How Good is the Standard Model?

Andrea Wulzer

ICREA

IFAE
Institut de Física d'Altes Energies

Based on:

D'Agnolo, AW, 2018
D'Agnolo, Grosso, Pierini, AW, Zanetti, 2019
D'Agnolo, Grosso, Pierini, AW, Zanetti, 2021
Letizia, Grosso, AW, et. al., 2022

Goodness of Fit

Statisticians formulate an interesting problem: **g.o.f.***

Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

***often question emerges after optimising distribution free parameters on the data, as a way to assess fit quality. But the problem is more general**

Goodness of Fit

Statisticians formulate an interesting problem: **g.o.f.**

Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

Answering is more **easy** the more **restrictive** assumptions we make on how the true distribution, if not R, can look like

But, more **partial** as well.

Goodness of Fit

Statisticians formulate an interesting problem: **g.o.f.**

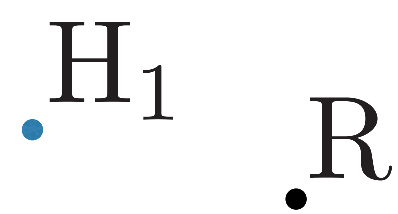
Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

Answering is more **easy** the more **restrictive** assumptions we make on how the true distribution, if not R, can look like

But, more **partial** as well.

Simple vs Simple hypothesis test



- Optimal approach provided by **Neyman–Pearson Lemma**
- Optimal answer to very specific question: **test has no or very limited power if truth ≠ H_1**

Goodness of Fit

Statisticians formulate an interesting problem: **g.o.f.**

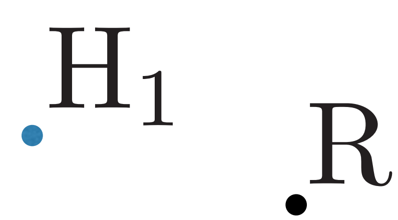
Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

Answering is more **easy** the more **restrictive** assumptions we make on how the true distribution, if not R, can look like

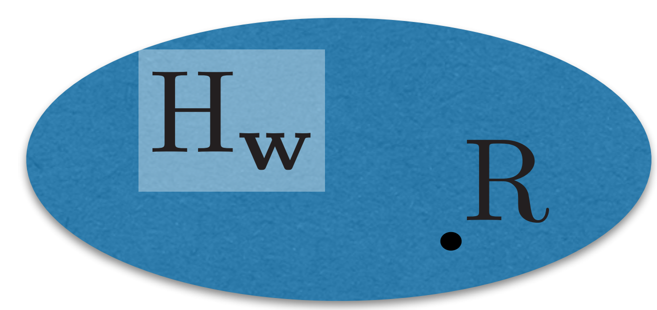
But, more **partial** as well.

Simple vs Simple hypothesis test



- Optimal approach provided by **Neyman–Pearson Lemma**
- Optimal answer to very specific question: **test has no or very limited power if truth ≠ H_1**

Simple vs Composite test



- No Optimal solution. But, **Likelihood Ratio** is **Good solution**
- Answers a more general question: **some power if truth is in H_w.** Generically, larger H_w = less power

The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

This is a tremendously hard gof problem!

BSM is tiny departure from SM, or large in tiny prob. region
Affecting few (unknown) observables over ∞ many we can measure

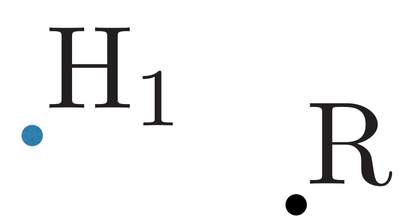
The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

This is a tremendously hard gof problem!

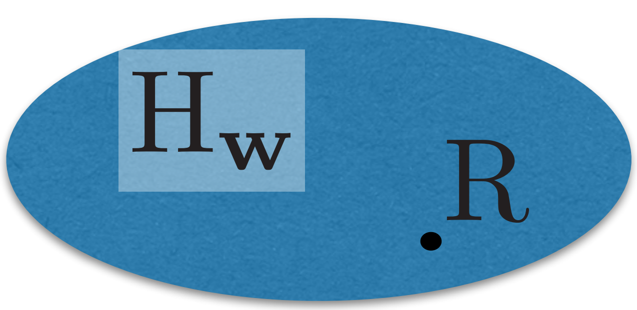
BSM is tiny departure from SM, or large in tiny prob. region
Affecting few (unknown) observables over ∞ many we can measure

Model-dependent
BSM searches



- Optimise sensitivity to **one specific BSM model**
- Fail to discover other models. **What if the right theoretical model is not yet formulated?**

Model-independent
searches

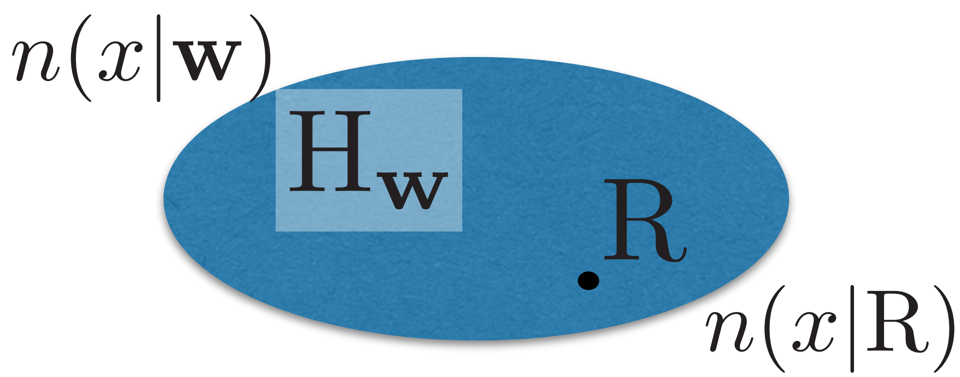


- Could reveal **truly unexpected** new physical laws.
- No hopes to find Optimal strategy. For a Good strategy, we need a **good choice of H_w.**

New Physics Learning Machine (NPLM)

Data: $\mathcal{D} = \{x_i\},\ i = 1, \ldots, N_{\mathcal{D}}$
I.i.d. measurements of, e.g., reconstructed particle momenta in a region of interest

$$n(x) = N\, P(x)$$
$$N = \int dx\, n(x)$$



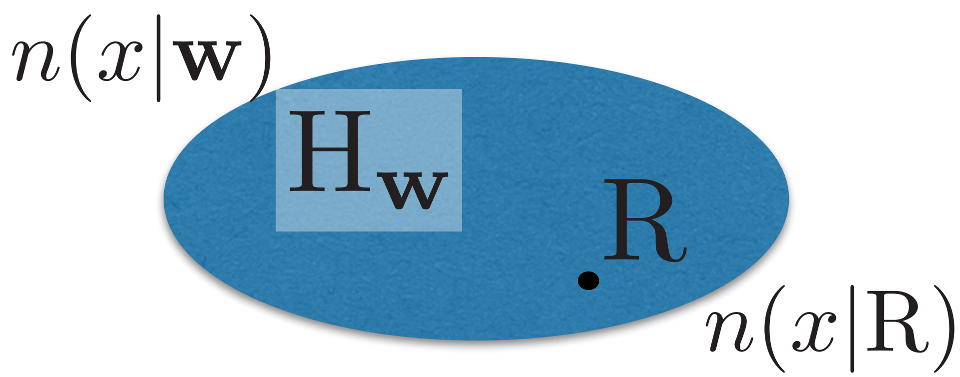
$$n(x|\mathbf{w}) = n(x|\mathrm{R})\, e^{f(x;\mathbf{w})}$$

$f(x;\mathrm{w})$ is **a neural network**, or other flexible functional approximant with good properties in many dimensions, like **kernels**

New Physics Learning Machine (NPLM)

Data: $\mathcal{D} = \{x_i\},\ i = 1, \dots, \mathcal{N}_{\mathcal{D}}$
I.i.d. measurements of, e.g., reconstructed particle momenta in a region of interest

$$n(x) = N\,P(x)$$
$$N = \int dx\, n(x)$$



$$n(x|\mathbf{w}) = n(x|\mathrm{R})\, e^{f(x;\mathbf{w})}$$

$f(x;\mathrm{w})$ is **a neural network**, or other flexible functional approximant with good properties in many dimensions, like **kernels**

Strategy is to evaluate the classical Likelihood Ratio test statistic

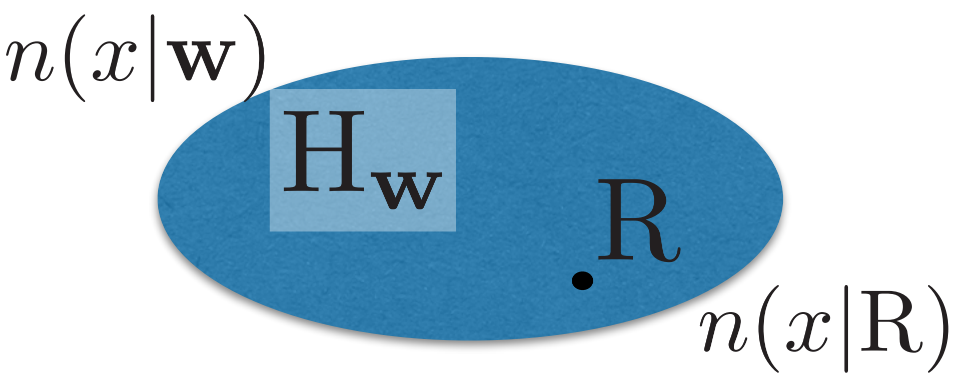
$$t(\mathcal{D}) = 2\,\log\frac{\max_{\mathbf{w}}[\mathcal{L}(\mathrm{H}_{\mathbf{w}}|\mathcal{D})]}{\mathcal{L}(\mathrm{R}|\mathcal{D})} = 2\,\max_{\mathbf{w}}\left\{\log\left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathrm{R})}}\prod_{i=1}^{\mathcal{N}_{\mathcal{D}}}\frac{n(x_i|\mathbf{w})}{n(x_i|\mathrm{R})}\right]\right\}$$

by **supervised training Data vs Reference** (background) sample.
Reference = artificial data distributed as predicted by the SM

New Physics Learning Machine (NPLM)

Data: $\mathcal{D}=\{x_i\},\ i=1,\ldots,\mathcal{N}_{\mathcal{D}}$
I.i.d. measurements of, e.g., reconstructed particle momenta in a region of interest

$$n(x)=N\,P(x)$$
$$N=\int dx\,n(x)$$



$$n(x|\mathbf{w})=n(x|\mathrm{R})\,e^{f(x;\mathbf{w})}$$

$f(x;\mathbf{w})$ is **a neural network**, or other flexible functional approximant with good properties in many dimensions, like **kernels**

Strategy is to evaluate the classical Likelihood Ratio test statistic

$$t(\mathcal{D})=2\,\log\frac{\max_{\mathbf{w}}[\mathcal{L}(\mathrm{H}_{\mathbf{w}}|\mathcal{D})]}{\mathcal{L}(\mathrm{R}|\mathcal{D})}=2\,\max_{\mathbf{w}}\left\{\log\left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathrm{R})}}\prod_{i=1}^{\mathcal{N}_{\mathcal{D}}}\frac{n(x_i|\mathbf{w})}{n(x_i|\mathrm{R})}\right]\right\}$$

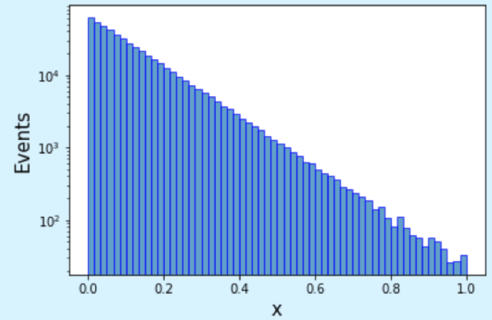
by **supervised training Data vs Reference** (background) sample.
Reference = artificial data distributed as predicted by the SM

By using a special loss function:

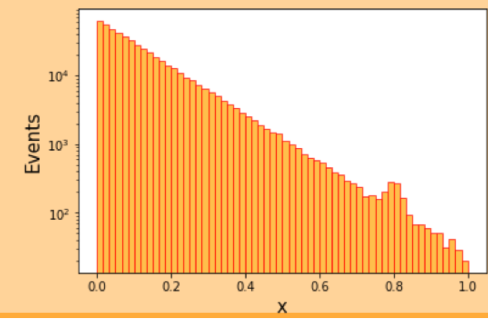
$$L[f]=\sum_{(x,y)}\left[(1-y)\frac{N(\mathrm{R})}{\mathcal{N}_{\mathcal{R}}}(e^{f(x)}-1)-y\,f(x)\right]\ \Rightarrow\ t(\mathcal{D})=-2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\,L[f(\,\cdot\,,\mathbf{w})]$$

INPUT

Reference sample (R)
label=0

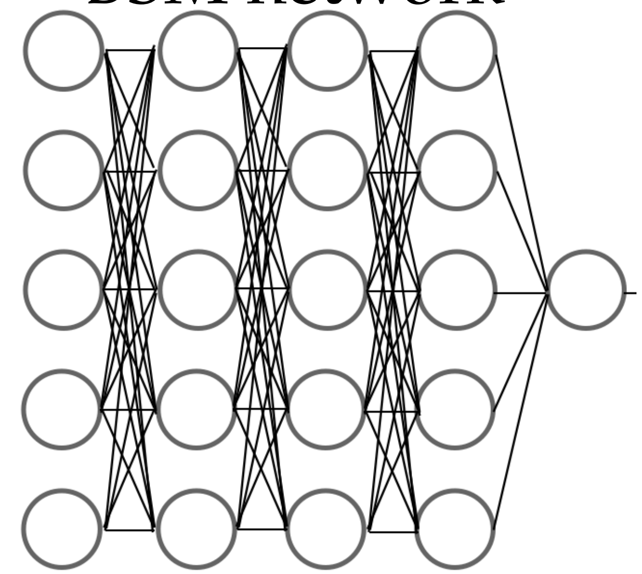


Data sample (D)
label=1



Unbinned training samples!

BSM network



NN training

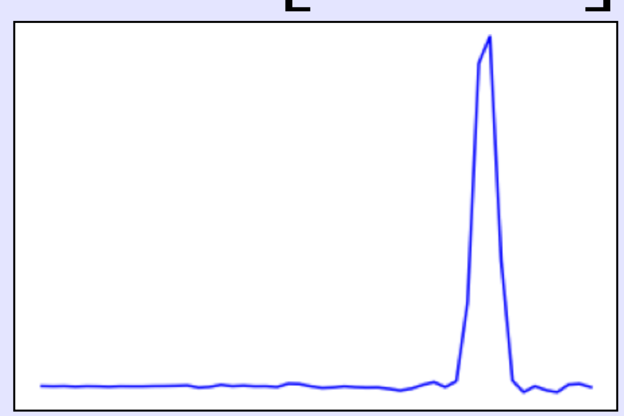
$\mathbf{w} \longrightarrow \hat{\mathbf{w}}$

OUTPUT

Single training

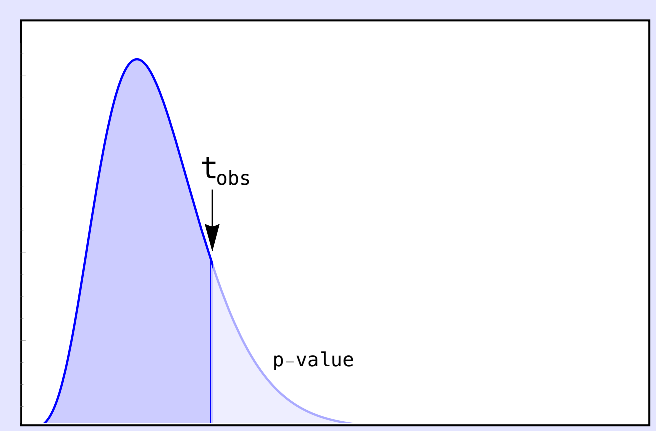
$$t(D) = -2\,L\left[f(x;\hat{\mathbf{w}})\right]$$

$$f(x;\hat{\mathbf{w}}) = \log\left[\frac{n(x\,|\,\mathrm{H}_{\hat{\mathbf{w}}})}{n(x\,|\,\mathrm{R}_0)}\right]$$



Many trainings (with pseudo-data)

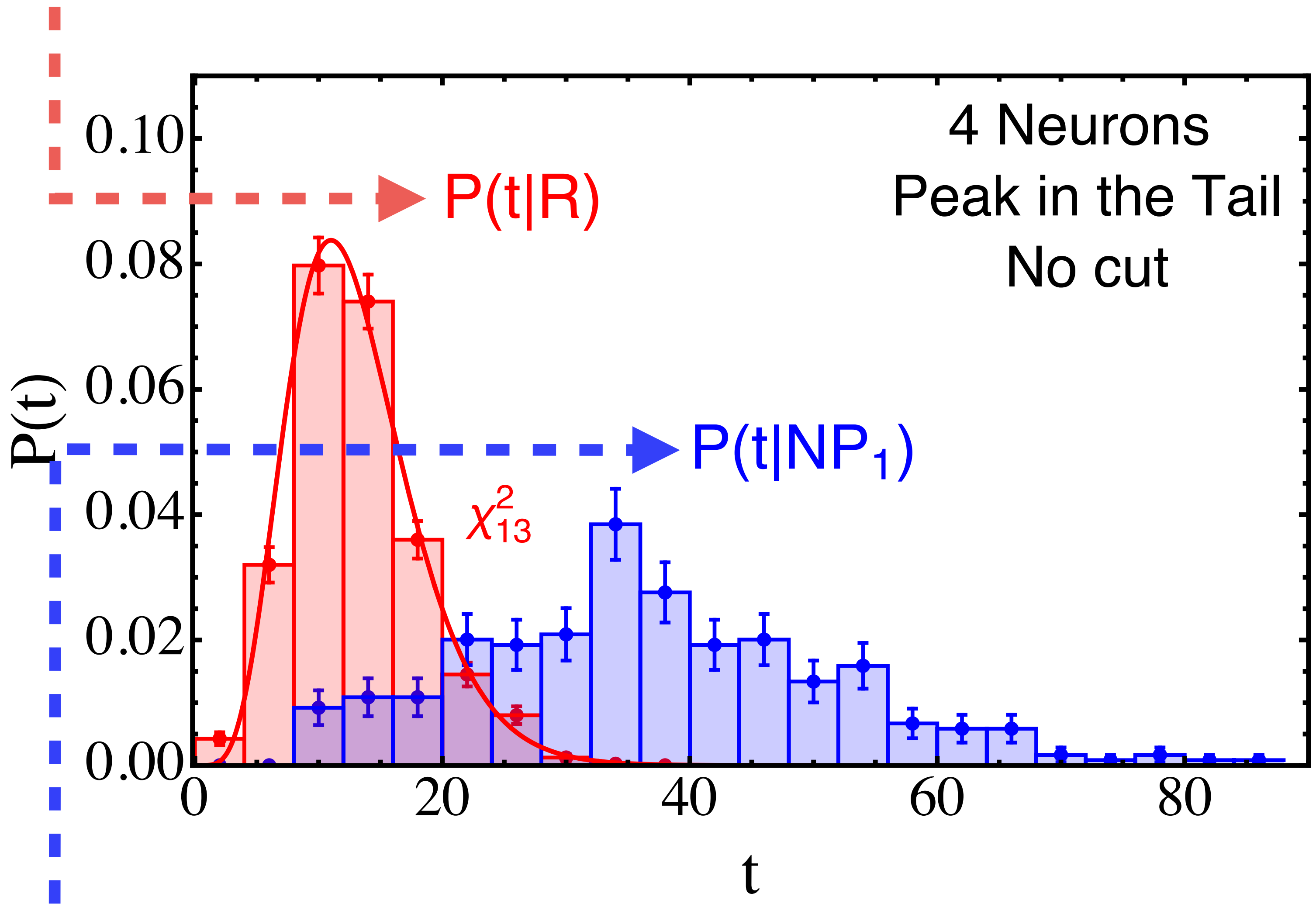
Empirical distribution of t
→ p-value for new datasets



Illustrating Performances

(Simple 1d example with exponential Reference)

Distribution of the test statistic “t” in Reference Hypothesis



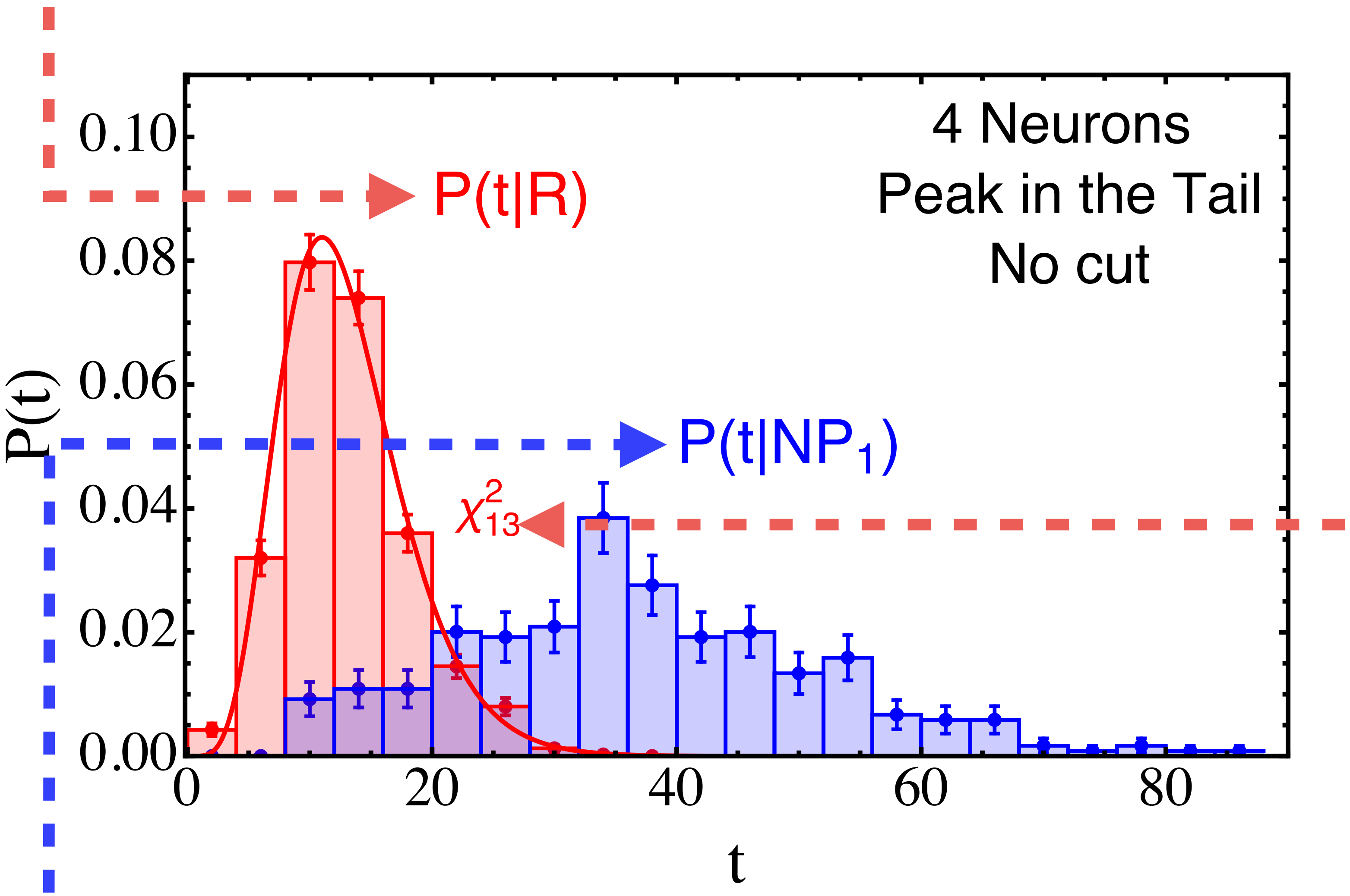
Distribution of “t” in one New Physics Model Hypothesis

t → p → Z-score (we use $Z = \Phi^{-1}(1-p)$)

Illustrating Performances

(Simple 1d example with exponential Reference)

Distribution of the test statistic "t" in Reference Hypothesis



Notice agreement with **Wilks' Formula:**

Sufficiently **regularised networks** found to **behave as if their number of d.o.f. was equal to number of parameters.**

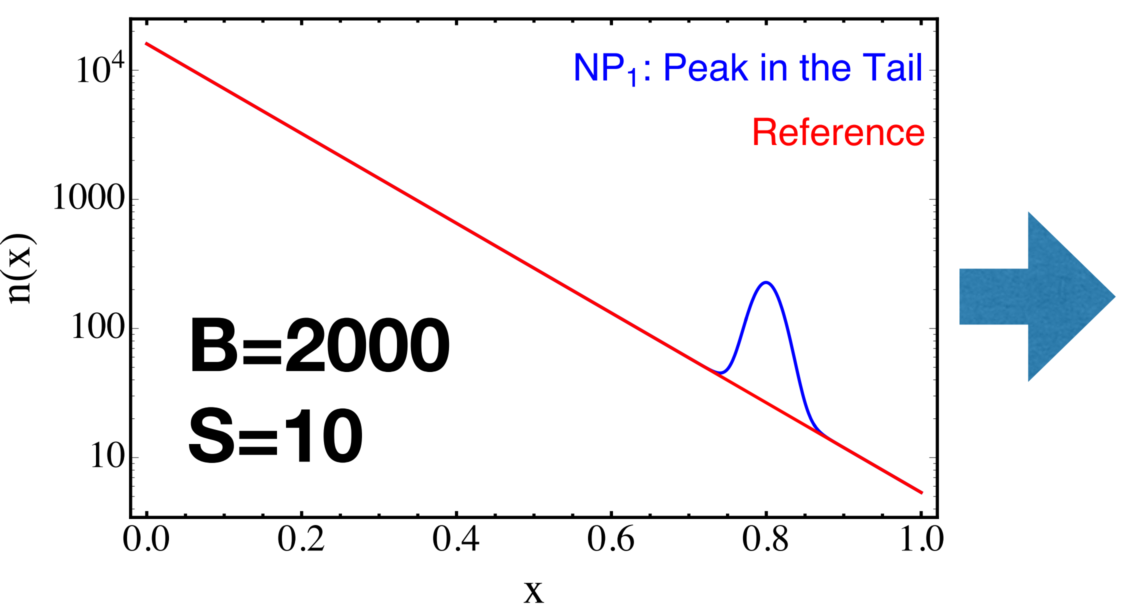
Theoretical reason mysterious

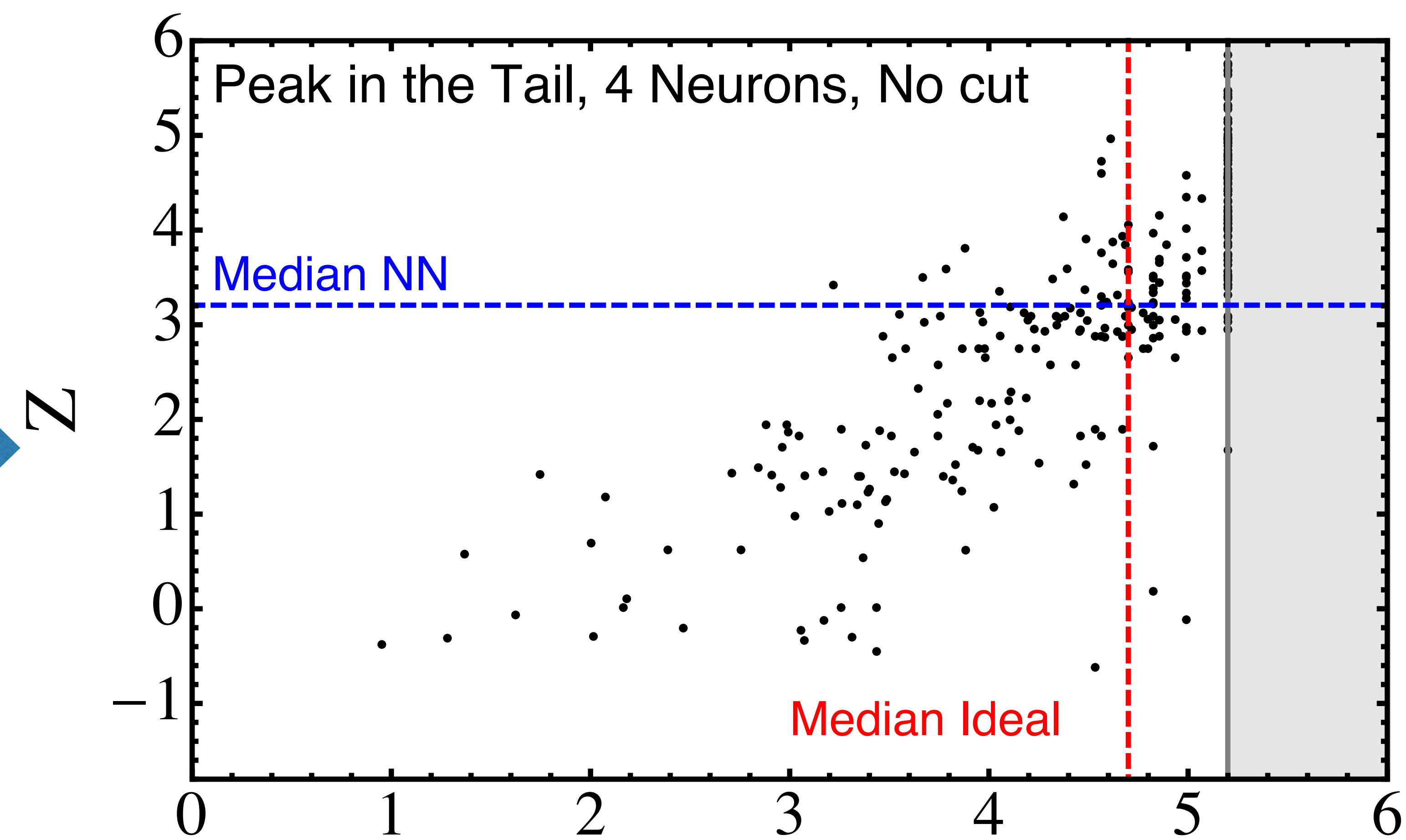
Distribution of "t" in one New Physics Model Hypothesis

t → p → Z-score (we use $Z = \Phi^{-1}(1-p)$)

Illustrating Performances

(Simple 1d example with exponential Reference)





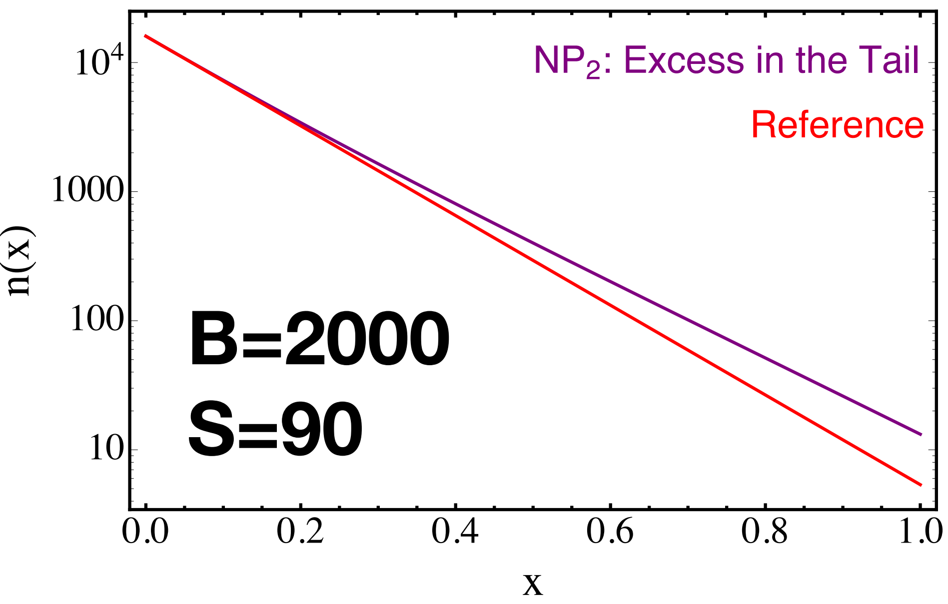
"Ideal Z-score": Z_{id}

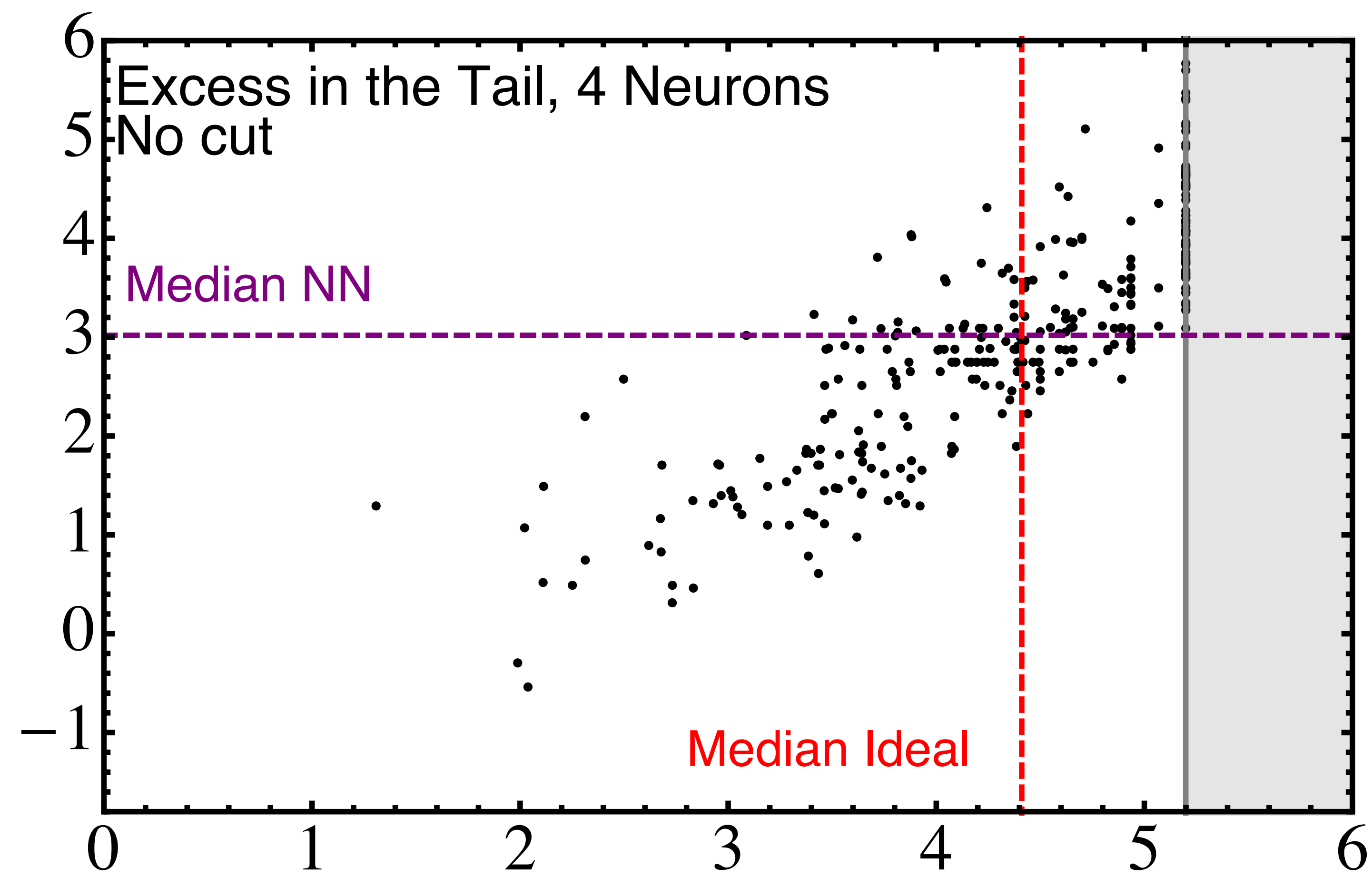
A "measure of dataset discrepancy"

(the Z-score of optimal test for NP1 model)

Illustrating Performances

(Simple 1d example with exponential Reference)





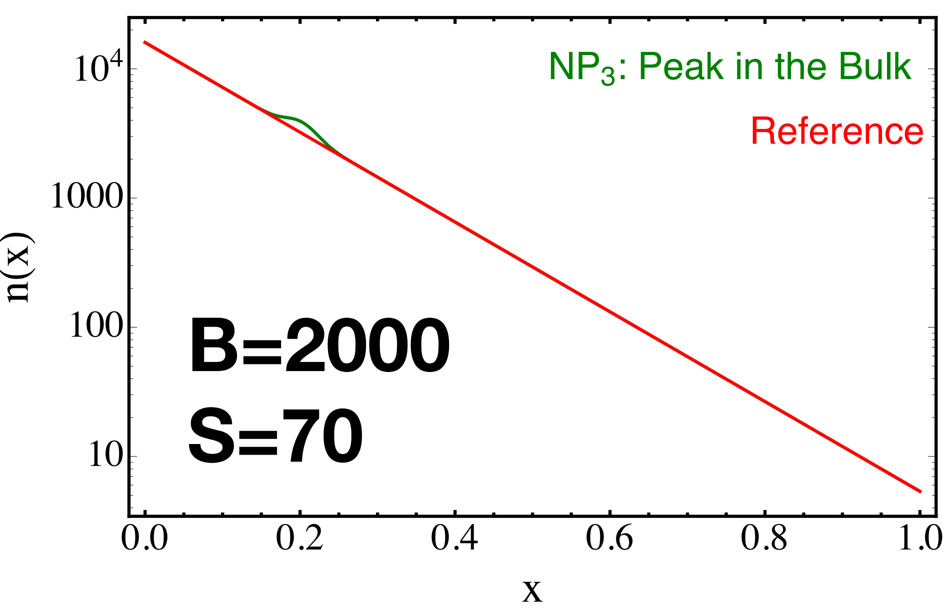
“Ideal Z-score”: $Z_{\rm id}$

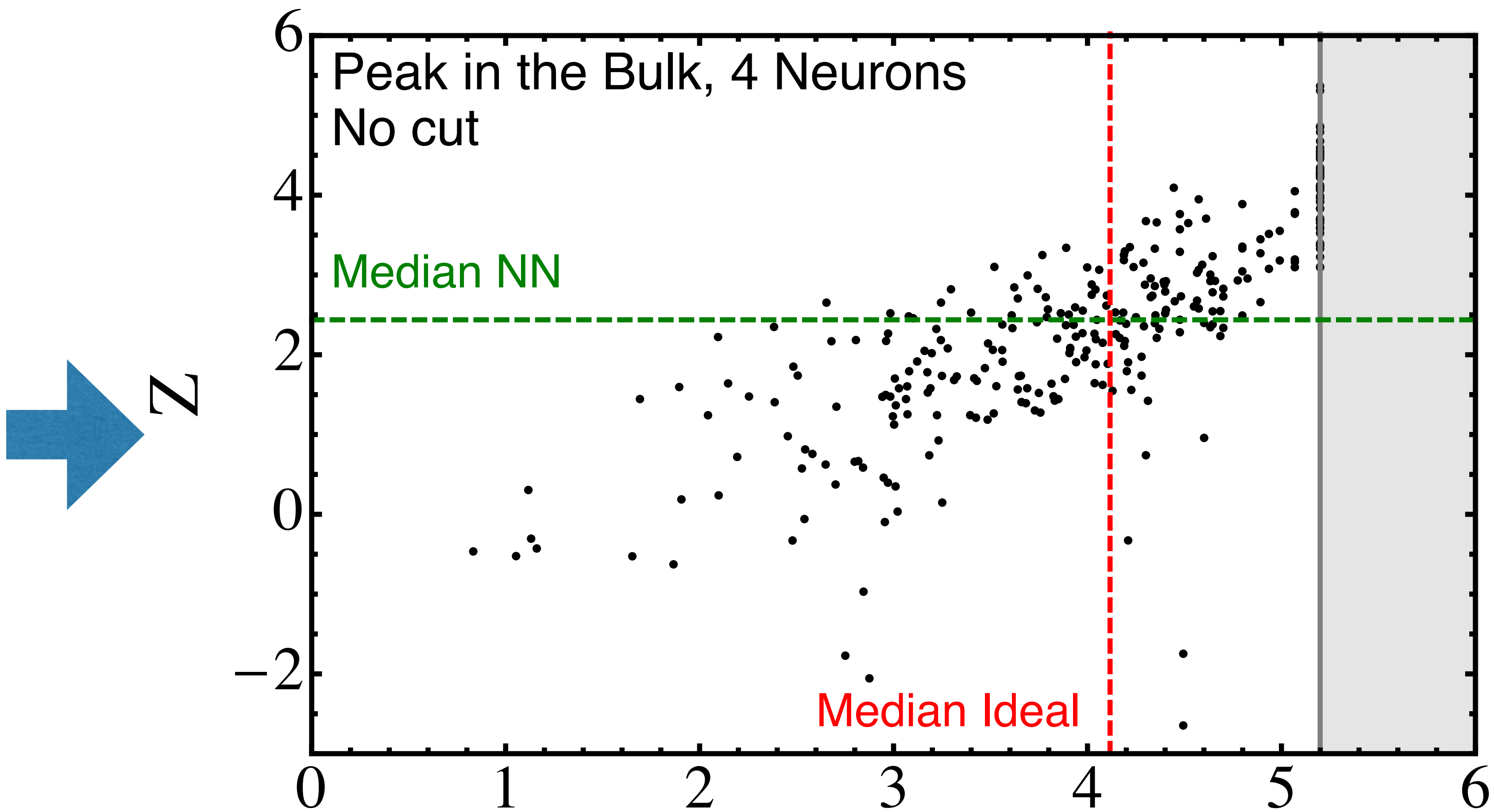
A “measure of dataset discrepancy”

(the Z-score of optimal test for NP2 model)

Illustrating Performances

(Simple 1d example with exponential Reference)





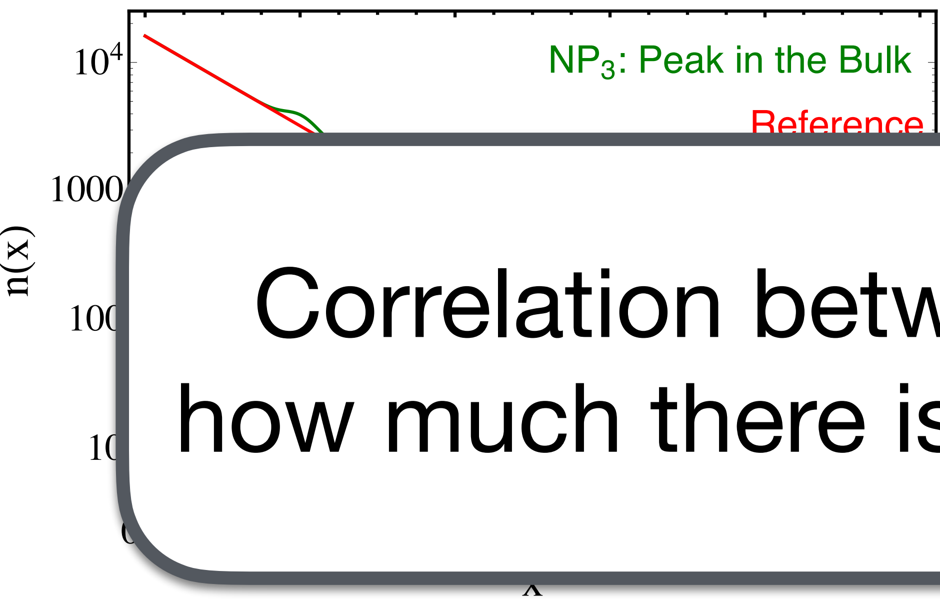
"Ideal Z-score": $Z_{\rm id}$

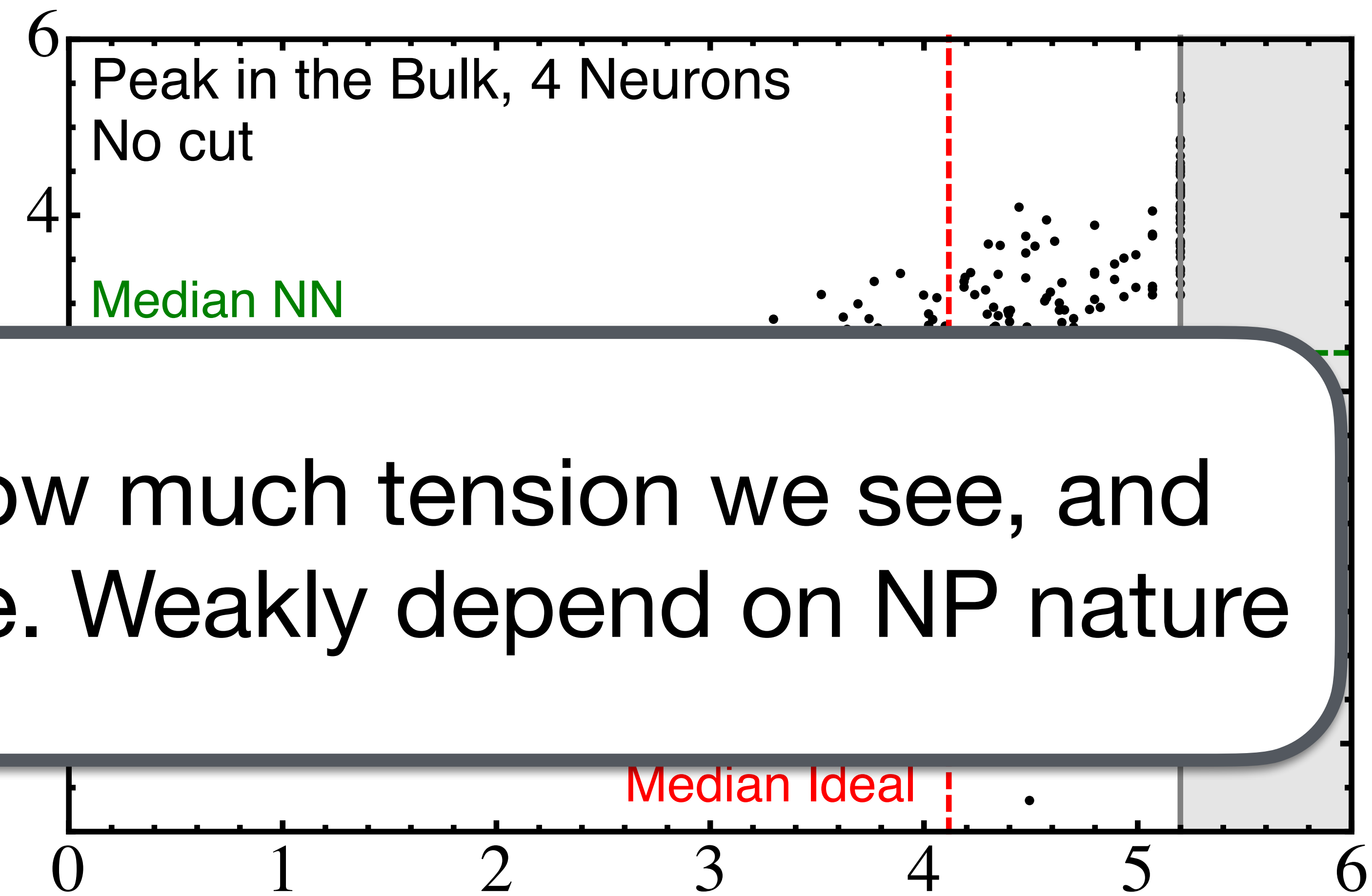
A "measure of dataset discrepancy"

(the Z-score of optimal test for NP3 model)

Illustrating Performances

(Simple 1d example with exponential Reference)





Correlation between how much tension we see, and how much there is to see. Weakly depend on NP nature

"Ideal Z-score": $Z_{\rm id}$

A "measure of dataset discrepancy"

(the Z-score of optimal test for NP3 model)

Imperfect Machine

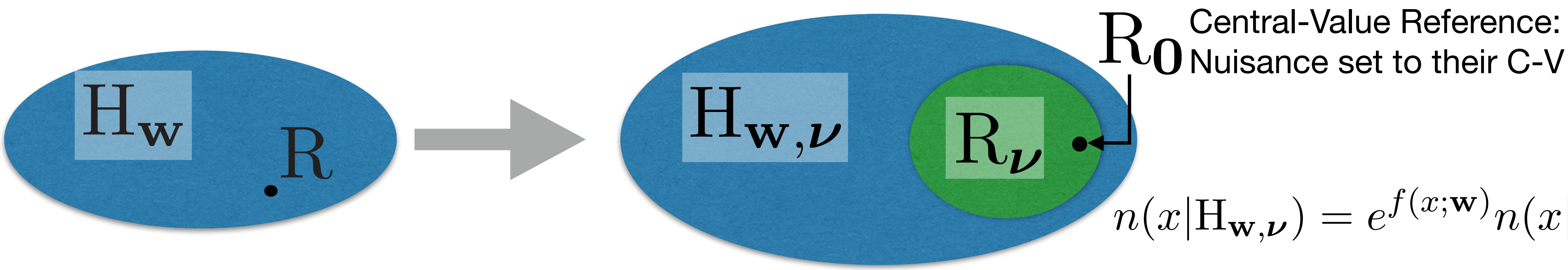
Reference Sample is an **imperfect** representation of SM

e.g., PDF/Lumi/Detector Modeling …

Imperfections are **Nuisance Parameters**

Constrained by **Auxiliary Measurements**

Define a **composite** Reference hypothesis



$$n(x|\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}) = e^{f(x;\mathbf{w})} n(x|\mathrm{R}_{\boldsymbol{\nu}})$$

Strategy conceptually unchanged. $t(\mathcal{D},\mathcal{A}) = 2\,\log \dfrac{\max\limits_{\mathbf{w},\boldsymbol{\nu}} \left[\mathcal{L}(\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}|\mathcal{D}) \cdot \mathcal{L}(\boldsymbol{\nu}|\mathcal{A})\right]}{\max\limits_{\boldsymbol{\nu}} \left[\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D}) \cdot \mathcal{L}(\boldsymbol{\nu}|\mathcal{A})\right]}$

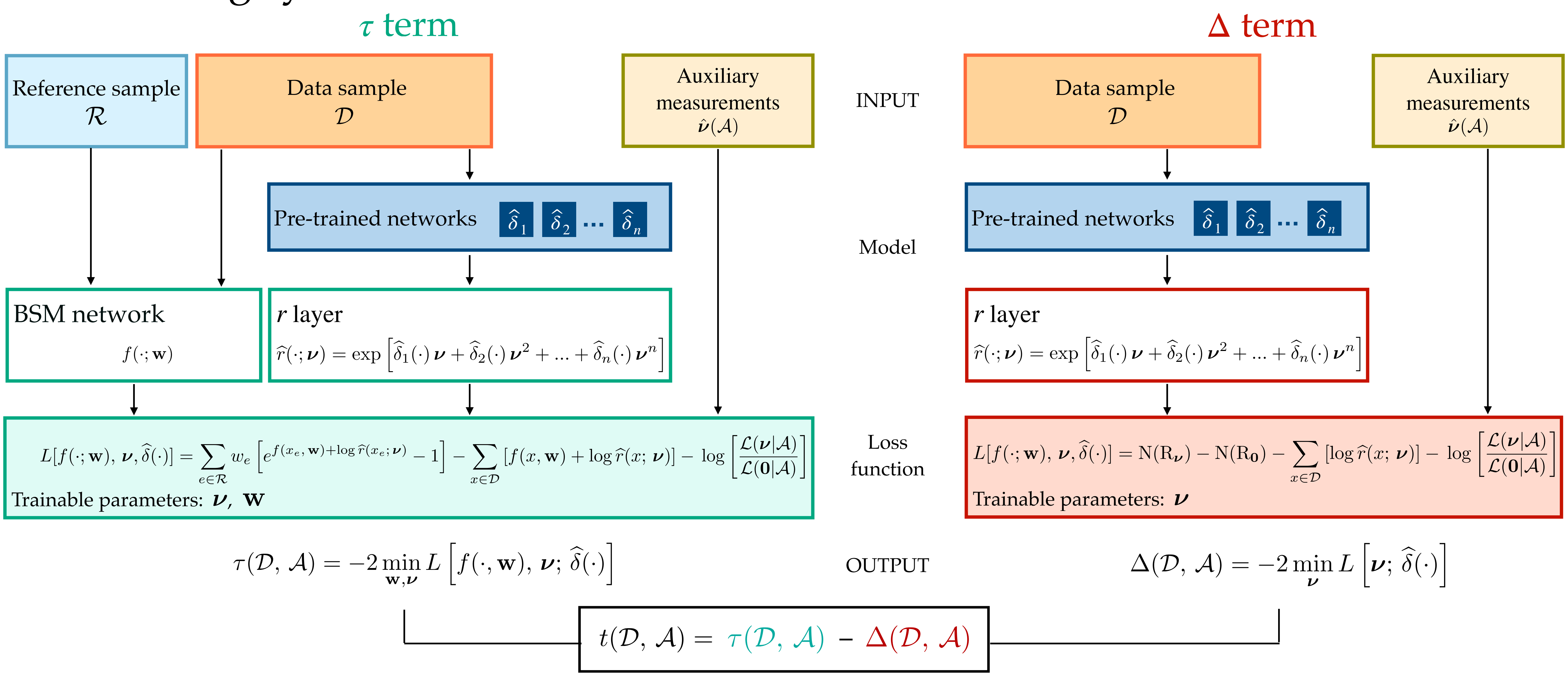
$$= 2\,\max_{\mathbf{w},\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{H}_{\mathbf{w},\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right] - 2\,\max_{\boldsymbol{\nu}} \log\left[\frac{\mathcal{L}(\mathrm{R}_{\boldsymbol{\nu}}|\mathcal{D})}{\mathcal{L}(\mathrm{R}_{\mathbf{0}}|\mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right] = \tau(\mathcal{D},\mathcal{A}) - \Delta(\mathcal{D},\mathcal{A})$$

Implementation slightly more complex

Imperfect Machine

New Physics Learning Machine (NPLM)

Including systematic uncertainties



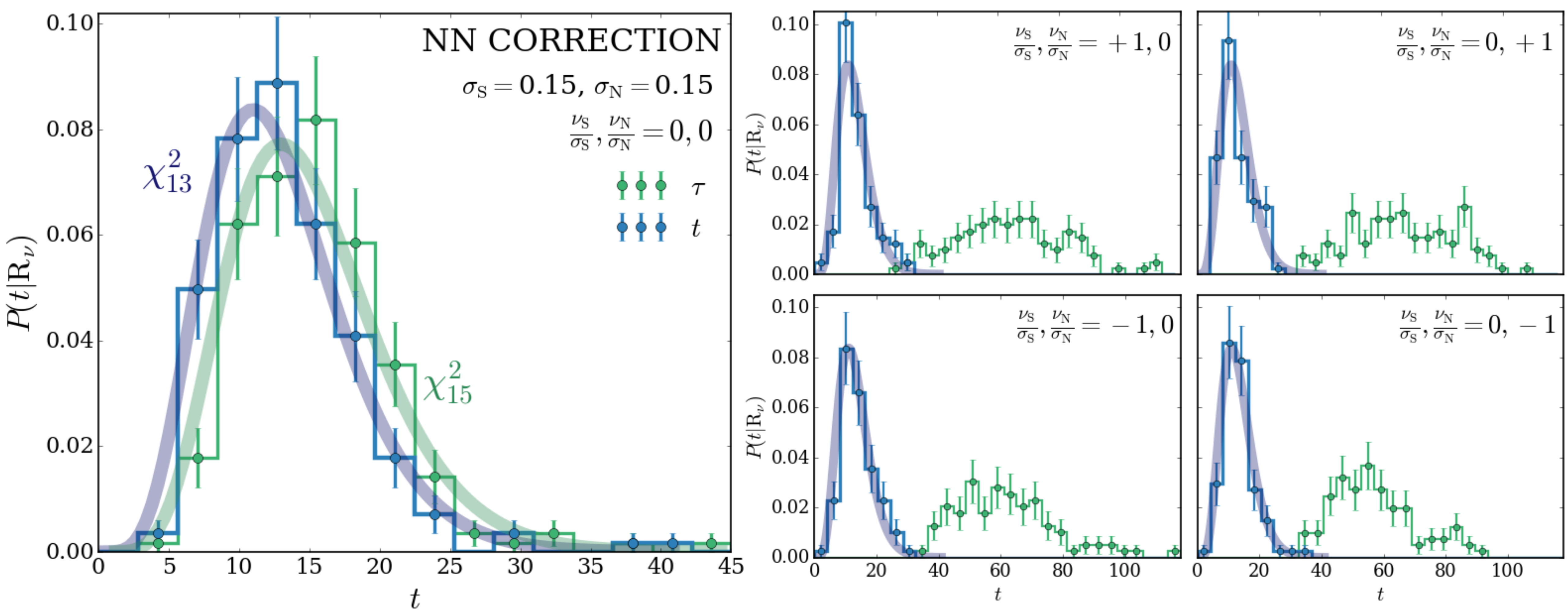
Courtesy of Gaia Grosso

An **Imperfect** Machine at Work

(Simple 1d example with exponential Reference)

Tau distribution distorted by non-central value nuisance

if not corrected, produces false positives



t = Tau-Delta independent of true nuisance value

this is essential for a feasible test

Towards LHC

Our proposed strategy is fully defined, including:

- Hyperparameters and regularisation selection
- Systematic approach to Reference mis-modelling

Validated on problems of realistic scale of complexity:

- 2-body final state with uncertainties (5D)
- ll+MET “SUSY” (8D)
- Heavy Higgs to WWbb (21D)

Towards LHC

Our proposed strategy is fully defined, including:

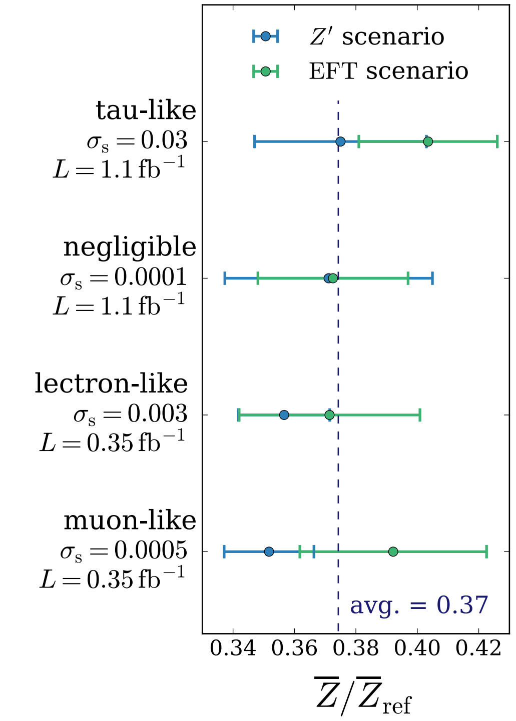
- Hyperparameters and regularisation selection
- Systematic approach to Reference mis-modelling

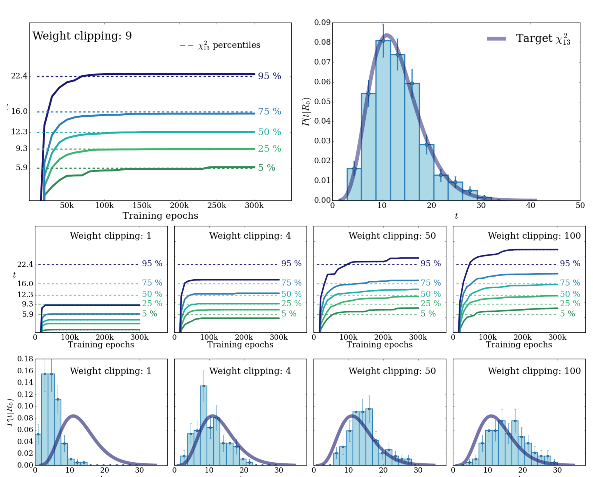
Validated on problems of realistic scale of complexity:

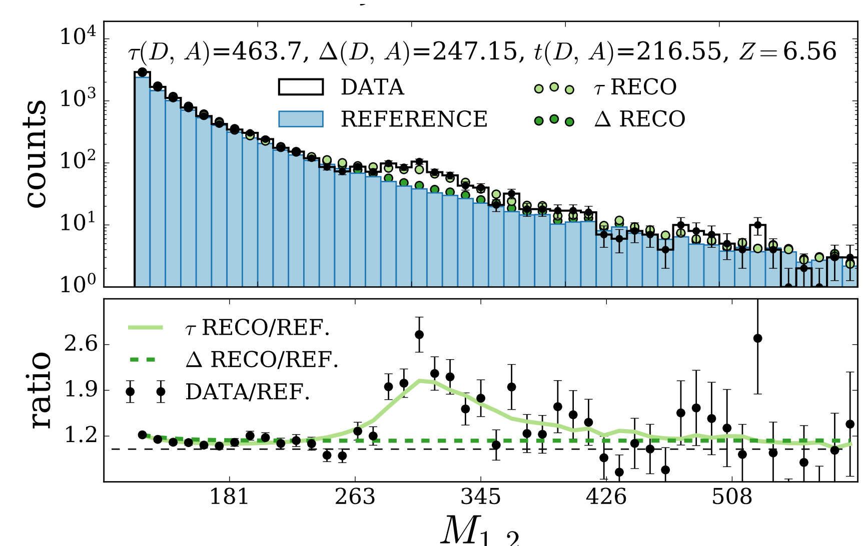
- 2-body final state with uncertainties (5D)
- ll+MET “SUSY” (8D)
- Heavy Higgs to WWbb (21D)

Results in summary:

- model-selection strategy converges
- sensitivity to resonant or non-resonant NP
- “uniform” response to NP of different nature
- trained network reconstruct NP







Outlook

Next step is **implementation** with true **LHC data.**

Open theoretical questions

- Why exactly we get chi-squared distributed “t”?
- Regularisation selects space of alternatives, where we are looking for NP
  A principled approach to regularisation and “reasonable” alternatives?
- …

Outlook

Next step is **implementation** with true **LHC data.**

Open theoretical questions

- Why exactly we get chi-squared distributed “t”?
- Regularisation selects space of alternatives, where we are looking for NP
  A principled approach to regularisation and “reasonable” alternatives?
- ...

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation/DQM
- Other GoF problems

First Real-Life Application?

[Grosso, Lai, Letizia, Pazzini, Rando, Wulzer, Zanetti, to appear]

*n*D DQM

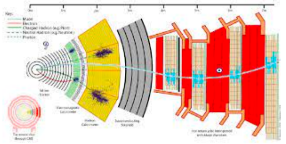
Online monitoring of a DT chamber:

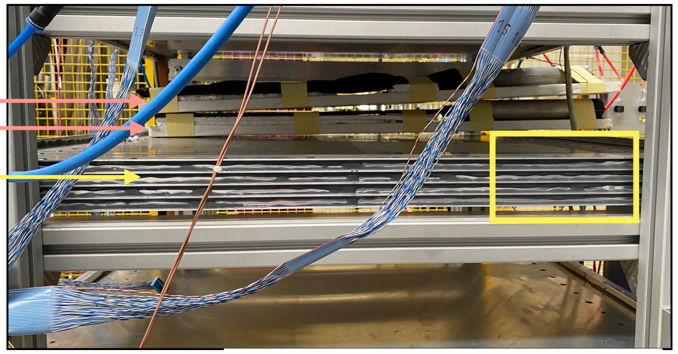
Setup (Legnaro INFN national laboratory):

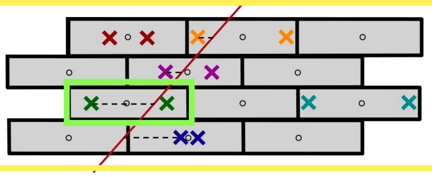
- 2 scintillators as signal trigger
- 1 drift tube chamber: 4 layers 16 wires each (16x4=64 wires)
- Source of signals: cosmic muons (triggered rate ~3 MHz)
- **Event**: muon track reconstructed interpolating 3/4 hits (one per layer)

Observables (6D problem):

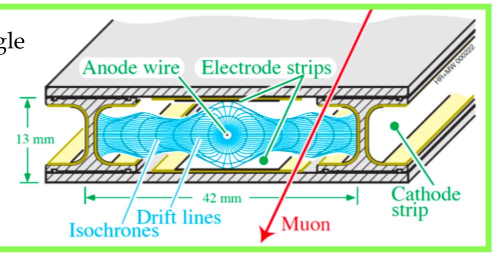
- 4 drift times $[t_{\text{drift},1}, t_{\text{drift},2}, t_{\text{drift},3}, t_{\text{drift},4}]$: time for the ionised electrons to reach the wire from the interaction point (v_{drift} = cm/s) .
- θ: reconstructed track angle
- N_{hits}: average number of hits per time window (“orbit”)







Sketch of a single chamber



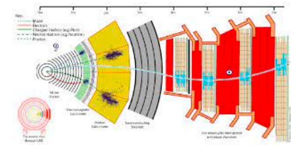
Dipartimento di Fisica e Astronomia Galileo Galilei
UNIVERSITÀ DEGLI STUDI DI PADOVA
UniGe | MaLGa

First Real-Life Application?

[Grosso, Lai, Letizia, Pazzini, Rando, Wulzer, Zanetti, to appear]

nD DQM

Online monitoring of a DT chamber:



Setup (Legnaro INFN national laboratory):

- 2 scintillators as signal trigger
- 1 drift tube chamber: 4 layers 16 wires each (16x4=64 wires)
- Source of signals: cosmic muons (triggered rate ~3 MHz)
- **Event**: muon track reconstructed interpolating 3/4 hits (one per layer)

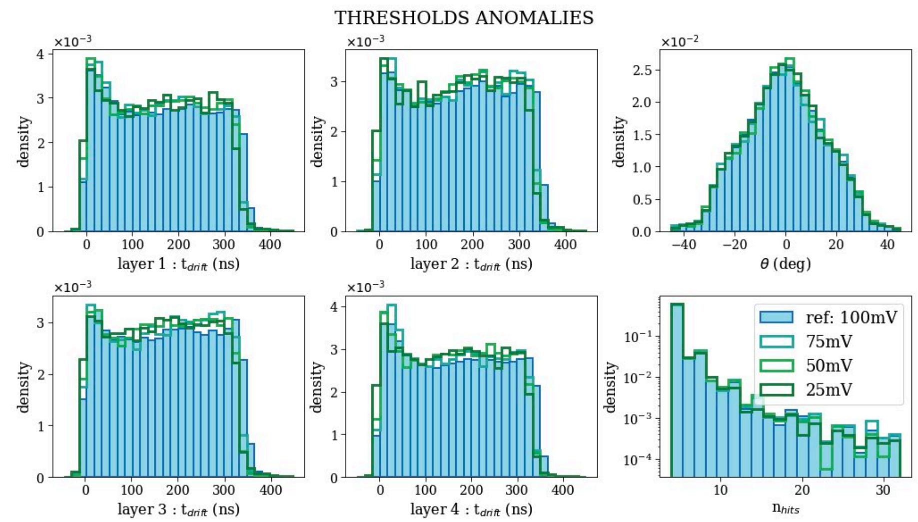
Observables (6D problem):

- 4 drift times [$t_{\text{drift},1}, t_{\text{drift},2}, t_{\text{drift},}$ electrons to reach the wire fror ($v_{\text{drift}} =$ cm/s) .
- θ: reconstructed track angle
- N_{hits}: average number of hits p

nD DQM

Online monitoring of a DT chamber:

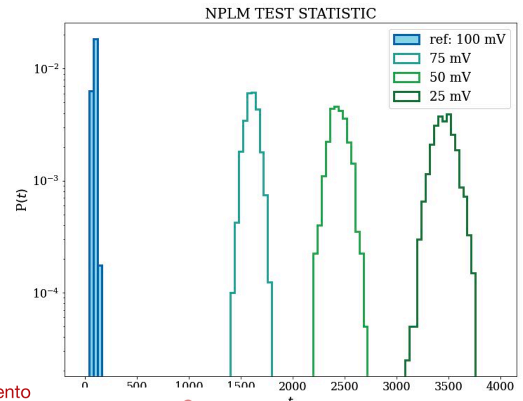
- **Reference sample:** long run in optimal conditions
- **Anomalous samples**: short runs acquired in presence of a controlled anomaly in the value of the **threshold tension** of the DT chamber



Distribution of the observables at different values of the threshold tension

- Result of the test statistics

Complete separation of the distributions!



NPLM with Falkon
$M = 50, \sigma = 4.84, \lambda = 10^{-7}$
N(D) = 5000
$N_{\text{ref}} = 200\,000$
Execution time: ~ 1.5 s

Outlook

Next step is **implementation** with true **LHC data.**

Open theoretical questions

- Why exactly we get chi-squared distributed “t”?
- Regularisation selects space of alternatives, where we are looking for NP
  A principled approach to regularisation and “reasonable” alternatives?
- …

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation/DQM
- Other GoF problems

When these techniques applied to real analyses, if truly powerful, we will discover mis-modelled backgrounds.

Outlook

Next step is **implementation** with true **LHC data.**

Open theoretical questions

- Why exactly we get chi-squared distributed “t”?
- Regularisation selects space of alternatives, where we are looking for NP
  A principled approach to regularisation and “reasonable” alternatives?
- ...

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation/DQM
- Other GoF problems

When these techniques applied to real analyses, if truly powerful, we will discover mis-modelled backgrounds.

But, maybe, New Physics as well !!