Equivariant Neural Networks for Robust CP Observables

2405.13524

ICNFP2024, Kolumbari, Crete

Dr. Pietro Vischia

work with Sergio Sánchez Cruz, Marina Kolosova, Clara Ramón Álvarez, and Giovanni Petrucciani

pietro.vischia@cern.ch

@pietrovischia

https://vischia.github.io/

Universidad de Oviedo

ICTEA

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

AGENCIA ESTATAL DE INVESTIGACIÓN

If you are reading this as a web page: have fun! If you are reading this as a PDF: please visit

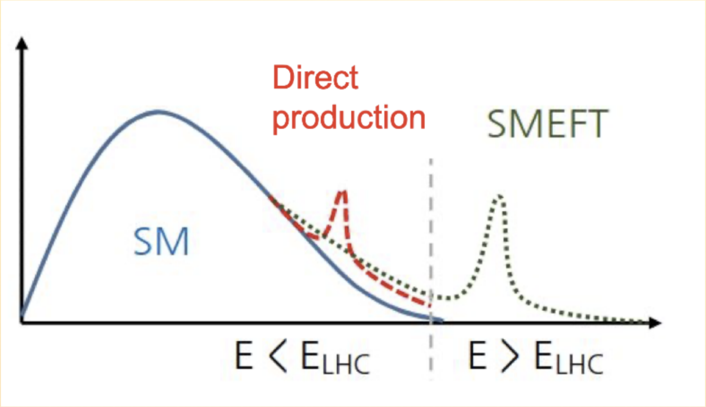
https://www.hep.uniovi.es/vischia/persistent/2024-09-04_EquivariantForCPAtICNFP2024_vischia.html

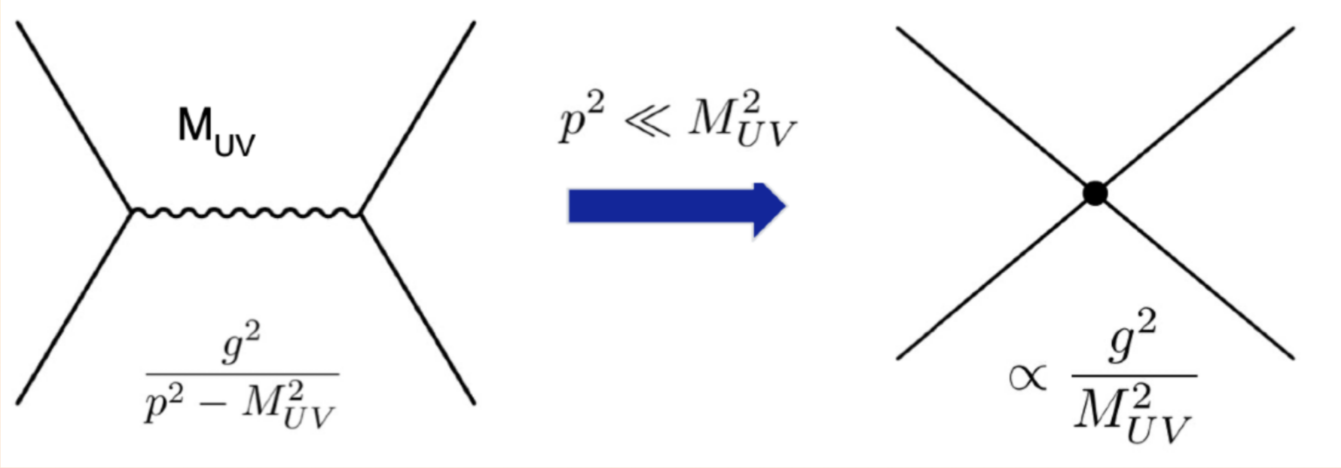
to get the version with working animations

SMEFT and CP Violation

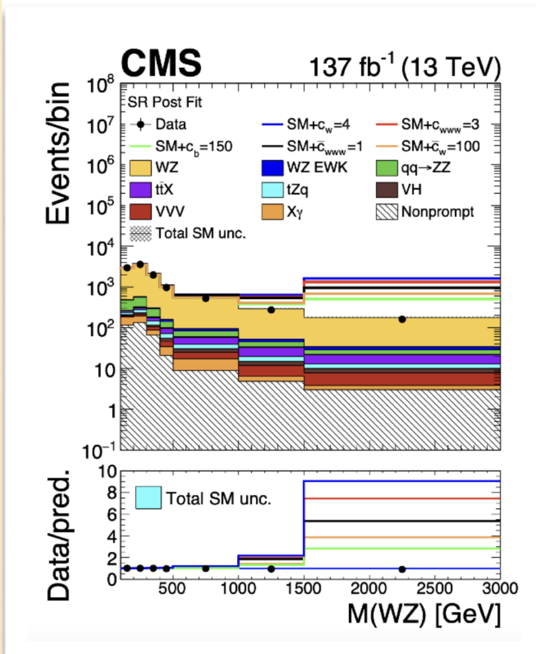
- SMEFT: standard model extended by postulating high-mass BSM particles
- 1350 CP-even operators, 1149 CP-odd operators

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{C_i}{\Lambda^2} \mathcal{O}_i^{(6)}$$

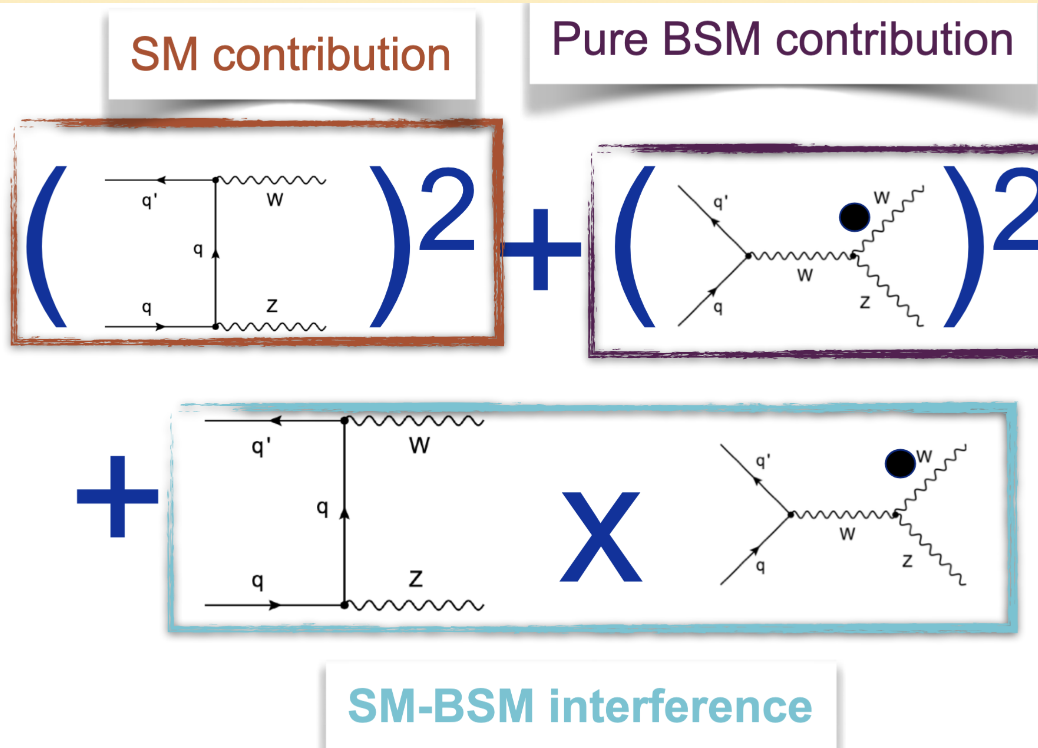




EFT Observables



JHEP 07 (2022) 032



CP-violating operators

- SM contribution: mostly CP-invariant
- Pure BSM contribution: CP-invariant e.g. in top/Higgs sectors
- SM-BSM interference: odd under CP tranformations
- Sensitivity to the interference given only by CP-odd observables. LHC cross section program insensitive.
- CP-odd observables are robust against signal mismodelling/background

Our Algorithm

- Build observables that are *equivariant with respect to CP symmetry
  - CP-invariant observables
  - discriminate between different SM backgrounds
  - discriminate between SM and quadratic terms or CP-even contribution
  - CP-odd observables
  - discriminate between signal-like and interference-like contributions
  - discriminate between interference-like and other SM backgrounds
- We fix $n_1 = 0$ and $n_2 = 1$, obtaining a single CP-odd observable
  - Can generalize to n_1 CP-invariant and n_2 CP-odd components

Our algorithm (reprise)

- A function $f : D \rightarrow R$ is odd under CP transformations if $f(CP(event)) = -f(event)$
  - Most general function satisfying this is $f(event) = g(event) - g(CP(event))$
  - We parameterize g using a neural network, training f to minimize a loss function
- Parameterizations of g: can be any parametric function, you don't strictly need a neural network
- Space of input features is fully general
  - Kinematics of set of particles, low- or high-level variables, particle set, graph network
  - Can also add features for background discrimination

Gutting the algo: the cost function

- Inductive bias (see the Machine Learning course!!) by learning the likelihood ratio
  - Method inspired by the SALLY procedure (Brehmer et al.)
  - Other loss functions can encode different properties (see recent example)
- Weighted simulations: $w(z) = w_{SM}(z) + c w_{int}(z) + c^2 w_{quad}(z)$
  - Weights are functions of parton level kinematics
- Intractable likelihood ratio:

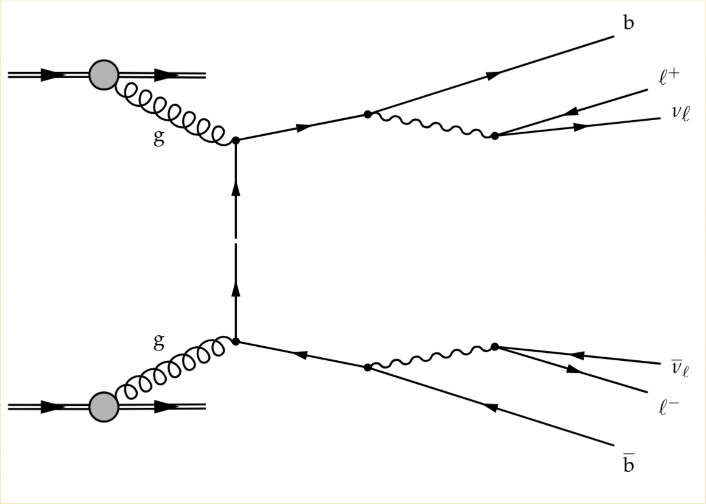
$$\frac{p(d, z|c_1}{p(d, z|c = 0)} = \frac{w_{SM} + c w_{int} + c^2 w_{quad}}{w_{SM}}$$

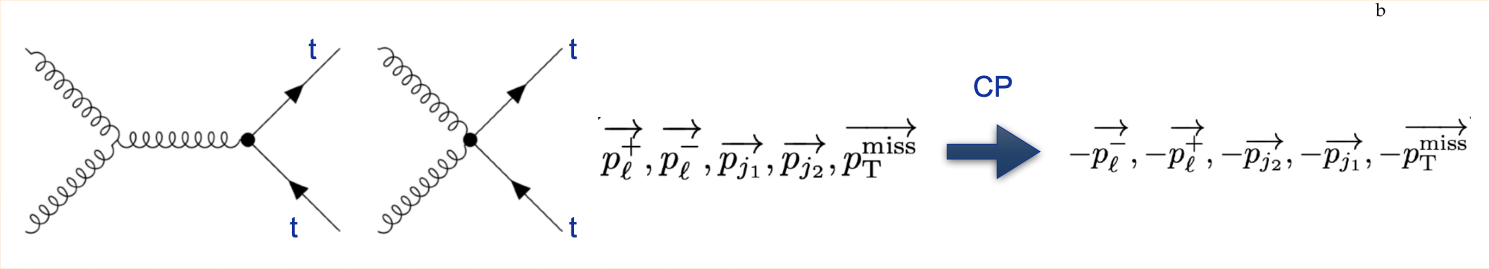
- The likelihood score at the SM point will be a sufficient statistic for small values of c
  - In the small-c regime, the linear component, describing the interference, is dominant
- Learn a surrogate model of the score

$$Loss = w_{SM}|f(d) - \frac{w_{int}(z)}{w_{SM}(d)}|^2$$

Use case: ttbar production

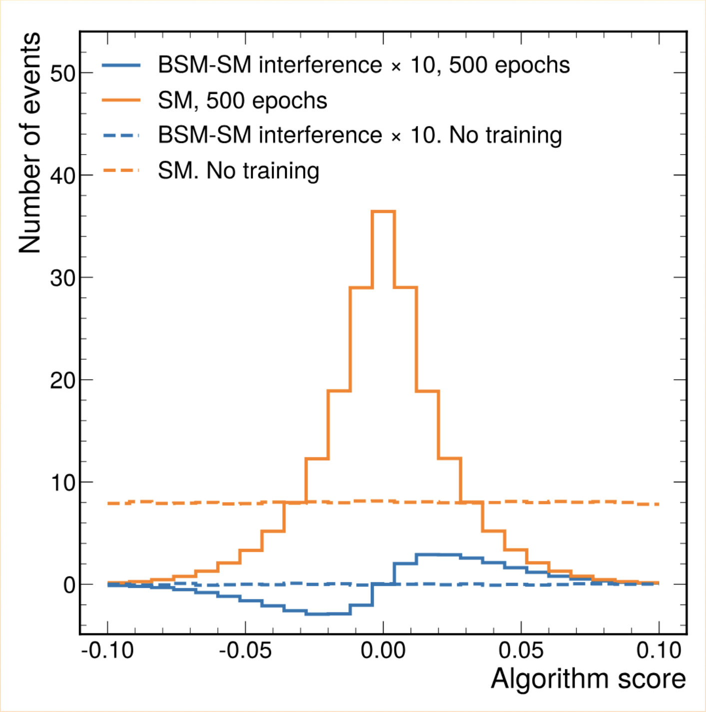
- Dileptonic final state
  - Semileptonic difficult, need to estimate jet charge (BSc thesis of Santiago Vila Domínguez)
- CP-violating chromoelectric dipole moment operator $g_s \frac{v}{\sqrt{2}}(\bar{t}\sigma^{\mu\nu}\gamma_5 T^A t)G^A_{\mu\nu}$





Use case: ttbar production

- The score after the training is CP-odd!
  - Symmetric for SM
  - Any SM-like mismodelling/background will be symmetric by construction!
  - Constructive/destructive interference pattern for positive/negative values
- Equivariance respected at all stages of training

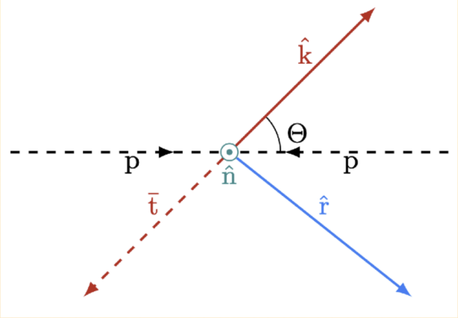


- The observable is robust even before training convergence

Use case: ttbar production

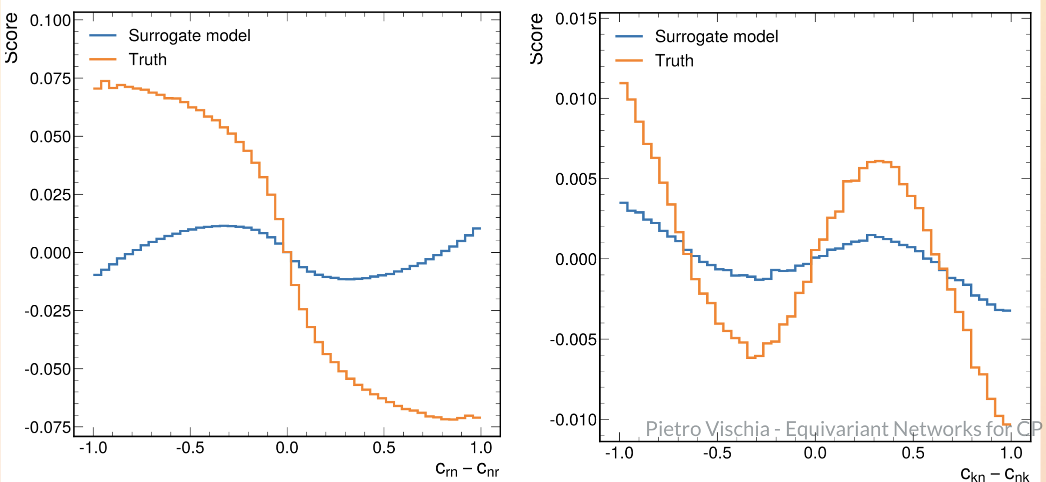
- Reweight events by the score, compare with parton-level CP-odd observables

- Reconstruct the ttbar system based on angles



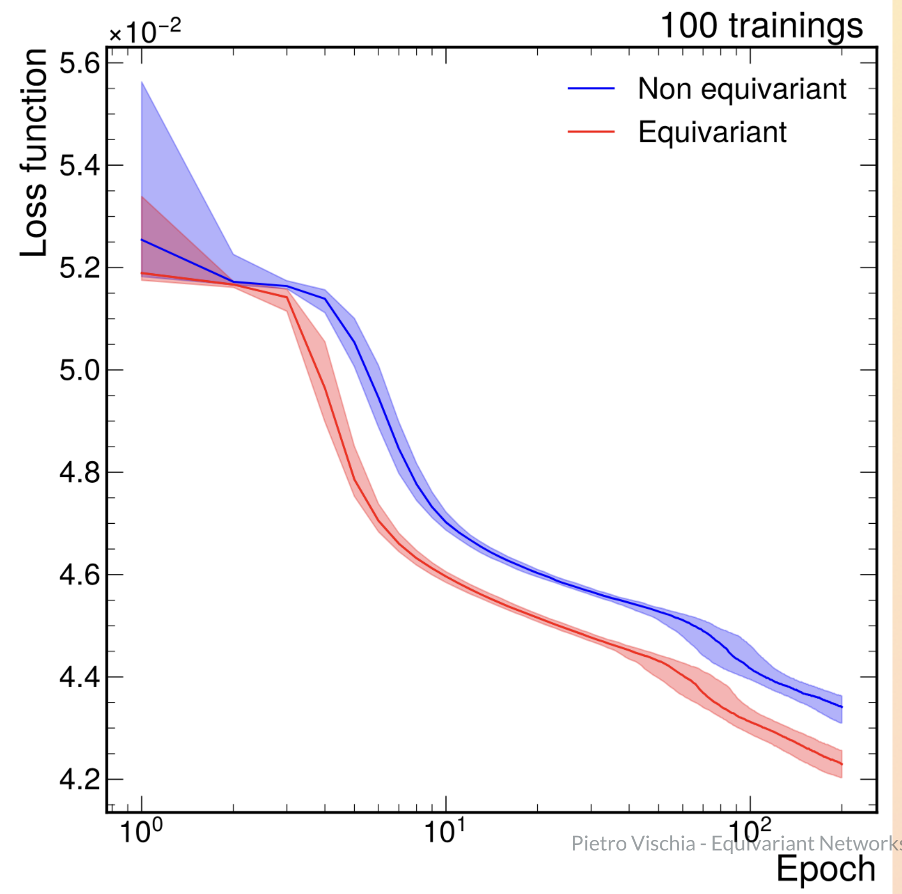
\begin{aligned}c{rn}-c{nr}&=cos(l{r}^{+})cos(l{r}^{-}-cos(l{n}^{+}))cos(l{r}^{-})\c{kn}-c{nk}&=cos(l{k}^{+})cos(l{n}^{-}-cos(l{n}^{+}))cos(l{k}^{-})\end{aligned}

- Limitation is the reconstruction of the ttbar system



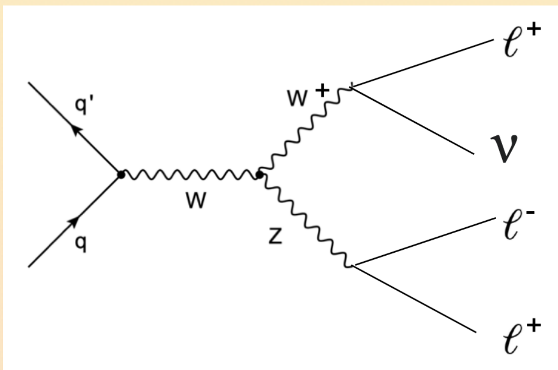
Use case: ttbar production

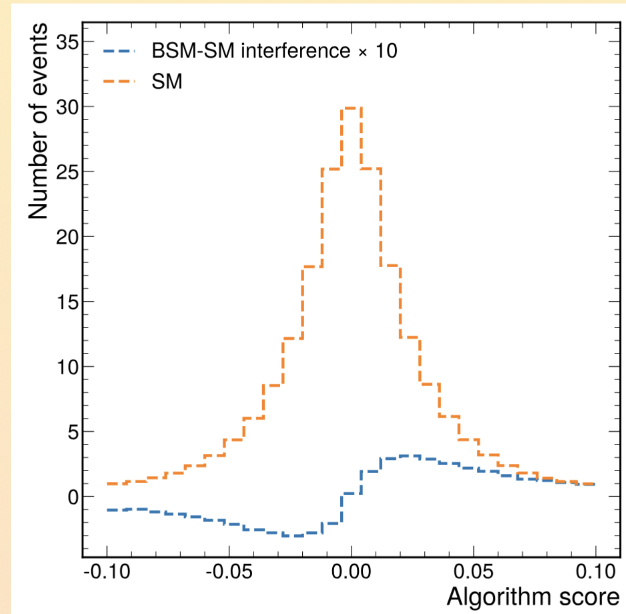
- Equivariance as inductive bias speeds up convergence
- Between **40% and 300% less iterations** needed to achieve the same loss value!!!



Use case: WZ production

- Trilepton final state
- CP-odd operator: $c_{\tilde{W}}$





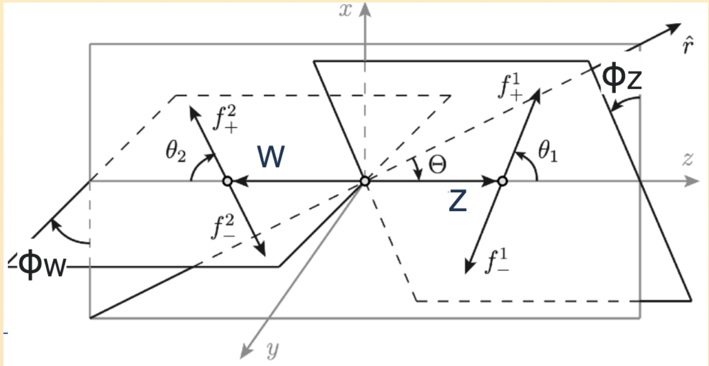
$$\overrightarrow{p^{\mathrm{Z}}_{\ell+}}, \overrightarrow{p^{\mathrm{Z}}_{\ell-}}, \overrightarrow{p^{\mathrm{W}}_{\ell}}, Q^{\mathrm{W}}, \overrightarrow{p^{\mathrm{miss}}_{\mathrm{T}}}$$

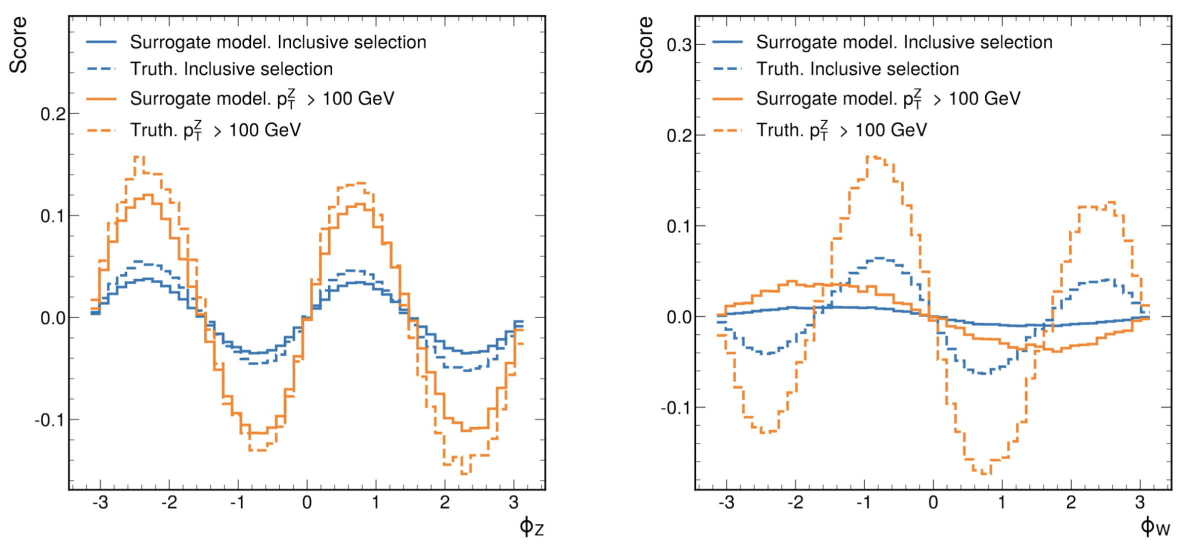
CP

$$-\overrightarrow{p^{\mathrm{Z}}_{\ell-}}, -\overrightarrow{p^{\mathrm{Z}}_{\ell+}}, -\overrightarrow{p^{\mathrm{W}}_{\ell}}, -Q^{\mathrm{W}}, -\overrightarrow{p^{\mathrm{miss}}_{\mathrm{T}}}$$

Use case: WZ production

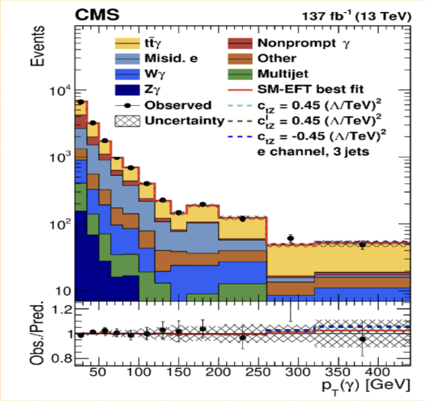
- Performance on parton-level observables **even better than dedicated observables!!!**
  - Can capture energy growth
  - Insensitivity to ϕ_W due to ambiguity in W decay reconstruction

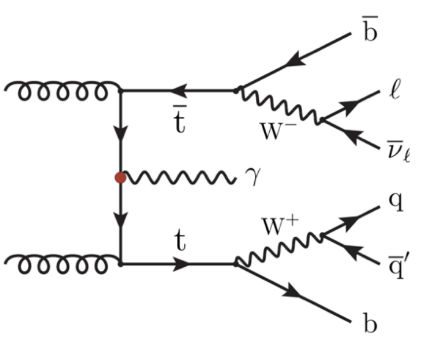




Use case: ttgamma production

- Single lepton channel, CP-odd operator c_{tZ^l}
- Literature mostly checks photon p_T, which is CP-even





$$\overrightarrow{p_\gamma}, \overrightarrow{p_\ell}, Q_\ell, \overrightarrow{p_{b_1}}, \overrightarrow{p_{b_2}}, \overrightarrow{p_{j_1}}, \overrightarrow{p_{j_2}}$$

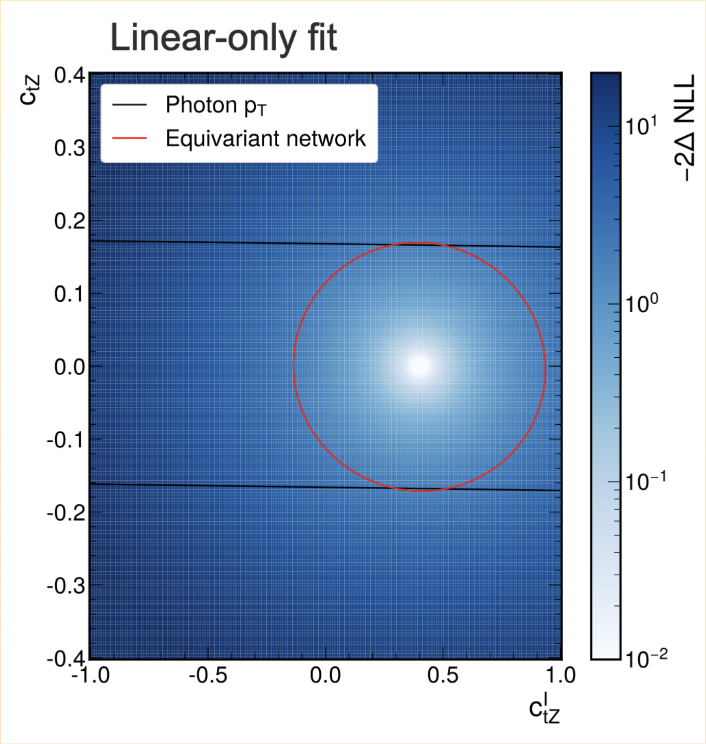
$$\Big\downarrow \text{ CP}$$

$$-\overrightarrow{p_\gamma}, -\overrightarrow{p_\ell}, -Q_\ell, -\overrightarrow{p_{b_2}}, -\overrightarrow{p_{b_1}}, -\overrightarrow{p_{j_2}}, -\overrightarrow{p_{j_1}}$$

Use case: ttgamma production

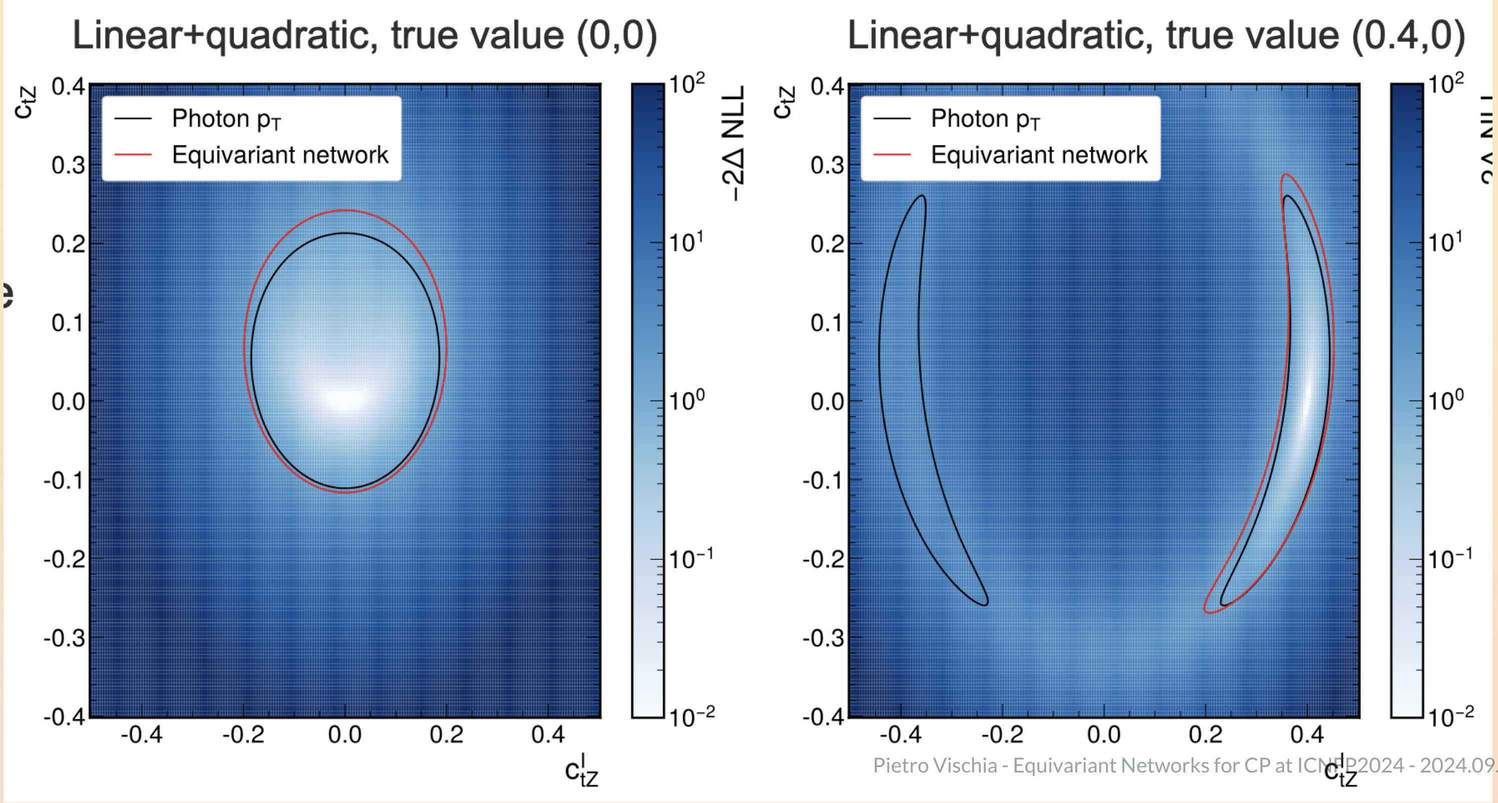
- Linear contribution constrainable only by our approach

- c_{tZ^I} (CP-odd): Comparison with photon p_T is damning (for the photon p_T, which is CP even)
- c_{tZ} (CP-even): similar sensitivity



Use case: ttgamma production

- Assuming the SM: same sensitivity
  - Our approach retains performance in CP-even observables!
- BSM cases: our approach disentangles the sign of c_{tZ^l} !!!
  - Equivariant training is superior, even if not trained for quadratic components!



Conclusions

- Implemented equivariant networks to obtain robust observables for CP violation
- Inductive bias encoded in the network structure
  - Robust regardless of convergence status
  - Training is faster than regular network
- Benchmarks: ttbar, WZ, ttgamma
  - Our approach is better than existing state-of-the-art observables
- Extensions under exploration
  - Maybe CP-invariant networks (to target CP-even observables)
- Already being employed for upcoming CMS analyses

Thank you!

—
Photo by me!