Probing BSM effects in $e^+e^- \to WW$ with machine learning

Shengdu Chai

Fudan University, Physics Department

Oct.12 2022

Current work with Lingfeng Li and Jiayin Gu

- How to probe beyond the standard model physics?
- Why do we study the process of diboson?
- Why it is necessary to use Machine Learning Method?

Big Picture

- Build large colliders $\rightarrow$ high energy $\rightarrow$ discover new particles!

- Build a larger collider?
- No guaranteed discovery!

Big Picture

- Build large colliders $\rightarrow$ high energy $\rightarrow$ discover new particles!

  $\searrow$

  do precision measurements $\rightarrow$ discover new physics indirectly!

- Build a larger collider?
- No guaranteed discovery!

- Higgs Factory! (CEPC,ILC,etc)
- Standard Model Effective Field Theory(SMEFT)

The Standard Model Effective Field Theory

- $[\mathcal{L}_{SM}] \leq 4$. Why?
  - Renormalizable
  - Higher dimensional operators are fine as long as we are happy with finite precision in perturbative calculation.
- Assuming Baryon and Lepton numbers are conserved,

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{\mathrm{c}_{\mathrm{i}}^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_j \frac{\mathrm{c}_{\mathrm{j}}^{(8)}}{\Lambda^4}\mathcal{O}_j^{(8)} + \cdots$$

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

The Standard Model Effective Field Theory

X^3		φ^6 and $\varphi^4 D^2$		$\psi^2\varphi^3$	
Q_G	$f^{ABC} G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	Q_φ	$(\varphi^\dagger\varphi)^3$	$Q_{e\varphi}$	$(\varphi^\dagger\varphi)(\bar l_p e_r \varphi)$
$Q_{\widetilde G}$	$f^{ABC} \widetilde G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	$Q_{\varphi\Box}$	$(\varphi^\dagger\varphi)\Box(\varphi^\dagger\varphi)$	$Q_{u\varphi}$	$(\varphi^\dagger\varphi)(\bar q_p u_r \widetilde\varphi)$
Q_W	$\varepsilon^{IJK} W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$	$Q_{\varphi D}$	$\left(\varphi^\dagger D^\mu\varphi\right)^\star \left(\varphi^\dagger D_\mu\varphi\right)$	$Q_{d\varphi}$	$(\varphi^\dagger\varphi)(\bar q_p d_r \varphi)$
$Q_{\widetilde W}$	$\varepsilon^{IJK} \widetilde W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$				

$X^2\varphi^2$		$\psi^2 X\varphi$		$\psi^2\varphi^2 D$	
$Q_{\varphi G}$	$\varphi^\dagger\varphi\, G^A_{\mu\nu} G^{A\mu\nu}$	Q_{eW}	$(\bar l_p \sigma^{\mu\nu} e_r)\tau^I \varphi W^I_{\mu\nu}$	$Q_{\varphi l}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar l_p \gamma^\mu l_r)$
$Q_{\varphi\widetilde G}$	$\varphi^\dagger\varphi\, \widetilde G^A_{\mu\nu} G^{A\mu\nu}$	Q_{eB}	$(\bar l_p \sigma^{\mu\nu} e_r)\varphi B_{\mu\nu}$	$Q_{\varphi l}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar l_p \tau^I \gamma^\mu l_r)$
$Q_{\varphi W}$	$\varphi^\dagger\varphi\, W^I_{\mu\nu} W^{I\mu\nu}$	Q_{uG}	$(\bar q_p \sigma^{\mu\nu} T^A u_r)\widetilde\varphi\, G^A_{\mu\nu}$	$Q_{\varphi e}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar e_p \gamma^\mu e_r)$
$Q_{\varphi\widetilde W}$	$\varphi^\dagger\varphi\, \widetilde W^I_{\mu\nu} W^{I\mu\nu}$	Q_{uW}	$(\bar q_p \sigma^{\mu\nu} u_r)\tau^I \widetilde\varphi\, W^I_{\mu\nu}$	$Q_{\varphi q}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar q_p \gamma^\mu q_r)$
$Q_{\varphi B}$	$\varphi^\dagger\varphi\, B_{\mu\nu} B^{\mu\nu}$	Q_{uB}	$(\bar q_p \sigma^{\mu\nu} u_r)\widetilde\varphi\, B_{\mu\nu}$	$Q_{\varphi q}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar q_p \tau^I \gamma^\mu q_r)$
$Q_{\varphi\widetilde B}$	$\varphi^\dagger\varphi\, \widetilde B_{\mu\nu} B^{\mu\nu}$	Q_{dG}	$(\bar q_p \sigma^{\mu\nu} T^A d_r)\varphi\, G^A_{\mu\nu}$	$Q_{\varphi u}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar u_p \gamma^\mu u_r)$
$Q_{\varphi WB}$	$\varphi^\dagger\tau^I\varphi\, W^I_{\mu\nu} B^{\mu\nu}$	Q_{dW}	$(\bar q_p \sigma^{\mu\nu} d_r)\tau^I \varphi\, W^I_{\mu\nu}$	$Q_{\varphi d}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar d_p \gamma^\mu d_r)$
$Q_{\varphi\widetilde WB}$	$\varphi^\dagger\tau^I\varphi\, \widetilde W^I_{\mu\nu} B^{\mu\nu}$	Q_{dB}	$(\bar q_p \sigma^{\mu\nu} d_r)\varphi\, B_{\mu\nu}$	$Q_{\varphi ud}$	$i(\widetilde\varphi^\dagger D_\mu\varphi)(\bar u_p \gamma^\mu d_r)$

$(\bar LL)(\bar LL)$		$(\bar RR)(\bar RR)$		$(\bar LL)(\bar RR)$	
Q_{ll}	$(\bar l_p \gamma_\mu l_r)(\bar l_s \gamma^\mu l_t)$	Q_{ee}	$(\bar e_p \gamma_\mu e_r)(\bar e_s \gamma^\mu e_t)$	Q_{le}	$(\bar l_p \gamma_\mu l_r)(\bar e_s \gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar q_p \gamma_\mu q_r)(\bar q_s \gamma^\mu q_t)$	Q_{uu}	$(\bar u_p \gamma_\mu u_r)(\bar u_s \gamma^\mu u_t)$	Q_{lu}	$(\bar l_p \gamma_\mu l_r)(\bar u_s \gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar q_p \gamma_\mu \tau^I q_r)(\bar q_s \gamma^\mu \tau^I q_t)$	Q_{dd}	$(\bar d_p \gamma_\mu d_r)(\bar d_s \gamma^\mu d_t)$	Q_{ld}	$(\bar l_p \gamma_\mu l_r)(\bar d_s \gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar l_p \gamma_\mu l_r)(\bar q_s \gamma^\mu q_t)$	Q_{eu}	$(\bar e_p \gamma_\mu e_r)(\bar u_s \gamma^\mu u_t)$	Q_{qe}	$(\bar q_p \gamma_\mu q_r)(\bar e_s \gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar l_p \gamma_\mu \tau^I l_r)(\bar q_s \gamma^\mu \tau^I q_t)$	Q_{ed}	$(\bar e_p \gamma_\mu e_r)(\bar d_s \gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar q_p \gamma_\mu q_r)(\bar u_s \gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar u_p \gamma_\mu u_r)(\bar d_s \gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar q_p \gamma_\mu T^A q_r)(\bar u_s \gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar u_p \gamma_\mu T^A u_r)(\bar d_s \gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar q_p \gamma_\mu q_r)(\bar d_s \gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar q_p \gamma_\mu T^A q_r)(\bar d_s \gamma^\mu T^A d_t)$

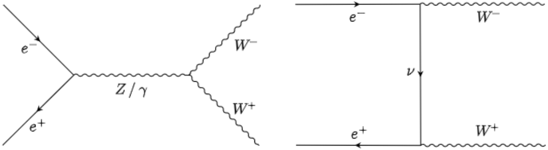
$(\bar LR)(\bar RL)$ and $(\bar LR)(\bar LR)$		B-violating	
Q_{ledq}	$(\bar l_p^j e_r)(\bar d_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(q_s^{\gamma j})^T C l_t^k\right]$
$Q_{quqd}^{(1)}$	$(\bar q_p^j u_r)\varepsilon_{jk}(\bar q_s^k d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{quqd}^{(8)}$	$(\bar q_p^j T^A u_r)\varepsilon_{jk}(\bar q_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(q_s^{\gamma m})^T C l_t^n\right]$
$Q_{lequ}^{(1)}$	$(\bar l_p^j e_r)\varepsilon_{jk}(\bar q_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{lequ}^{(3)}$	$(\bar l_p^j \sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar q_s^k \sigma^{\mu\nu} u_t)$		

- Write down all D6 operators, eliminate redundant ones via field redefinition, integration by parts, equations of motion...
- 59 operators (76 parameters) for 1 generation, or 2499 parameters for 3 generations. [arXiv:1008.4884] Grzadkowski, Iskrzyński, Misiak, Rosiek,[arXiv:1312.2014] Alonso, Jenkins, Manohar, Trott.

Why Diboson

- Diboson is an important part of the precision measurement program
- Connected to the higgs couplings in the SMEFT frame
- Can be measured very well at Higgs factories

EFT Parameterization

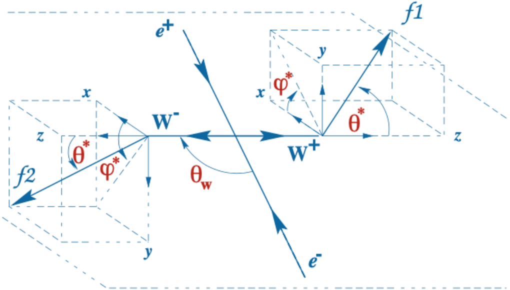


- Focusing on tree-level CP-even dimension-6 contributions
- $e^-e^+ \to WW$ can be parameterized by

$$\delta g_{1,Z},\ \delta\kappa_\gamma,\ \lambda_Z,\ \delta g^{ee}_{Z,L},\ \delta g^{ee}_{Z,R},\ \delta g^{e\nu}_{W},\ \delta m_W$$

- m_W is better constrained, so we can simply set $\delta m_W = 0$

$e^- e^+ \to WW$ with Histogram



- The TGCs are sensitive to the differential distributions
  - One could do a fit to the binned distributions of all angles.
  - Not the most efficient way of extracting information.
  - Correlations among angles are sometimes ignored.

$e^-e^+ \to WW$ with Optimal Observable

- What are Optimal Observables?

  Diehl, M., Nachtmann, O., 1994. Zeitschrift Für Physik C Part Fields 62, 397–411.

  - In the limit of large statistics (everything is Gaussian) and small parameters (linear contribution dominates), the best possible reaches can be derived analytically!

    $$\frac{d\sigma}{d\Omega} = S_0 + \Sigma_i S_{1,i} g_i, \qquad c_{ij}^{-1} = \int d\Omega \frac{S_{1,i} S_{1,j}}{S_0} \cdot \mathcal{L}$$

$e^- e^+ \to WW$ with Optimal Observable

- What are Optimal Observables?

  Diehl, M., Nachtmann, O., 1994. Zeitschrift Für Physik C Part Fields 62, 397–411.

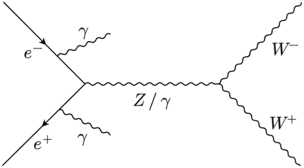
  - In the limit of large statistics (everything is Gaussian) and small parameters (linear contribution dominates), the best possible reaches can be derived analytically!

    $$\frac{d\sigma}{d\Omega} = S_0 + \Sigma_i S_{1,i} g_i, \qquad c_{ij}^{-1} = \int d\Omega \frac{S_{1,i} S_{1,j}}{S_0} \cdot \mathcal{L}$$

- The optimal observable is a function of 5 angles and is given by $\mathcal{O}_i = \frac{S_{1,i}}{S_0}$

Systematic Effects

- Initial state radiation



- Jet smearing
- Detect effects
  - final state jets can not be distinguished
  - neutrino cannot be directly measured
- They are systematic effects

Systematic Effects

- In simulation, systematic effects can't be ignored

Systematic Effects

- In simulation, systematic effects can't be ignored
- Analytical methods become more difficult and time consuming when we include more realistic effects.
- Naively applying optimal observables could lead to a bias

Systematic Effects

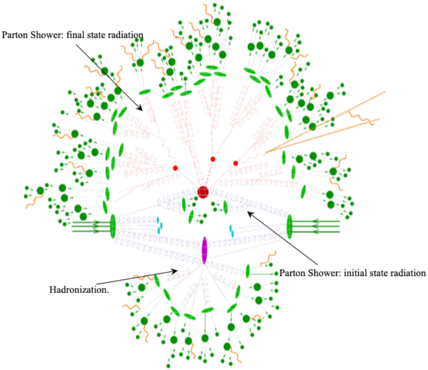
- In simulation, systematic effects can't be ignored
- Analytical methods become more difficult and time consuming when we include more realistic effects.
- Naively applying optimal observables could lead to a bias
- New method in need

Likelihood Inference

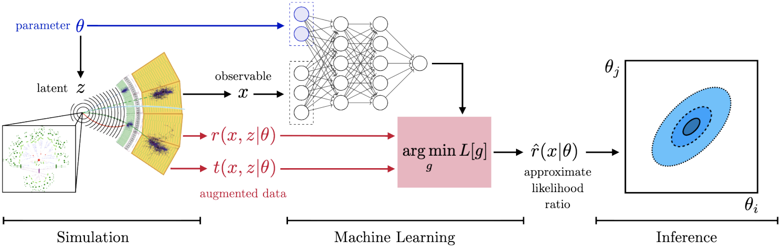
- Neyman-Pearson lemma says the best statistics to test new physics is the likelihood ratio given data x and theory parameters θ_1 and θ_0

$$r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

- The key thing is $r(x|\theta_0, \theta_1)$
- Analytical methods always computational consuming and ignore systematic effects



Likelihood Inference



- Johann Brehmer, etc develop new simulation-based inference techniques that are tailored to the structure of particle physics processes.[arXiv:1805.00013]Brehmer, J, Cranmer, K, Louppe, G, Pavez, J
- Machine Learning method can extract more information from x to predict the likelihood ratio

Particle-Physics Structure

- The likelihood function can be written as

$$p(x|\theta) = \int dz \; p(x, z|\theta) = \int dz \; p(x|z)p(z|\theta)$$

- Here $p(z|\theta) = 1/\sigma(\theta) d\sigma/dz$ is the parton level density distribution.
- $p(x|z)$ describes the probabilistic evolution from the parton-level four-momenta to observable particle properties

$$p(x|z) = \int dz_d \int dz_s \int dz \; p(x|z_d)p(z_d|z_s)p(z_s|z)$$

Particle Structure

- We can extract more information from the simulator by defining joint likelihood ratio and joint score

$$r(x, z|\theta_0, \theta_1) = \frac{p(x|z)p(z|\theta_0)}{p(x|z)p(z|\theta_1)} = \frac{p(z|\theta_0)}{p(z|\theta_1)}$$

$$\alpha(x, z|\theta_0, \theta_1) = \nabla_{\theta_0} r(x, z|\theta_0, \theta_1)|_{\theta_0=\theta_1}$$

- The loss function is

$$\mathcal{L}[\hat{g}(x)] = \int dxdz\ p(x, z|\theta)|g(x, z) - \hat{g}(x)|^2$$

- The loss function is minimized when $g(x, z) = \hat{g}(x)$

ML Algorithm: ALICE

- Approximate likelihood with improved crossentropy estimator
- Directly predict the likelihood ratio
- Loss function $\mathcal{L}$ is

$$\mathcal{L}(\hat{s}) \propto \sum_{x}[s(x,z|\theta_0,\theta_1)\log(\hat{s}(x)) + (1 - s(x,z|\theta_0,\theta_1))\log(1 - \hat{s}(x))]$$

- Here $s(x,z|\theta_0,\theta_1) = \frac{1}{1+r(x,z|\theta_0,\theta_1)}$
- $\hat{r}(x|\theta_0,\theta_1)$ can be reconstructed by $\hat{s}(x) = \frac{1}{1+\hat{r}(x|\theta_0,\theta_1)}$

ML Algorithm: SALLY

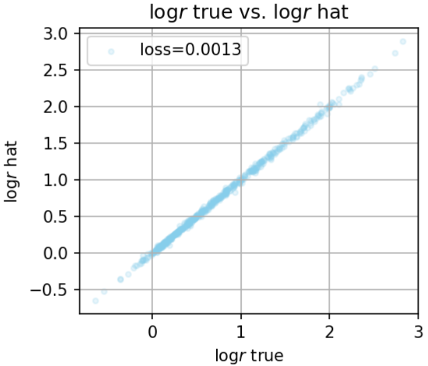
- Score approximates likelihood locally
- likelihood ratio can also be parameterized by Wilson coefficients

$$\hat{r}(x, \theta) = 1 + \sum_i \hat{\alpha}_i(x)\theta_i$$

- And we can predict α_i term as well
- Loss function $\mathcal{L}$ is

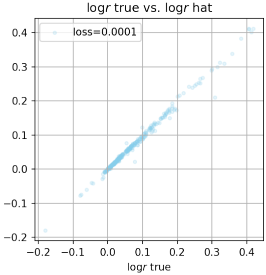
$$\mathcal{L} \propto \sum_i |\hat{\alpha}_i(x) - \alpha_i(x, z|\theta_0, \theta_1)|^2$$

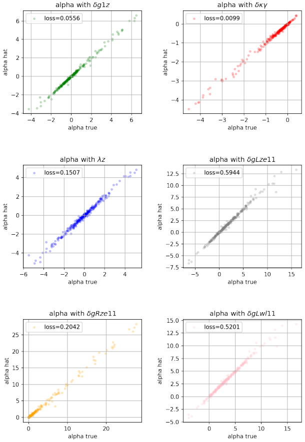
Prediction of Likelihood Ratio:ALICE



- ALICE method offers a precise way to predict the likelihood ratio directly.

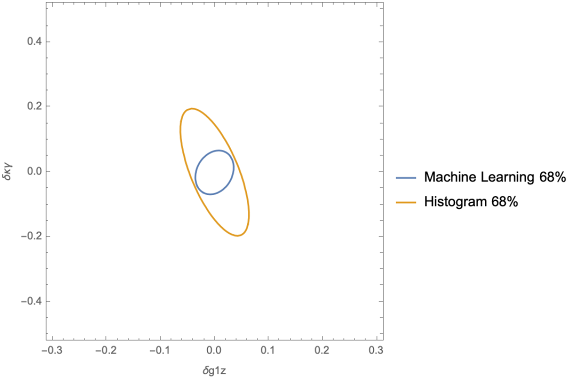
Prediction of Likelihood Ratio:SALLY





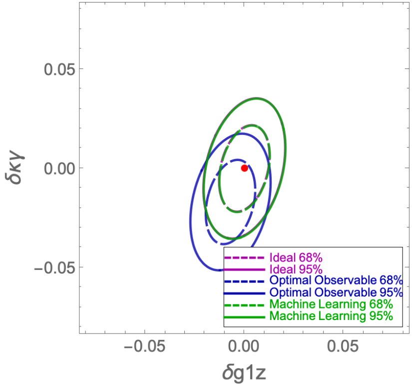
- We can construct the $\hat{r}(x, \theta)$ by predicting the alpha term and give an analytical expression of $\hat{r}(x, \theta)$

Estimation of the Boundary:Compared with Histogram



- no bias
- precise bounds along individual directions
- weak constraints in other directions

Estimation of the Boundary:Compared with OO



- Once you get the $\hat{r}(x|\theta)$, $\chi^2 = -2\sum_i \log(\hat{r}(x_i|\theta))$
- The χ^2 analysis shows that ML method can correct the large bias and give a strong constrain on the model parameters.

Conclusion

- Future colliders will generate large amount of data, ML will benefit it a lot
- By machine learning, we can construct 6D likelihood ratio to improve the global fit result
- Machine Learning can easily take care of systematic effects as long as the MC simulation is accurate.
- Machine learning is (likely to be) the future

Thanks!

Backup Slides: $e^- e^+ \to WW$ parameterization

$$
\begin{aligned}
\mathcal{L}_{TGC} = & \, igs_{\theta_w} A^{\mu}(W^{-\nu} W^{+}_{\mu\nu} - W^{+\nu}_{\mu\nu}) \\
& + ig(1+\delta g_1^Z) c_{\theta_w} Z^{\mu}(W^{-\nu} W^{+}_{\mu\nu} - W^{+\nu}_{\mu\nu}) \\
& + ig[(1+\delta\kappa_Z) c_{\theta_w} Z^{\mu\nu} + (1+\delta\kappa_\gamma) s_{\theta_w} A^{\mu\nu}] W^-_\mu W^+_\nu \\
& + \frac{ig}{m_W^2}(\lambda_Z c_{\theta_w} Z^{\mu\nu} + \lambda_\gamma s_{\theta_w} A^{\mu\nu} W^{-\rho}_{\nu} W^{+}_{\rho\mu})
\end{aligned}
$$

- Imposing Gauge invariance one obtains $\delta\kappa_Z = \delta g_{1,Z} - t^2_{\theta_w}\delta\kappa_\gamma$ and $\lambda_Z = \lambda_\gamma$
- $\delta g_{1,Z}, \ \delta\kappa_\gamma, \ \lambda_Z, \ \delta g^{ee}_{Z,L}, \ \delta g^{ee}_{Z,R}, \ \delta g^{e\nu}_{W}, \ \delta m_W$