

From Optimal Observables to Machine learning in EFT analyses of $e^+e^- \rightarrow W^+W^-$

Jiayin Gu

Fudan University

Mini-workshop on WWdiff focus topic
June 25, 2024



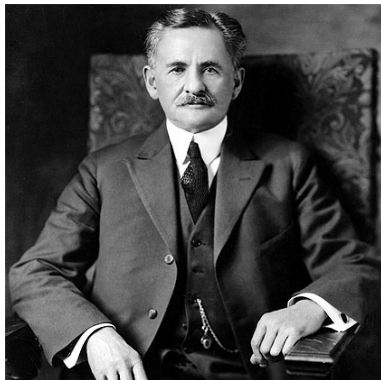
Why SMEFT at future lepton colliders?

- ▶ **Build large colliders** → go to high energy → discover new particles!
- ▶ Higgs and nothing else?
- ▶ What's next?
 - ▶ Build an even larger collider (~ 100 TeV)?
 - ▶ No guaranteed discovery!

Why SMEFT at future lepton colliders?

- ▶ **Build large colliders** → go to high energy → discover new particles!
- ▶ Higgs and nothing else?
- ▶ What's next?
 - ▶ Build an even larger collider (~ 100 TeV)?
 - ▶ No guaranteed discovery!
- ▶ **Build large colliders** → do precision measurements → probe new physics!
 - ▶ Higgs factory! (HL-LHC, or a future lepton collider)
 - ▶ Many other precision measurements! (Z, W, top, ...)
 - ▶ **S**tandard **M**odel **E**ffective **F**ield **T**heory (model independent approach)

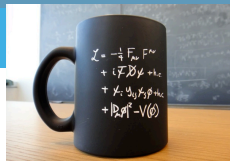
To summarize in one sentence...



“Our future discoveries must be looked for in the sixth place of decimals.”

— Albert A. Michelson

The Standard Model Effective Field Theory



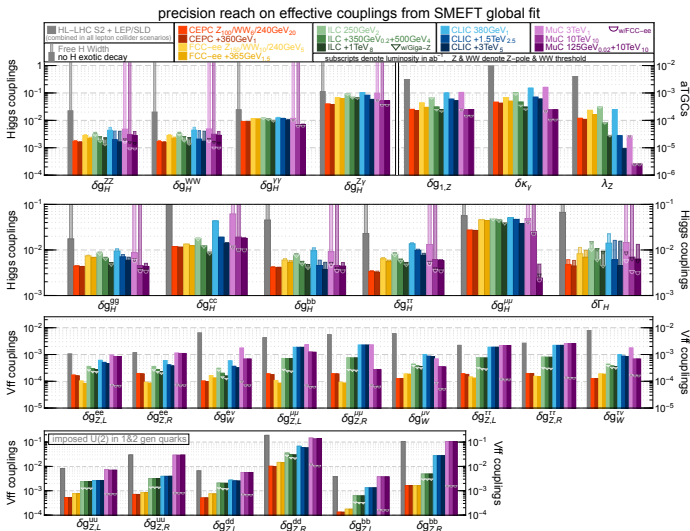
- ▶ $[\mathcal{L}_{\text{SM}}] \leq 4$. Why?
 - ▶ **Bad things happen when we have non-renormalizable operators!**
 - ▶ Everything is fine as long as we are happy with finite precision in perturbative calculation.
- ▶ **d=5:** $\frac{c}{\Lambda} LLHH \sim \frac{c\nu^2}{\Lambda} \nu\nu$, Majorana neutrino mass.
- ▶ Assuming Baryon and Lepton numbers are conserved,

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_j \frac{c_j^{(8)}}{\Lambda^4} \mathcal{O}_j^{(8)} + \dots$$

- ▶ If $\Lambda \gg v, E$, then **SM + dimension-6 operators** are sufficient to parameterize the physics around the electroweak scale.

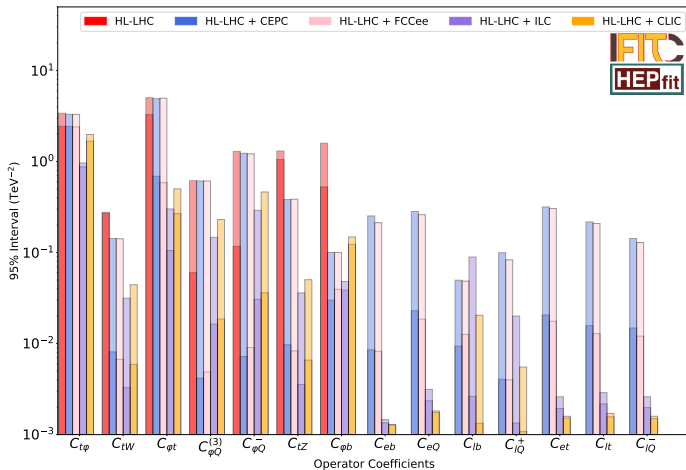
Higgs + EW, Results from the Snowmass 2021 (2022) study

[2206.08326] de Blas, Du, Grojean, JG, Miralles, Peskin, Tian, Vos, Vryonidou



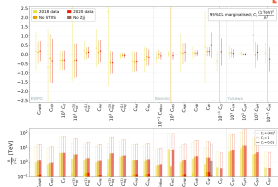
Top operators with $e^+e^- \rightarrow t\bar{t}$

[2206.08326] de Blas, Du, Grojean, JG, Miralles, Peskin, Tian, Vos, Vryonidou

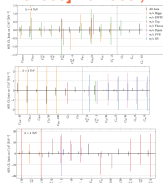


Many studies on SMEFT global fits!

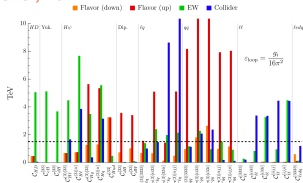
[2012.02779] Ellis, Madigan, Mimasu, Sanz, You



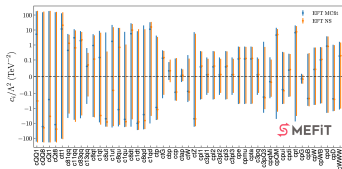
[2311.04963] Bartocci, Biekötter, Hurth



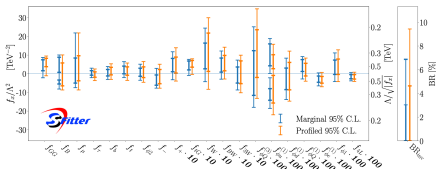
[2311.00020] Allwicher, Cornella, Isidori, Stefaneke



[2105.00006] The SMEFIT Collaboration



[2208.08454] Brivio, Bruggisser, Elmer, Geoffroy, Luchmann, Plehn



Machine learning in SMEFT analyses

Machine learning is not physics!



past

真香！ (Delicious!)



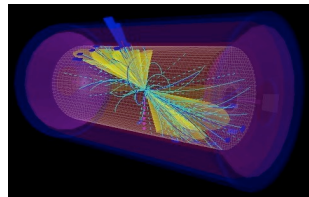
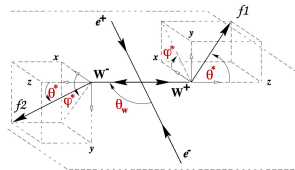
now

- ▶ [2401.02474] Shengdu Chai, JG, Lingfeng Li on $e^+ e^- \rightarrow W^+ W^-$.
- ▶ **Many studies!**
 - ▶ [1805.00013, 1805.00020] Brehmer, Cranmer, Louppe, Pavez,
 - [2007.10356] Chen, Glioti, Panico, Wulzer ($pp \rightarrow ZW$),
 - [2211.02058] Ambrosio, Hoeve, Madigan, Rojo, Sanz ($pp \rightarrow tt, pp \rightarrow hZ$),
 -

Why Machine learning in SMEFT analyses?

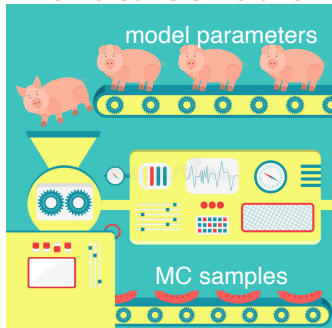
- ▶ In many cases, the new physics contributions are sensitive to the differential distributions.
 - ▶ $e^+e^- \rightarrow W^+W^- \rightarrow 4f \Rightarrow 5$ angles
 - ▶ $e^+e^- \rightarrow t\bar{t} \rightarrow bW^+\bar{b}W^- \rightarrow 6f \Rightarrow 9$ angles
 - ▶ How to extract information from the differential distribution?
 - ▶ If we have the full knowledge of $\frac{d\sigma}{d\Omega} \Rightarrow$ matrix-element method, optimal observables...

- ▶ The ideal $\frac{d\sigma}{d\Omega}$ we can calculate is not the $\frac{d\sigma}{d\Omega}$ that we actually measure!
 - ▶ detector acceptance, measurement uncertainties, ISR/beamstrahlung ...
 - ▶ In practice we only have **MC samples**, not analytic expressions, for $\frac{d\sigma}{d\Omega}$.

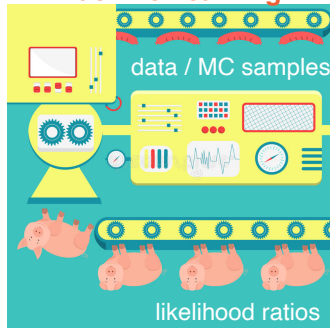


The “inverse problem”

Monte Carlo simulation



Machine Learning

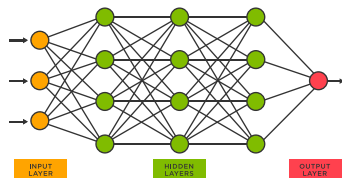


- ▶ **Forward:** From model parameters we can calculate the ideal $\frac{d\sigma}{d\Omega}$, simulate complicated effects and produce MC samples.
- ▶ **Inverse:** From data / MC samples, how do we know the model parameters?
- ▶ With **Neural Network** we can (in principle) reconstruct $\frac{d\sigma}{d\Omega}$ (or likelihood ratios) from MC samples.

A rough sketch

- ▶ We have a theory (SMEFT) that gives a differential cross section $\frac{d\sigma}{d\Omega}$ which is a function of the parameters of interest \mathbf{c} (Wilson coefficients).
- ▶ For simplicity, let's ignore the total rate and focus on $\frac{1}{\sigma} \frac{d\sigma}{d\Omega} \equiv p(\mathbf{x}|\mathbf{c})$, i.e. it's a probability density function of the observables \mathbf{x} .
- ▶ Define the likelihood function $\mathcal{L}(\mathbf{c}|\mathbf{x}) \equiv p(\mathbf{x}|\mathbf{c})$. For a sample of N events, maximizing the total likelihood $\prod_{i=1}^N \mathcal{L}(\mathbf{c}|\mathbf{x}_i)$ (or the log likelihood) gives the best estimator for \mathbf{c} . (matrix-element method)
- ▶ For two model points \mathbf{c}_0 and \mathbf{c}_1 , the likelihood ratio $r(\mathbf{x}|\mathbf{c}_0, \mathbf{c}_1) = \frac{p(\mathbf{x}|\mathbf{c}_0)}{p(\mathbf{x}|\mathbf{c}_1)}$ provides the optimal statistical test (Neyman–Pearson lemma).
 - ▶ We usually set \mathbf{c}_1 to be SM.

A rough sketch

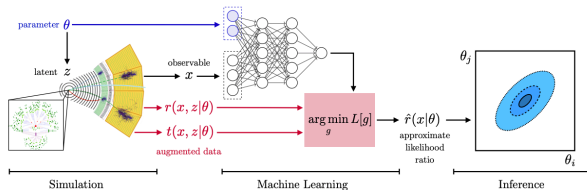


- ▶ We do not know $p(x|c)$ or $r(x|c_0, c_1)$, but we can use **neural network** to construct an estimator $\hat{r}(x|c_0, c_1)$ and a **loss function(al)** $L(\hat{r})$ which is minimized when $\hat{r} = r$.
- ▶ By minimizing $L(\hat{r})$ with respect to \hat{r} we can find the true r in the ideal limit (large sample, perfect training).
- ▶ There are many ways to construct a loss function(al)....
- ▶ With additional assumptions on how $\frac{d\sigma}{d\Omega}$ depends on \mathbf{c} (i.e., a linear or a quadratic relation), we only need to train a finite number of times to obtain an estimator $\hat{r}(x|c_0, c_1)$ for any \mathbf{c}_0 .

Particle physics structure

- ▶ One could make use of **latent variable “z”** (the parton level analytic result for $\frac{d\sigma}{d\Omega}$) to increase the performance of ML.

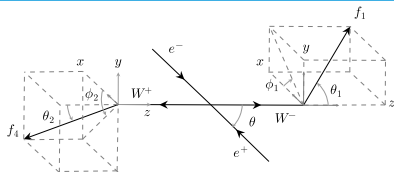
[1805.00013, 1805.00020] Brehmer, Cranmer, Louppe, Pavez



- ▶ Assuming linear dependences $\frac{d\sigma}{d\Omega} = S_0 + \sum_i S_{1,i} c_i$, there is a method called **SALLY** (Score approximates likelihood locally).
 - ▶ In this case, for each parameter we only need to train once to obtain $\alpha_i \equiv \frac{S_{1,i}}{S_0}$. (It is basically the ML version of **Optimal Observables**.)
 - ▶ We can calculate the “ideal” $\alpha(z)$ which will help us train the actual $\alpha(x)$.

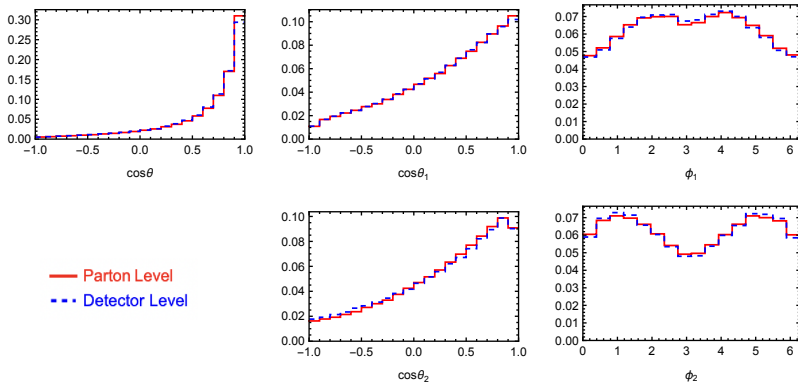
$$L[\hat{\alpha}(x)] = \sum_{x_i, z_i \sim \text{SM}} |\alpha(z_i) - \hat{\alpha}(x_i)|^2.$$

The ML analysis of $e^+e^- \rightarrow W^+W^-$



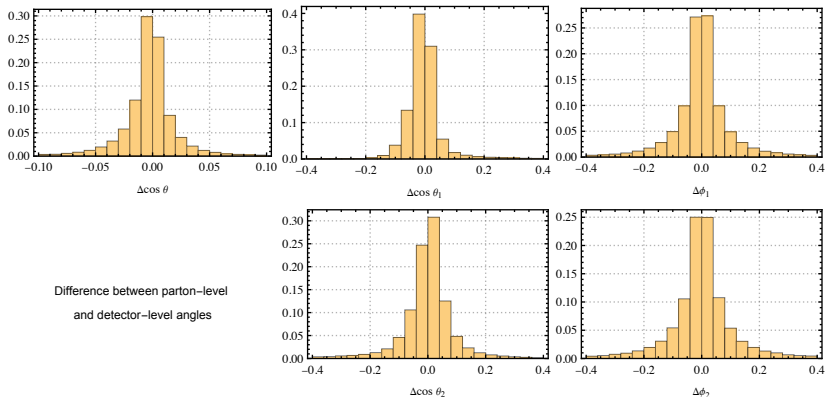
- ▶ $e^+e^- \rightarrow W^+W^-$, 240 GeV, unpolarized beams, semileptonic channel.
- ▶ Training sample: 2×10^6 events. Validation sample: 5×10^5 events.
 - ▶ This is much smaller than the actual data set ($\sim 10^8$ events) we will have!
- ▶ MadGraph/Pythia/Delphes, ILD-like detector card, IRS implemented.
- ▶ Background: $e^+e^- \rightarrow ZZ \rightarrow jj\ell^+\ell^-$ with a missing lepton.
- ▶ Inputs: particle 4 momenta + 5 reconstructed angles.
- ▶ Fully connected neural network (FCNN), 9 layers and 200 nodes each layer.
- ▶ Average over 8 NN models to reduce systematics from training.

The parton-level and detector-level distributions



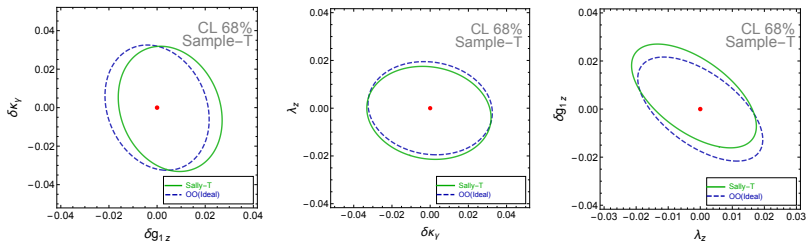
- ▶ Here the decay angles for the hadronic W are “unfolded.”
- ▶ In the Analysis, the hadronic decay angles are folded.

The parton-level and detector-level distributions



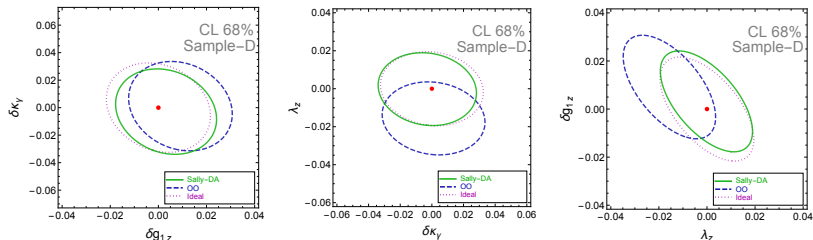
- ▶ The difference between detector-level and parton-level values for each angle.

3-aTGC fit, truth-level sample



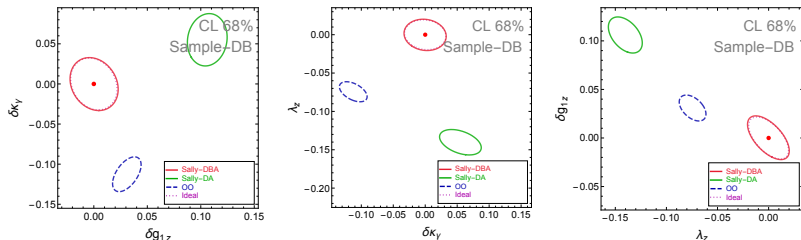
- ▶ The results are scaled to 10^4 events.
- ▶ At the truth level, Optimal Observables (OO) gives the ideal results by construction.
- ▶ Machine learning suffers from imperfect training and has no advantage.

3-aTGC fit, detector-level sample



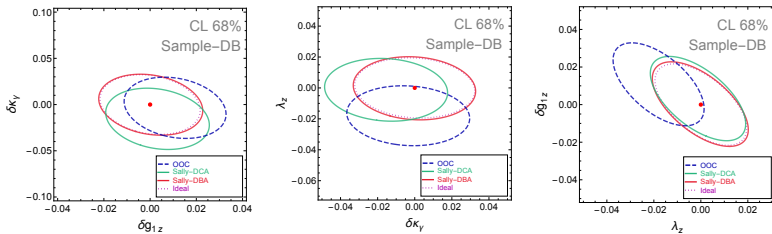
- ▶ Naively applying truth-level optimal observables to detector-level samples could lead to a large bias!
- ▶ ML model trained on detector-level samples (Sally-DA) automatically take care of the detector effects and are more robust.

3-aTGC fit, detector-level sample with background



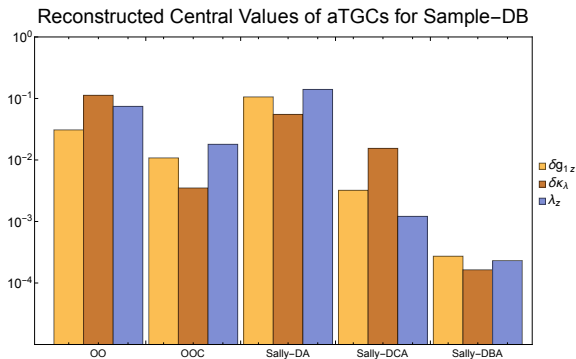
- ▶ 10% ZZ background. A large bias can be introduced if we failed to take account of it!
- ▶ SALLY-DBA: trained with both signal and backgrounds with the correct weighting to reconstruct the $\hat{\alpha}(x)$ for the combined differential cross section.

Further comparisons between methods



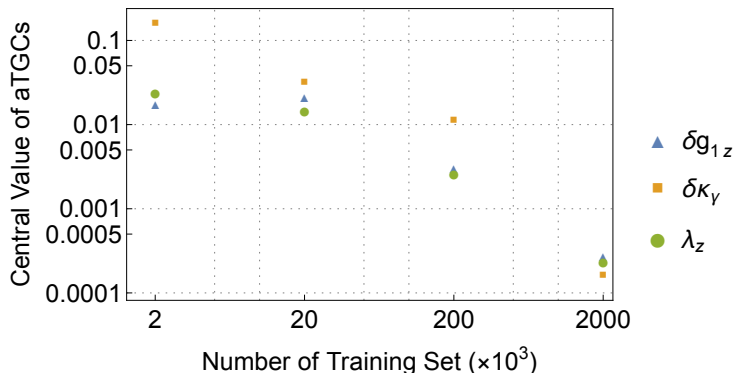
- ▶ OOC, Sally-DCA: Optimal observables and Sally-DA combined with a classifier.

Comparison of biases (reconstructed central values)



- For the detector-level sample with background, Sally-DBA has the least bias.

bias vs. training sample size



- ▶ The current bias is still unacceptable for future colliders with $\sim 10^8$ WW events.
- ▶ Hopefully with more computing resources in the future, the bias can be reduced to the desired level.

Conclusion

- ▶ **We have no idea what is the new physics beyond the Standard Model.**
- ▶ **One important direction to move forward is to do precision measurements of the Standard Model processes.**
 - ▶ HL-LHC is ok, but a future lepton collider is better!
 - ▶ SMEFT is a good theory framework (but is not everything).
- ▶ **Machine learning is (likely to be) the future!**
 - ▶ High precision \Rightarrow high demand on reducing biases/systematics to the same level.
 - ▶ ML helps take care of the detector/background effects.

Conclusion



- ▶ **When will Machine take over?**
 - ▶ Before or after a future lepton collider is built?
- ▶ **Many more studies to do!**
 - ▶ Di-leptonic & fully hadronic channels.
 - ▶ Other processes, *e.g.* $e^+e^- \rightarrow t\bar{t}$ (current work with other students),
 - ▶ In reality, MC simulation does not perfectly describe data...

backup slides

SMEFT parameterization of $e^+ e^- \rightarrow W^+ W^-$

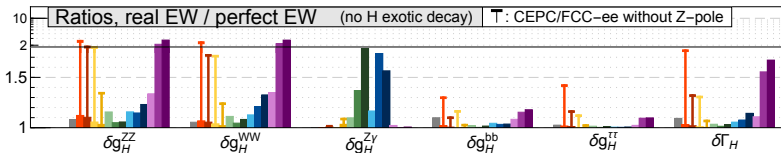
Tree-level dim-6 CP-even operators: 6 parameters (excluding modifications in m_W):

$$\delta g_{1Z}, \quad \delta \kappa_\gamma, \quad \lambda_Z, \quad \delta g_W^\ell, \quad \delta g_{Z,L}^e, \quad \delta g_{Z,R}^e. \quad (1)$$

$$\begin{aligned} \mathcal{L}_{\text{TGC}} = & ie(W_{\mu\nu}^+ W^{-\mu} - W_{\mu\nu}^- W^{+\mu})A^\nu + ie(1 + \delta \kappa_\gamma)A^{\mu\nu} W_\mu^+ W_\nu^- \\ & + igc_w \left[(1 + \delta g_{1Z})(W_{\mu\nu}^+ W^{-\mu} - W_{\mu\nu}^- W^{+\mu})Z^\nu + (1 + \delta g_{1Z} - \frac{s_w^2}{c_w^2} \delta \kappa_\gamma) Z^{\mu\nu} W_\mu^+ W_\nu^- \right] \\ & + \frac{ig\lambda_Z}{m_W^2} (s_w W_\mu^{+\nu} W_\nu^{-\rho} A_\rho^\mu + c_w W_\mu^{+\nu} W_\nu^{-\rho} Z_\rho^\mu), \end{aligned} \quad (2)$$

$$\begin{aligned} \mathcal{L}_{\text{Vff}} = & -\frac{g}{\sqrt{2}}(1 + \delta g_W^\ell) [W_\mu^+ \bar{\nu}_L \gamma^\mu e_L + \text{h.c.}] \\ & -\frac{g}{c_w} Z_\mu \left[\bar{e}_L \gamma^\mu \left(-\frac{1}{2} + s_w^2 + \delta g_{Z,L}^e\right) e_L + \bar{e}_R \gamma^\mu (s_w^2 + \delta g_{Z,R}^e) e_R \right] + \dots, \end{aligned} \quad (3)$$

Impacts of (lack of) the Z-pole run



- ▶ Without good Z-pole measurements, the $eeZh$ contact interaction may have a significant impact on the Higgs coupling determination.
- ▶ Current (LEP) Z-pole measurements are not good enough for CEPC/FCC-ee Higgs measurements!
 - ▶ **A future Z-pole run is important!**
- ▶ Linear colliders suffer less from the lack of a Z-pole run. **(Win Win!)**



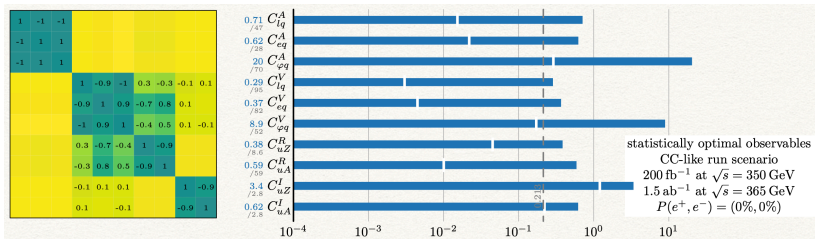
Probing Top operators with $e^- e^+ \rightarrow t \bar{t}$

[arXiv:1807.02121] Durieux, Perelló, Vos, Zhang

$$\begin{aligned}
 O_{\varphi q}^1 &\equiv \frac{y_t^2}{2} \bar{q} \gamma^\mu q \quad \varphi^\dagger i \overleftrightarrow{D}_\mu \varphi, & O_{uG} &\equiv y_t g_s \bar{q} T^A \sigma^{\mu\nu} u \quad \epsilon \varphi^* G_{\mu\nu}^A, \\
 O_{\varphi q}^3 &\equiv \frac{y_t^2}{2} \bar{q} \tau^I \gamma^\mu q \quad \varphi^\dagger i \overleftrightarrow{D}_\mu^I \varphi, & O_{uW} &\equiv y_t g_W \bar{q} \tau^I \sigma^{\mu\nu} u \quad \epsilon \varphi^* W_{\mu\nu}^I, \\
 O_{\varphi u} &\equiv \frac{y_t^2}{2} \bar{u} \gamma^\mu u \quad \varphi^\dagger i \overleftrightarrow{D}_\mu \varphi, & O_{dW} &\equiv y_t g_W \bar{q} \tau^I \sigma^{\mu\nu} d \quad \epsilon \varphi^* W_{\mu\nu}^I, \\
 O_{\varphi ud} &\equiv \frac{y_t^2}{2} \bar{u} \gamma^\mu d \quad \varphi^T \epsilon i D_\mu \varphi, & O_{uB} &\equiv y_t g_Y \bar{q} \sigma^{\mu\nu} u \quad \epsilon \varphi^* B_{\mu\nu},
 \end{aligned}$$

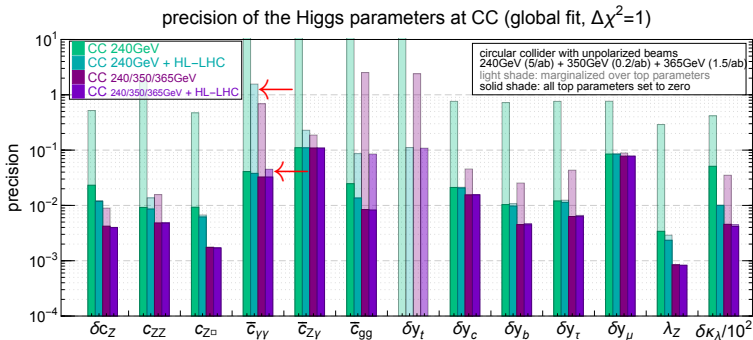
$$\begin{aligned}
 O_{lq}^1 &\equiv \frac{1}{2} \bar{q} \gamma_\mu q \quad \bar{l} \gamma^\mu l, \\
 O_{lq}^3 &\equiv \frac{1}{2} \bar{q} \tau^I \gamma_\mu q \quad \bar{l} \tau^I \gamma^\mu l, \\
 O_{lu} &\equiv \frac{1}{2} \bar{u} \gamma_\mu u \quad \bar{l} \gamma^\mu l, \\
 O_{eq} &\equiv \frac{1}{2} \bar{q} \gamma_\mu q \quad \bar{e} \gamma^\mu e, \\
 O_{eu} &\equiv \frac{1}{2} \bar{u} \gamma_\mu u \quad \bar{e} \gamma^\mu e,
 \end{aligned}$$

- ▶ Also need to include **top dipole** interactions and **$eett$** contact interactions!
- ▶ Hard to resolve the **top couplings** from **$4f$** interactions with just the 365 GeV run.
 - ▶ Can't really separate $e^+ e^- \rightarrow Z/\gamma \rightarrow t \bar{t}$ from $e^+ e^- \rightarrow Z' \rightarrow t \bar{t}$.
 - ▶ Is that a big deal?



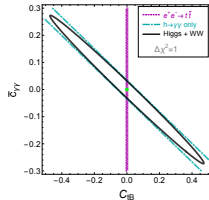
Top operators in loops (Higgs processes)

[1809.03520] G. Durieux, JG, E. Vryonidou, C. Zhang



- ▶ $O_{IB} = (\bar{Q}\sigma^{\mu\nu}t)\tilde{\varphi}B_{\mu\nu} + h.c.$ is not very well constrained at the LHC, and it generates dipole interactions that contributes to the $h\gamma\gamma$ vertex.

- ▶ Deviations in $h\gamma\gamma$ coupling \Rightarrow run at ~ 365 GeV to confirm?



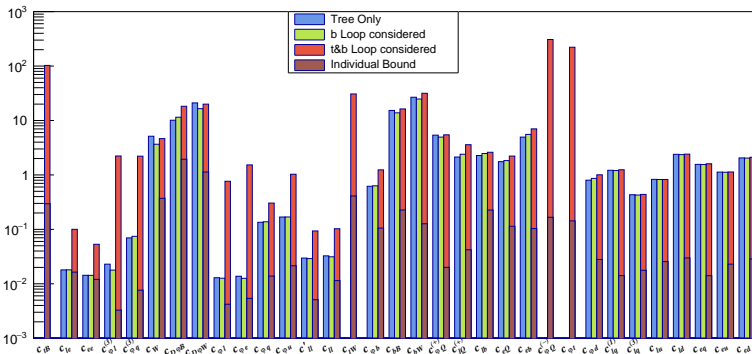
Top operators in loops (current EW processes)

[2205.05655] Y. Liu, Y. Wang, C. Zhang, L. Zhang, JG

	Experiment	Observables
Low Energy	CHARM/CDHS/ CCFR/NuTeV/ APV/QWEAK/ PVDIS	Effective Couplings
Z-pole	LEP/SLC	$\frac{\text{Total decay width } \Gamma_Z}{\text{Hadronic cross-section } \sigma_{had}}$ $\frac{\text{Ratio of decay width } R_f}{\text{Forward-Backward Asymmetry } A_{FB}^f}$ $\frac{\text{Polarized Asymmetry } A_f}{\text{Total decay width } \Gamma_W}$
W-pole	LHC/Tevatron/ LEP/SLC	$\frac{W \text{ branching ratios } Br(W \rightarrow l\nu_l)}{\text{Mass of W Boson } M_W}$ $\frac{\text{Hadronic cross-section } \sigma_{had}}{\text{Ratio of cross-section } R_f}$
$ee \rightarrow qq$	LEP/TRISTAN	$\frac{\text{Forward-Backward Asymmetry for } b/c}{\text{cross-section } \sigma_f} A_{FB}^f$
$ee \rightarrow ll$	LEP	$\frac{\text{Forward-Backward Asymmetry } A_{FB}^f}{\text{Differential cross-section } \frac{d\sigma_f}{d\cos\theta}}$
$ee \rightarrow WW$	LEP	$\frac{\text{cross-section } \sigma_{WW}}{\text{Differential cross-section } \frac{d\sigma_{WW}}{d\cos\theta}}$

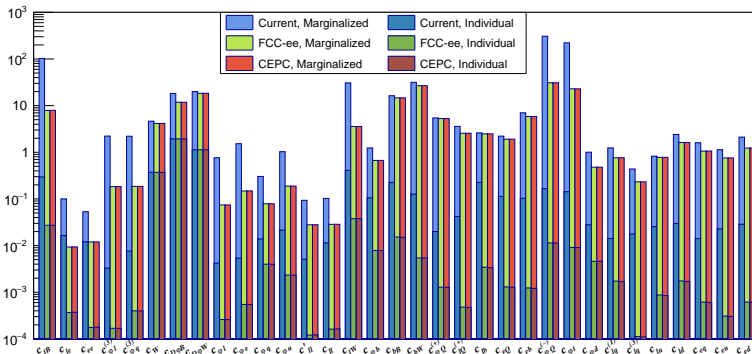
- ▶ Top operators (1-loop) + EW operators (tree, including bottom dipole operators)
- ▶ $e^+e^- \rightarrow f\bar{f}$ at different energies, $e^+e^- \rightarrow W^+W^-$.

Top operators in loops (current EW processes)



► Good sensitivities, but too many parameters for a global fit...

Top operators in loops (future EW processes)



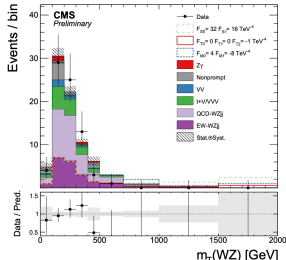
- ▶ Good sensitivities, but too many parameters for a global fit...
- ▶ It shows the importance of directly measuring $e^+e^- \rightarrow t\bar{t}$.

Probing dimension-8 operators?

- ▶ The dimension-8 contribution has a large energy enhancement ($\sim E^4/\Lambda^4$)!
- ▶ It is difficult for LHC to probe these bounds.
 - ▶ Low statistics in the high energy bins.
 - ▶ Example: Vector boson scattering.
 - ▶ $\Lambda \lesssim \sqrt{s}$, the EFT expansion breaks down!
- ▶ Can we separate the dim-8 and dim-6 effects?
 - ▶ Precision measurements at several different \sqrt{s} ?
(A **very** high energy lepton collider?)
 - ▶ Or find some special process where dim-8 gives the leading new physics contribution?

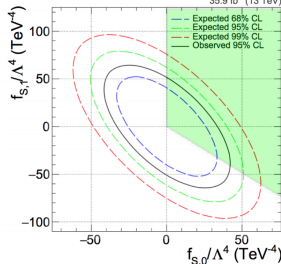
CMS-PAS-SMP-18-001

35.9 fb⁻¹ (13 TeV)



positivity bounds from 1902.08977 Bi, Zhang, Zhou

35.9 fb⁻¹ (13 TeV)



The diphoton channel [arXiv:2011.03055] Phys.Rev.Lett. 129, 011805, JG, Lian-Tao Wang, Cen Zhang

- ▶ $e^+e^- \rightarrow \gamma\gamma$ (or $\mu^+\mu^- \rightarrow \gamma\gamma$), SM, non-resonant.
- ▶ Leading order contribution: **dimension-8 contact interaction**.
($f^+f^- \rightarrow \bar{e}_L e_L$ or $e_R \bar{e}_R$)

$$\mathcal{A}(f^+f^- \gamma^+ \gamma^-)_{\text{SM+d8}} = 2e^2 \frac{\langle 24 \rangle^2}{\langle 13 \rangle \langle 23 \rangle} + \frac{a}{v^4} [13][23] \langle 24 \rangle^2.$$

- ▶ Can probe dim-8 operators (and their positivity bounds) at a **Higgs factory** (~ 240 GeV)!

