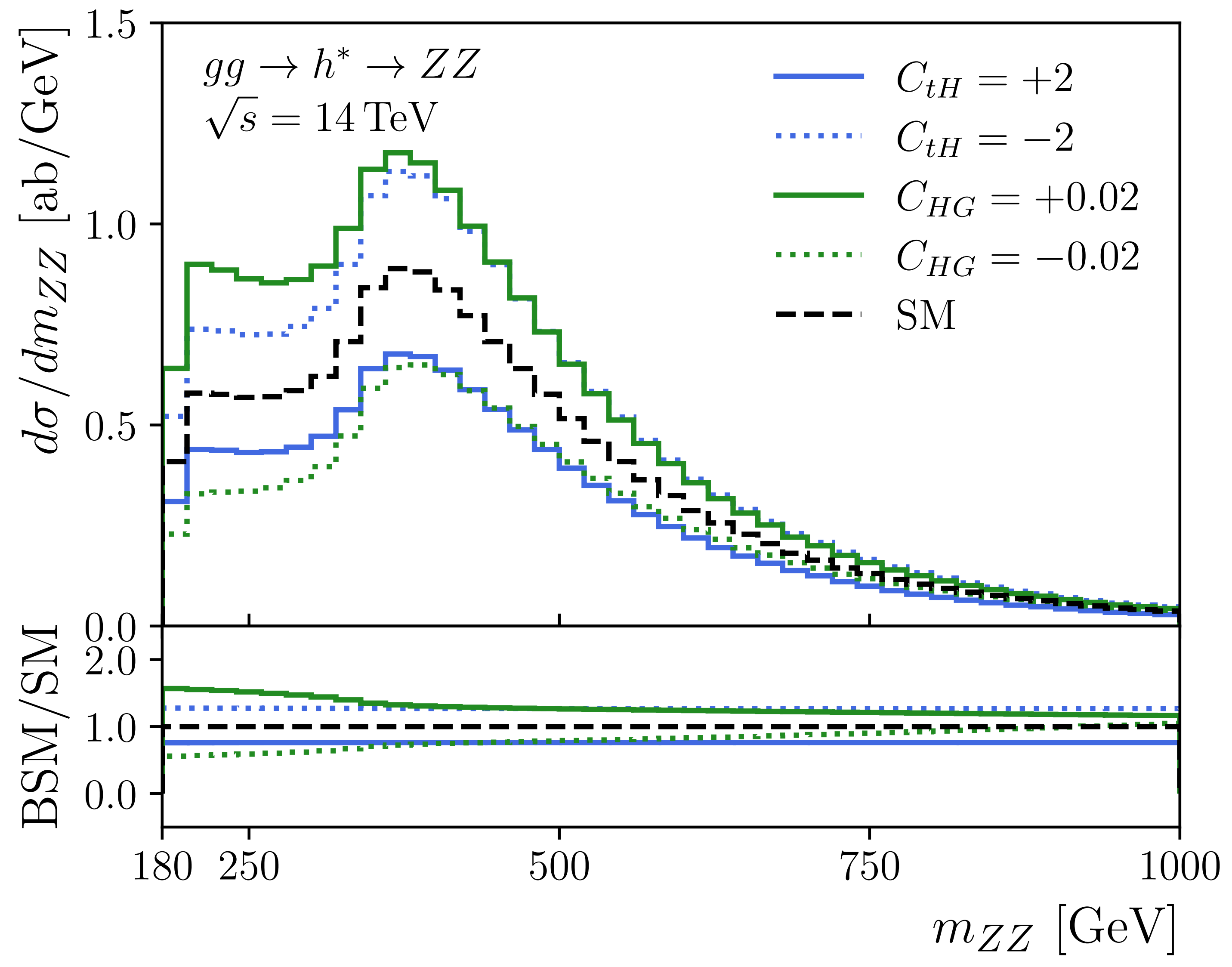
$$Q_{tH} = \left(H^\dagger H\right) \bar{q}\widetilde{H}t \Rightarrow \kappa_t = 1 - \frac{v}{\sqrt{2}m_t}\frac{v^2}{\Lambda^2} C_{tH}$$

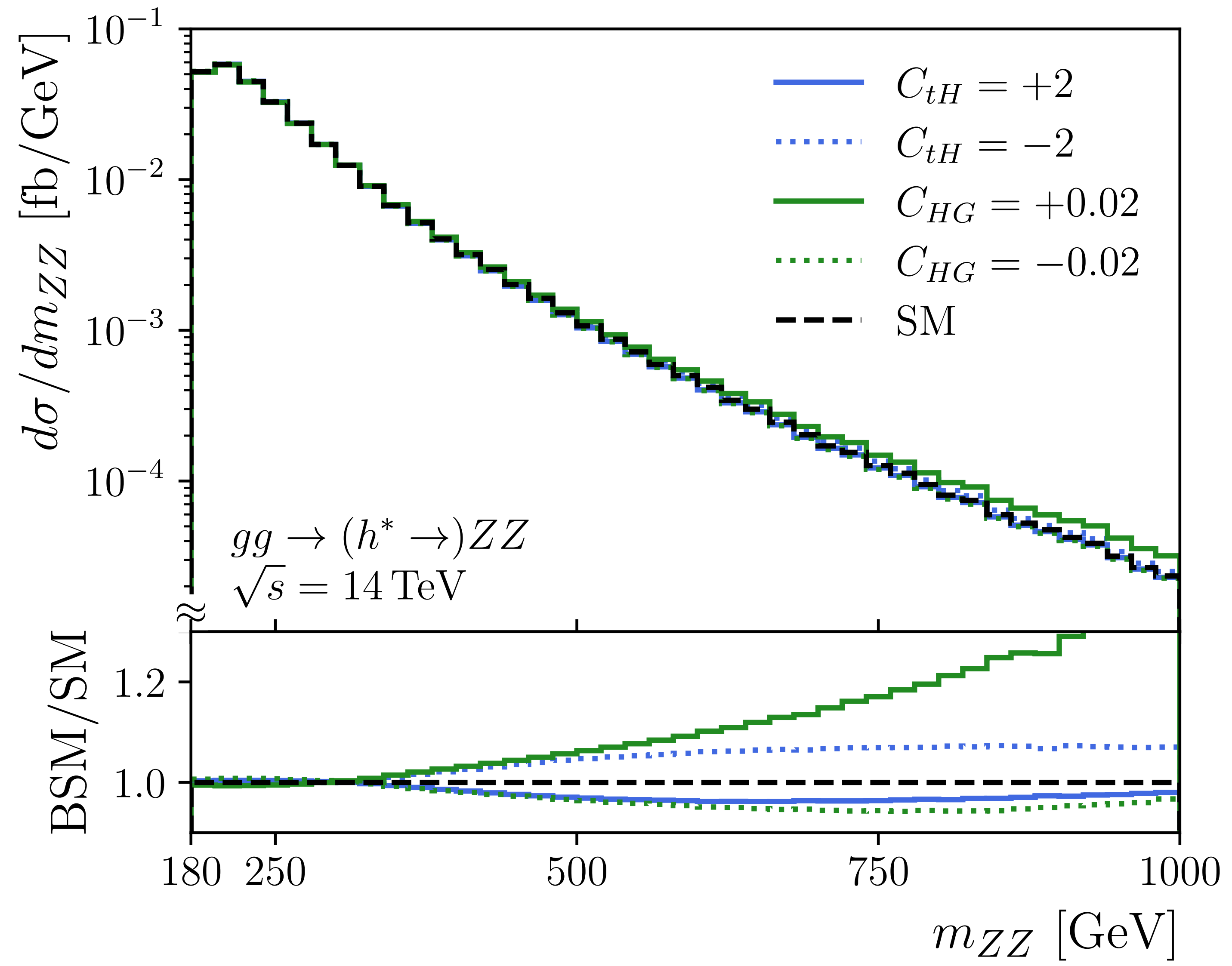
$$Q_{HG} = \left(H^\dagger H\right) G^a_{\mu\nu}G^{a,\mu\nu} \Rightarrow \kappa_g = 1 + \frac{8\pi}{\alpha_s}\frac{v^2}{\Lambda^2} C_{HG}$$

  - Appears as one-loop/tree-level contributions.
  - Enhancement of "tails" — typical of SMEFT expansions.



Predictions obtained from (modified) MCFM 10.3 (to be made public)





2

Neural simulation-based inference

How can we probe these effects (as best as possible)?

Analysis event selection

- Considering signal(+background+interference) $gg \to (h^* \to)ZZ$ and $q\bar{q} \to ZZ, q\bar{q} \to W^+W^-$ backgrounds.
- Considering both 4ℓ and $2\ell 2\nu$ channels.

Event selection	Cross section [fb] $gg \to h^* \to ZZ$	$gg \to (h^* \to)\ ZZ$	$q\bar{q} \to ZZ$	$q\bar{q} \to W^+W^-$
$pp \to 4\ell$				
$p_T^{\ell_{1,2,3,4}} > 20, 15, 10, 7\,\text{GeV}$ $70\,\text{GeV} < m_{\ell\ell} < 110\,\text{GeV}$ $m_{ZZ} > 180\,\text{GeV}$	0.3099(4)	5.656(5)	44.87(3)	—
$pp \to 2\ell 2\nu$				
$p_T^{\ell_{1,2}} > 30, 20\,\text{GeV}$ $80\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$ $\Delta R_{\ell\ell} < 2$ $E_T^{\text{miss}} > 60\,\text{GeV}$ $m_T^{ZZ} > 250\,\text{GeV}$	1.4006(9)	5.046(7)	44.49(6)	1.47(1)

- Goal: perform NSBI analysis to obtain BSM sensitivity from $pp \to ZZ$.

Inference at LHC

Maximum Likelihood: Estimate parameters behind hypothesized probability density, given observed data.

$$\mathcal{L}(\theta|\mathcal{D}) = \mathrm{Pois}(n; \nu(\theta)) \prod_{i \leq n} p\left(x_i|\theta\right)$$

Neyman-Pearson Lemma: Likelihood ratio is the most powerful test statistic.

$$\Lambda(x) = \frac{\mathcal{L}(\theta_1|x)}{\mathcal{L}(\theta_2|x)}$$

Task: Compute the expected probability (ratio).

$$\frac{p(x|\theta_1)}{p(x|\theta_2)} = ?$$

Neural simulation-based inference in a nutshell

- **Parameters of interest** (θ) determine hypotheses.
  - e.g. EFT Wilson coefficients.

- Hypotheses initially predict probability distribution of **latent variables** (z).
  - e.g. Matrix element of hard-scatter process.

(Simulation-based)

- A simulator generates samples over distributions of **experimental observables** (x).
  - e.g. PDFs, parton shower & hadronization, detector effects.

$$x \sim p(x|\theta) = \frac{1}{\sigma(\theta)} \int dz\, p_{\text{sim}}(x|z) |\mathcal{M}(z|\theta)|^2$$

- Use neural network(s) to estimate the probability density (ratio) of observation under different hypotheses.

- MLE → Find the hypothesis which maximizes NN estimates.

(Neural inference)

- When feasible, NN training can be augmented by information from simulation chain.

Neural simulation-based inference in a nutshell

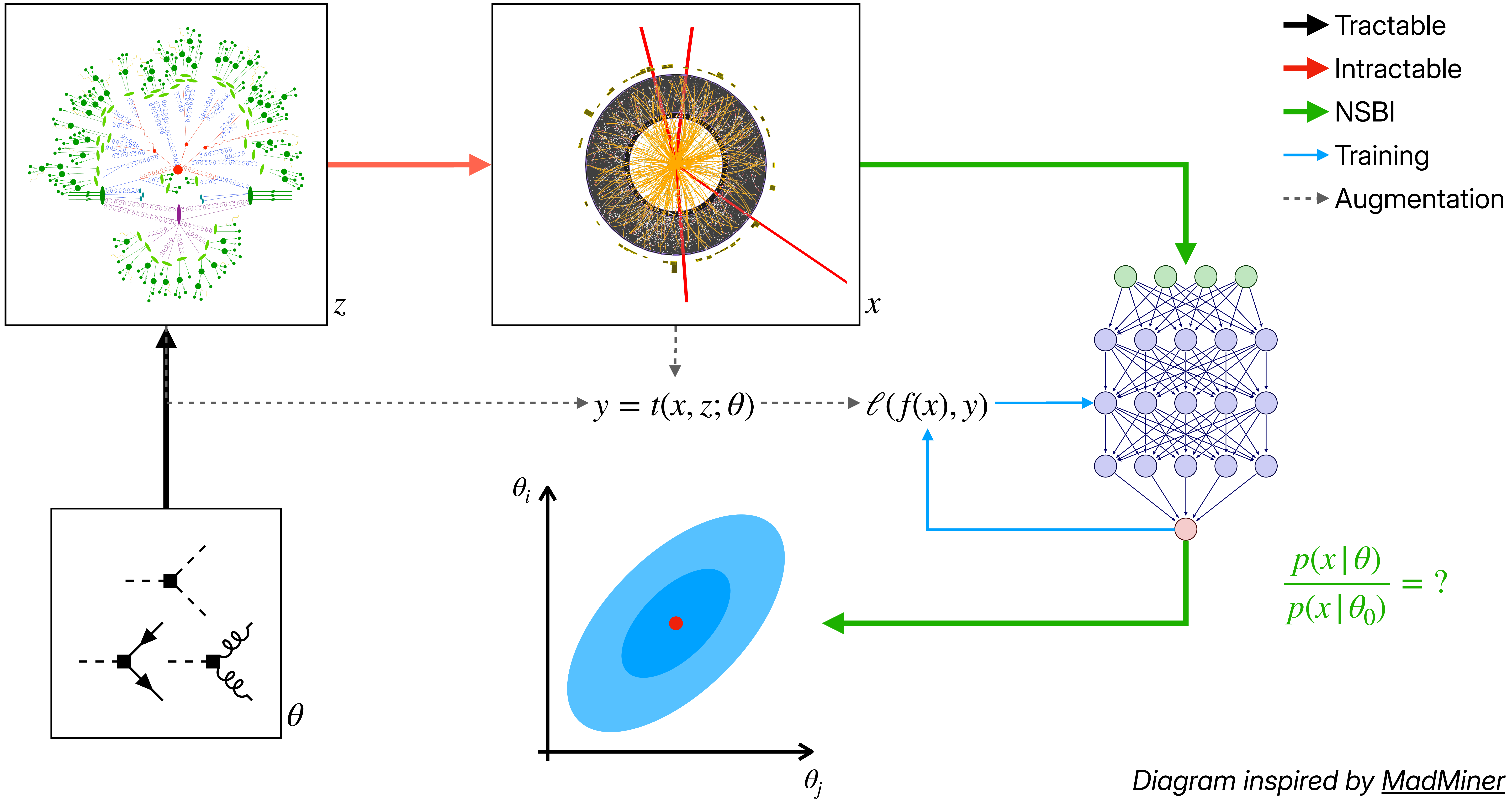


Diagram inspired by MadMiner

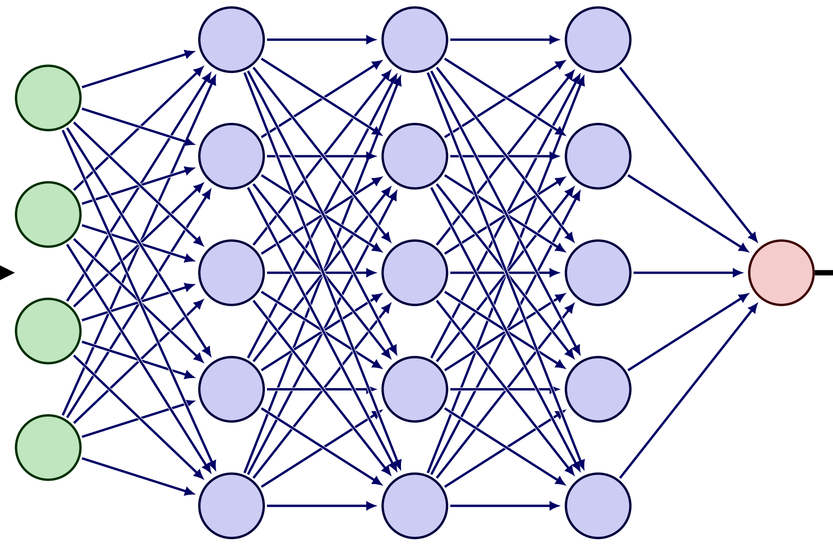
NSBI methods: classifier vs. regression

Classify between *balanced* samples

$$r(x; q, g) \equiv \frac{q(x)}{g(x)}$$

$x \sim q(x), y = 1$

$x \sim g(x), y = 0$



$$s^*(x) = \frac{q(x)}{q(x) + g(x)}$$

$$r(x; q, g) \approx \frac{s(x)}{1 - s(x)}$$

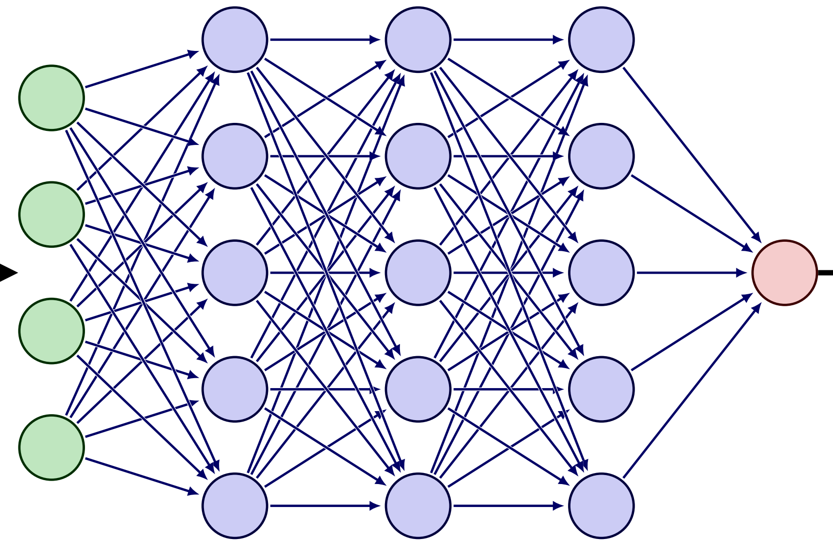
Known as "Likelihood ratio trick"

Regress to the *joint* probability ratio

$$g(x, z|\theta) \equiv p_{\text{sim}}(x|z)|\mathcal{M}(z;\theta)|^2/\sigma(\theta)$$

$$R(x, z; \theta) \equiv \frac{\sigma_g(\theta) g(x, z|\theta)}{\sigma_g(0) g(x, z|0)} = (\cdots)$$

$x \sim g(x|0)$



$$a_i^*(x) = (\cdots) = \frac{\partial R(x; \theta/0)}{\partial \theta_i}$$

$$R(x; \theta) \approx 1 + a_i(x)\theta_i + a_{ij}(x)\theta_i\theta_j + (\cdots)$$

$(\cdots)$ For identical initial & final states, this is ratio of squared MEs!

$(\cdots)$ Targeting joint ratios converges to observable ratio!

$(\cdots)$ Taylor expansion is exact at finite order for SMEFT!

(More info on $\cdots$'s in backup)

NSBI methods: classifier vs. regression

Classify between *balanced* samples

$$r(x; q, g) \approx \frac{s(x)}{1 - s(x)}$$

- **Only requires sampling from hypotheses.** 👍
- Large training data requirements & slower convergence. 👎
- Need to sample from multiple hypotheses. 👎
- Will inter/extrapolate hypotheses not seen during training. 👎

Regress to the *joint* probability ratio

$$R(x; \theta) \approx 1 + a_i(x)\theta_i + a_{ij}(x)\theta_i\theta_j + (\cdots)$$

- Joint probability ratio not always available in simulation. 👎
- **Faster training convergence.** 👍
- **Only need to sample from one hypothesis.** 👍
- **Can enforce physics structure through custom targets.** 👍

NSBI methods: classifier vs. regression

Classify between *balanced* samples

$$r(x; q, g) \approx \frac{s(x)}{1 - s(x)}$$

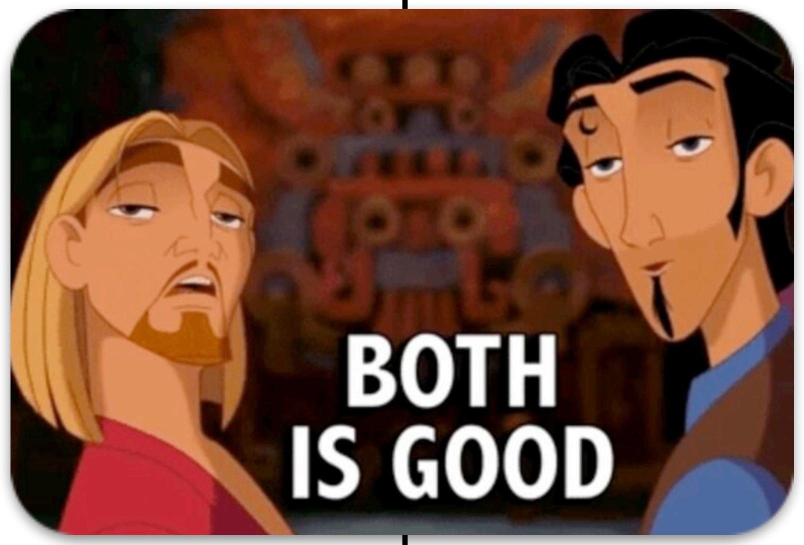
- **Only requires sampling from hypotheses.** 👍
- Large training data requirements & slower convergence. 👎
- Need to sample from multiple hypotheses. 👎
- Will inter/extrapolate hypotheses not seen during training. 👎

Regress to the *joint* probability ratio

$$R(x; \theta) \approx 1 + a_i(x)\theta_i + a_{ij}(x)\theta_i\theta_j + (\cdots)$$

- Joint probability ratio not always available in simulation. 👎
- **Faster training convergence.** 👍
- **Only need to sample from one hypothesis.** 👍
- **Can enforce physics structure through custom targets.** 👍

Which method to use?



NN training setup

Input features	4ℓ	$\left\{p_\ell = \left(p_T^\ell, \eta_\ell, \phi_\ell, E_\ell\right)\right\}_{\ell=1,2,3,4}$		4ℓ: Full final state reconstruction.
	$2\ell 2\nu$	$\left\{p_\ell = \left(E_\ell, p_T^\ell, \eta_\ell, \phi_\ell\right)\right\}_{\ell=1,2}, \ \vec{E}_T^{\text{miss}} = \left(E_T^{\text{miss}}, \phi_{\text{miss}}\right)$		$2\ell 2\nu$: Missing transverse energy from neutrinos.
Numerator hypothesis		$gg \to (h^* \to) ZZ$	$q\bar{q} \to ZZ, q\bar{q} \to W^+W^-$	
Denominator hypothesis		SM $gg \to (h^* \to) ZZ$ process		Common denominator
Parameter dependence		$C = \{C_H, C_{tH}, C_{HG}\}$	—	
Probability ratio		$R\,(x; C)$	$r\,(x; q\bar{q}, gg)$	
Training target		$a_{i,ij,ijk,ijkl}(x)$	$s(x)$	
Output activation		linear	sigmoid	Classifier: Binary Cross Entropy
Loss function		MSE	BCE	Regression: Mean Squared Error
Number of layers times nodes		20×100		
Batch size		1024		
Learning rate		$\leq 10^{-3}$		
Number of epochs		≤ 300		

Probability (ratio) mixture model

- Probability mixture of $pp = gg + q\bar{q}$ process:

$$p(x|\theta) = \frac{\sigma_g(\theta) g(x|\theta) + \sigma_q q(x)}{\sigma_g(\theta) + \sigma_q}$$

- Take ratio over a common denominator hypothesis:

$$\frac{p(x|\theta)}{g(x|0)} = \frac{\sigma_g(\theta) g(x|\theta)/g(x|0) + \sigma_q q(x|\theta)/g(x|0)}{\sigma_g(\theta) + \sigma_q}$$

- Estimate sub-terms via classifier/regression:

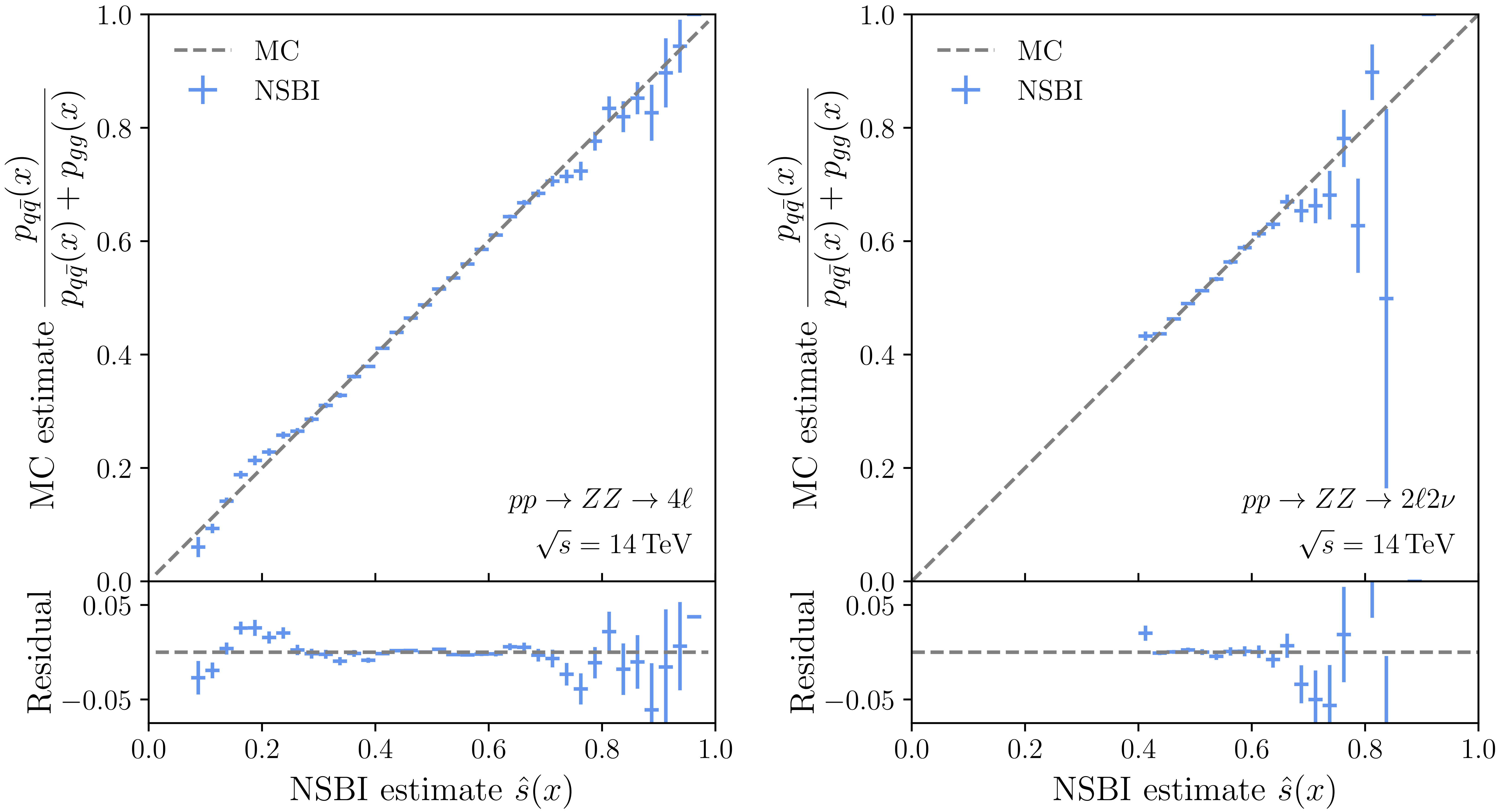
$$\frac{p(x|\theta)}{g(x|0)} = \frac{\sigma_g(0) R(x;\theta) + \sigma_q r(x;q,g)}{\sigma_g(\theta) + \sigma_q}$$

- Cancel out denominator for statistical tests:

$$\frac{p(x|\theta)}{p(x|0)} = \frac{\sigma_g(0) + \sigma_q}{\sigma_g(\theta) + \sigma_q} \frac{\sigma_g(0) R(x;\theta) + \sigma_q r(x;q,g)}{\sigma_g(0) + \sigma_q r(x;q,g)}$$

SBI diagnostics: calibration closure

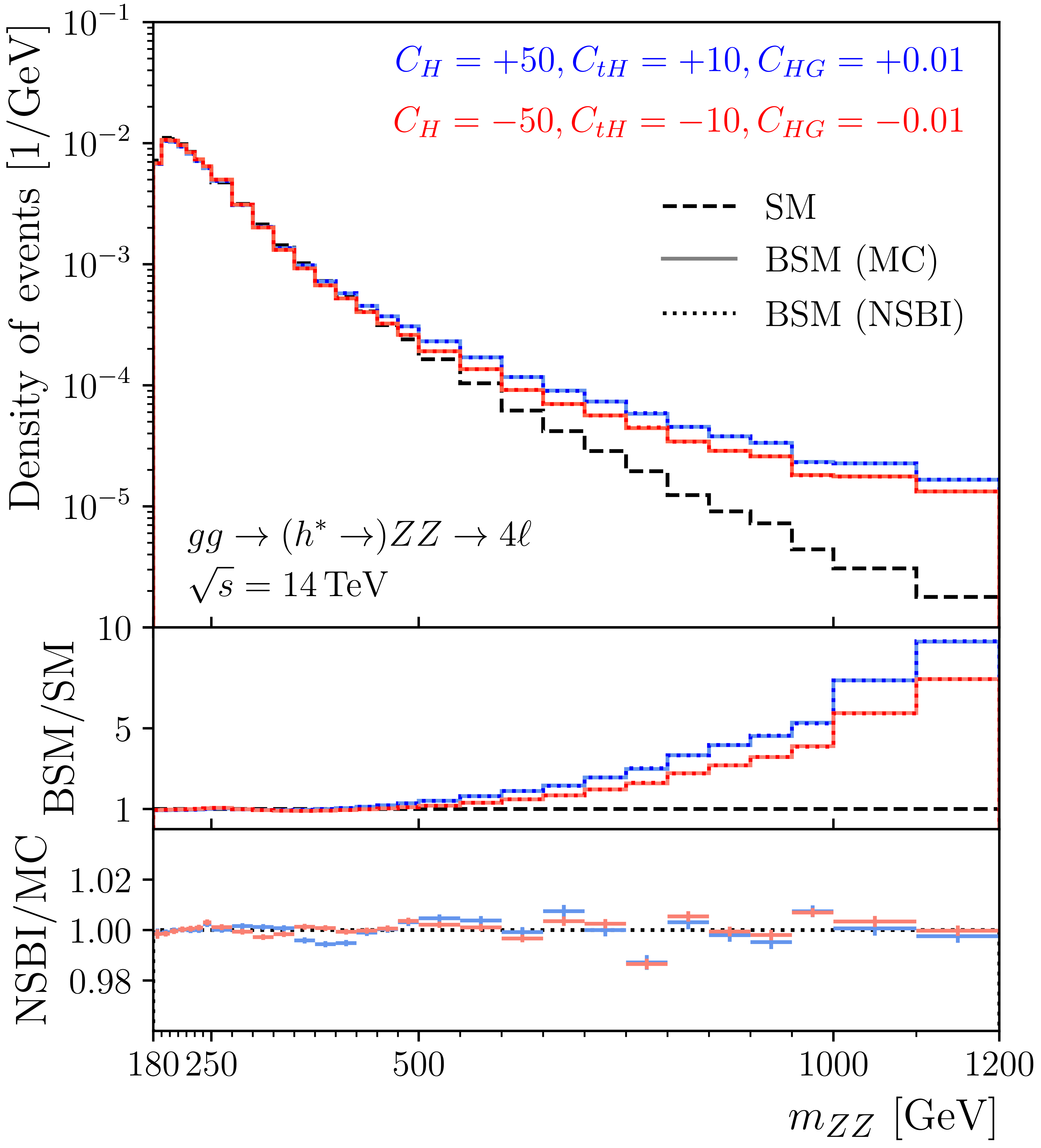
- NSBI estimate: classifier decision output.
- MC estimate: fraction of events belonging to numerator hypothesis out of samples drawn from the balanced hypothesis.

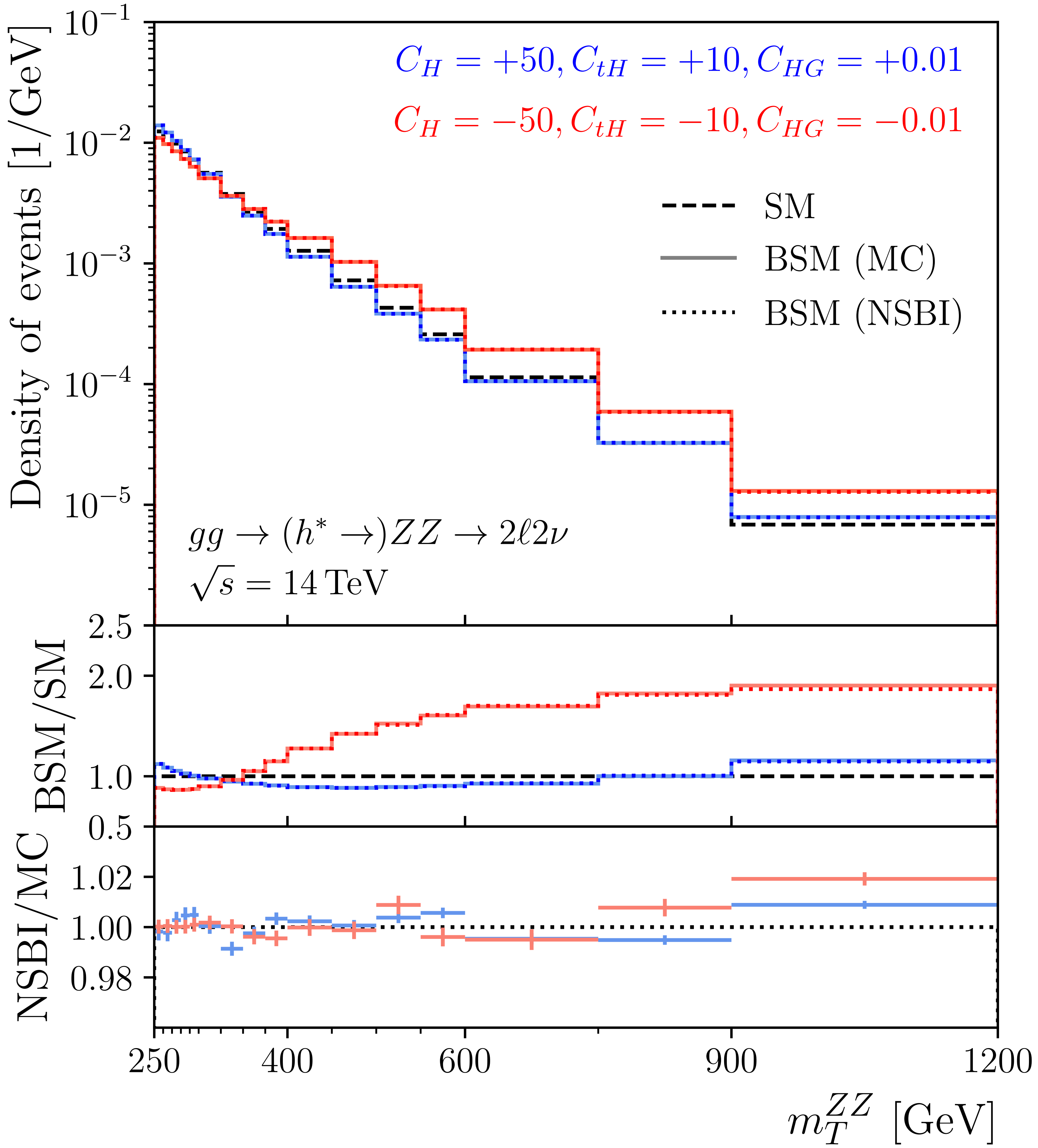


- NSBI provides a robust estimate of the probability ratios between hypotheses.
- NNs can be further improved through hyper-parameter optimization & ensembling.

SBI diagnostics: reweighting closure

- Probability ratio between hypotheses ⇔ Reweighing between hypotheses.



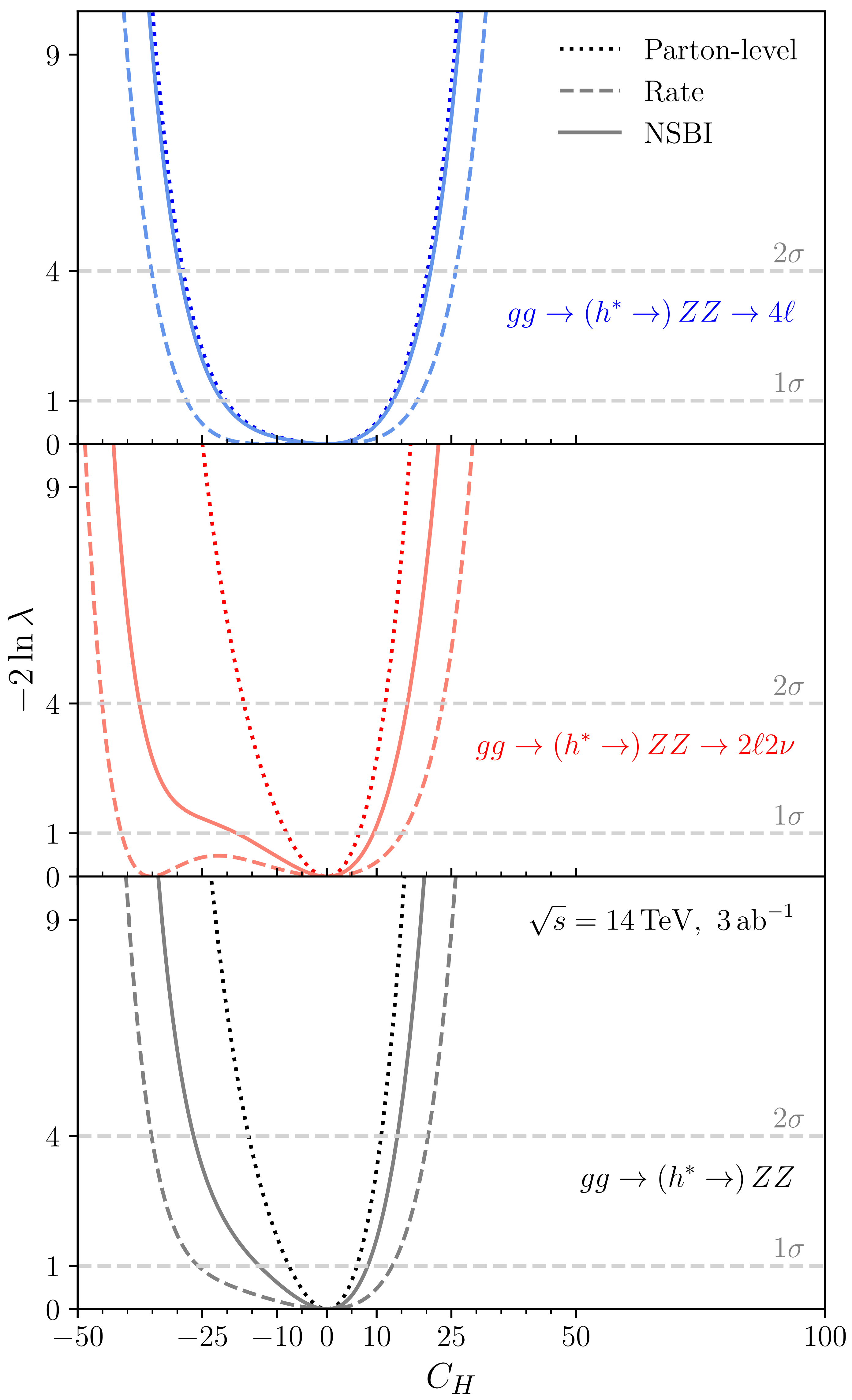


- Reminder: m_{ZZ}, m_T^{ZZ} not training inputs.
- Trained on parton-level targets, converges to observable-level!

3

Results

Results: $gg \to (h^* \to)ZZ$

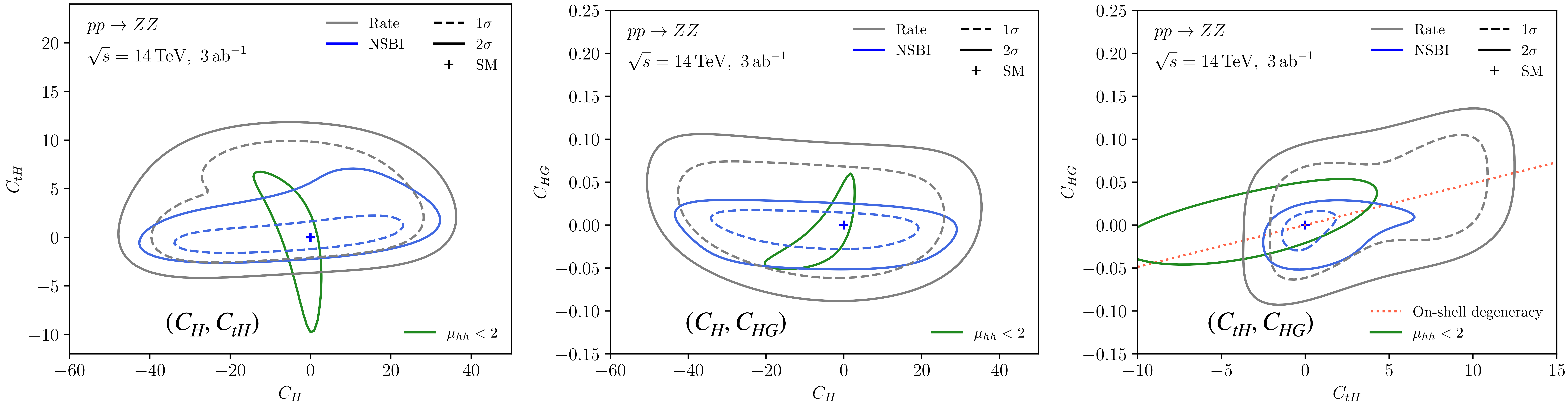


- Showing SM-expected limits at HL-LHC.
- Parton-level limits show theoretical best limits achievable.
- Under ideal conditions, i.e. $p(x\,|\,z) \to \delta(x - z)$, NSBI approaches optimal sensitivity.

- Incomplete observable space (neutrinos) leads to inevitable loss of sensitivity.
  - Nevertheless achieves sensitivity gain over naive rate analysis.
  - Comment: NSBI could also be applied to $h \to W^+W^-$ channel!

- $4\ell + 2\ell 2\nu$ combined limits: $-5.5 < \kappa_\lambda < 13.5$
  - Significantly weaker than ATLAS Run 2 on-shell single/di-Higgs bounds.
  - But this assumes all other Higgs bosons couplings remain at SM…

Results: $pp \to ZZ$



- Two-at-a-time simultaneous constraints on SMEFT operators.
  - Also shown: corresponding limits from $\mu_{hh} < 2$ (ATLAS HL-LHC prospects).
  - Also shown: on-shell single-Higgs degeneracy in (C_{tH}, C_{HG}).
- **Orthogonality in exclusions among off-shell & di-Higgs measurements.**
  - **Expected to be relevant for global SMEFT constraints.**

Conclusions

- Rich BSM phenomenology in off-shell Higgs boson production at the LHC.
  - Modifications to self-coupling modifications distinguishable from other effects.

- Applied multi-prong NSBI approach to likelihood ratio estimate.
  - Benchmarked performance with (in)complete $(2\ell 2\nu)4\ell$ observables.

- Off-shell offers complementary BSM probes to on-shell single-Higgs & di-Higgs measurements.

- Future work/collaboration:
  - Release of source code + datasets, tutorials.
  - Scalable NSBI infrastructure with HEP-CCE (Dennis Bollweg).

Thank you for your attention!

Backup

Full $pp \to ZZ$ calibration closure