Parametrized classifiers for optimal EFT sensitivity

Giuliano Panico

Università di Firenze and INFN Firenze

Machine Learning at GGI — Firenze — 8/9/2022

based on S. Chen, A. Glioti and A. Wulzer 2007.10356 [hep-ph]
+ work in progress

Fundamental physics at colliders

The main goal of the collider program is to deepen our knowledge of fundamental physics

Our current knowledge about the properties and dynamics of elementary particles is encoded in the **Standard Model** (**SM**)

It agrees with what we measured at colliders so far with amazing accuracy
… but there are many hints that it is not the ultimate theory

In practical terms, the collider physics program aims at
testing the SM

looking for its possible **failures** → evidence of **New Physics** (BSM)

Testing the SM

Complementarity

devising different strategies to test the SM predictions
and to cover different types of new physics

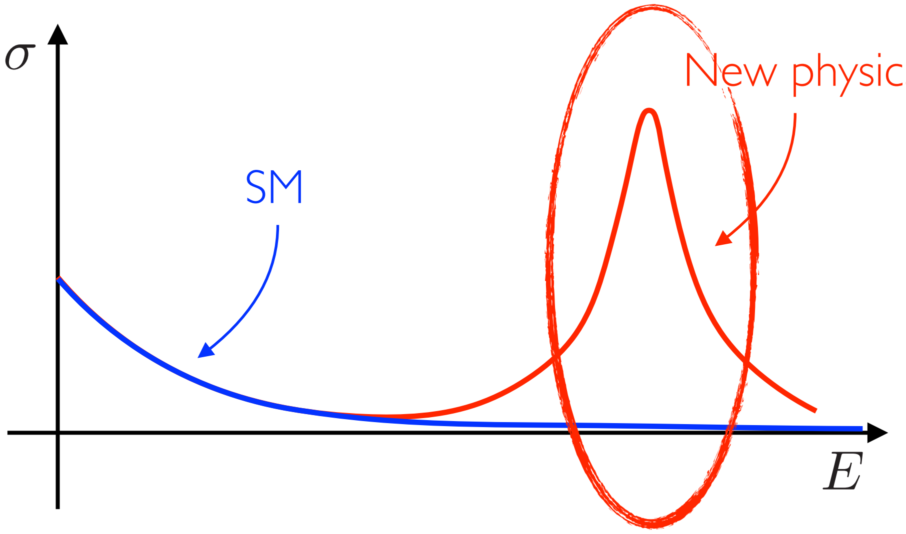
Optimality

improve and optimize the new-physics probes to achieve better sensitivity

How to look for new physics

Direct searches:

look for signals of production
of new particles

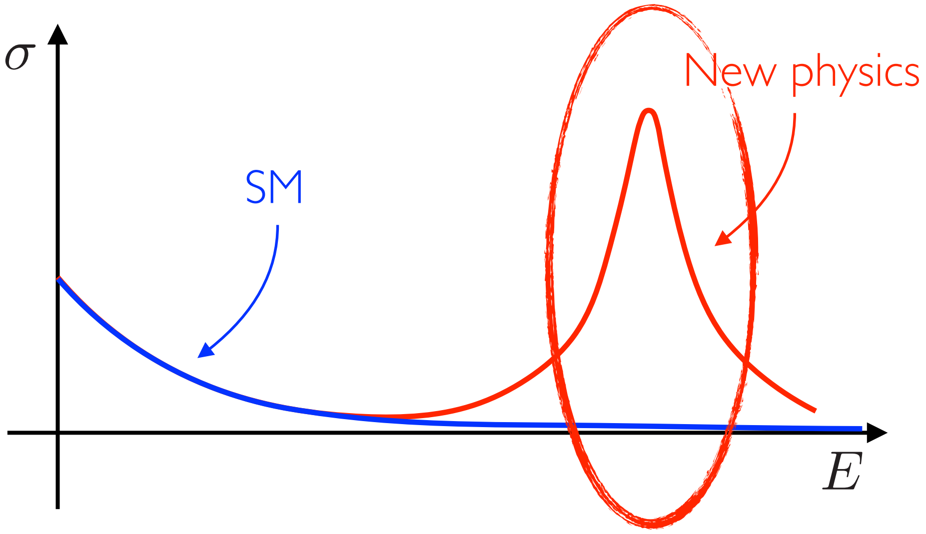


- resonant effects in kinematic distributions
- "bump" on top of a smooth SM background
  (that can be often extracted from the data)

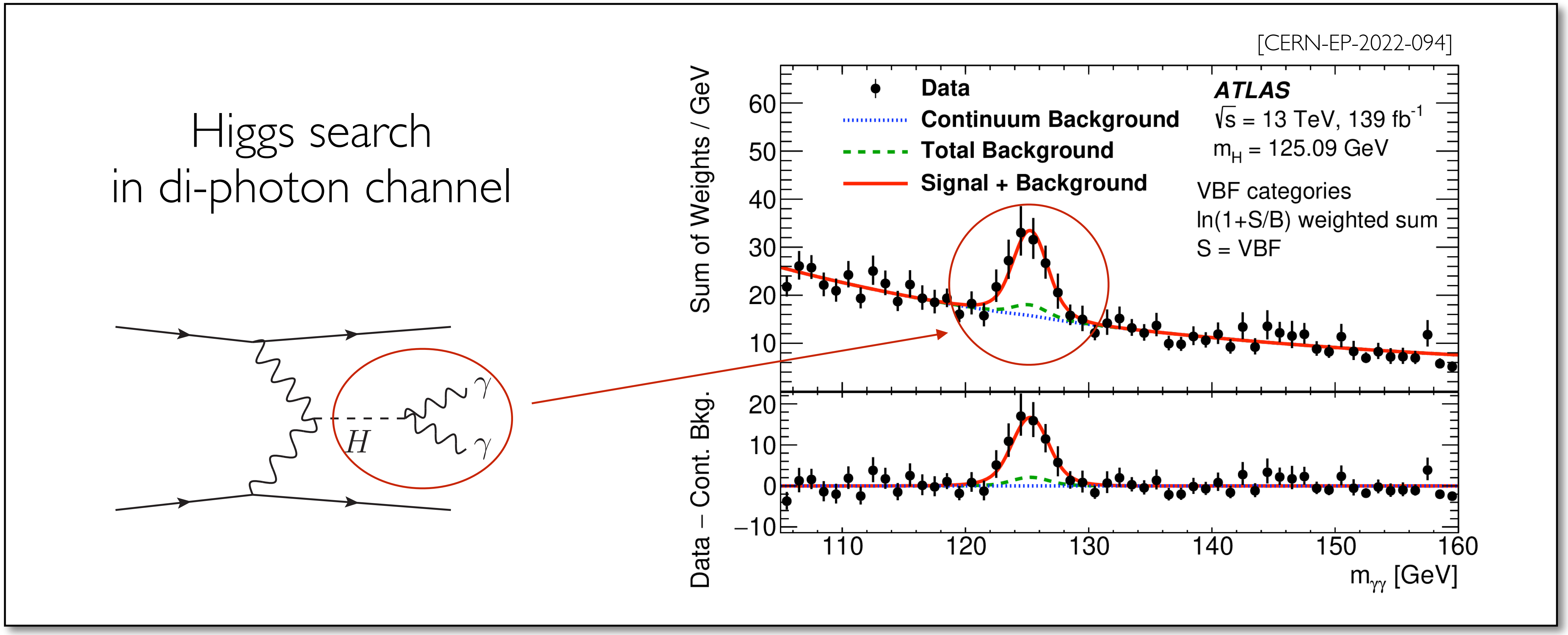
How to look for new physics

Direct searches:

look for signals of production
of new particles



- resonant effects in kinematic distributions
- "bump" on top of a smooth SM background
(that can be often extracted from the data)



Higgs search
in di-photon channel

[CERN-EP-2022-094]

How to look for new physics

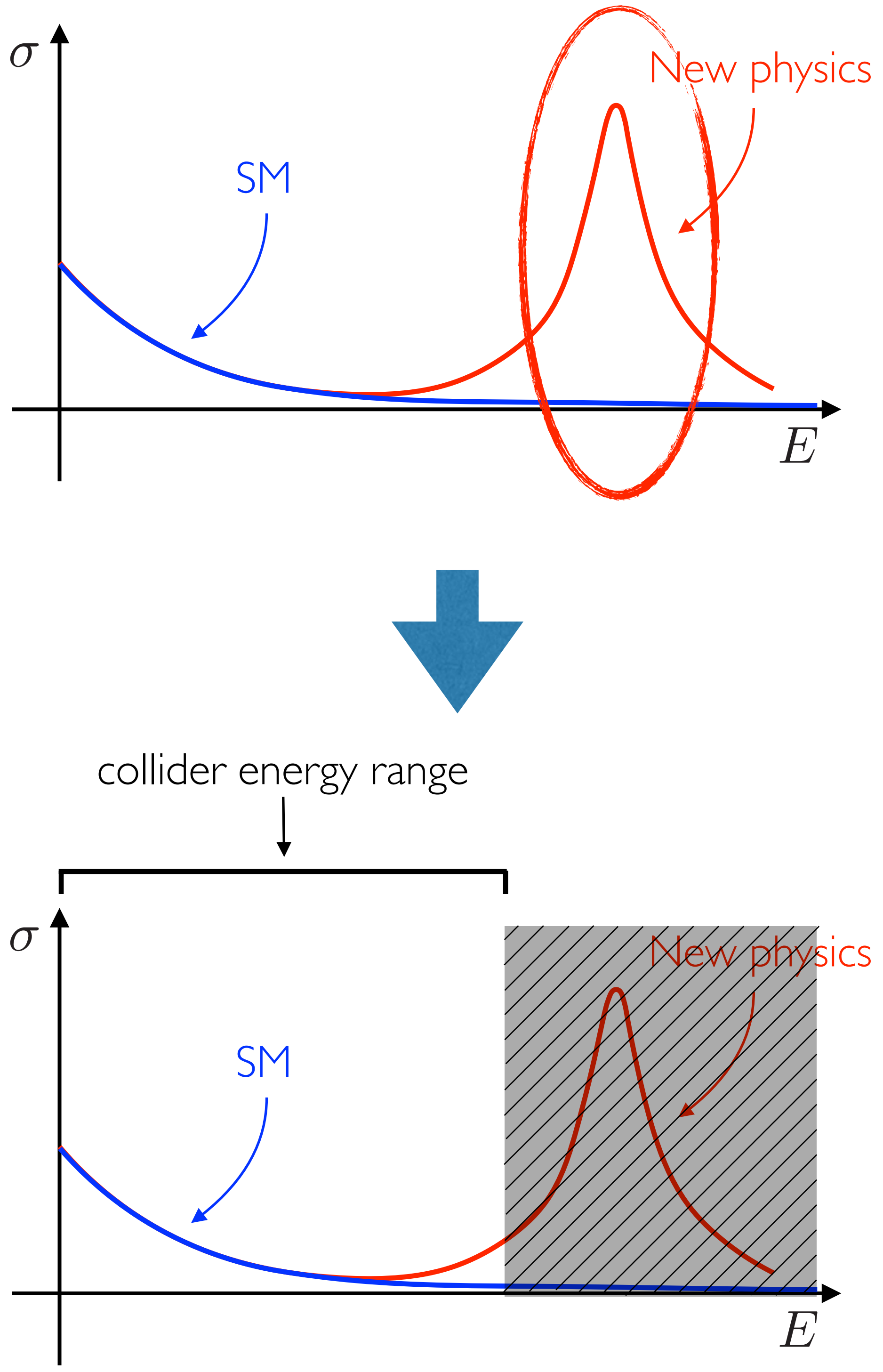
<u>Direct searches:</u>

look for signals of production
of new particles

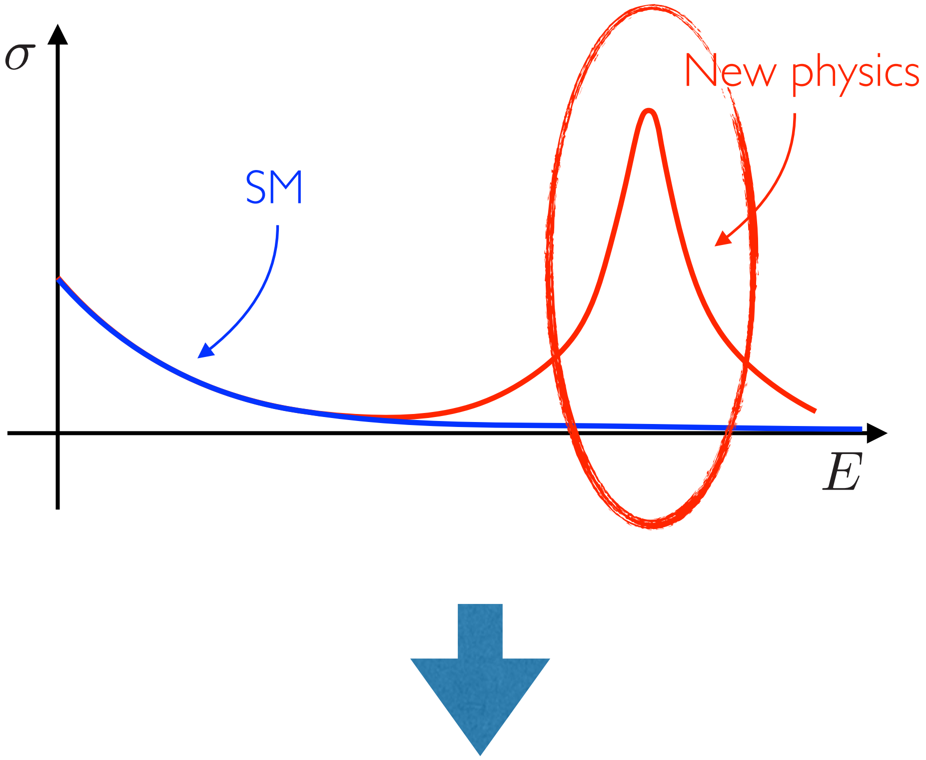
- resonant effects in kinematic distributions
- "bump" on top of a smooth SM background
(that can be often extracted from the data)

Limitations:

- new particle must be resonantly produced and must decay to reconstructable final state
- limited by collider energy range



How to look for new physics



Direct searches:

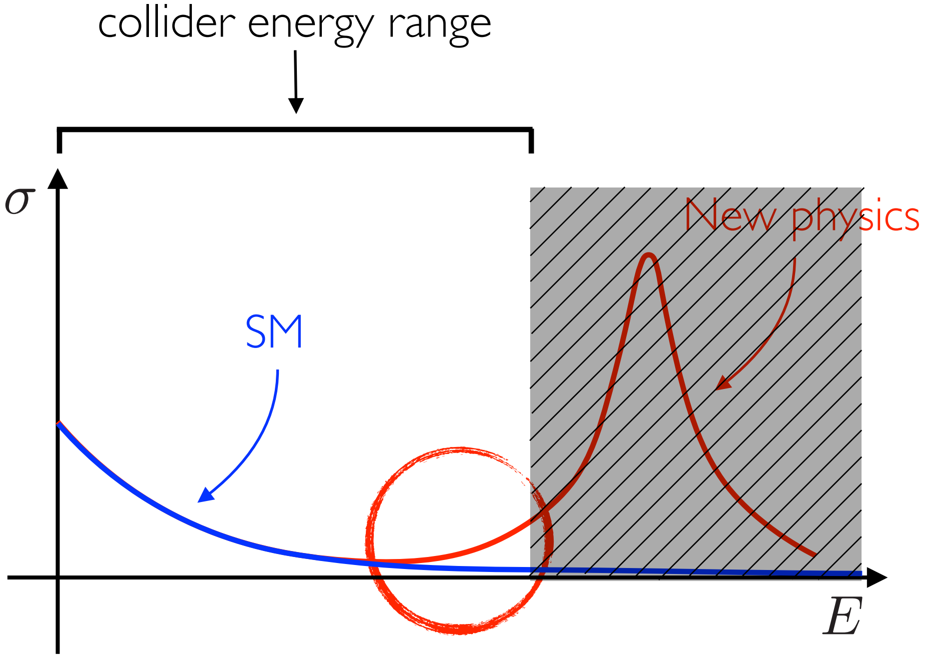
look for signals of production
of new particles

- resonant effects in kinematic distributions
- "bump" on top of a smooth SM background
(that can be often extracted from the data)

Looking for the tail: Indirect searches

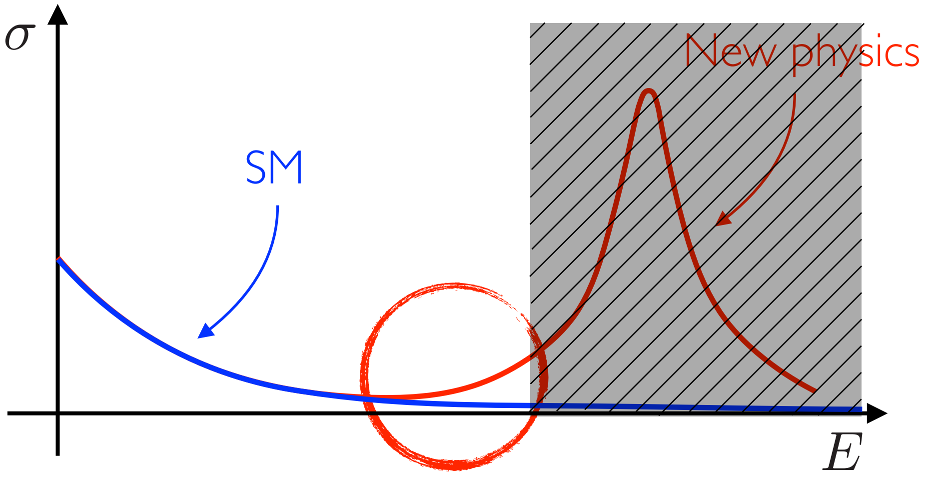
even if we can not directly produce
the new particles,
we can test their **indirect effects**

- LEP data at 200 GeV tested new particles with masses up to 3 TeV !



Tails are "universal"

Indirect searches have important advantages



"universality"

- deviations from SM exhibit small number of behaviors dictated by symmetries

"model independence"

- captures a huge class of new-physics models

"ubiquity"

- deviations are present also in channels with non-resonant new physics production
- can often be seen also in channels where the final state can not be fully reconstructed

The EFT approach

The **Effective Field Theory** (**EFT**) description can be obtained
"integrating out" heavy particles

- interactions mediated by heavy particles are replaced by new contact interactions

$$\sum_d \sum_i \frac{d_i}{\Lambda^{d-2}} \mathcal{O}_i^{dim-d}$$

The EFT approach

The **Effective Field Theory** (**EFT**) description can be obtained
"integrating out" heavy particles

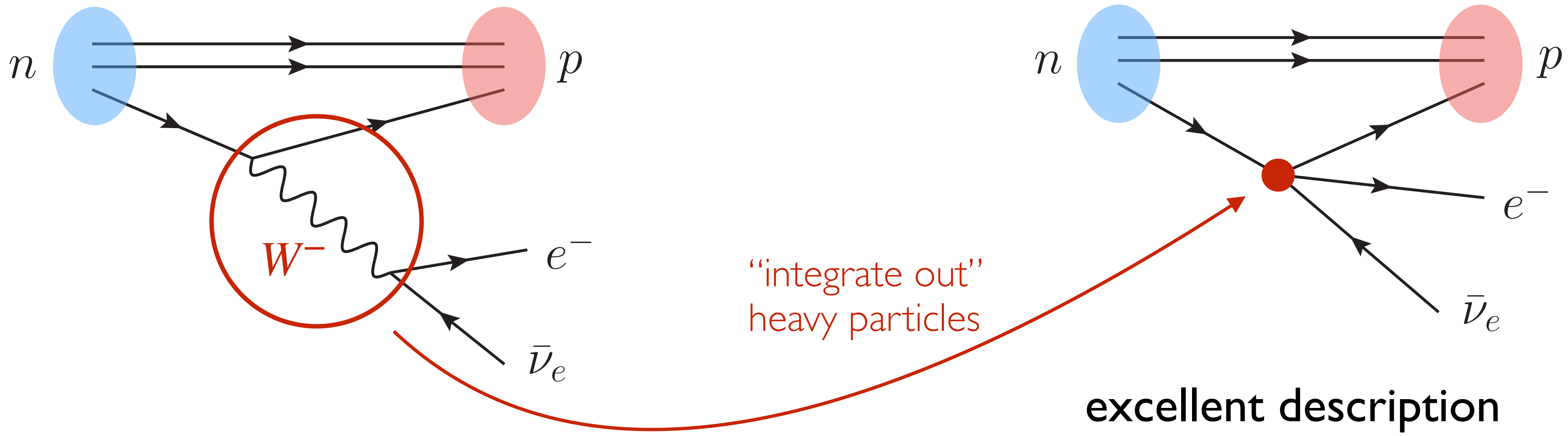
- interactions mediated by heavy particles are replaced by new contact interactions

Electroweak theory

(Fundamental)

Fermi theory

(Effective Field Theory)



excellent description
at low energy,
but with **limited range of validity**

$$E \ll \underset{\substack{\nwarrow \\ \text{cut-off}}}{\Lambda} \sim m_W = 80\ \text{GeV}$$

The EFT approach

The **Effective Field Theory** (**EFT**) description can be obtained
“integrating out” heavy particles

- interactions mediated by heavy particles are replaced by new contact interactions

Rules of the game:

- **only SM particles** appear in the EFT
- **SM symmetries** are respected

→ “small” number of new terms

Terms in the EFT Lagrangian can be organized in an expansion in E/Λ

$$\mathscr{L}_{SMEFT} = \mathscr{L}_{SM} + \frac{1}{\Lambda^2}\mathscr{L}_6 + \frac{1}{\Lambda^4}\mathscr{L}_8 + \cdots$$

energy dimension ≤ 4 dim = 6 dim = 8

The EFT approach

The **Effective Field Theory** (**EFT**) description can be obtained "integrating out" heavy particles

- interactions mediated by heavy particles are replaced by new contact interactions

Rules of the game:

- **only SM particles** appear in the EFT
- **SM symmetries** are respected

→ "small" number of new terms

Terms in the EFT Lagrangian can be organized in an expansion in E/Λ

$$\mathscr{L}_{SMEFT} = \mathscr{L}_{SM} + \underbrace{\frac{1}{\Lambda^2}\mathscr{L}_6}_{\text{leading corrections}} + \frac{1}{\Lambda^4}\mathscr{L}_8 + \cdots$$

The challenges of indirect searches

Performing indirect searches is a challenging task
that requires several key ingredients

- Accurate theoretical knowledge of the SM and BSM predictions
  (i.e. small theoretical systematic uncertainty)

  → needed to compare theoretical expectation with the experimental data

- Accurate experimental measurements
  (i.e. small experimental systematic and statistical uncertainty)

  → in many cases we expect small deviations with respect to the SM

- Use of effective search strategies and optimized statistical analysis

Optimal tests of new physics

The **differential distribution** contain the maximal information about a process

$$d\sigma(x; C)$$

measurable
kinematic quantities

new-physics
parameters

- basis to perform optimal statistical tests (eg. likelihood ratio test)

Optimal tests of new physics

The **differential distribution** contain the maximal information about a process

$$d\sigma(x; C)$$

measurable kinematic quantities

new-physics parameters

- basis to perform optimal statistical tests (eg. likelihood ratio test)

How to determine the theoretical kinematic distributions?

- not known in analytic form
- only available knowledge are Monte Carlos event samples following $d\sigma(x; C)$
    - "latent" Monte Carlo variables z do not coincide with measurable quantities x

$$\begin{tikzcd} z \arrow[r, bend left] & x \arrow[l, bend left, "\times"] \end{tikzcd}$$

    - higher-order effects generate "unphysical" events (negative weights)
    - showering, hadronization, detector effects not known "analytically"

A simple analysis strategy

Simplest approach: exploit partial kinematic information

- keep only few kinematic variables and ‘ignore’ the others
- reconstruct the distributions through binning

A simple analysis strategy

Simplest approach: exploit partial kinematic information

- keep only few kinematic variables and 'ignore' the others
- reconstruct the distributions through binning

Example: **di-lepton production**

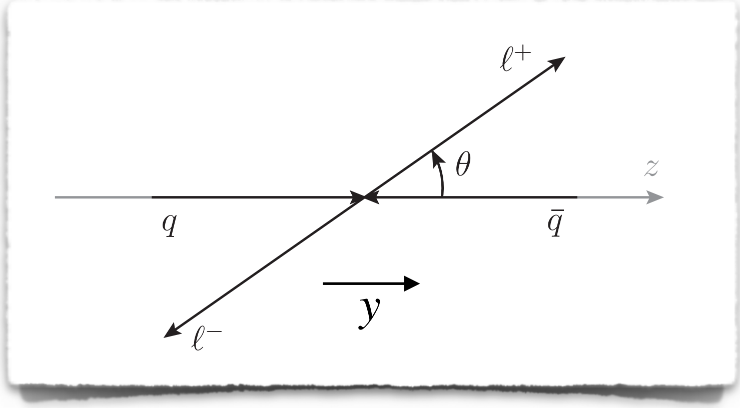
$$pp \to \ell^+ \ell^-$$

three kinematic variables

invariant mass $m_{\ell^+\ell^-}$

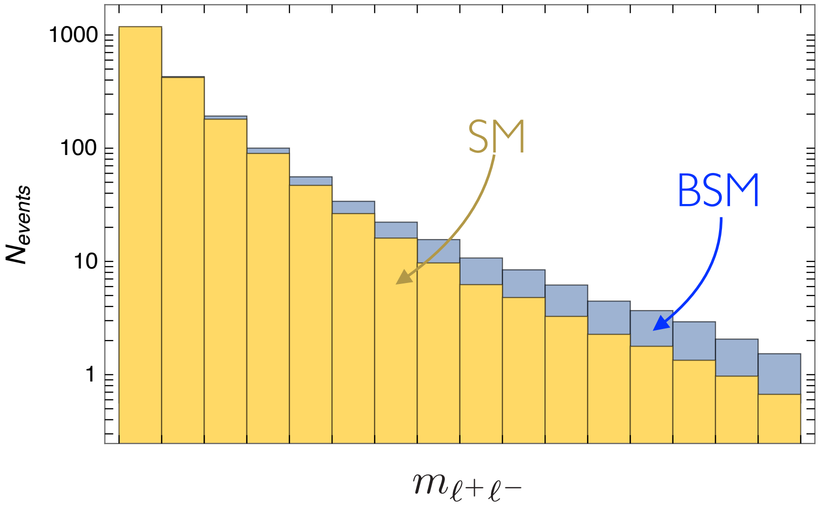
scattering angle θ

c.o.m. rapidity y



- can focus only on invariant mass
- distribution reconstructed with simple 1-dimensional binning

$$d\sigma(m_{\ell^+\ell^-}; C)$$



A simple analysis strategy

Simplest approach: exploit partial kinematic information

- keep only few kinematic variables and 'ignore' the others
- reconstruct the distributions through binning

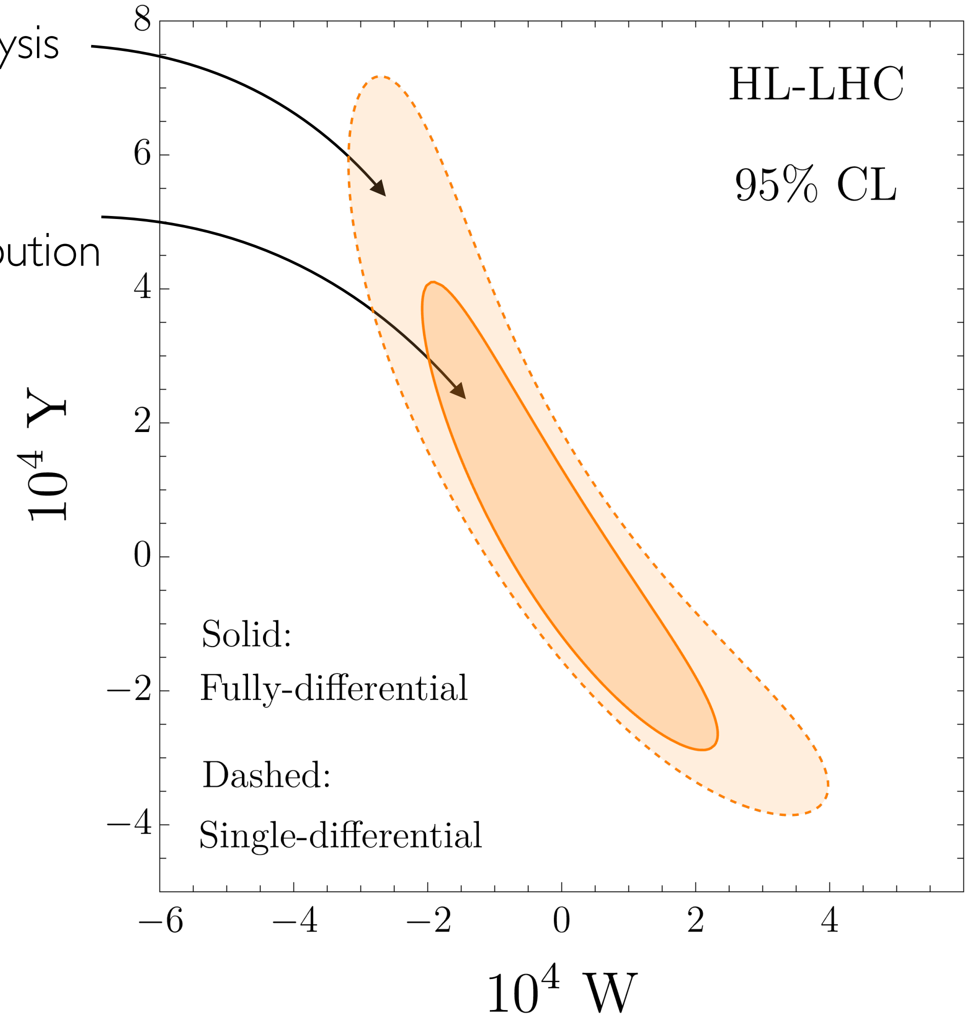
Example: **di-lepton production**

$$pp \to \ell^+\ell^-$$

[G.P., L. Ricci, A. Wulzer '21]

1-dim analysis

analysis with
full kinematic distribution



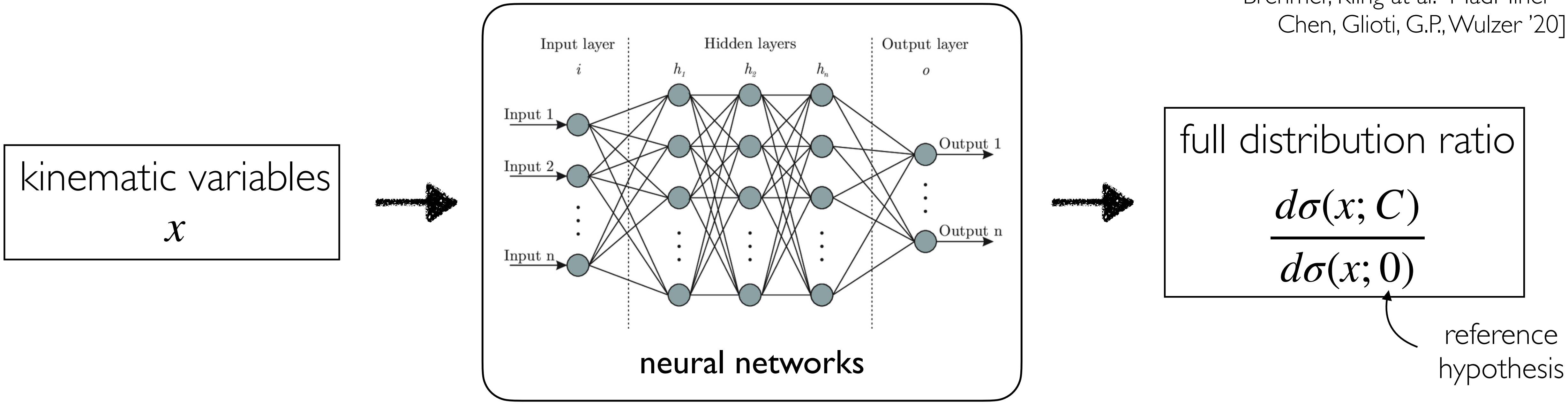
Big loss in sensitivity!

A Machine Learning approach

Full distributions through ML

[Baldi, Cranmer et al. '16;
Brehmer, Cranmer, Louppe, Pavez '18;
Cranmer, Pavez, Louppe '18;
Stoye, Brehmer at al. '18;
Brehmer, Louppe at al. '18;
Brehmer, Cranmer at al. '18;
Brehmer, Kling at al. "MadMiner" '19;
Chen, Glioti, G.P., Wulzer '20]

Basic idea: reconstruct $d\sigma(x; C)$ with Neural Networks



kinematic variables
x

neural networks

full distribution ratio

$$\frac{d\sigma(x; C)}{d\sigma(x; 0)}$$

reference
hypothesis

- fully differential (analytic) result in all measurable quantities
- obtained with a relatively small amount of Monte Carlo data
- systematically improvable
  - with more data, reacher NN structure, …
  - with more accurate Monte Carlo samples (eg. higher-order effects, backgrounds, …)

A Simple Classifier

A **binary classifier** can be used to reconstruct the distribution ratio from Monte Carlo data

- two samples, following new physics ($C = \overline{C}$) and reference ($C = 0$) distributions

$$\mathcal{S}_{\overline{C}} = \{x_i \sim d\sigma(x; \overline{C})\} \qquad \mathcal{S}_0 = \{x_i \sim d\sigma(x; 0)\}$$

notice that $\overline{C}$ is fixed

- binary classifier (eg. with quadratic loss)

$$L = \sum_{x_i \in \mathcal{S}_{\overline{C}}} [NN(x_i) - 1]^2 + \sum_{x_i \in \mathcal{S}_0} [NN(x_i)]^2$$

- in the infinite training sample limit

$$\frac{\delta L}{\delta NN} = 0 \quad \longrightarrow \quad NN(x) = \frac{d\sigma(x; \overline{C})}{d\sigma(x; \overline{C}) + d\sigma(x; 0)} \quad \longleftrightarrow \quad \frac{d\sigma(x; \overline{C})}{d\sigma(x; 0)} = \frac{NN(x)}{1 - NN(x)}$$

✦ weighted samples can be treated in an analogous way introducing weights in L

Application to WZ production

$$\boxed{pp \to W^{\pm}Z \to (\ell^{\pm}\nu)(\ell^{+}\ell^{-})}$$

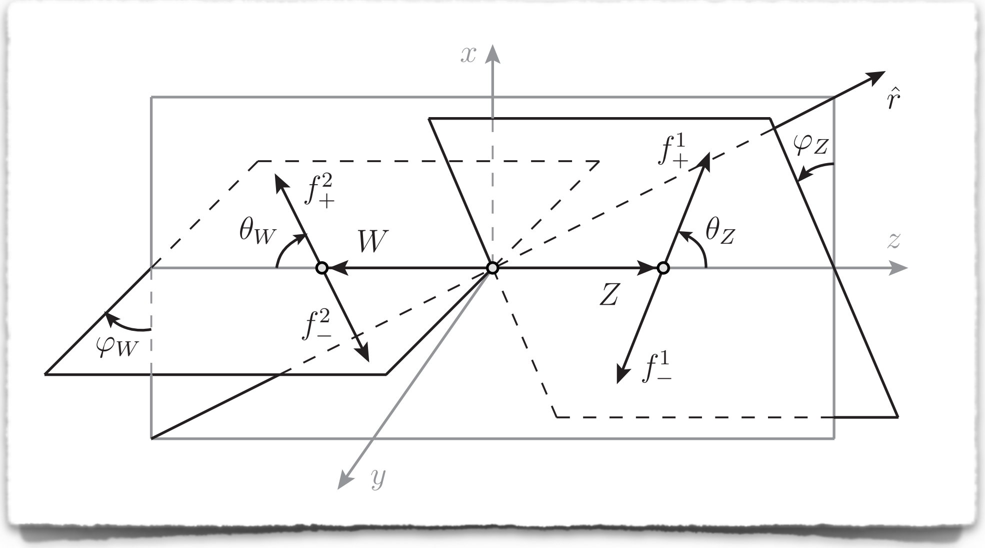
Final state described
by 6 kinematic variables!

invariant mass m_{WZ}

scattering angle θ

W decay angles θ_W, ϕ_W

Z decay angles θ_Z, ϕ_Z



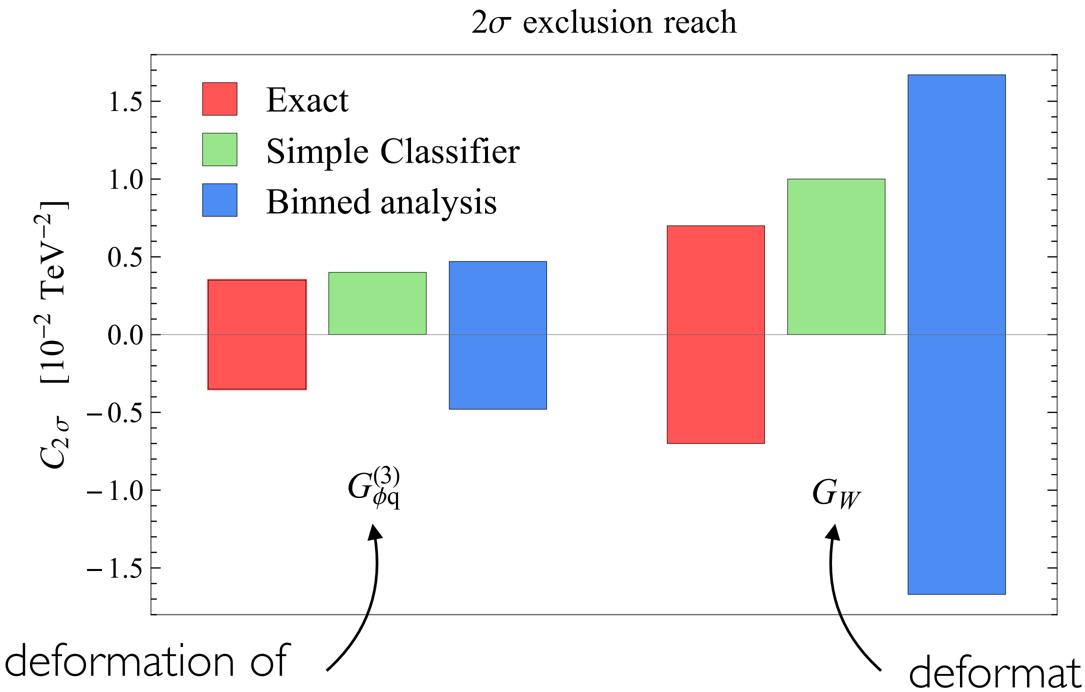
- standard binned analysis can not take into account all kinematic variables (at most two or three)

→ important features characterizing new-physics distributions are lost

→ huge loss in sensitivity

Simple Classifier performance

- The Simple Classifier approach works, but there is still some gap



deformation of W,Z-quark couplings

deformation of ZWW coupling

Technical details

- fully connected feed-forward NN {6, 32, 32, 1}
- sigmoid activation
- 500k + 500k training events

(also checked: {6, 4x32, 1}, {6, 6x32, 1}; 3M + 3M training events; ReLU activation functions; but not much improvement)

Drawbacks:

- must be trained for ‘every’ value of $\overline{C}$ (with new Monte Carlo sample!)
- becomes inefficient for small values of $\overline{C}$
  (differential distribution very close to reference, large amount of training data are needed to reconstruct the ratio)

Joining Machine Learning with Theory

The Quadratic Classifier

Theory fixes the structure of the differential distribution

$$d\sigma(x;C) = d\sigma(x;0)\,\underbrace{\left[(1 + C\,\alpha(x))^2 + C^2\,\beta^2(x)\right]}_{\text{positive \textbf{quadratic polynomial} in } C}$$

We can exploit this information to solve the drawbacks of the Simple Classifier approach

The Quadratic Classifier

Theory fixes the structure of the differential distribution

$$d\sigma(x;C) = d\sigma(x;0)\underbrace{\Big[(1+C\,\alpha(x))^2 + C^2\,\beta^2(x)\Big]}$$

positive **quadratic polynomial** in C

- we use a standard binary classifier loss

$$L = \sum_{\{C_i\}} \left\{ \sum_{x_i \in \mathcal{S}_0} [F(x_i;C_i) - 1]^2 + \sum_{x_i \in \mathcal{S}_{C_i}} [F(x_i;C_i)]^2 \right\}$$

- but the distribution ratio is parametrized in terms of **two neural networks**

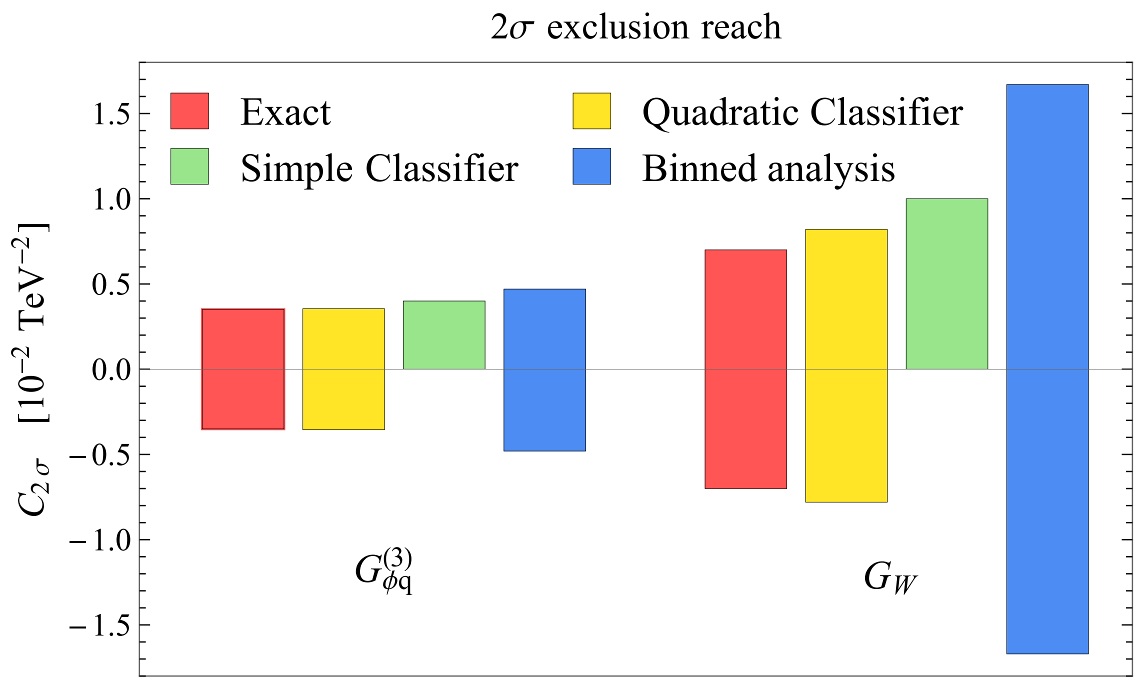
$$F(x;C) = \frac{1}{1 + (1 + C\,NN_\alpha(x))^2 + C^2\,NN_\beta^2(x)}$$

- training data must include different values of C

$$NN_\alpha(x) \rightarrow \alpha(x) \qquad\qquad NN_\beta(x) \rightarrow \beta(x)$$

The Quadratic Classifier

- The Quadratic Classifier provides a significantly better performance



Technical details

- fully connected feed-forward NN {10, 32, 32 32, 32, 1}
- ReLU activation
- 12 x 500k training events

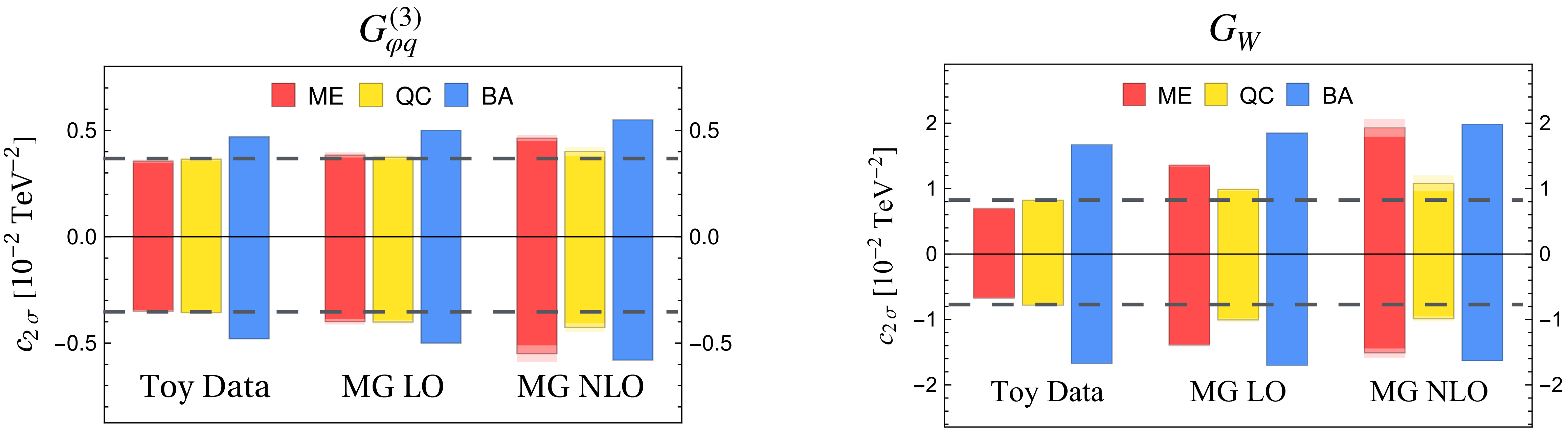
(Adam, 10^4 training epochs,
5 hours w/ pytorch
on NVIDIA GeForce GTX 1070)

- a single training can reconstruct the distribution ratio for any C
- training with "large" values of C avoids small differences from reference

→ limit $C \to 0$ properly reconstructed

The Quadratic Classifier

Additional effects can be included by changing the training data



- realistic Monte Carlo data can be used (eg. MadGraph @ LO or @ NLO)
- performance remains very stable

[for applications to LHC analyses
see talk by R. Schöfbeck]

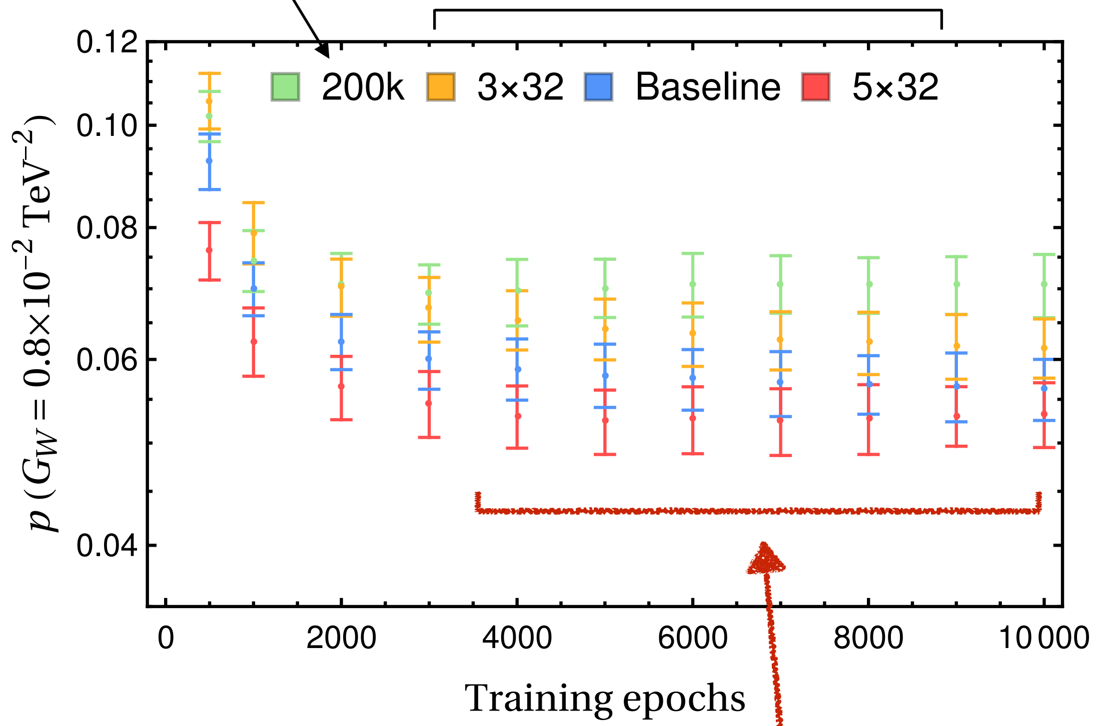
The Quadratic Classifier validation

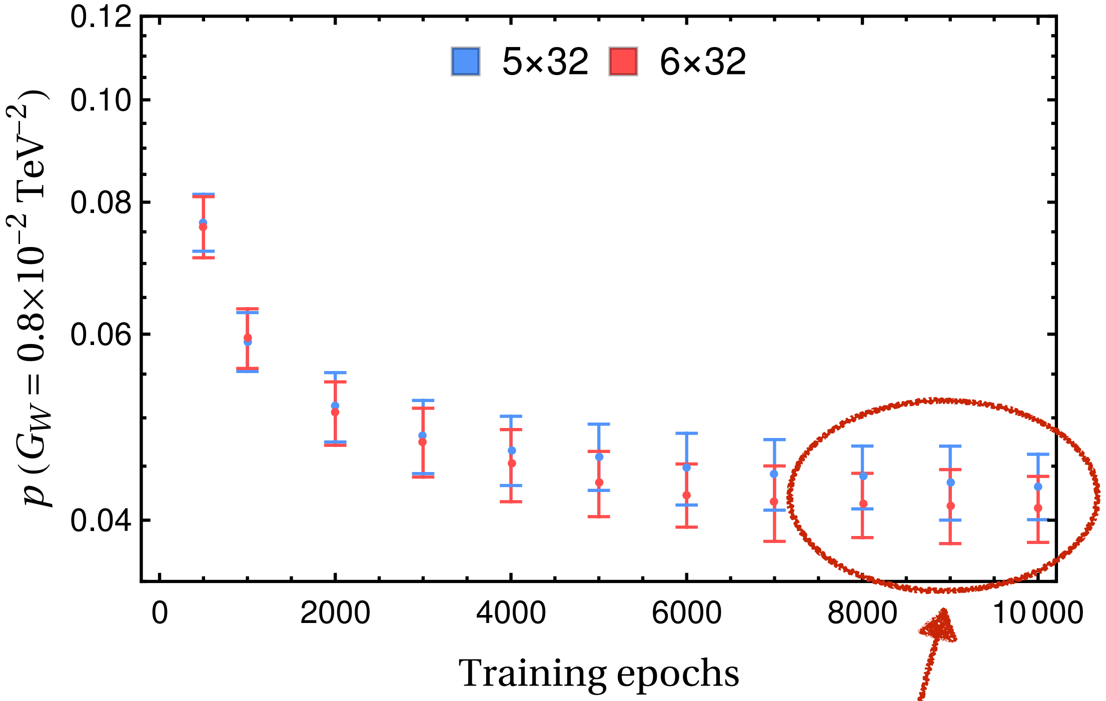
Validation for change in NN size and training data size

Baseline NN (4 × 32)
12 × 200k training events

12 × 500k training events

12 × 3M training events





good convergence

improvement in bounds
with larger training set /
bigger NN

Conclusions and Outlook

Conclusions

Machine learning provides new tools to optimize the sensitivity in model-independent new-physics searches at colliders

Key ingredient: reconstruction of differential distributions from MC data

‘Minimal’ ML approach: **Simple binary Classifier**

- fair performance
- some drawbacks (lack of embedded theory knowledge)

Improved ML approach: **Quadratic Classifier**

- directly embeds theory knowledge (analytic dependence on parameters)
- only one training needed to test different new-physics parameters

Outlook

Further developments:

- ✦ Simultaneous treatment of **many new-physics deformations**

  [see also talk by J. Ter Hoeve]

- ✦ Inclusion of **systematic errors** (eg. pdf errors)

- ✦ Exploitation of **event ‘reweighting’** to improve performance
  - faster generation of training data
  - better NN reconstruction (significantly smaller training sets needed)