hep-ph/2405.15847

Constraining the Higgs Potential Shape with Machine Learning

Radha Mastandrea

In collaboration with Benjamin Nachman and Tilman Plehn

ML4Jets
07/11/2024

Our goal: measure the Higgs potential $V(\phi)$

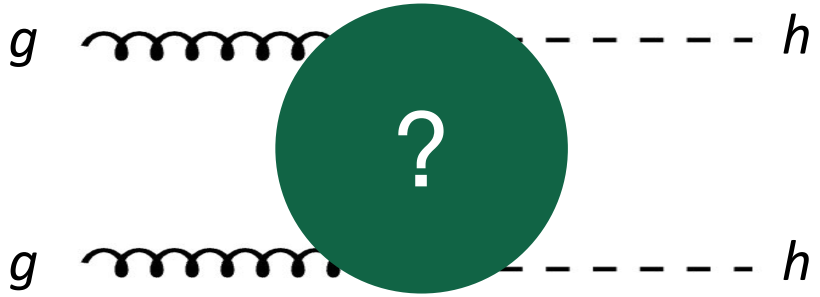
$$-m^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 + \sum_i \frac{c_i}{\Lambda^{d_i-4}}\mathcal{O}_i$$

SM higgs

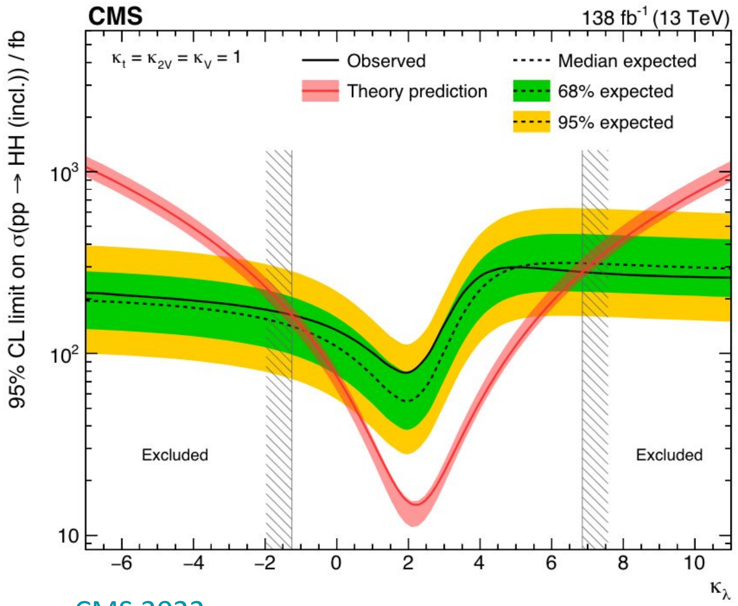
BSM Higgs

(SMEFT parameterization)

Concretely, we want to observe **dihiggs production** at the LHC

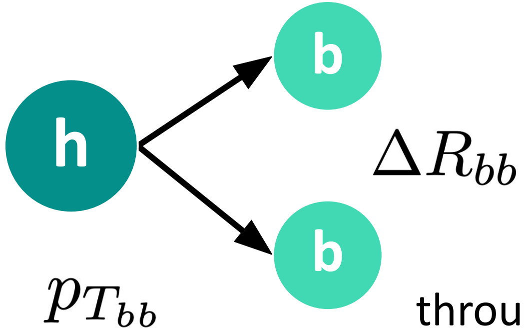


Previous work: cut and count

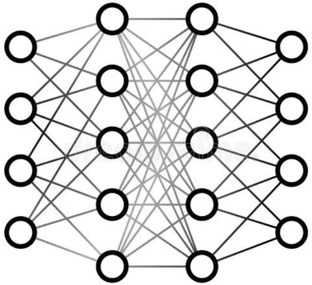


CMS 2022

This work: + kinematic information



through



The test statistic *q* now includes shape information

$$q(c|D) = q_{\text{rate}}(c|D) + q_{\text{shape}}(c|D)$$

$(c_\phi,\, c_{\phi d},\, c_{t\phi})$ Data

Likelihood ratio of
data | BSM hypothesis
to
data | SM hypothesis

The test statistic *q* now includes shape information

$$q(c|D) = \boxed{q_{\text{rate}}(c|D)} + q_{\text{shape}}(c|D)$$

$(c_\phi,\ c_{\phi d},\ c_{t\phi})$ Data

Likelihood ratio of
data | BSM hypothesis
to
data | SM hypothesis

Likelihood ratio for
Poisson events
(cut and count)

The test statistic *q* now includes shape information

$$q(c|D) = \boxed{q_{\text{rate}}(c|D)} + \boxed{q_{\text{shape}}(c|D)}$$

$(c_\phi,\, c_{\phi d},\, c_{t\phi})$ Data

Likelihood ratio of
data | BSM hypothesis
to
data | SM hypothesis

Likelihood ratio for
Poisson events
(cut and count)

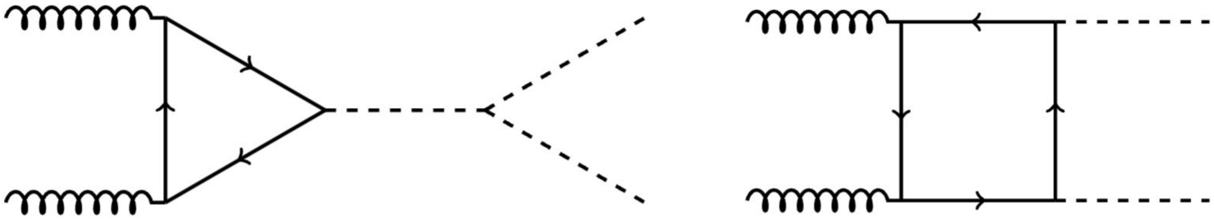
Likelihood ratio for shapes of
kinematic distributions
(machine learning)

SMEFT Operators in Detail

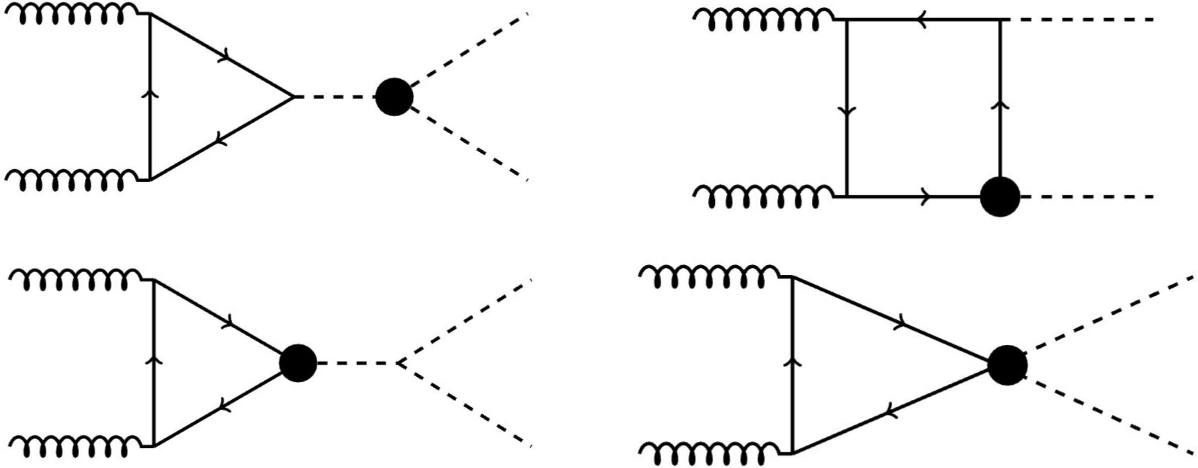
Symbol	Operator	Meaning
$\mathcal{O}_\phi$	$(\phi^\dagger\phi - \frac{v^2}{2})^3$	trilinear coupling
$\mathcal{O}_{\phi d}$	$\partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi)$	dynamical coupling
$\mathcal{O}_{t\phi}$	$(\phi^\dagger\phi - \frac{v^2}{2})\overline{Q}t\tilde{\phi} + \mathrm{h.c.}$	top-Yukawa coupling

These operators result in a zoo of diagrams…

SM diagrams



SMEFT diagrams

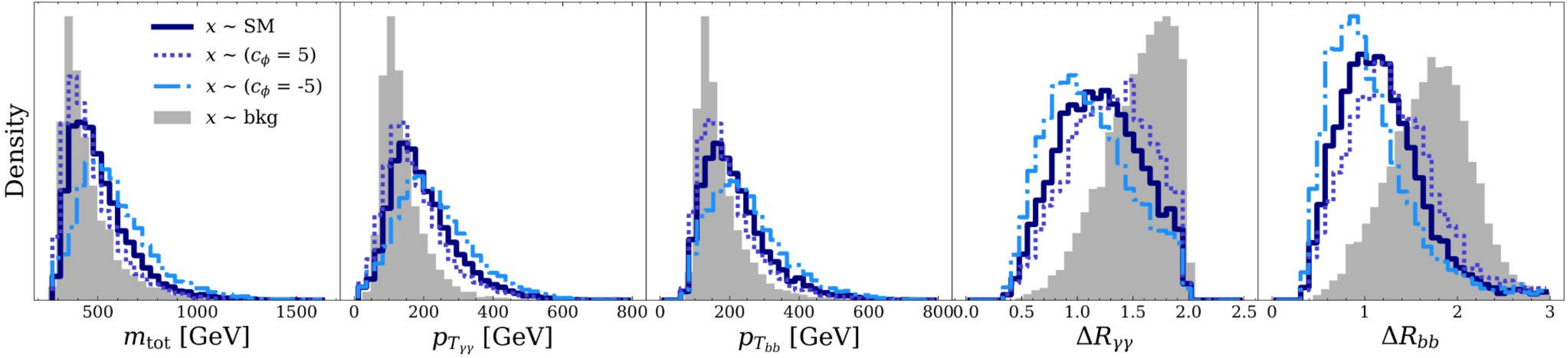


…whose inclusion changes the shapes of kinematic features

Production channel: $hh \rightarrow bb\gamma\gamma$

Simulation pipeline: MadGraph (SMEFT@NLO model) → Pythia → Delphes

Collider setup: HL-LHC (14 TeV, 3/ab), S/B ≈ 0.3



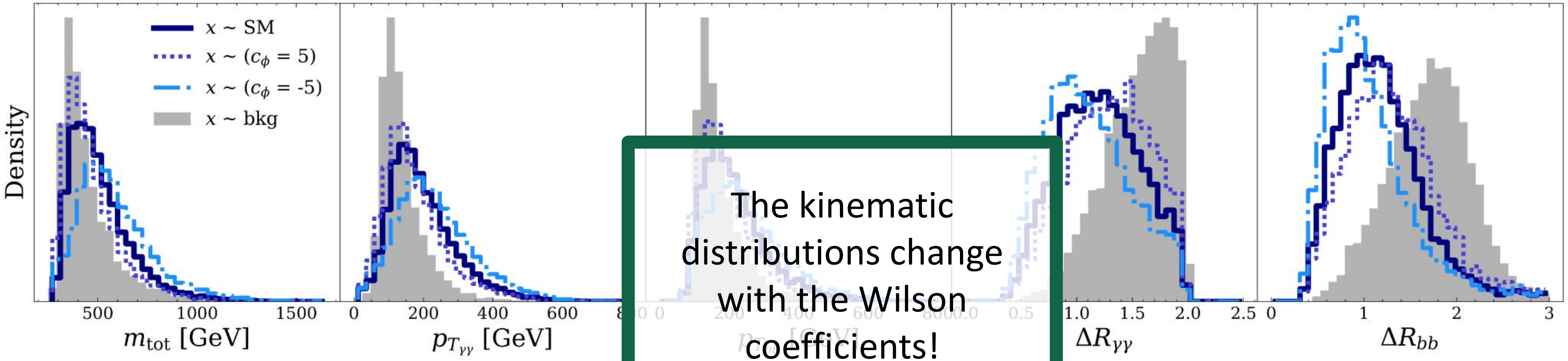
**see backups for full background breakdown

…whose inclusion changes the shapes of kinematic features

Production channel: $hh \rightarrow bb\gamma\gamma$

Simulation pipeline: MadGraph (SMEFT@NLO model) → Pythia → Delphes

Collider setup: HL-LHC (14 TeV, 3/ab), S/B ≈ 0.3

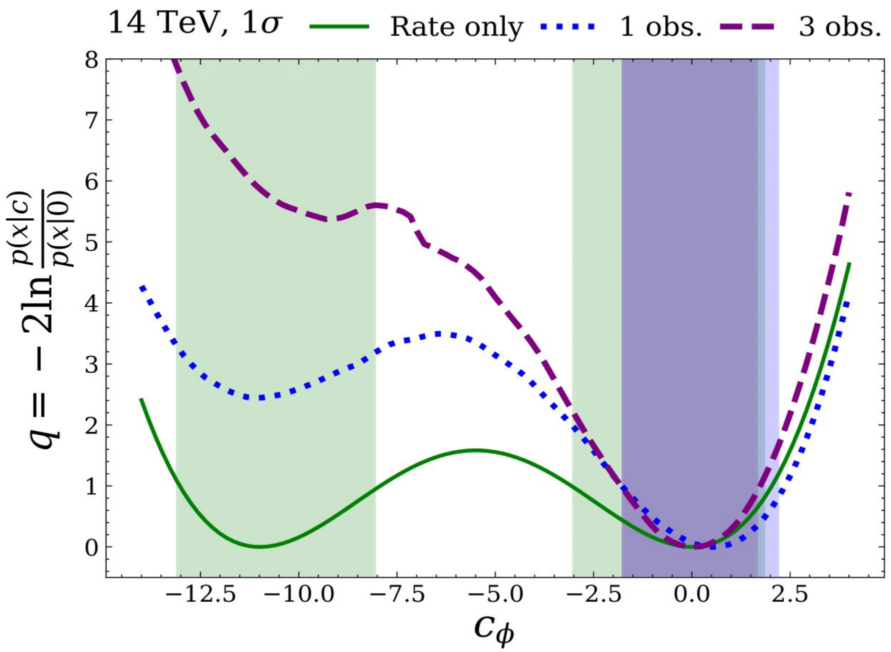


The kinematic distributions change with the Wilson coefficients!

**see backups for full background breakdown

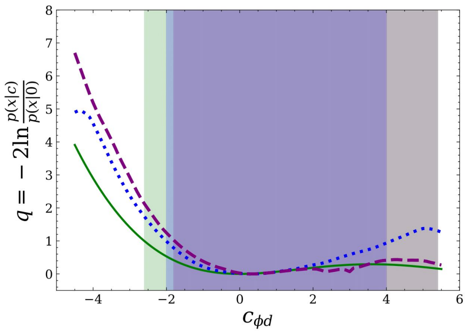
1D coefficient recovery on a SM test set

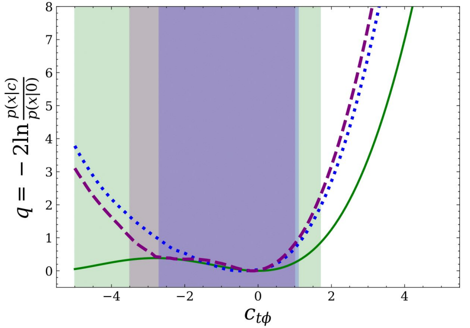
Shape-informed likelihood ratios allow for tighter confidence intervals than in the rate-only case – more information is better!



1 obs: m_{tot}

3 obs: $+\,p_{T_{bb}}, p_{T_{\gamma\gamma}}$

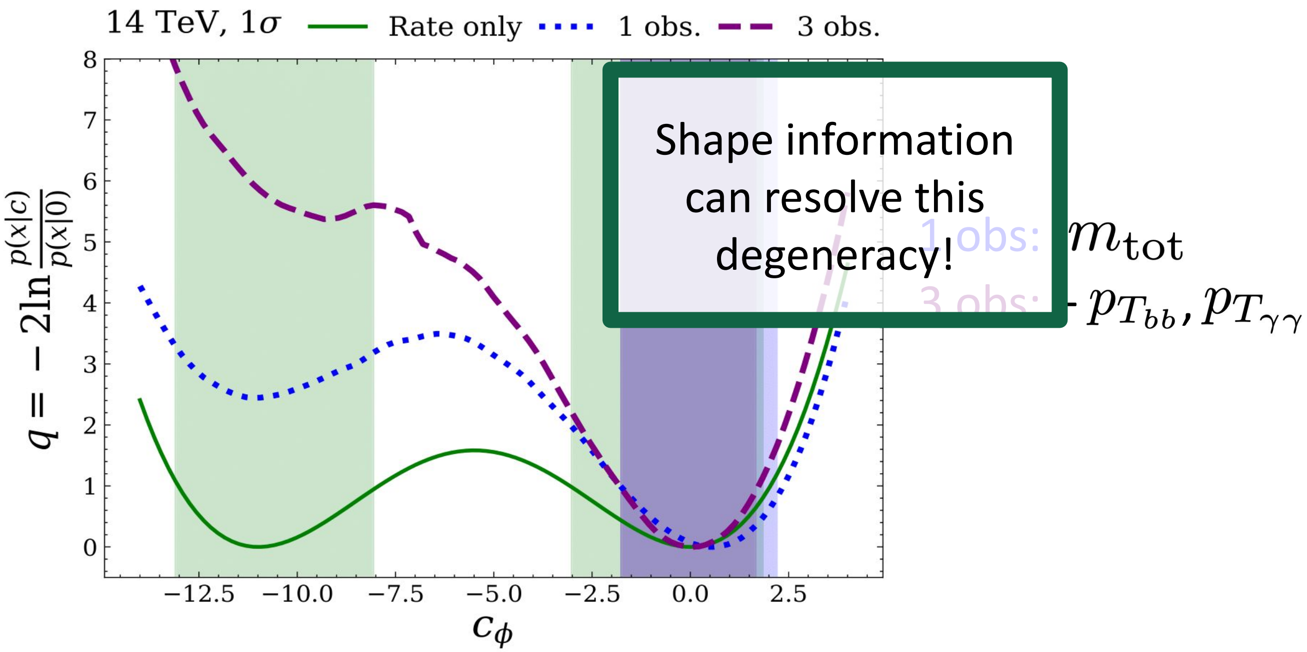




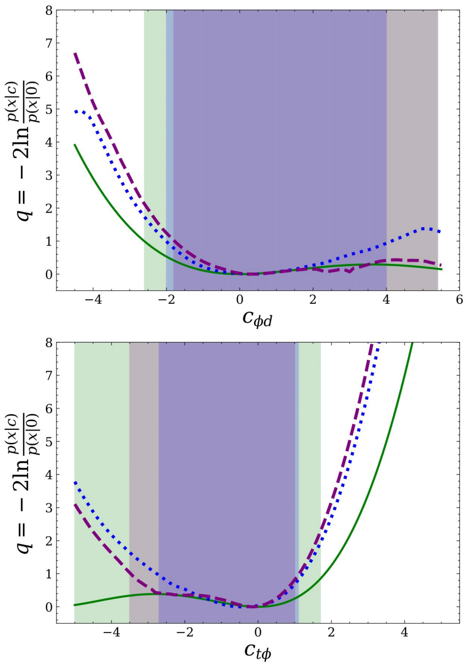
**see backups for FCC-hh projections

1D coefficient recovery on a SM test set

Shape-informed likelihood ratios allow for tighter confidence intervals than in the rate-only case – more information is better!



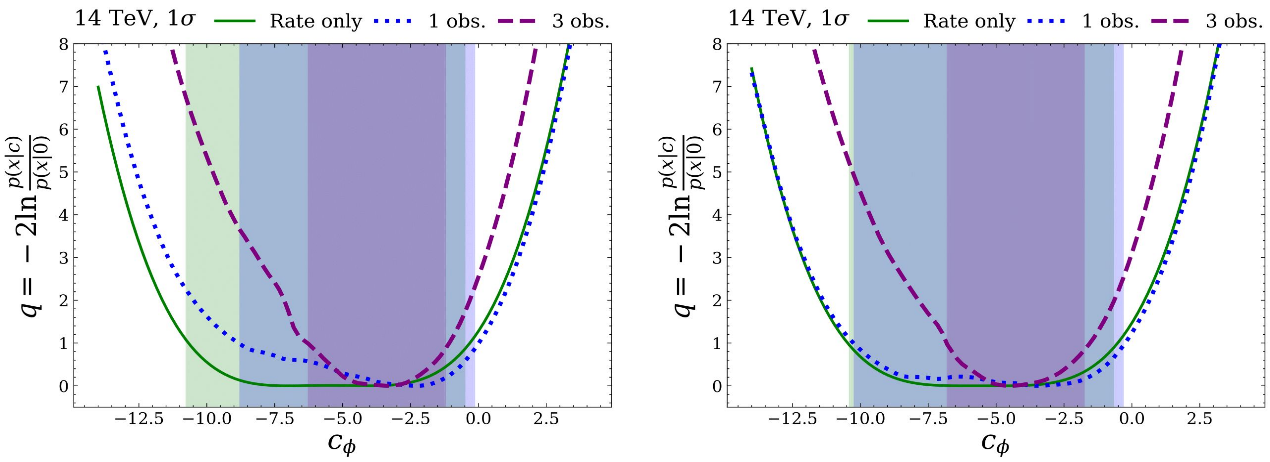
Shape information can resolve this degeneracy!



**see backups for FCC-hh projections

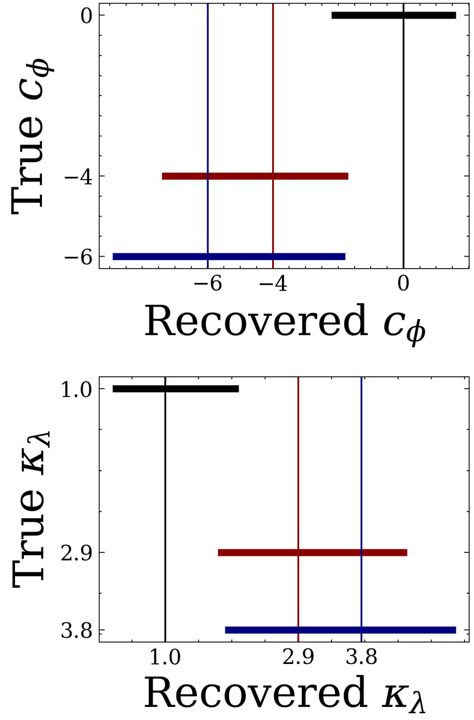
1D coefficient recovery on BSM test sets

Assuming a BSM test set (truth), we construct confidence intervals as before.
Sensitivity seems best at the SM.



c_ϕ = -4

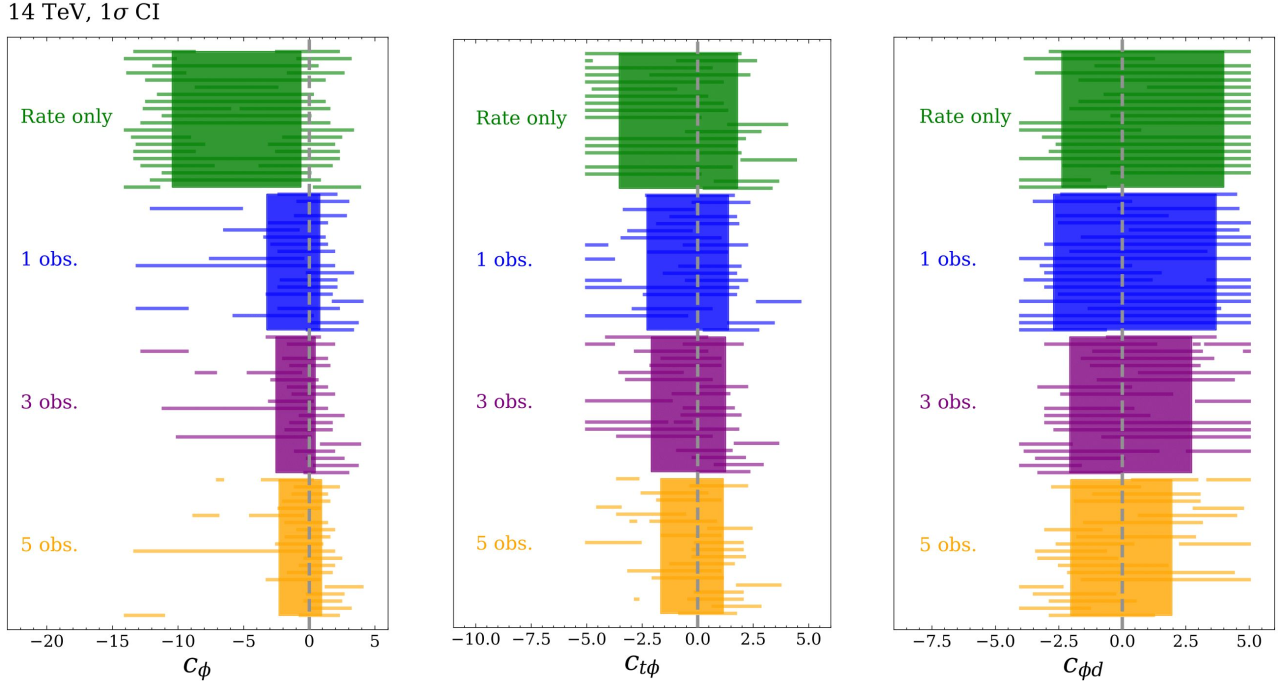
c_ϕ = -6



3 observable results ⟶

1D coefficient recovery on many test sets

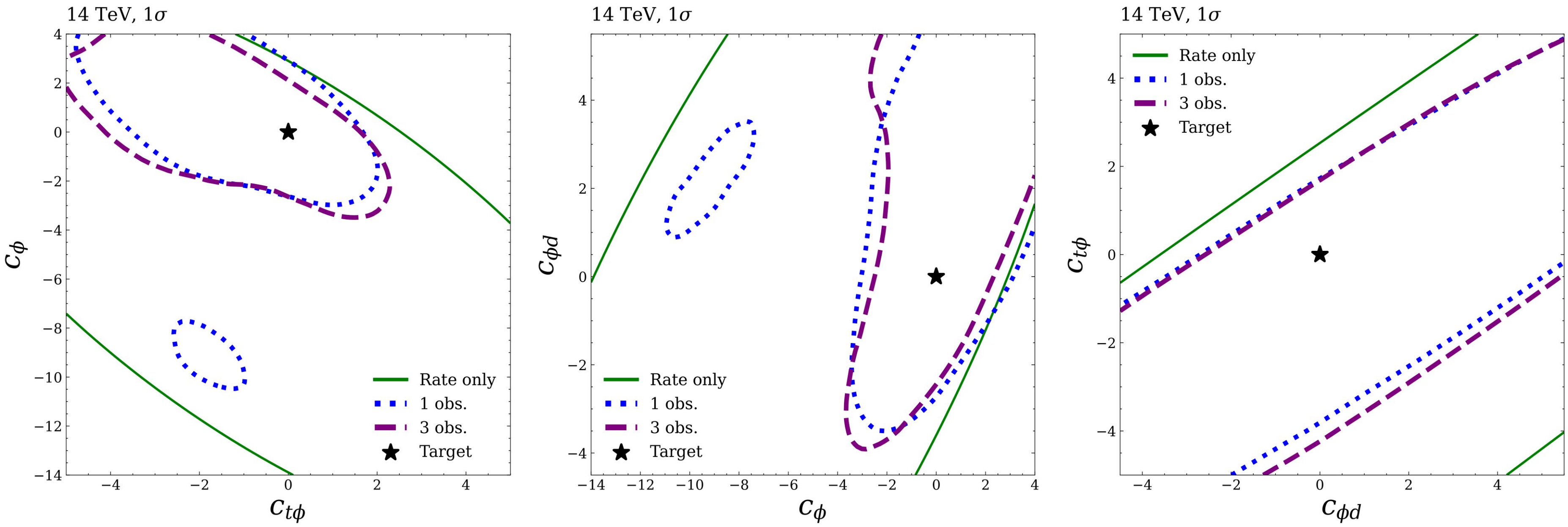
Repeating the analysis on many test set instantiations, we see more reliable and precise results by including more information in the likelihood ratio.



The intervals are *over-conservative*

2D coefficient recovery on a SM test set

Shape-informed likelihood ratios are also superior for constraining 2 parameters at once.



**see backups for FCC-hh projections

Summary

Adding shape information of kinematic observables to cut-and-count analyses can greatly improve their constraining power

Future investigations

❖ More realistic background modeling (with uncertainties)

❖ Balancing signal-enhancing cuts and event preservation for classifier testing

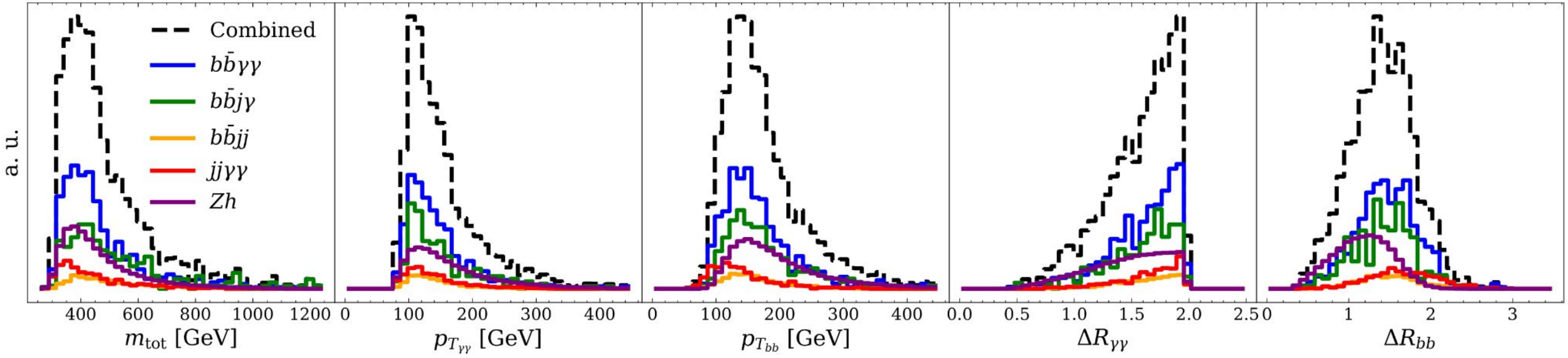
❖ Expanding the Wilson coefficient space

Backup slides

SMEFT Operators in Detail

Operator	Explicit form
$\mathcal{O}_{\phi}$	$(\phi^\dagger\phi - \frac{v^2}{2})^3$
$\mathcal{O}_{\phi d}$	$\partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi)$
$\mathcal{O}_{\phi D}$	$(\phi^\dagger D^\mu\phi)^\dagger(\phi^\dagger D_\mu\phi)$
$\mathcal{O}_{t\phi}$	$(\phi^\dagger\phi - \frac{v^2}{2})\overline{Q}t\tilde{\phi} + \text{h.c.}$
$\mathcal{O}_{\phi G}$	$(\phi^\dagger\phi - \frac{v^2}{2})G_A^{\mu\nu}G_{\mu\nu}^A$
$\mathcal{O}_{tG}$	$ig_s(\overline{Q}\tau^{\mu\nu}T_A t)\tilde{\phi}G_{\mu\nu}^A + \text{h.c.}$

Relevant background processes for *hh* production



In this project, we use the *bbγγ* process as a proxy for all backgrounds

Background feature shapes do not vary with the SMEFT coefficients

The cut flow and event yields

	HL-LHC, 14TeV, $3ab^{-1}$				Future Collider, 100 TeV, $30ab^{-1}$			
	Signal		Background		Signal		Background	
	Events	Retention	Events	Retention	Events	Retention	Events	Retention
Start	257	100%	–	–	89,604	100%	–	–
+ tagging & efficiencies	95	37.1%	4.65×10^4	100%	29,600	33.0%	5.16×10^6	100%
+ kinematic cuts	49	18.9%	1.43×10^4	30.8%	11,100	12.3%	1.58×10^6	30.6%
+ m_h windows	15	5.89%	4.09×10^2	0.88%	3,950	4.40%	4.02×10^4	0.78%
+ angular cuts	13	4.92%	4.37×10^1	0.094%	3,600	4.02%	4.34×10^3	0.084%

$S/B \approx 0.30$

$S/B \approx 0.83$

Mixture models

To learn the likelihood ratio of *mixture models*, e.g.

data = signal (sig) + (bkg)

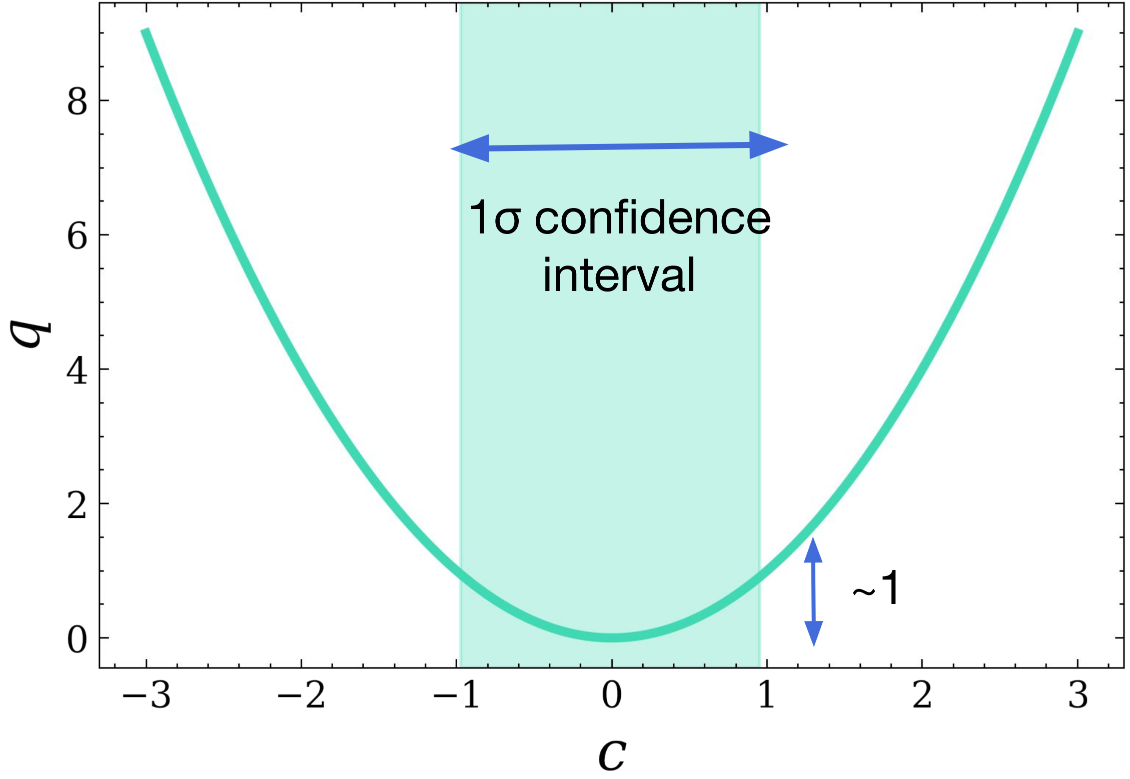
it is useful to decompose the *ratio of mixtures* into *sums of ratios of components*

$$\frac{\text{data 1}}{\text{data 0}} = \left(\frac{\text{sig 0}}{\text{sig 1}} + \frac{\text{bkg 0}}{\text{sig 1}} \right)^{-1} + \left(\frac{\text{sig 0}}{\text{bkg 1}} + \frac{\text{bkg 0}}{\text{bkg 1}} \right)^{-1}$$

These component ratios are much easier for classifiers to learn!

Constructing confidence intervals from q

$$q(c|D) = -2\ln\left(\frac{p(D|c)}{p(D|c=0)}\right)$$



Neural networks and training

Architecture: Dense Neural Networks with 2 layers of 32 nodes

Training: batch size of 1024, weight decay 1e-4, learning rate of 1e-3 that reduces if the validation loss stagnates. Train until validation loss stagnates for 20 epochs.

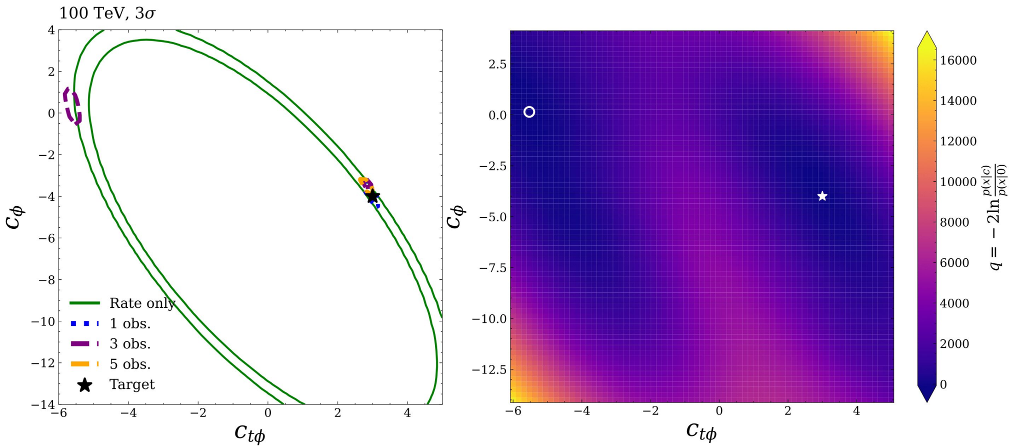
Train-val split: 80-20

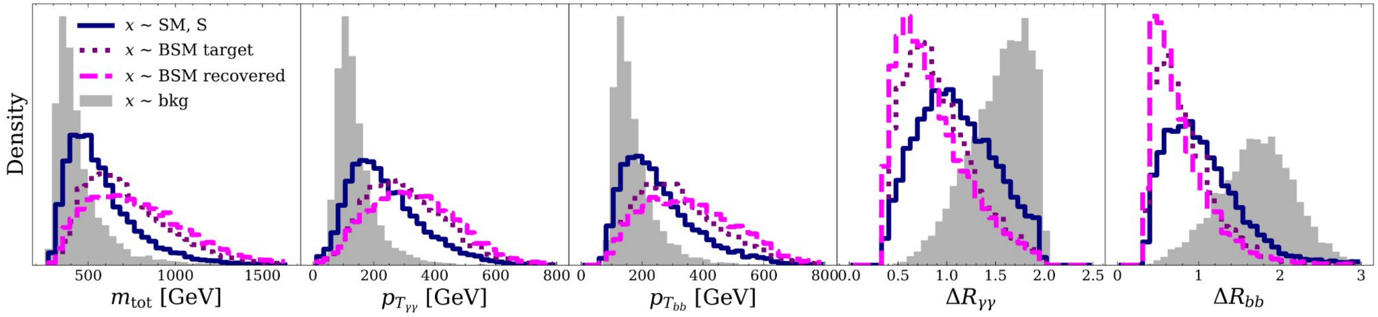
To mitigate the stochastic nature of network training, we ***ensemble*** *the outputs of five networks with different initial random seeds.*

In 2+ dimensions, coefficient recovery is harder

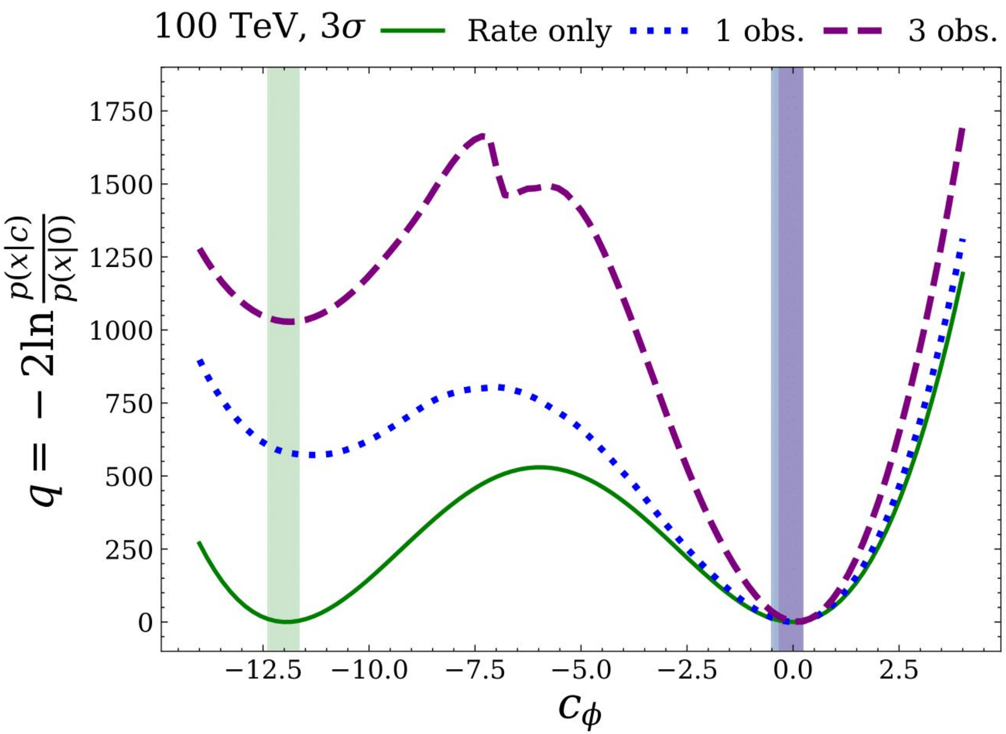
Different BSM values of the coupling can lead to similar kinematic distributions, so the loss landscape becomes multimodel!

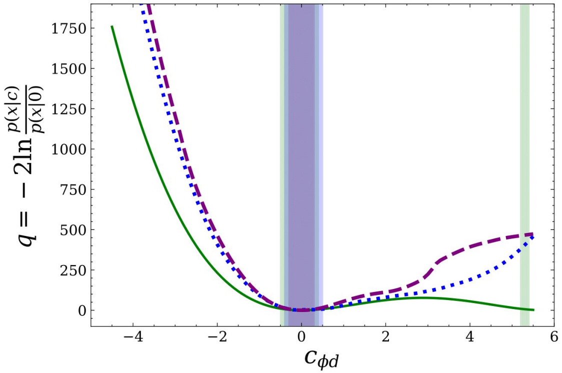
Note that the degeneracy is resolved by using all 5 features.

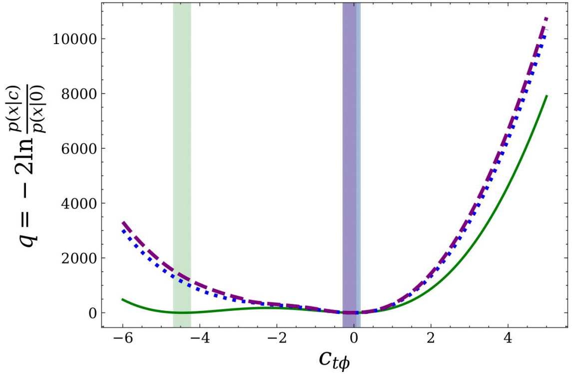




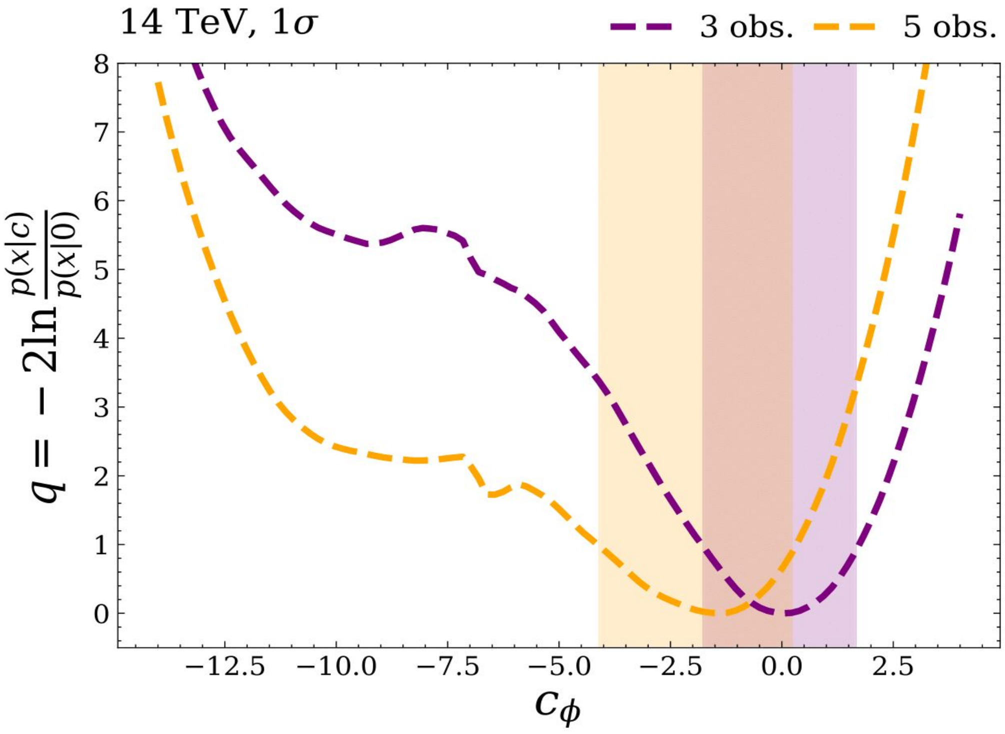
100 TeV Results: 1D

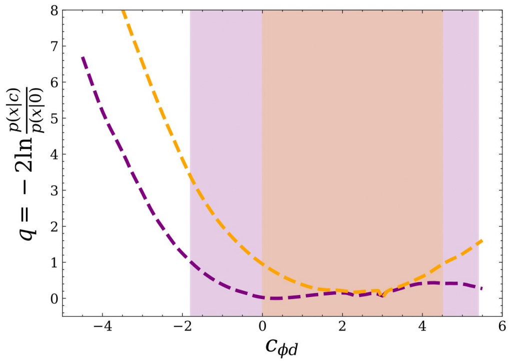


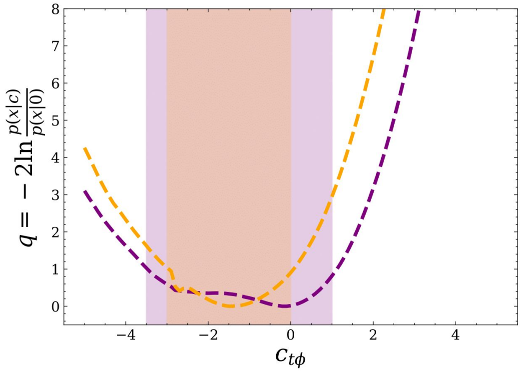




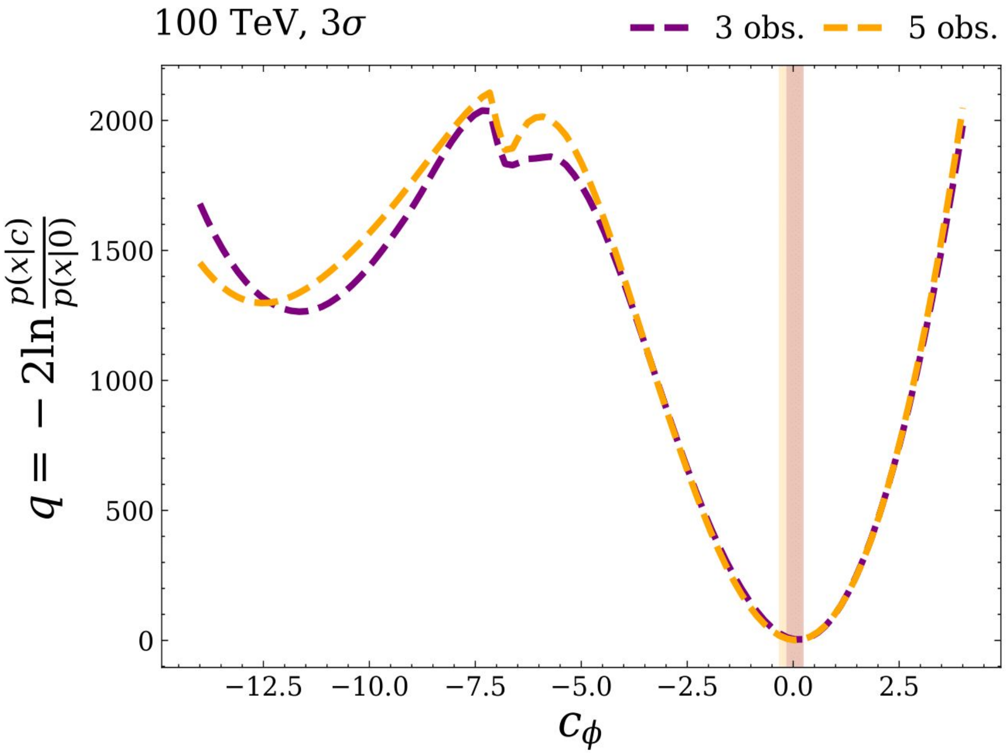
14 TeV Results: 1D, 5 observables

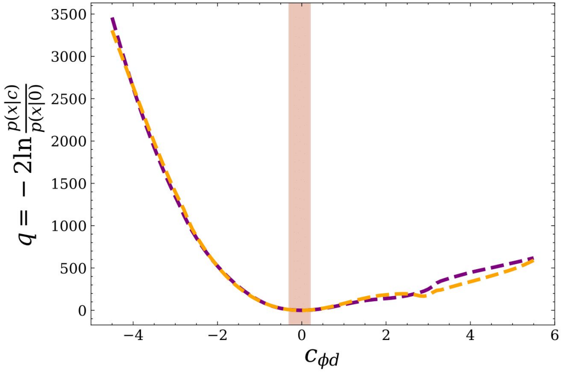


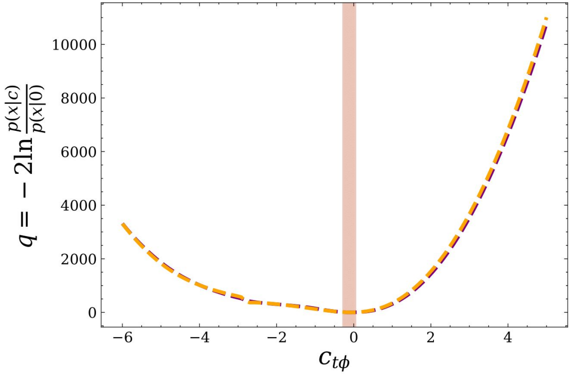




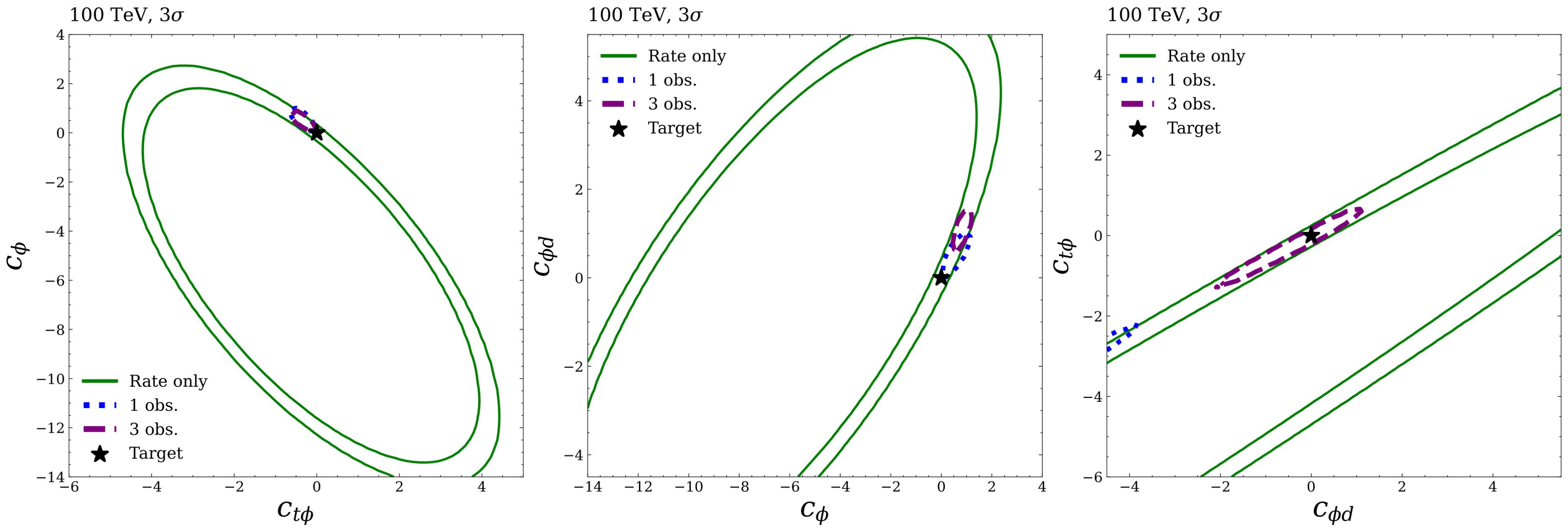
100 TeV Results: 1D, 5 observables



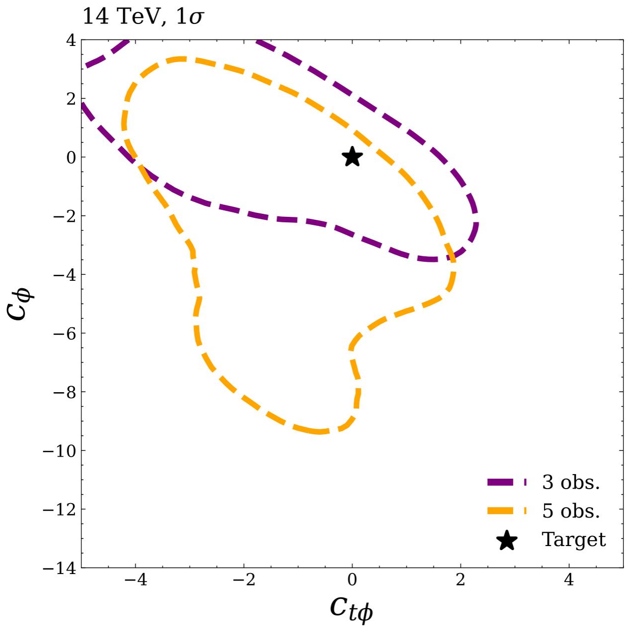


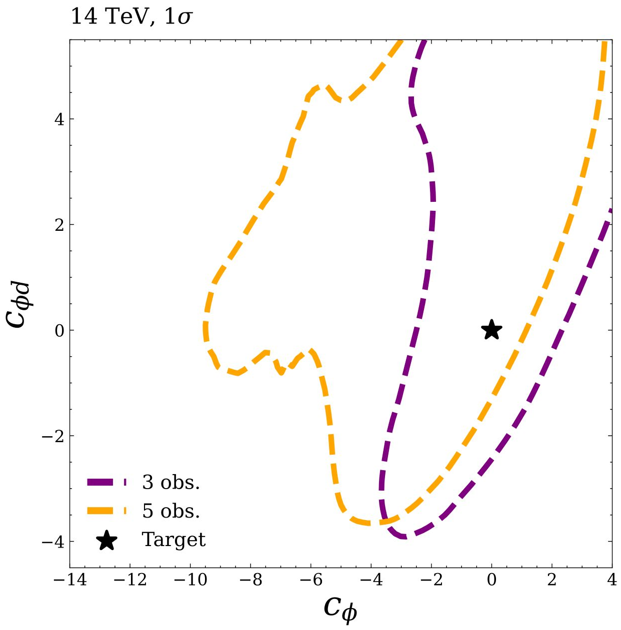


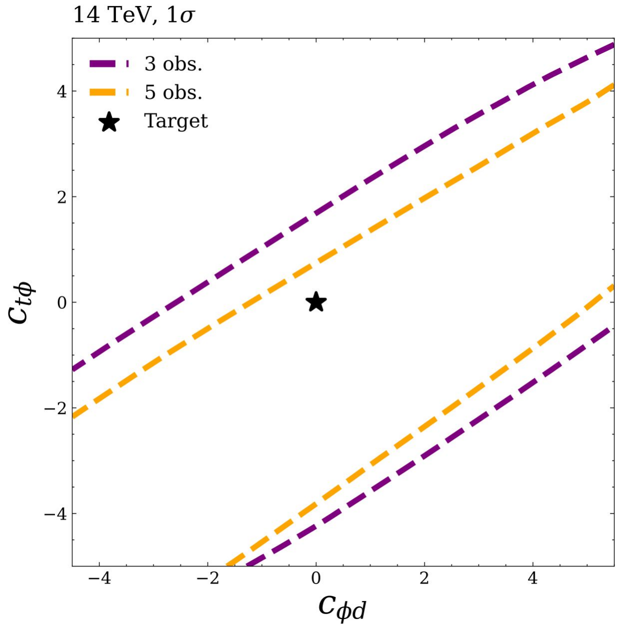
100 TeV Results: 2D



14 TeV Results: 2D, 5 observables







100 TeV Results: 2D, 5 observables