Digital twins, inductive bias, and symmetries

The case for obtaining robust observables by enforcing symmetries in neural networks

Phys. Rev. D 110, 096023 (2024)
NPTwins2024, Genova, Italia

Dr. Pietro Vischia
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/2024-12-17_InductiveBiasAndEquivariantNetworksAtNPTwinsGenova_vischia.html

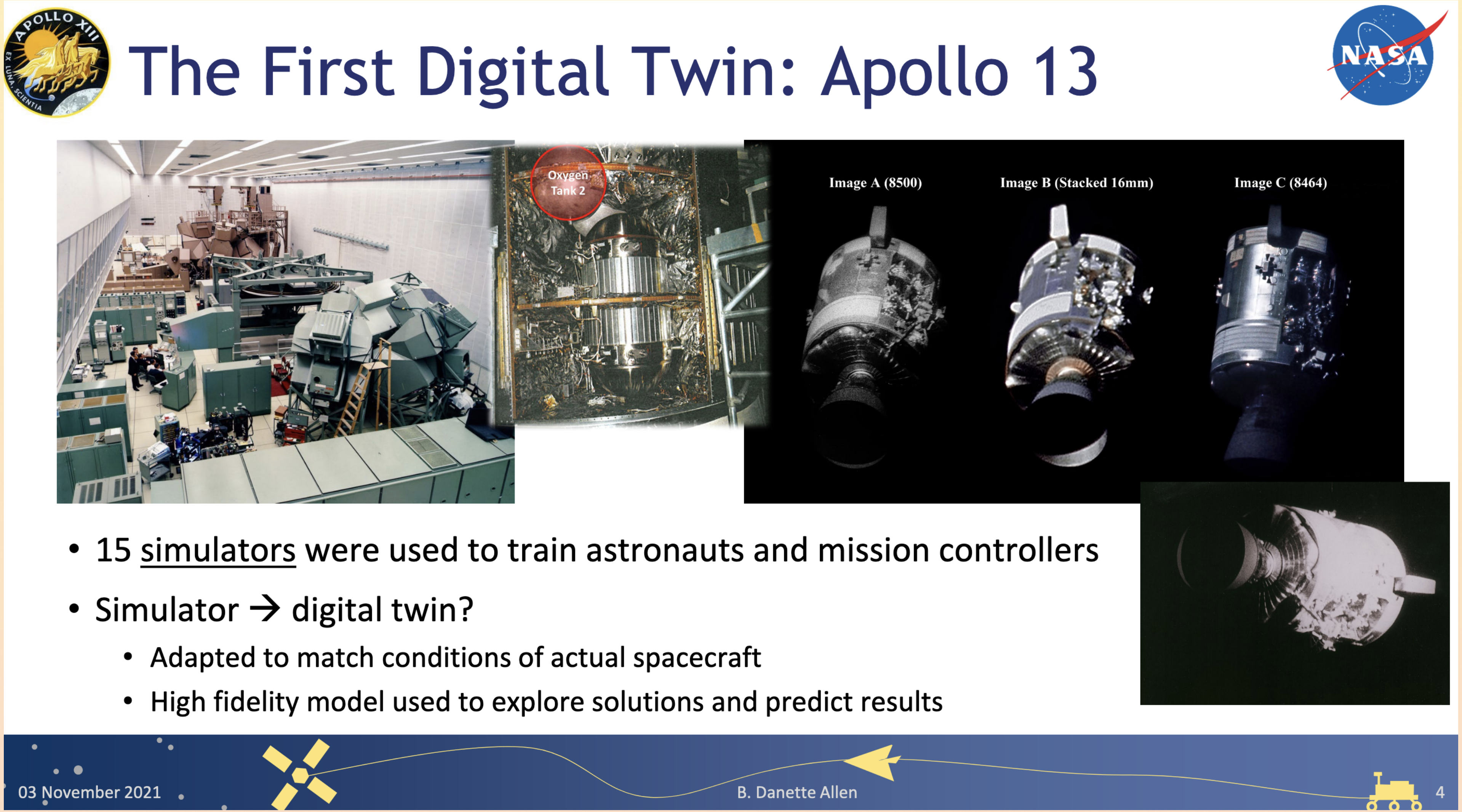
to get the version with working animations

Digital Twins, today's buzzword

A digital twin is a virtual representation of an object or system designed to reflect a physical object accurately. It spans the object's lifecycle, is updated from real-time data and uses simulation, machine learning and reasoning to help make decisions.

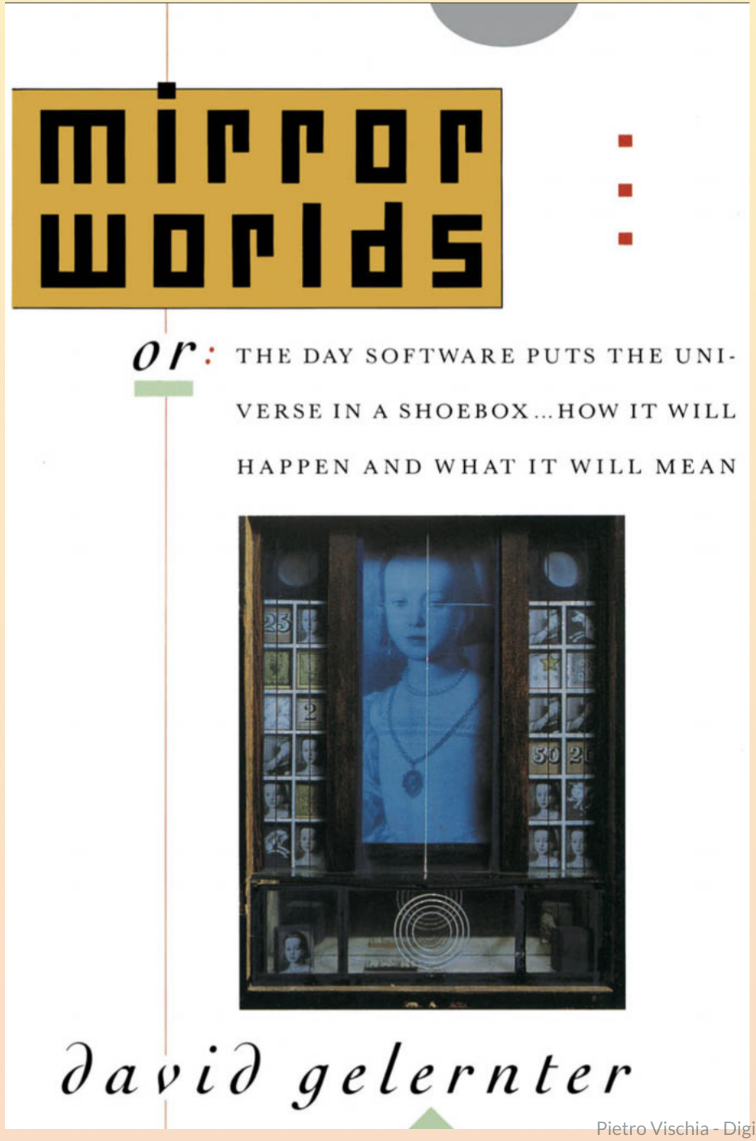
—
From IBM

1960s: Digital Twins, a rather old buzzword



From NASA

1991: digital worlds

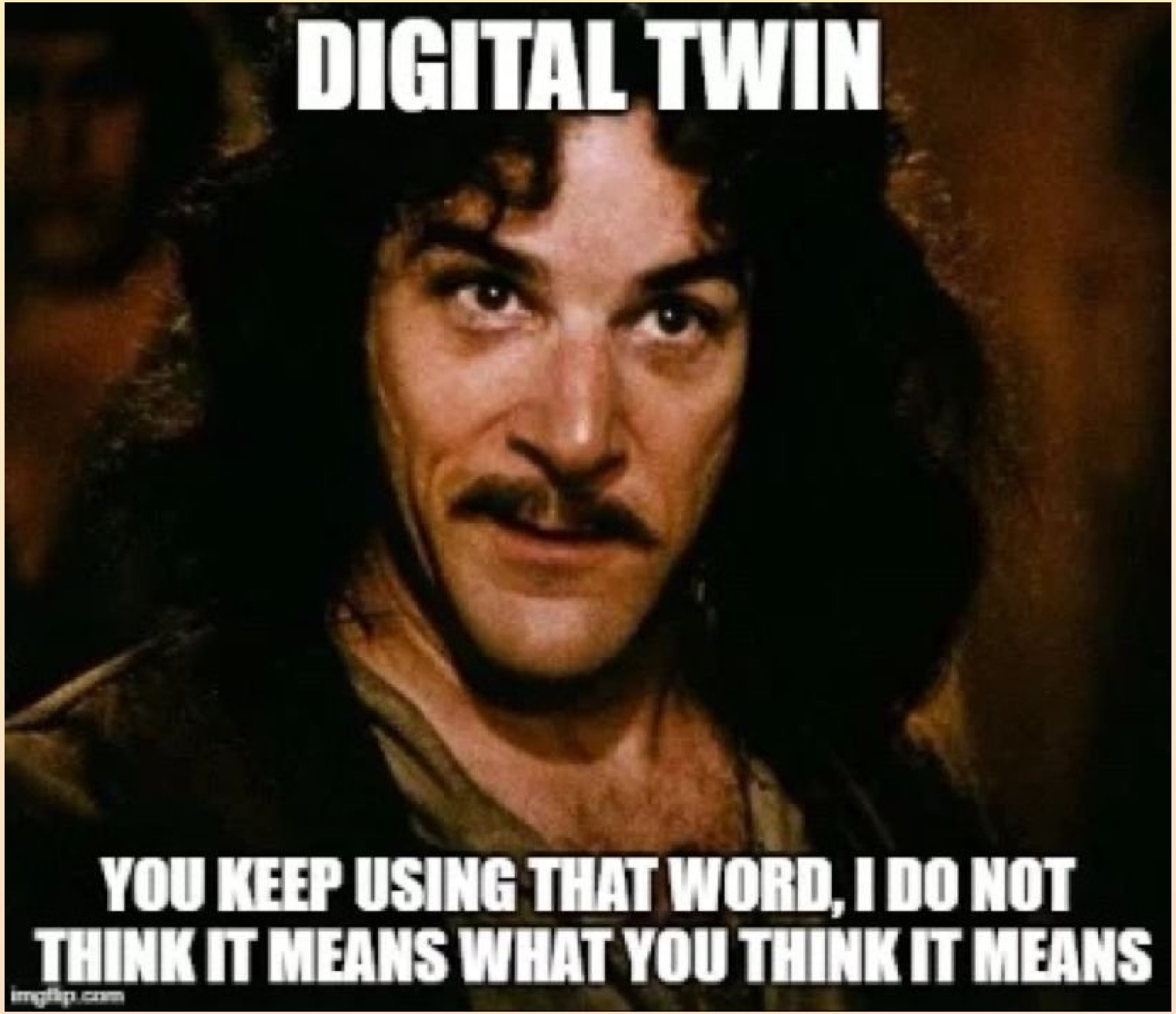


—
Curiously, the author was almost killed by the Unabomber in 1993 with an explosive package

1993: "Digital twin"

Figure 4. Single frame along animation path

In HEP, people use it as a quasi-synonim of "simulation"



—
Image from hardyards

...except maybe the virtuous collider community

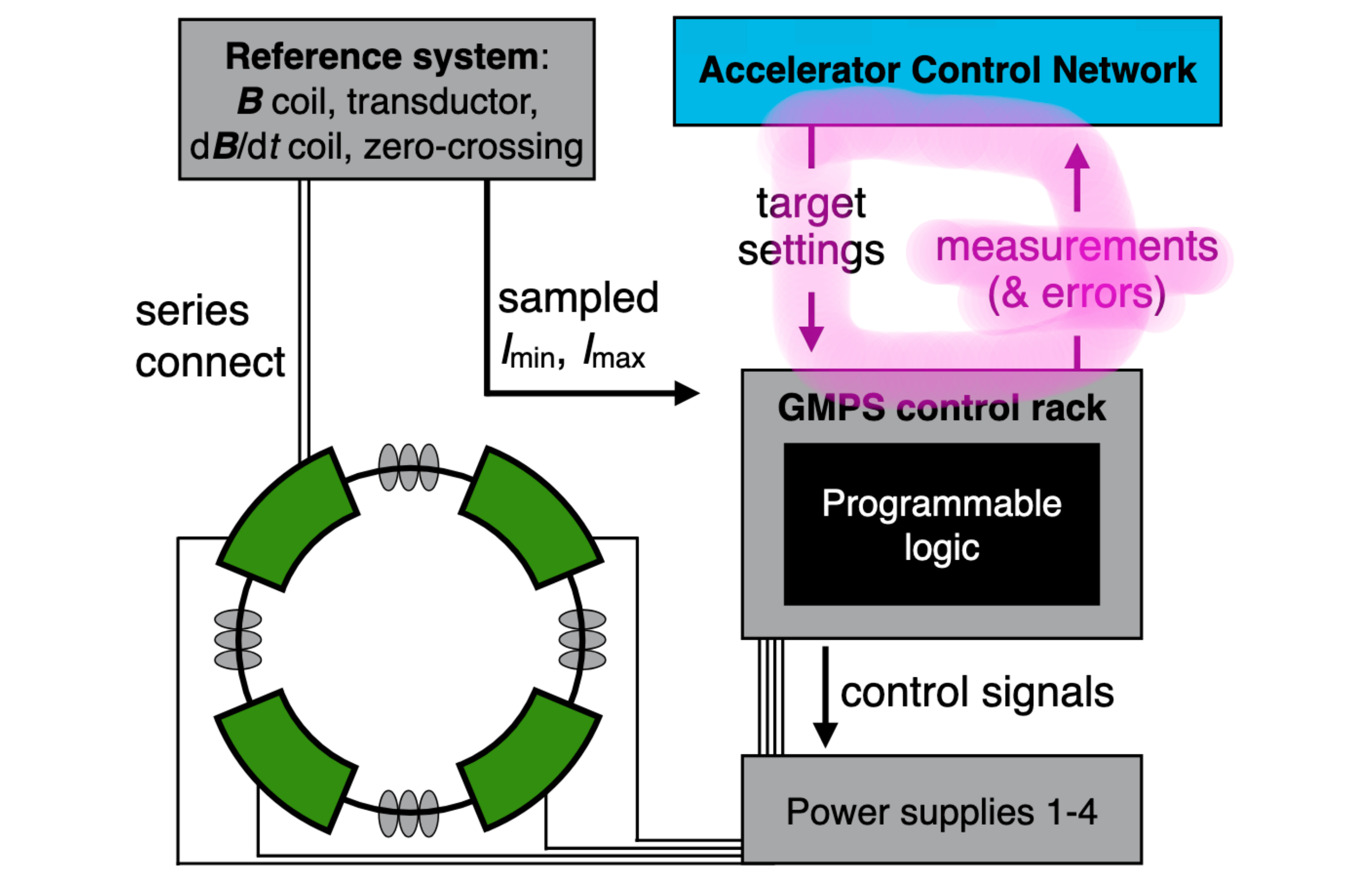


Figure 1: Overview of current PID-GMPS control system.

—
Image from 2105.12847

A Digital twin is not just a simulation!

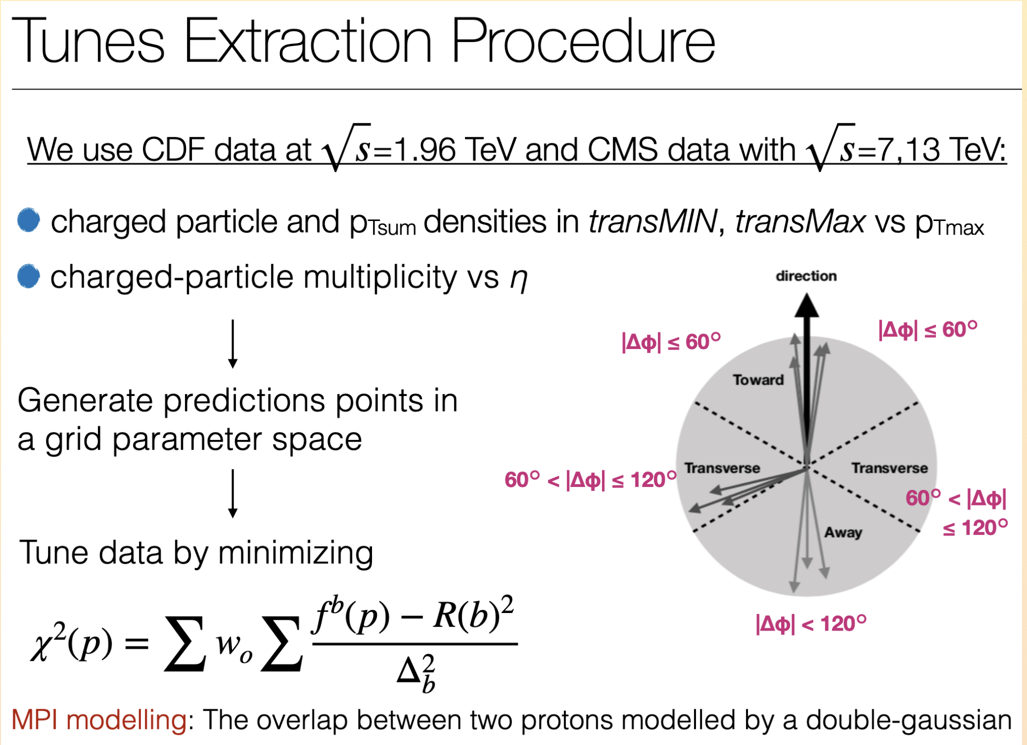
A digital twin is a virtual representation of an object or system designed to reflect a physical object accurately. It spans the object's lifecycle, is updated from real-time data and uses simulation, machine learning and reasoning to help make decisions.

—
From IBM

To be honest, we kind of always did Digital Twins

- Although kind of the "poor man's" version of it
  - Tune once every data taking era or even less
  - Monte Carlo for a given data taking era usually tuned on the previous data taking era

No real-time!!!



—
Image from Vieri Candelise's talk on tunes

Digital twins for HEP today and tomorrow

- Accelerator control: pretty natural use of digital twins (see e.g. 2105.12847)

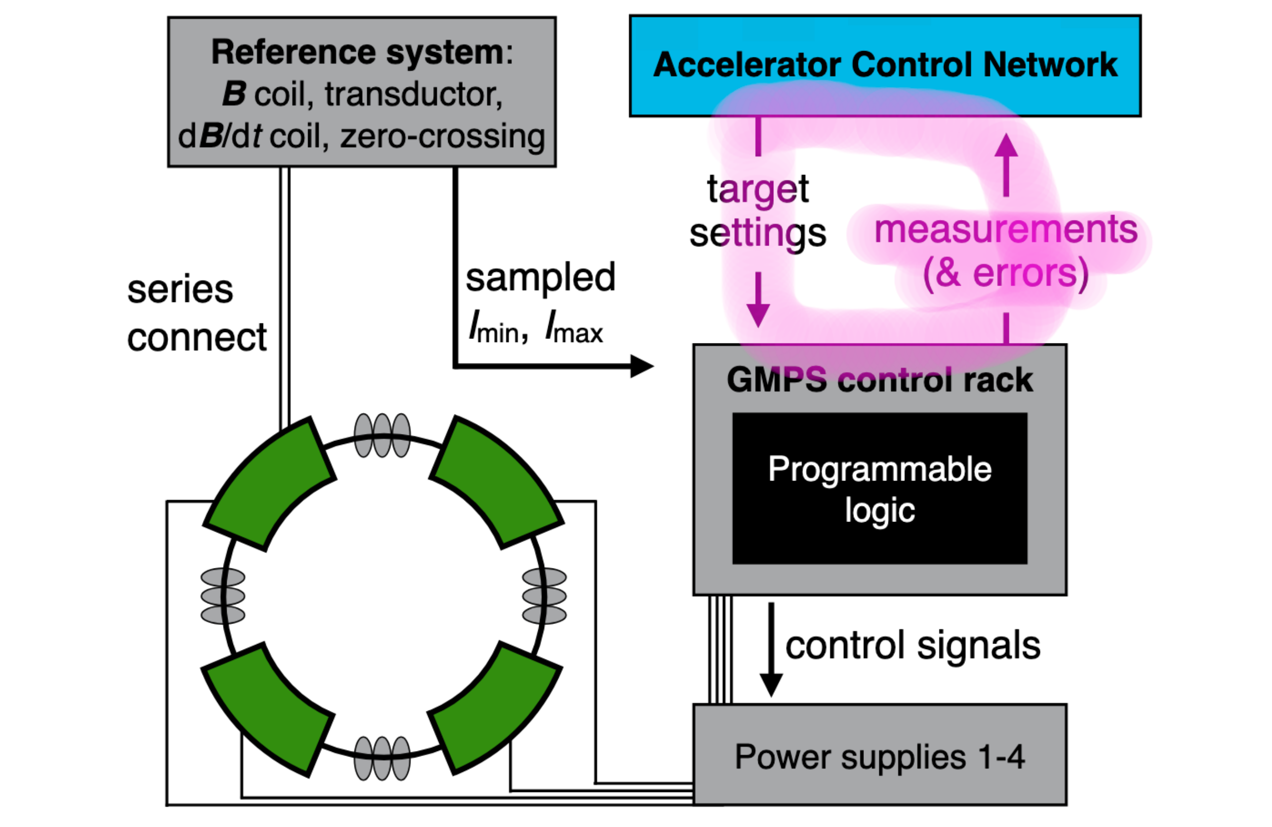


Figure 1: Overview of current PID-GMPS control system.

Digital twins for HEP today and tomorrow

- Fast simulation, even when end-to-end or surrogates (e.g. FlashSim) are not digital twins
  - They lack the realtime retuning feedback component

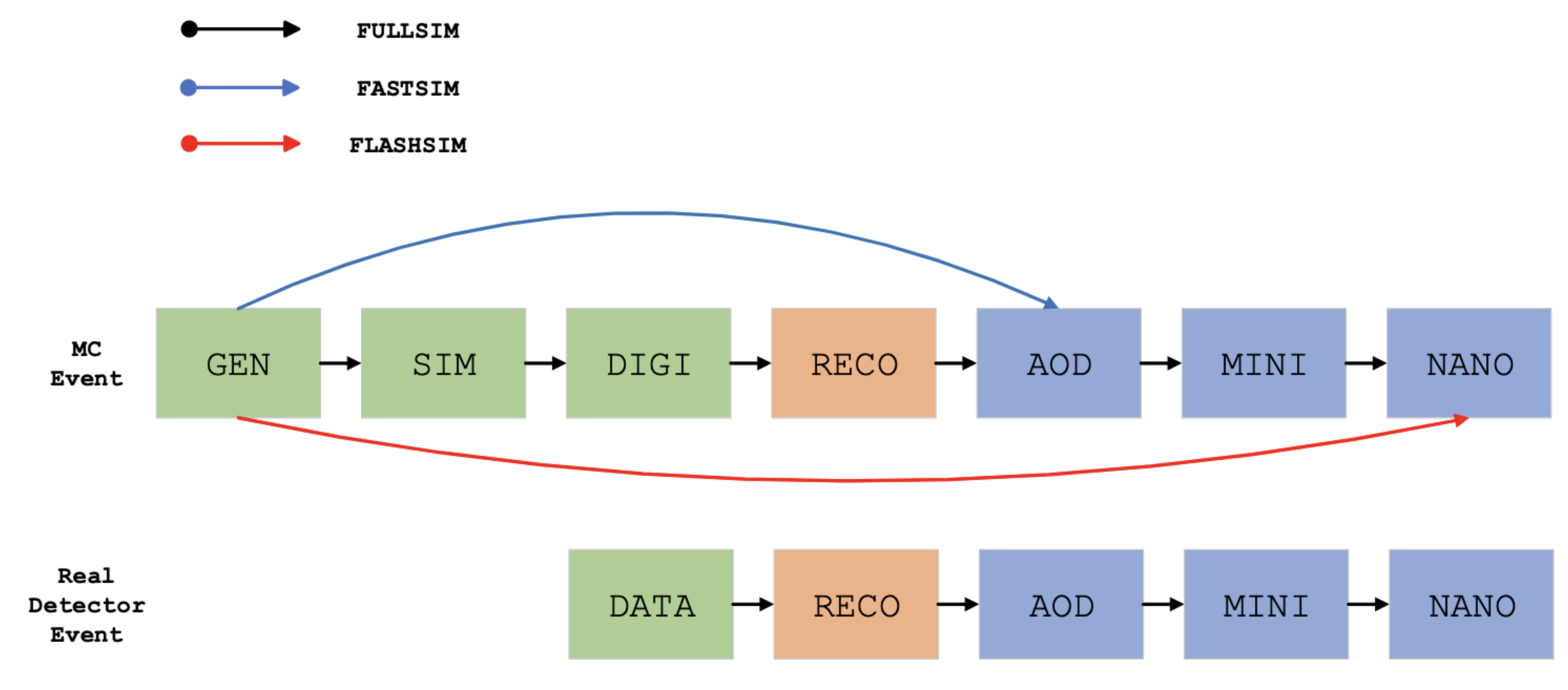


Figure 1: A comparison between the steps of the various simulation frameworks in CMS.

Digital twins for HEP: what is needed?

- ✔ *"A digital twin is a virtual representation of an object or system designed to reflect a physical object accurately."*
- ✔ *"It spans the object's lifecycle,"*
- ✖ *"is updated from real-time data"*
- ✔ *"and uses simulation, machine learning and reasoning"*
- ✖ *"to help make decisions."*

Real-time update

- *"Updating"* probably can be rephrased as building a full pipeline where the parameters of the simulation get updated
  - Via gradient descent, or unsupervised methods (Bayesian optimization, reinforcement learning, etc.)
- Akin to the experiment design of MODE (see next talk by Tommaso)
- The more information you plug into the AI, the less it has to figure it out itself
  - E.g. encode physical symmetries into the AI: inductive bias

MODE

Machine-learning Optimized Design of Experiments

Towards real-time Monte Carlo tunes?

- Can we parameterize our simulation in such a way that the whole simulation gets updated real-time while new data are taken?
- Typically the answer is "no" if we think of the usual stochastic Monte Carlos
- A generative surrogate may have the accuracy and speed needed to be subject to such an update
    - Sample from surrogate
    - Take data
    - Update surrogate parameters based on the agreement between data and surrogate
    - Sample from surrogate, rinse and repeat

Bias: a blessing and a curse

- Inductive bias: a blessing

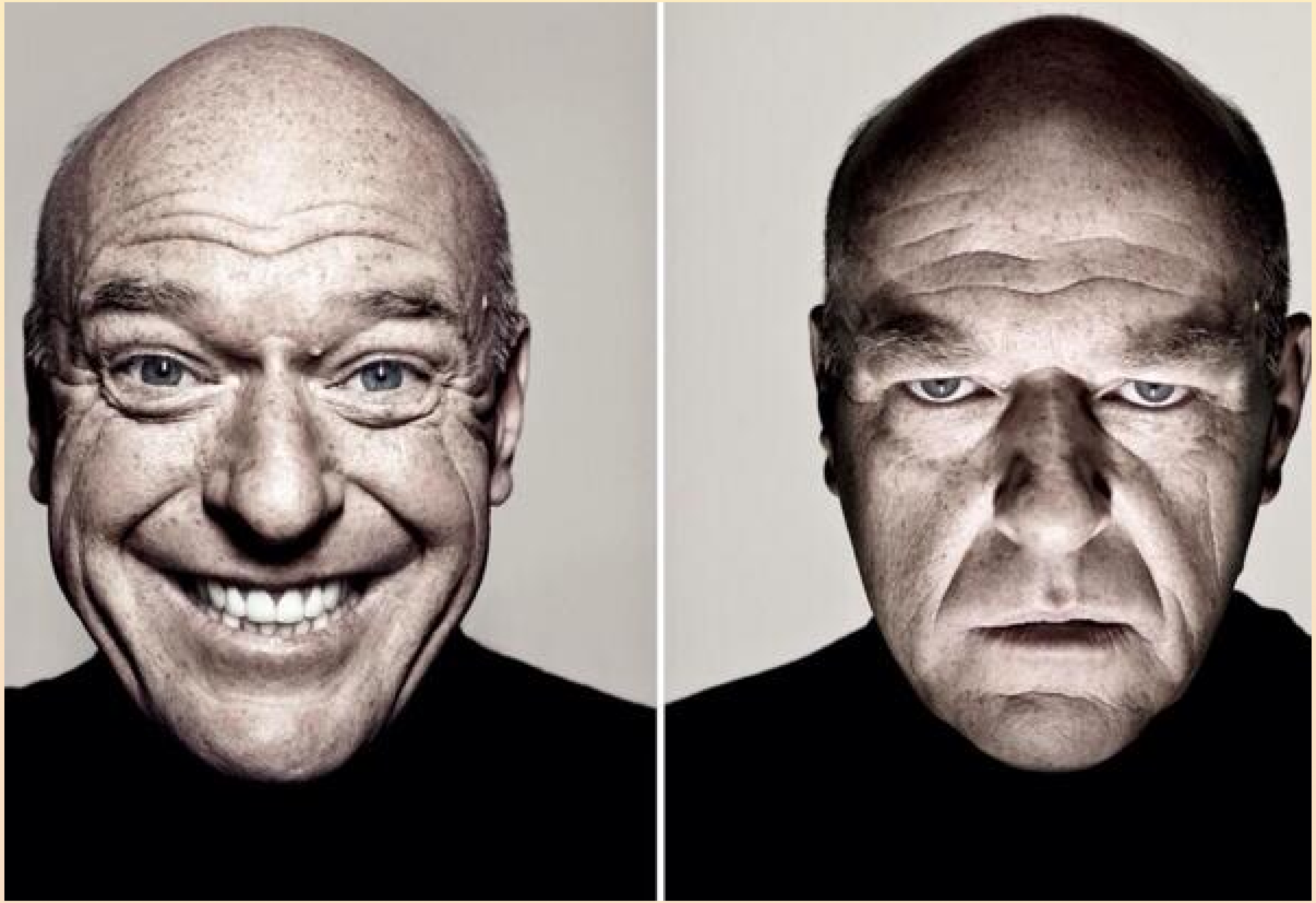


Image by Dean Norris himself

Symmetries in AI models

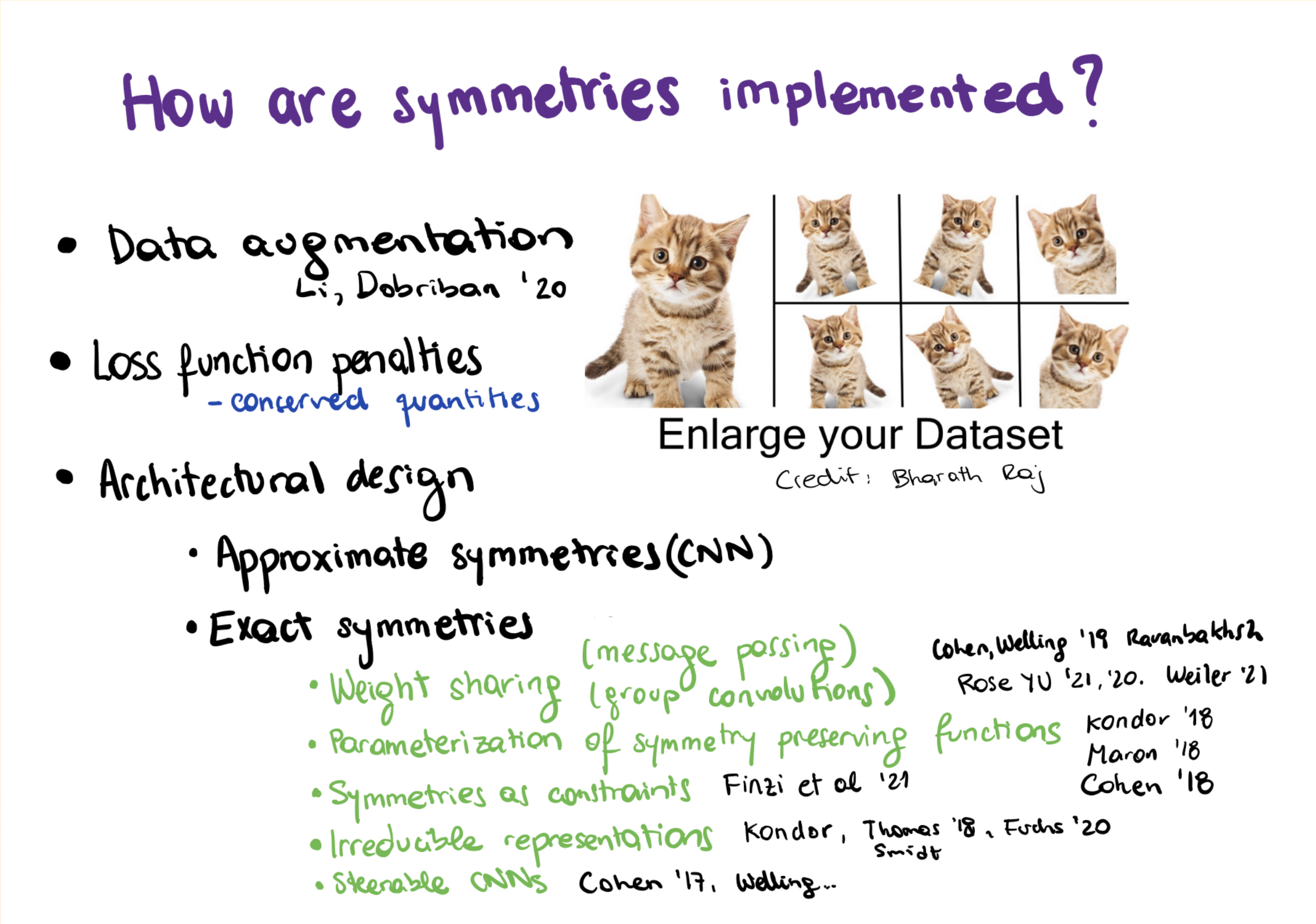


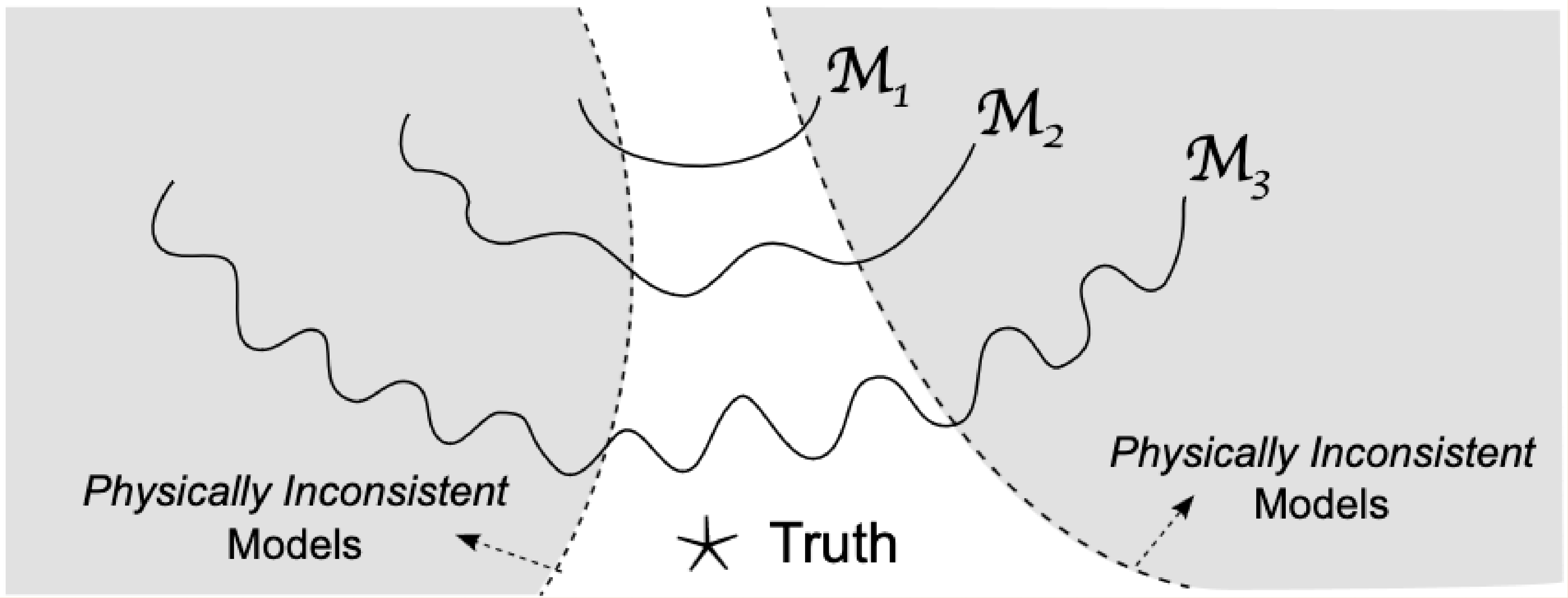
Illustration by Soledad Villar

Plug the physics into the AI: constraints

$$\hat{y} = f(\mathbf{x}, \theta)$$

- Encode physics knowledge (e.g. inconsistency of models) inside the loss function as a penalty term

$$\mathbf{J}(\mathbf{w}) = Loss(y, \hat{y}) + \lambda ||\mathbf{w}||_2^2 + \gamma \Omega(\hat{y}, \Phi)$$



—
Figure from 10.1109/TKDE.2017.2720168

Plug the physics into the AI: network structure

- Equivariance under group transformation can e.g. enforced by convolutional layers
- Some implementations available in pytorch

$$\begin{tikzcd}
L_{V_1}(\mathcal{X}_1) \arrow[r, "\mathbb{T}_g"] \arrow[d, "\phi"] & L_{V_1}(\mathcal{X}_1) \arrow[d, "\phi"] \\
L_{V_2}(\mathcal{X}_2) \arrow[r, "\mathbb{T}'_g"] & L_{V_2}(\mathcal{X}_2)
\end{tikzcd}$$

—
Figure from 10.1109/TKDE.2017.2720168

Plug the physics into the AI: impose output transformations

Phys. Rev. D 110, 096023 (2024)

OPEN ACCESS

Equivariant neural networks for robust CP observables

Sergio Sánchez Cruz[1,*], Marina Kolosova[2], Clara Ramón Álvarez[3], Giovanni Petrucciani[1], and Pietro Vischia[3]

Show more

Phys. Rev. D **110**, 096023 – **Published 26 November, 2024**

DOI: https://doi.org/10.1103/PhysRevD.110.096023

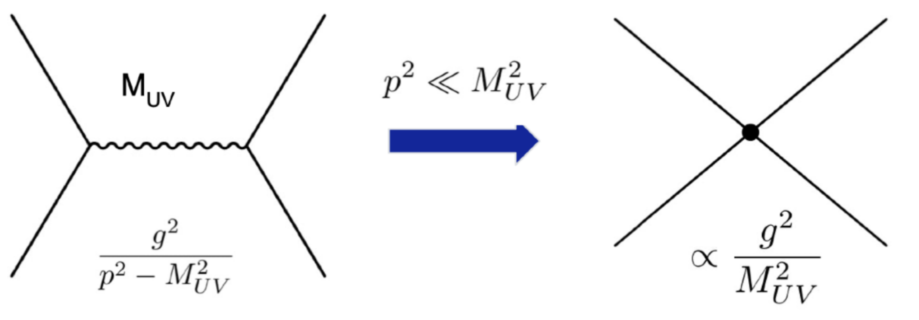
Where is new physics?

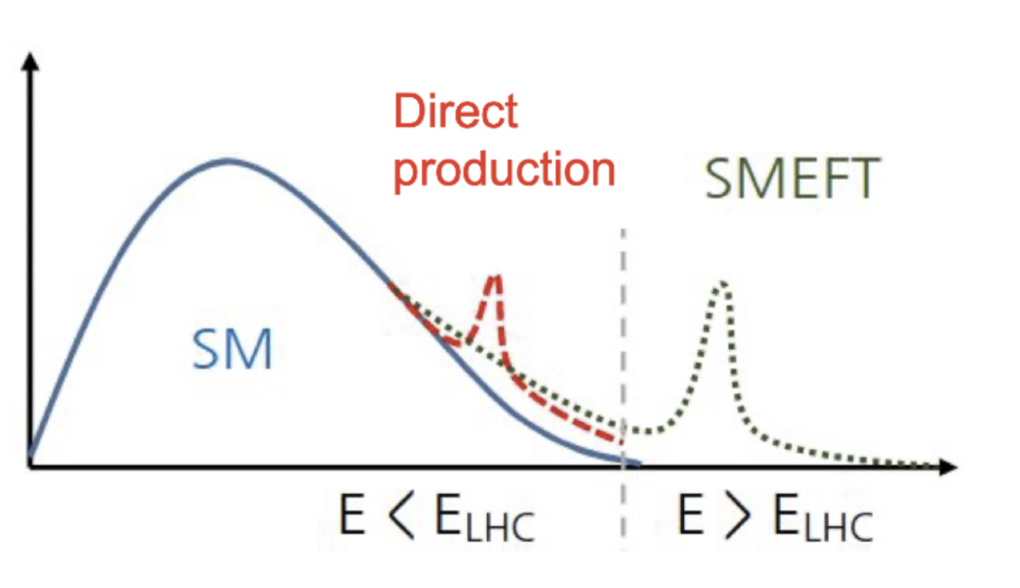
Not in resonances, at least so far

SMEFT and CP Violation Phys. Rev. D 110, 096023 (2024)

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

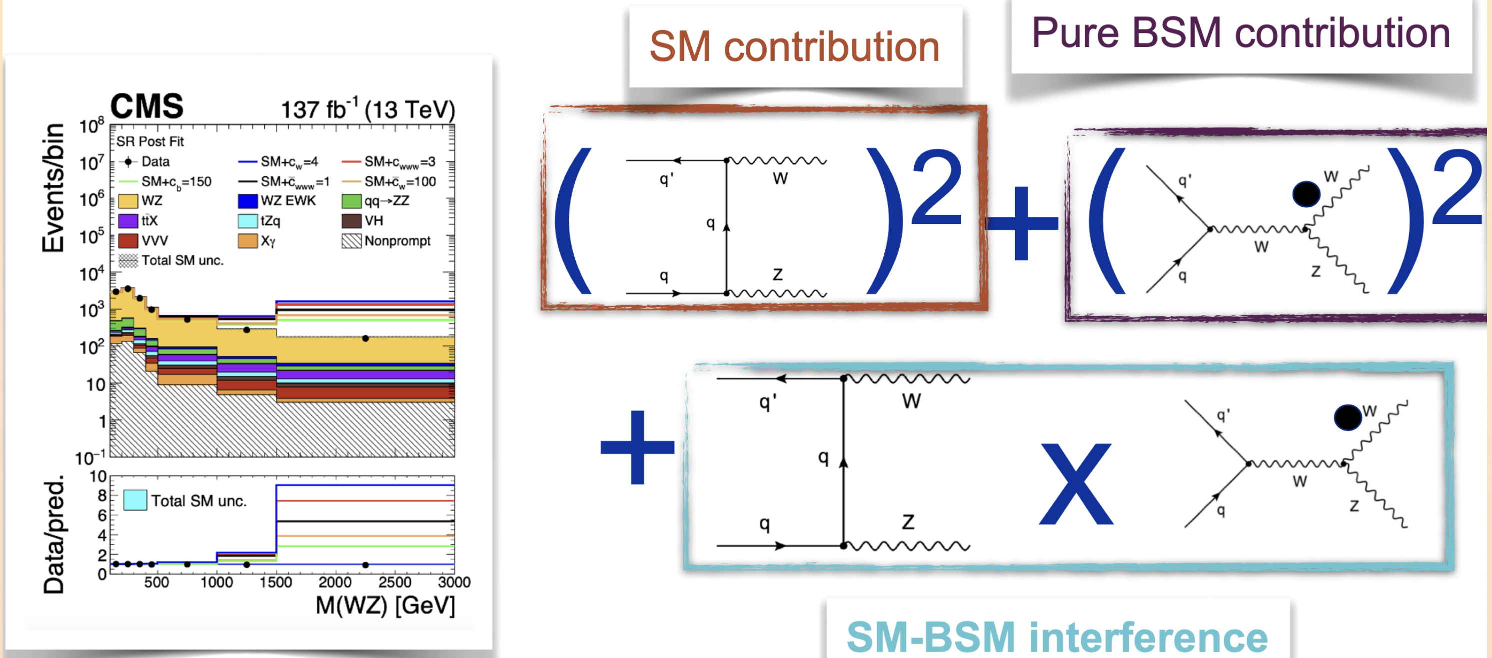
$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{C_i}{\Lambda^2}\mathcal{O}_i^{(6)}$$





EFT Observables

- 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



[JHEP 07 (2022) 032]

Our Algorithm

- Build n_1 CP-invariant observables
  - discriminate between different SM backgrounds
  - discriminate between SM and quadratic terms or CP-even contribution

- Build n_2 CP-odd observables
  - discriminate between signal-like and interference-like contributions
  - discriminate between interference-like and other SM backgrounds

- Obtain a single CP-odd observable by fixing $n_1 = 0$ and $n_2 = 1$ (can generalize to n_1 CP-invariant and n_2 CP-odd components)

- A function $f : D \to R$ is odd under CP transformations if $f(CP(event)) = -f(event)$
- The 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
  - A neural network is not strictly needed, can be any parametric function with enough capacity
- 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: $w(z) = w_{SM}(z) + cw_{int}(z) + c^2 w_{quad}(z)$
  - Weights are functions of parton level kinematics
- Intractable likelihood ratio:

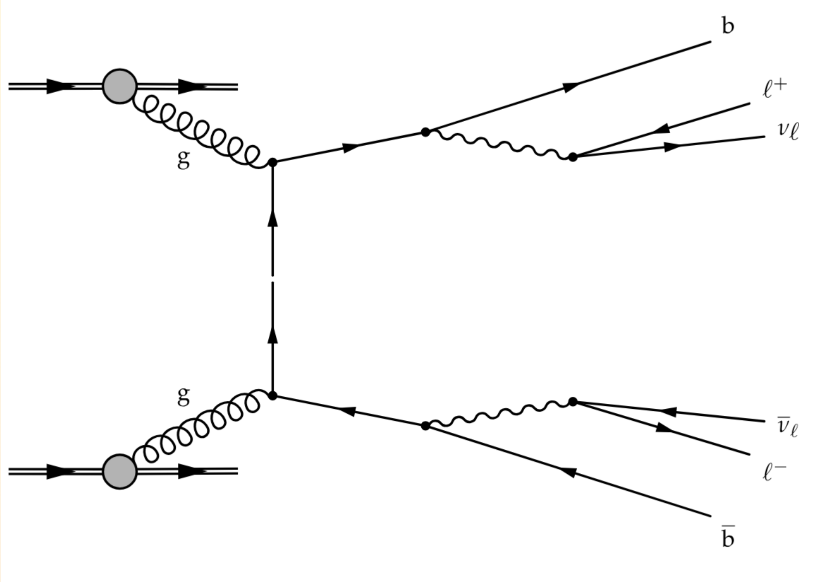
$$\frac{p(d, z|c_1)}{p(d, z|c=0)} = \frac{w_{SM} + cw_{int} + c^2 w_{quad}}{w_{SM}}$$

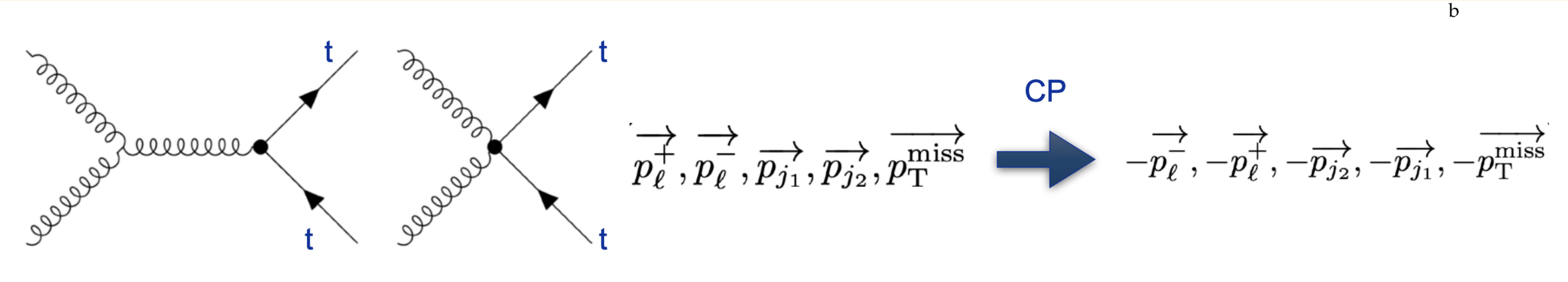
- The likelihood score at the SM point will be a sufficient statistic for small values of c
  - In the small-c regime, the linear component, describing the interference, is dominant
- Learn a surrogate model of the score

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

Use case: ttbar production

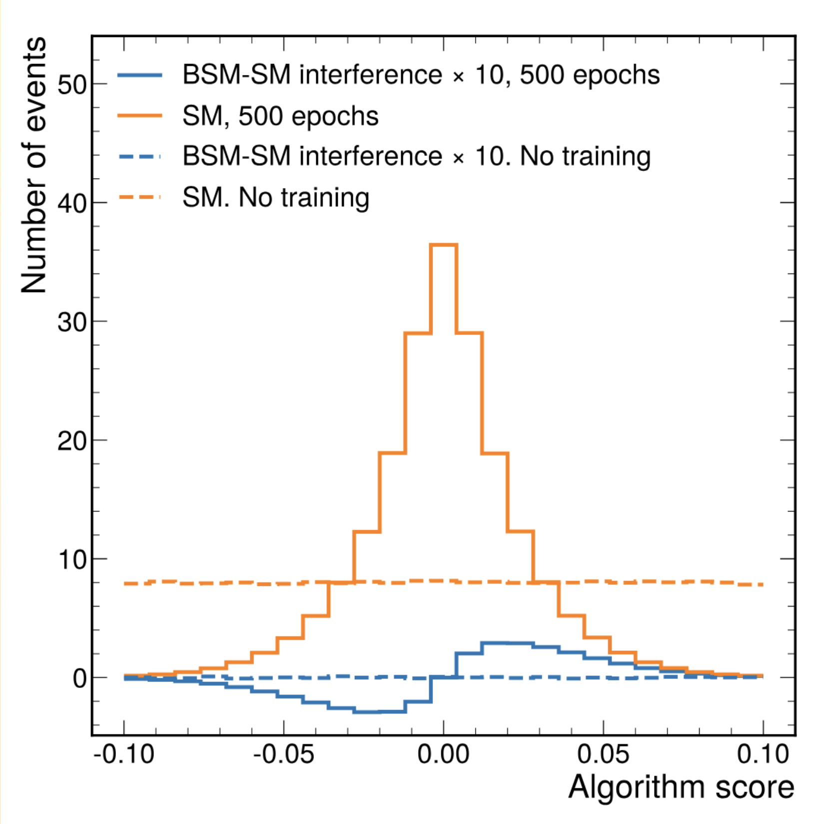
- Dileptonic final state
  - Semileptonic difficult, need to estimate jet charge (BSc thesis of Santiago Vila Domínguez)
- CP-violating chromoelectric dipole moment operator $g_s \frac{v}{\sqrt{2}} (\bar{t}\sigma^{\mu\nu}\gamma_5 T^A t) G^A_{\mu\nu}$





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



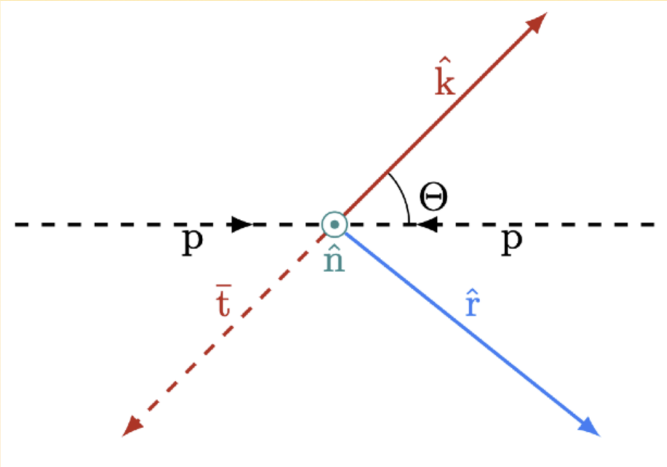
- The observable is robust even before training convergence

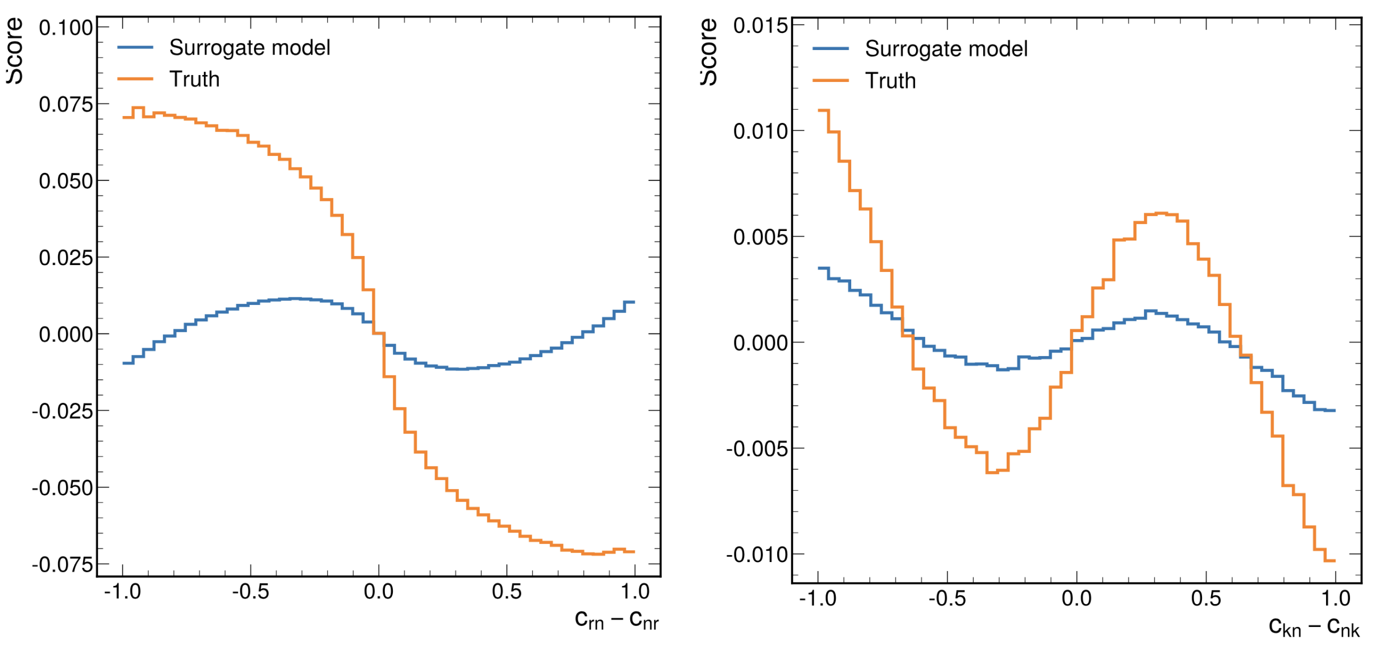
Use case: ttbar production

- Reweight events by the score, compare with parton-level CP-odd observables
- Reconstruct the ttbar system based on angles

$$c_{rn} - c_{nr} = cos(l_r^+)cos(l_r^-) - cos(l_n^+)cos(l_r^-)$$
$$c_{kn} - c_{nk} = cos(l_k^+)cos(l_n^-) - cos(l_n^+)cos(l_k^-)$$

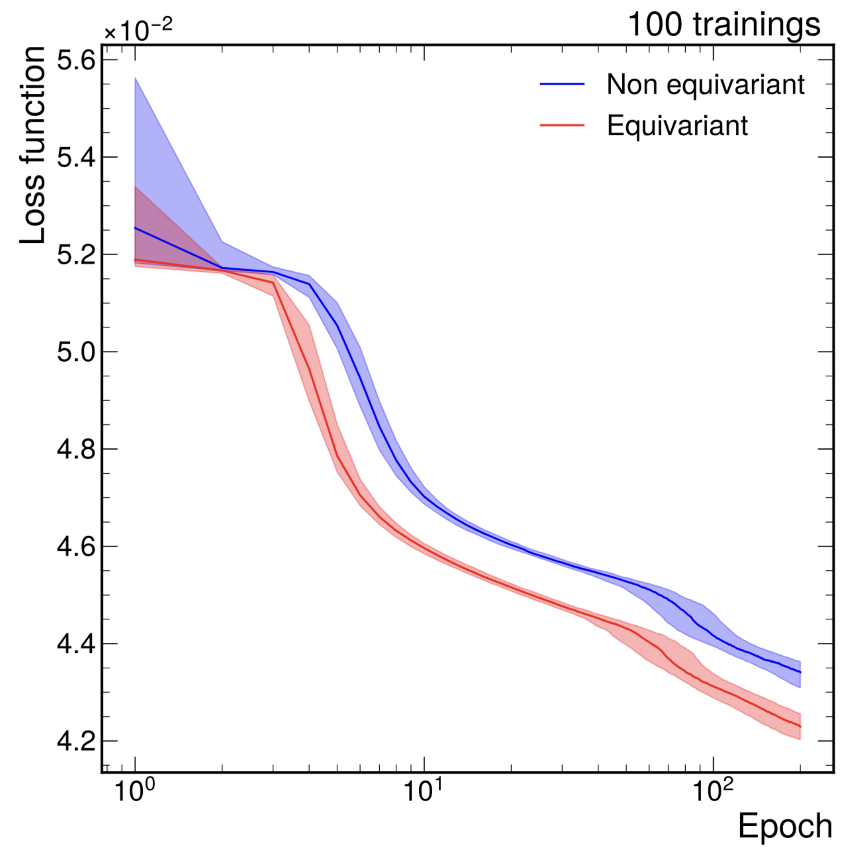
- Limitation is the reconstruction of the ttbar system





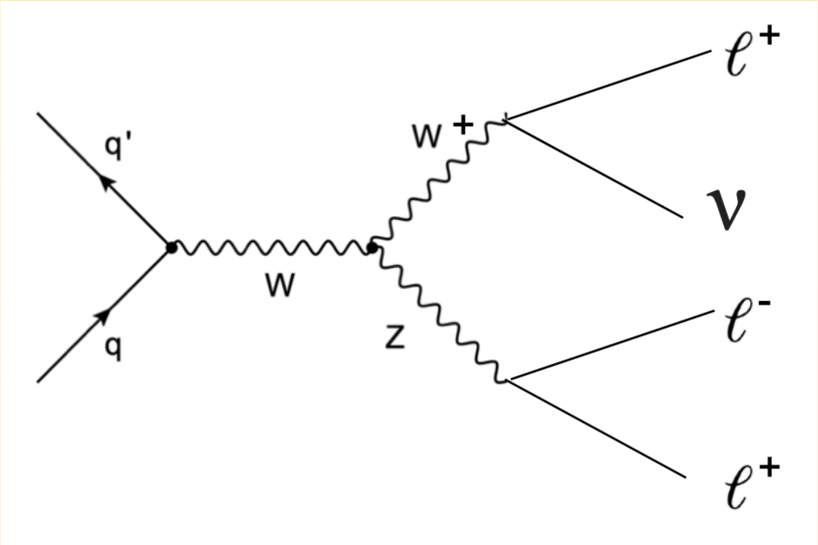
Use case: ttbar production

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

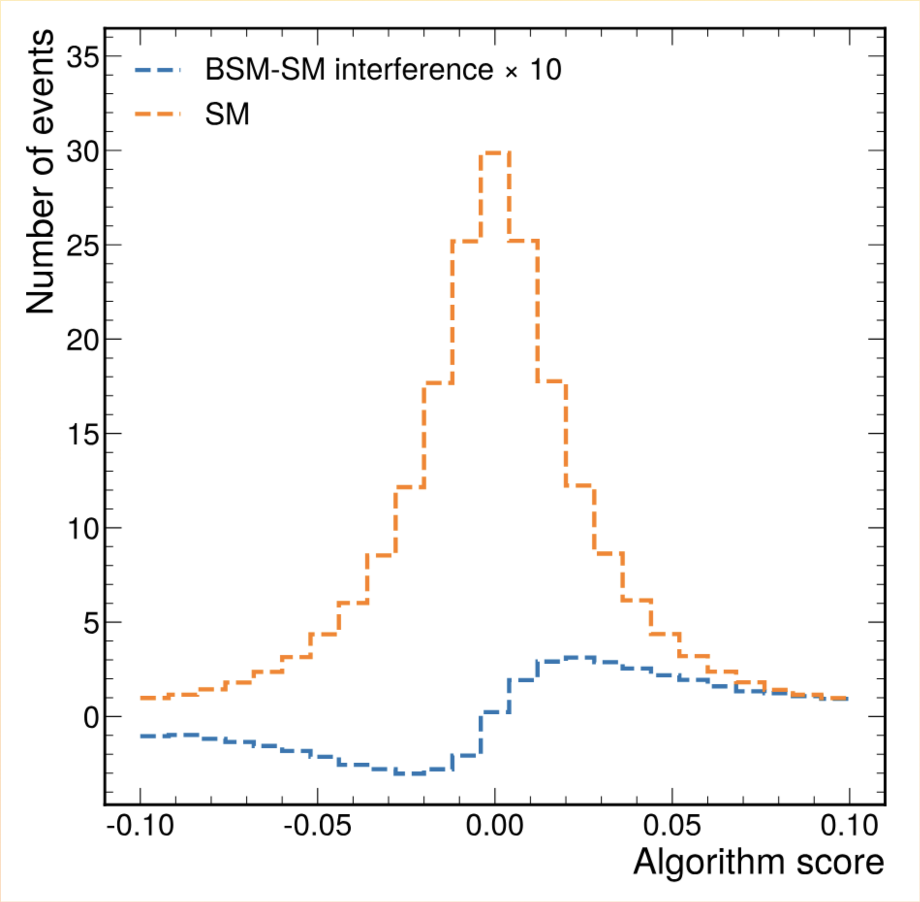


Use case: WZ production

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

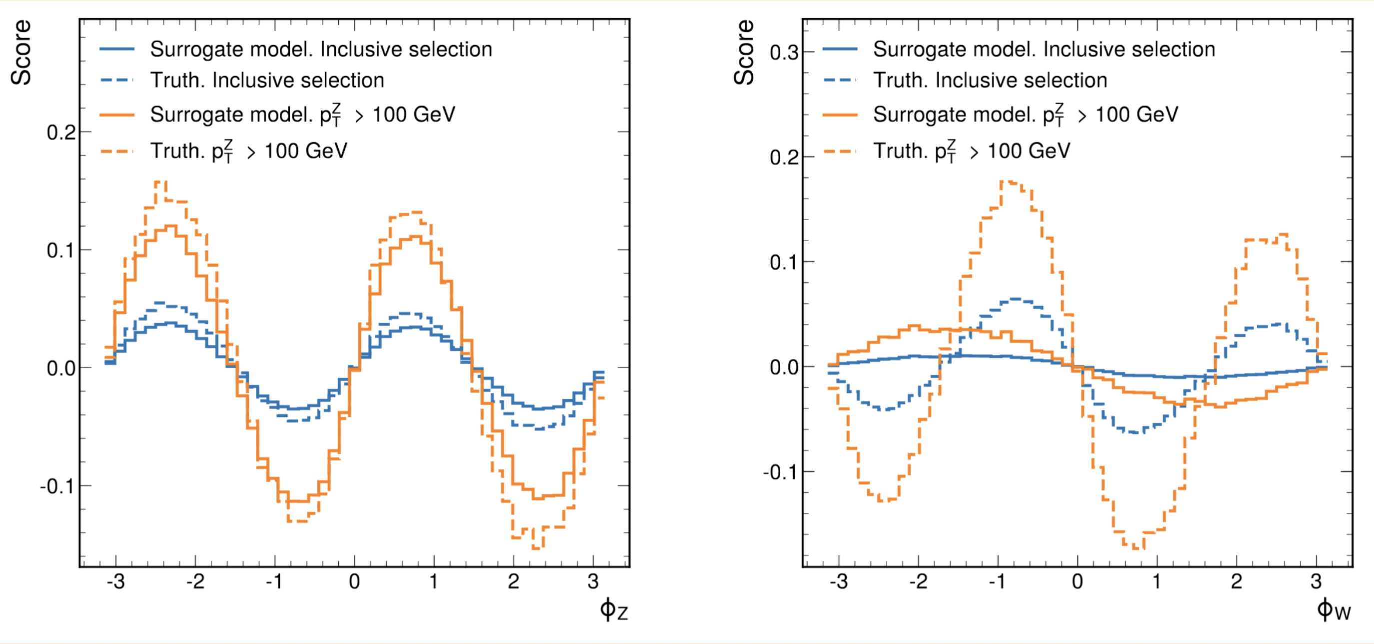


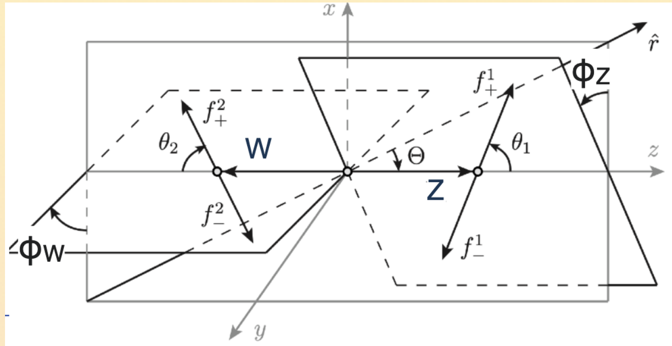
$$\overrightarrow{p^{\mathrm{Z}}_{\ell+}}, \overrightarrow{p^{\mathrm{Z}}_{\ell-}}, \overrightarrow{p^{\mathrm{W}}_{\ell}}, Q^{\mathrm{W}}, \overrightarrow{p^{\mathrm{miss}}_{\mathrm{T}}} \xrightarrow{\mathrm{CP}} -\overrightarrow{p^{\mathrm{Z}}_{\ell-}}, -\overrightarrow{p^{\mathrm{Z}}_{\ell+}}, -\overrightarrow{p^{\mathrm{W}}_{\ell}},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



Use case: WZ production

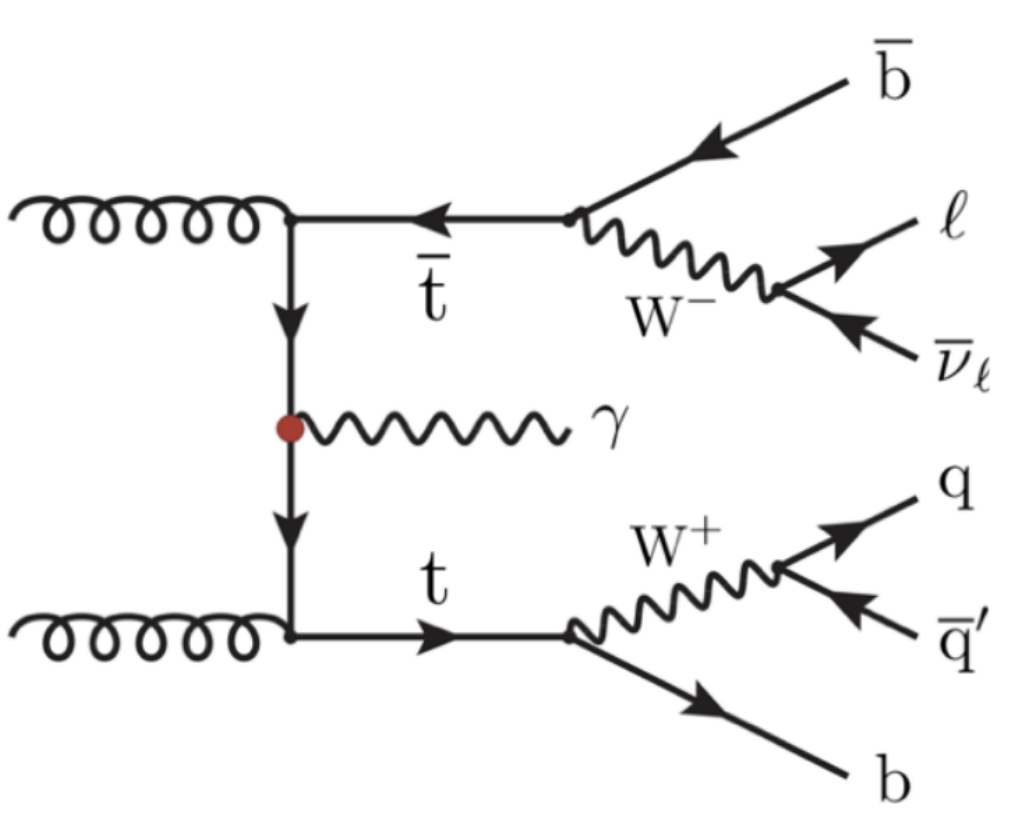
- Performance on parton-level observables **even better than dedicated observables!!!**
  - Can capture energy growth
  - Insensitivity to ϕ_W due to ambiguity in W decay reconstruction

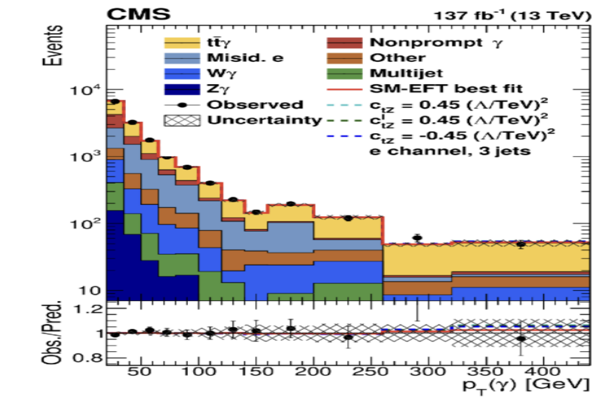




Use case: ttgamma production

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





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

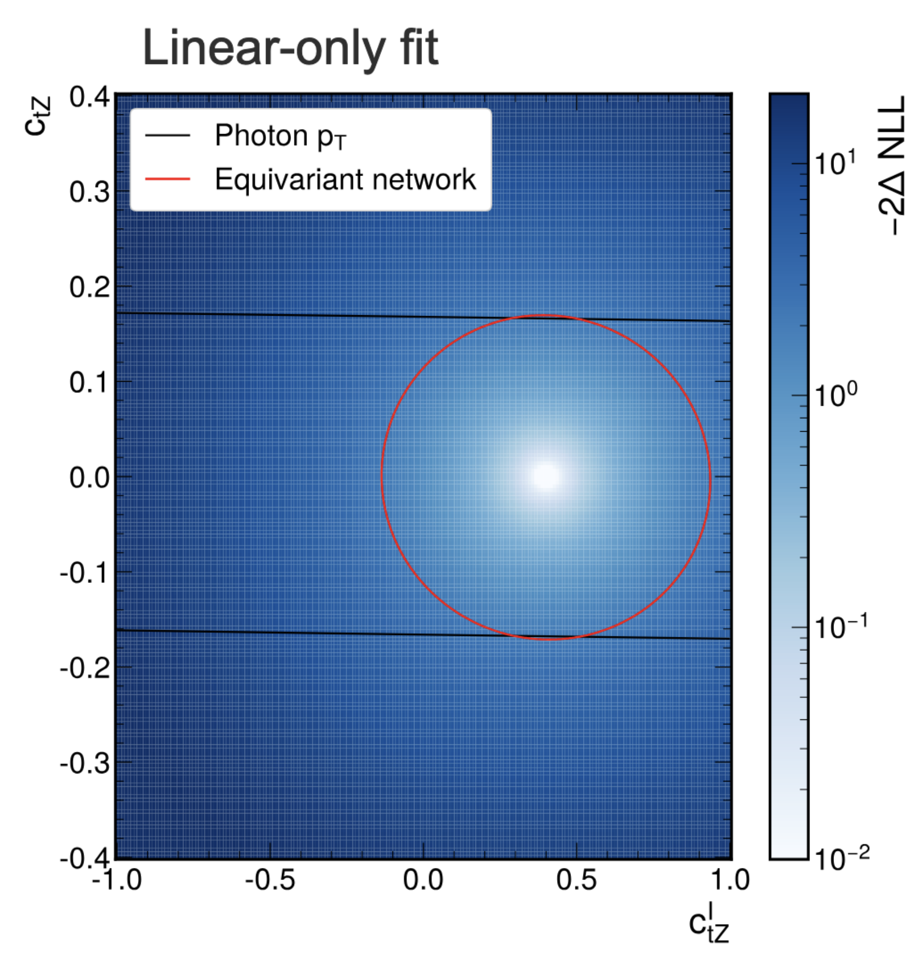
$$\Big\downarrow \text{CP}$$

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

Use case: ttgamma production

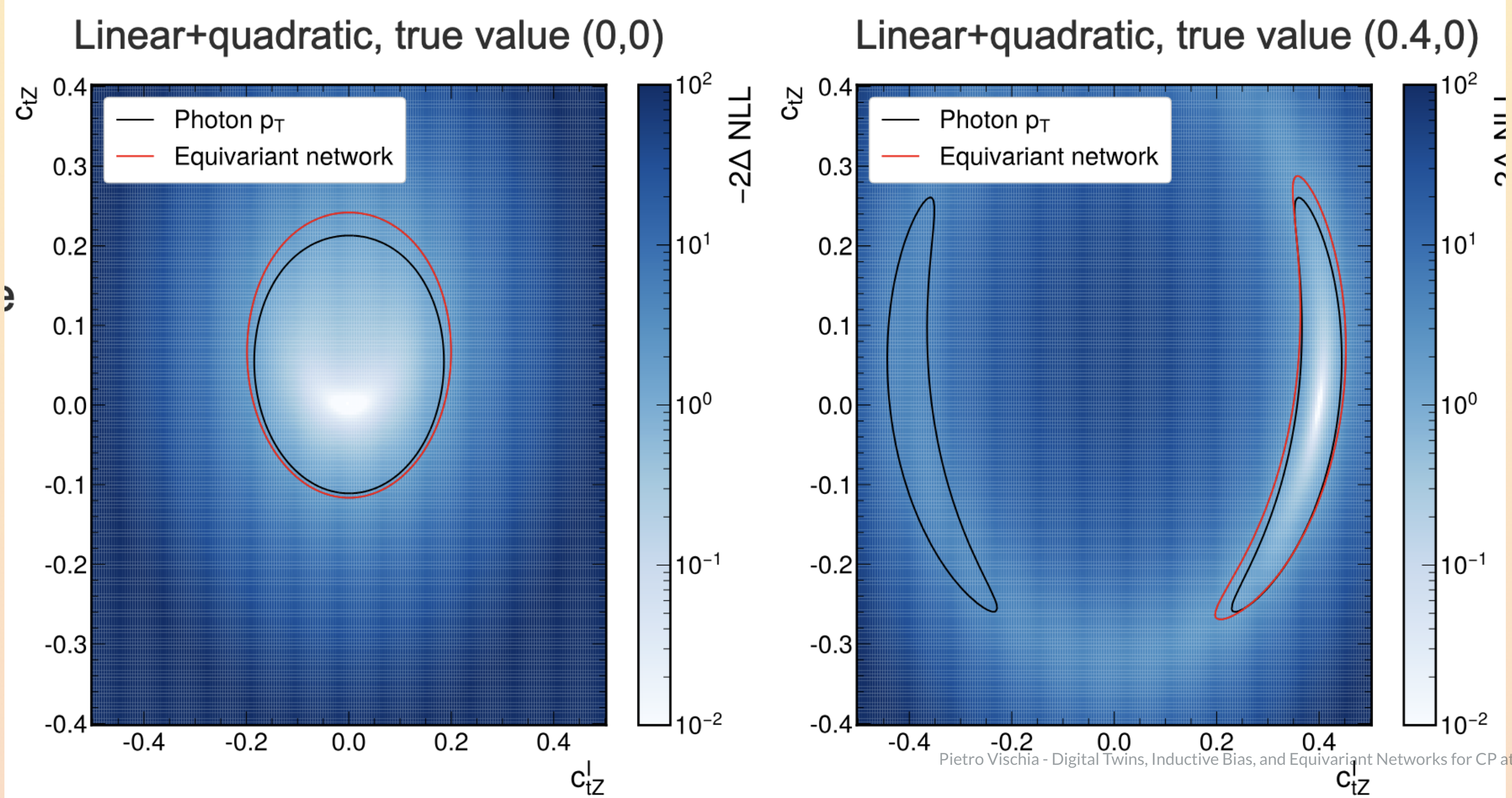
- Linear contribution constrainable only by our approach

- c_{tZ^l} (CP-odd): Comparison with photon p_T is damning (for the photon p_T, which is CP even)
- c_{tZ} (CP-even): similar sensitivity
- With poisson-based likelihoods, systematic uncertainties will be highly suppressed (they are mostly CP-even)



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

- Digital twins are not "just simulations"
- Can we use generative AI to go towards real-time simulation tunes?
- Inductive bias must inform our models
- Implemented equivariant networks to obtain robust observables for CP violation
  - 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
- Check the paper out at Phys. Rev. D 110, 096023 (2024)!!!

Thank you!

—
Photo by me!