ttH(bb) at CMS

Joosep Pata
supervised by Prof. G. Dissertori,
G. Kasieczka(1) and L. Bianchini (2)

HiggsTools Final Meeting
September 12, 2017

1. now at Hamburg
2. now at Pisa

Outline

- What makes ttH(bb) interesting?
- Results from CMS and discussion
- Status and ongoing work
- **additional**: what I did at Lingvist Technologies

ETH Zürich

higgstools

ttH(bb) motivation

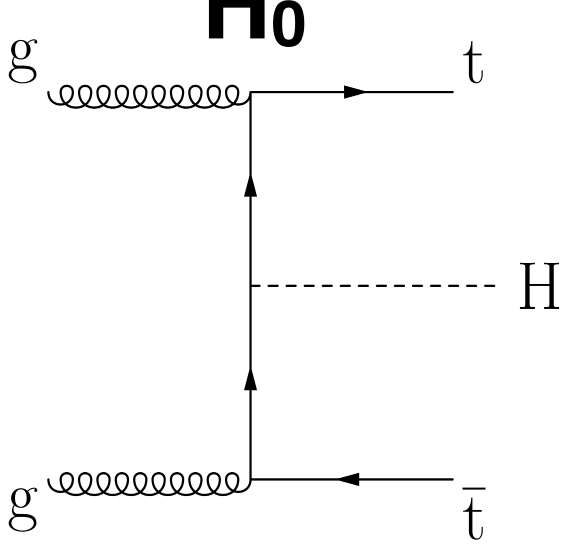
- Large top mass → Yukawa coupling y_t near unity
- Measure H⇔t coupling directly instead of through loops
- H→bb has the highest branching ratio
- **Goal:** measure ttH production cross-section in Run II
- ... and thus constrain the fundamental SM parameter y_t

Challenges

- large background: mainly tt+jets production
    - σ(tt+jets) = **830 pb** vs. σ(tt+Hbb) = **0.3 pb**
- **heavy flavour (tt+bb)**: irreducible, multiscale process, difficulties in modelling, combinatorial self-background for H(bb)

both have the same final state

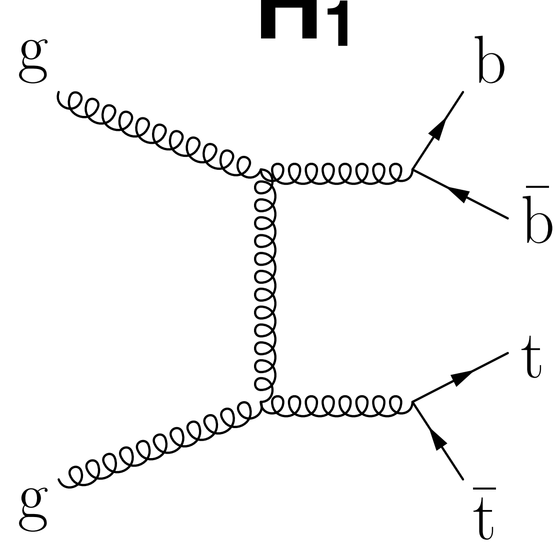
H_0



signal

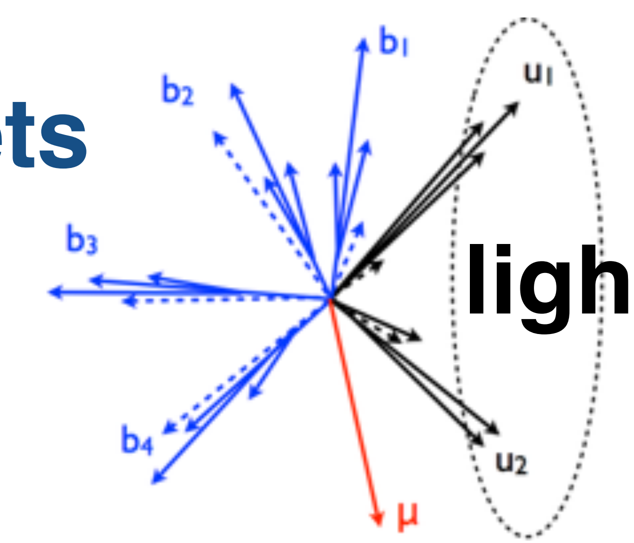
or

H_1



background

b-jets

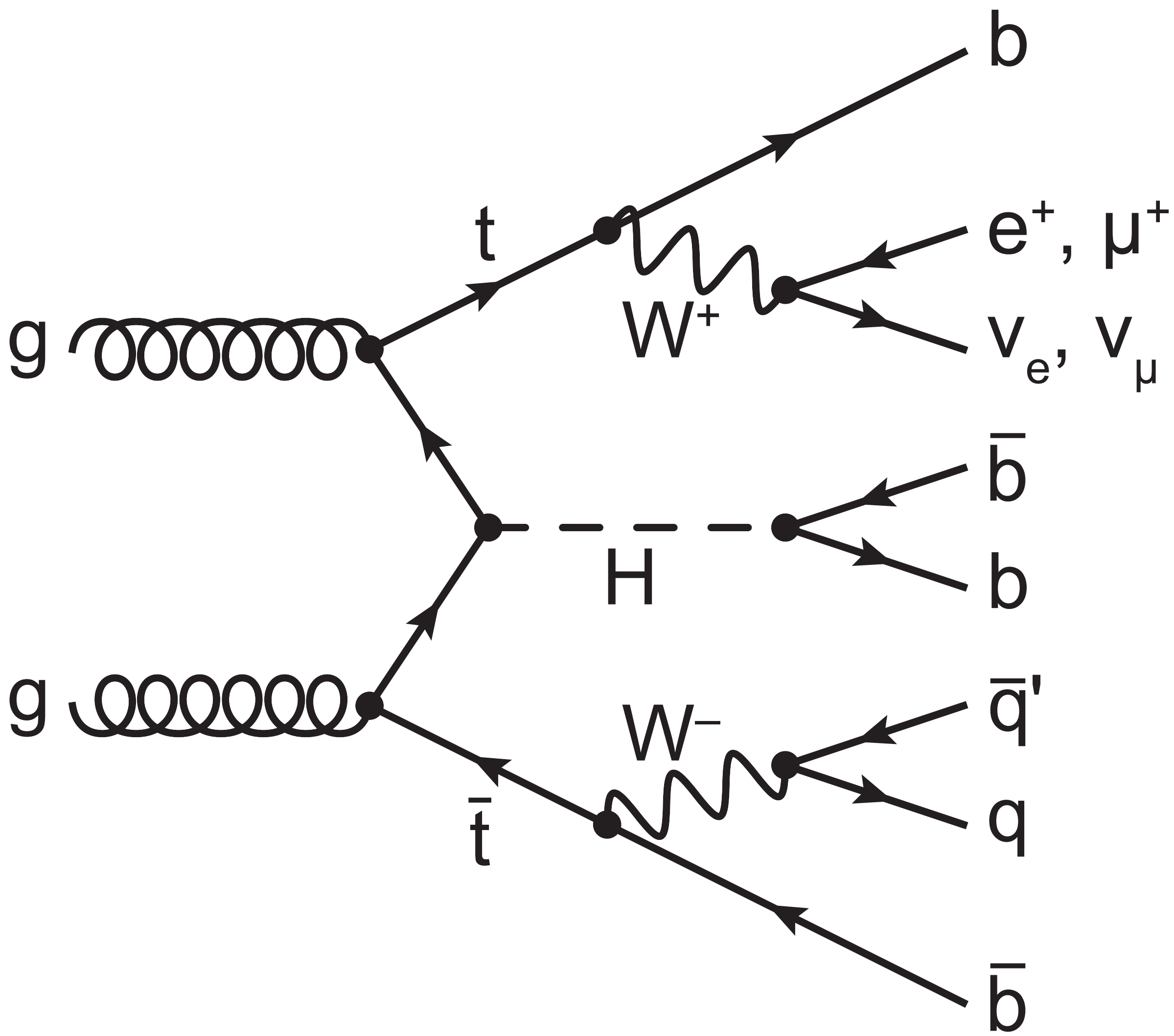


light jets

lepton

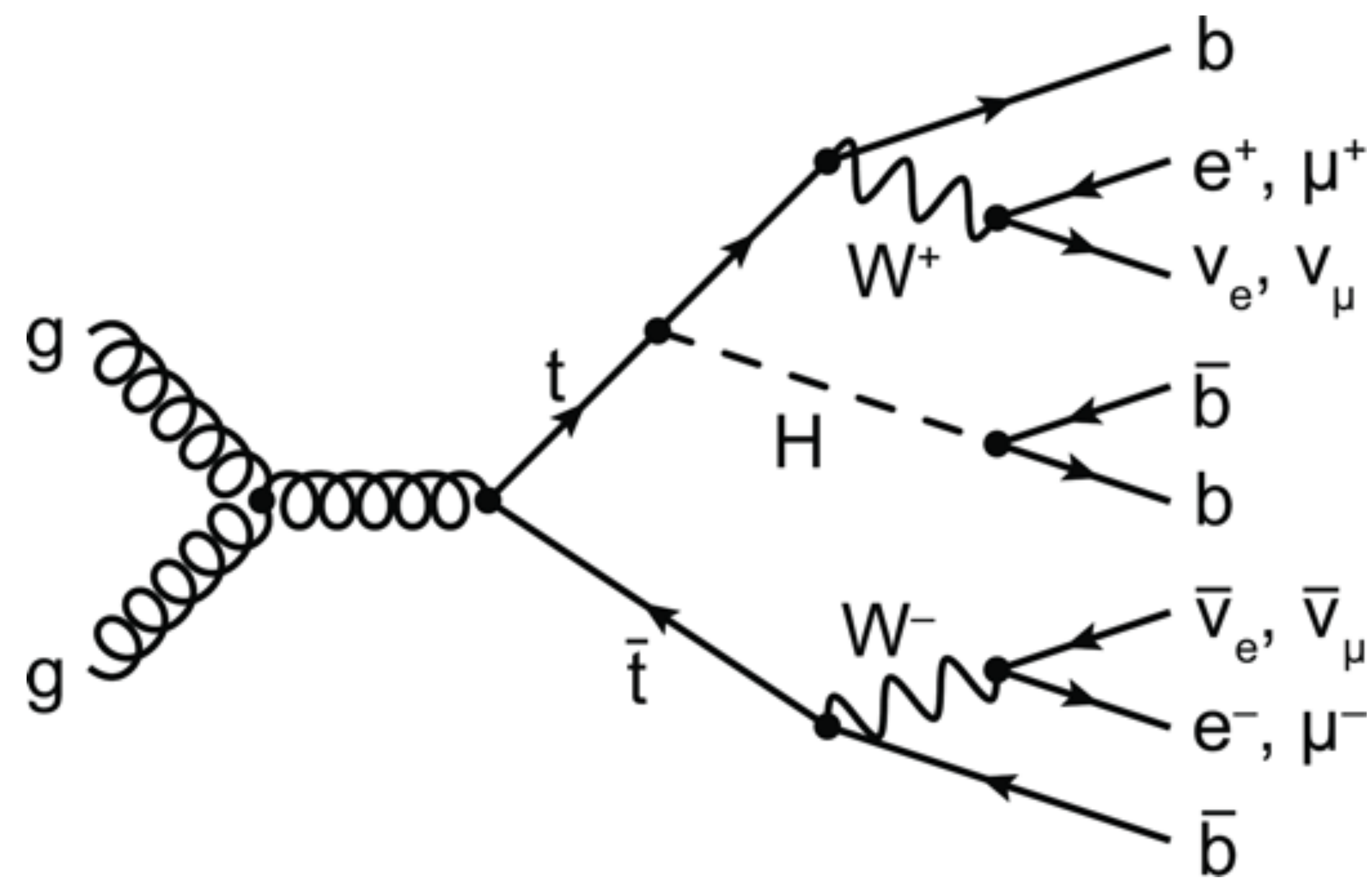
Final state

semileptonic



1 lepton, 6 jets, 4 b-tags

dileptonic



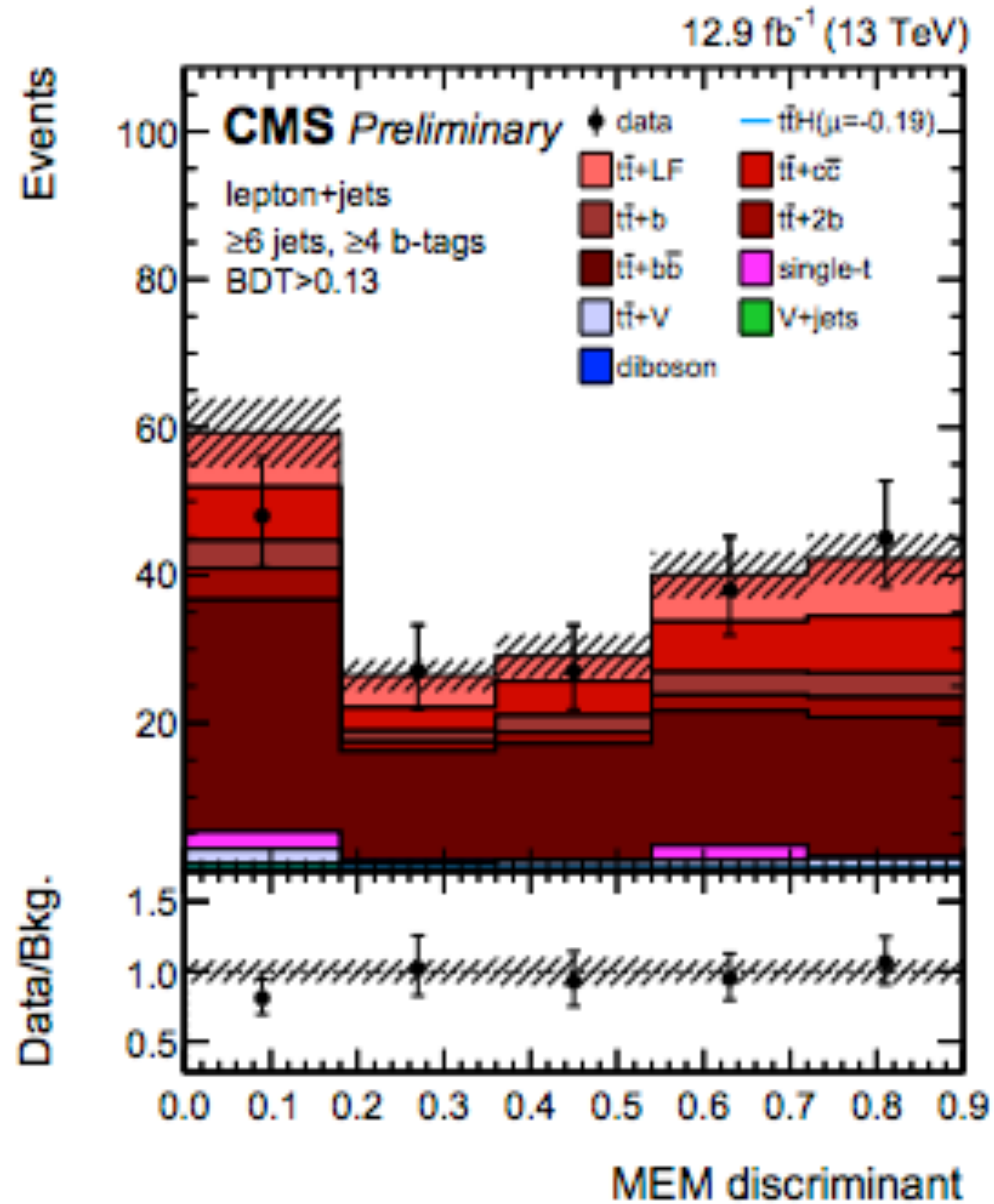
2 leptons, 4 jets, 4 b-tags

Complicated multi-particle final state!

No clear resonant H(bb) peak.

Key points

- tt→dilepton, lepton+jets topologies, ~10 jet/b-tag multiplicity categories
- 4b final state: extensive b-tagging
- Major experimental uncertainties: jet energy corrections, b-tagging
- Theoretical uncertainty: tt+bb (heavy flavour) modelling!
- use multivariate classifiers against tt+bb: MEM, BDT, DNN...



MEM

Directly compute theory-motivated event likelihood from kinematics ***y*** only.

Observed ***y*** under hypothesis ***θ***.

$$P(\boldsymbol{y},\boldsymbol{\theta}) = \sum_{k=1}^{N_a} \int \frac{\mathrm{d}x_1 \mathrm{d}x_2}{2 x_1 x_2 s} \int \prod_{i=1}^{n} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i}$$
$$\times\, \delta^4\left(q_1 + q_2 - \sum_{i=1}^{n} p_i\right)$$
$$\times\, g(x_1) g(x_2)$$
$$\times\, \mathcal{R}(\tilde{\boldsymbol{\rho}}_T, \boldsymbol{\rho}_T)$$
$$\times\, |\mathcal{M}_{\boldsymbol{\theta}}(q_1, q_2, p_1, \ldots, p_n)|^2$$
$$\times\, W(\boldsymbol{y}, \boldsymbol{p})$$

all jet-to-parton association combinations

integrate over parton-level momenta ***p*** (unlike MELA)

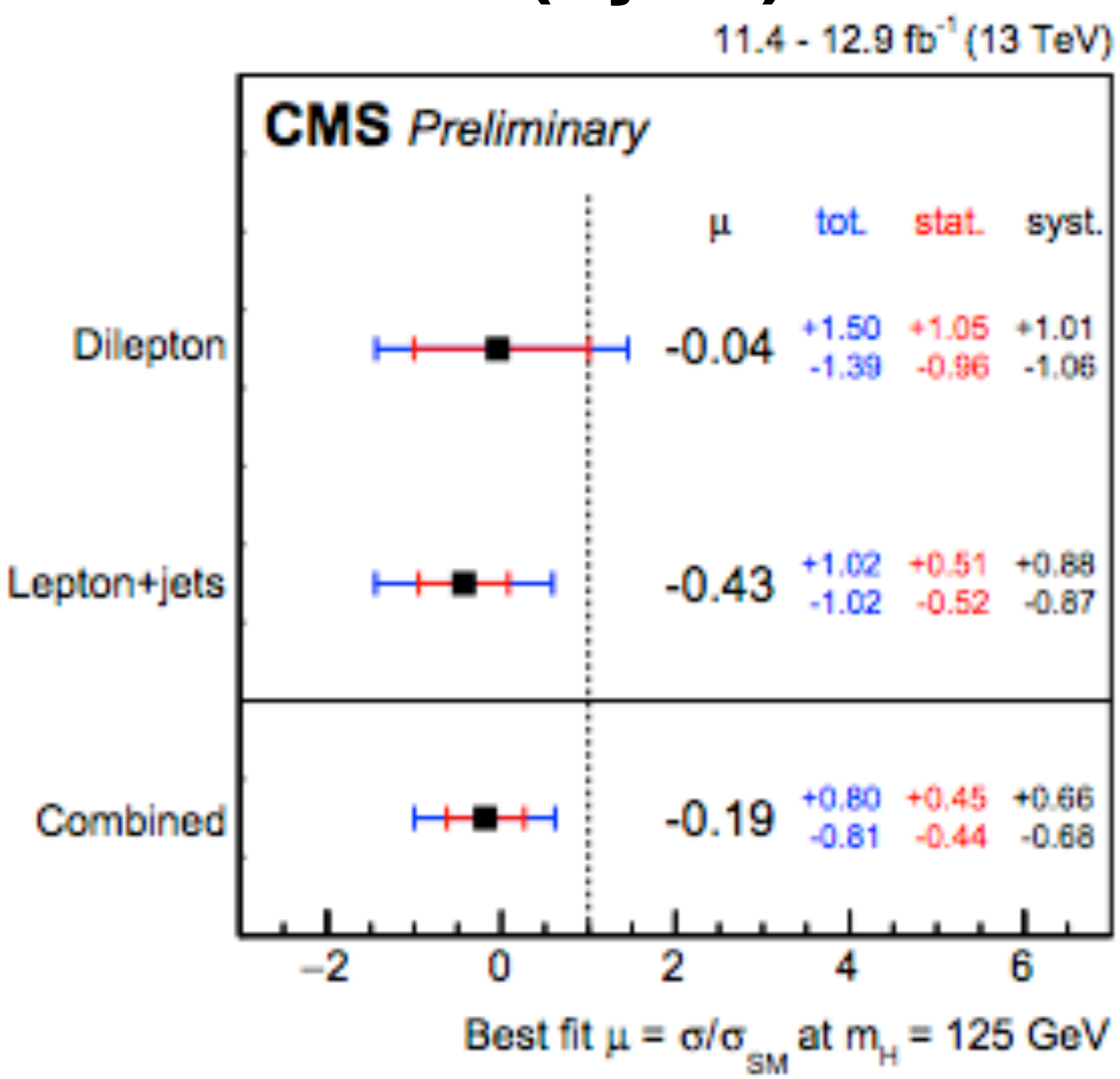
detector transfer function

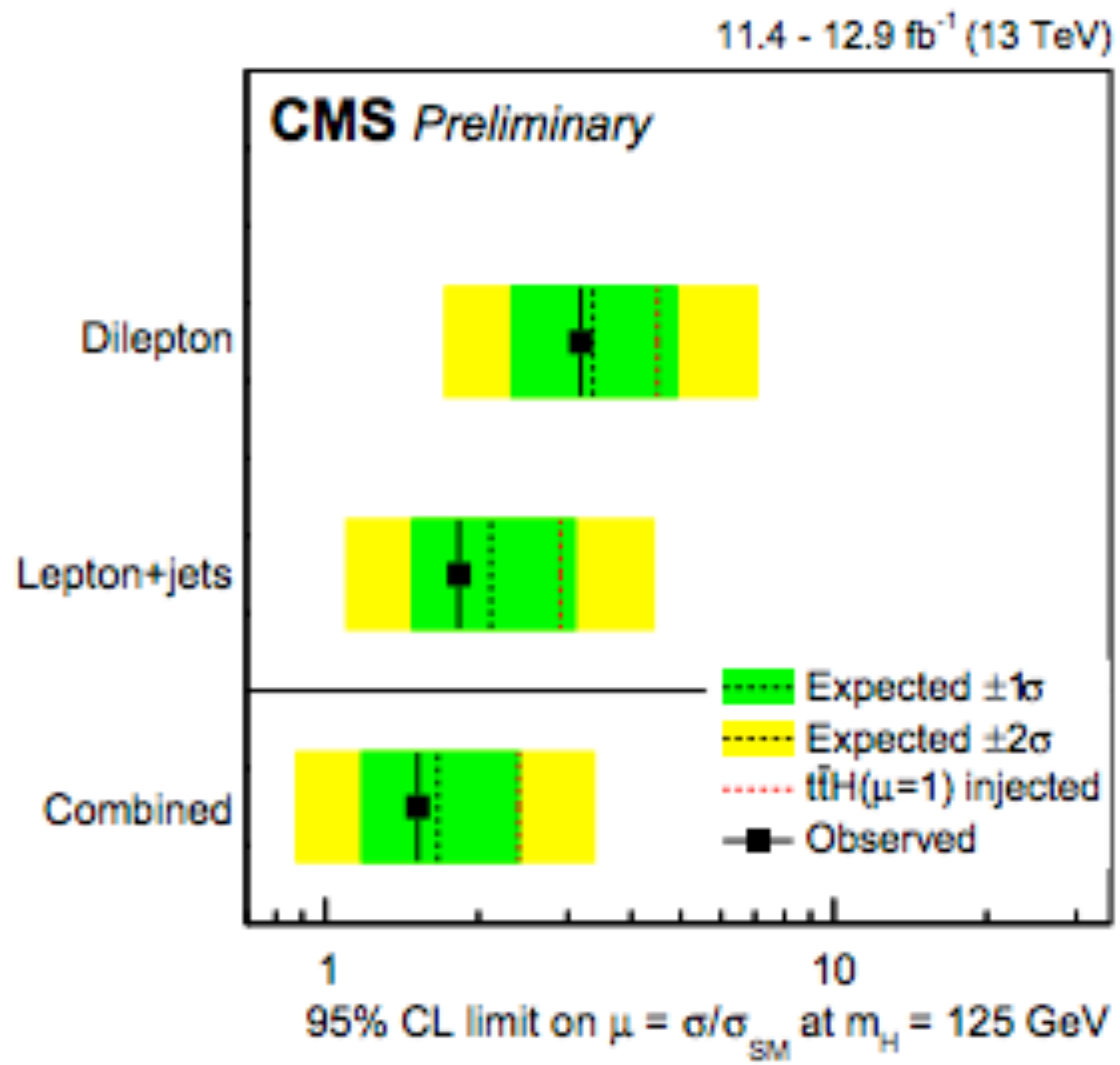
LO amplitude (signal or background)

Run II CMS

- CMS preliminary analysis with 12.9/fb [HIG-16-038]

  - μ < 1.5 (1.7), best fit μ = -0.19 ± 0.45 (stat.) ± **0.68 (syst.)**

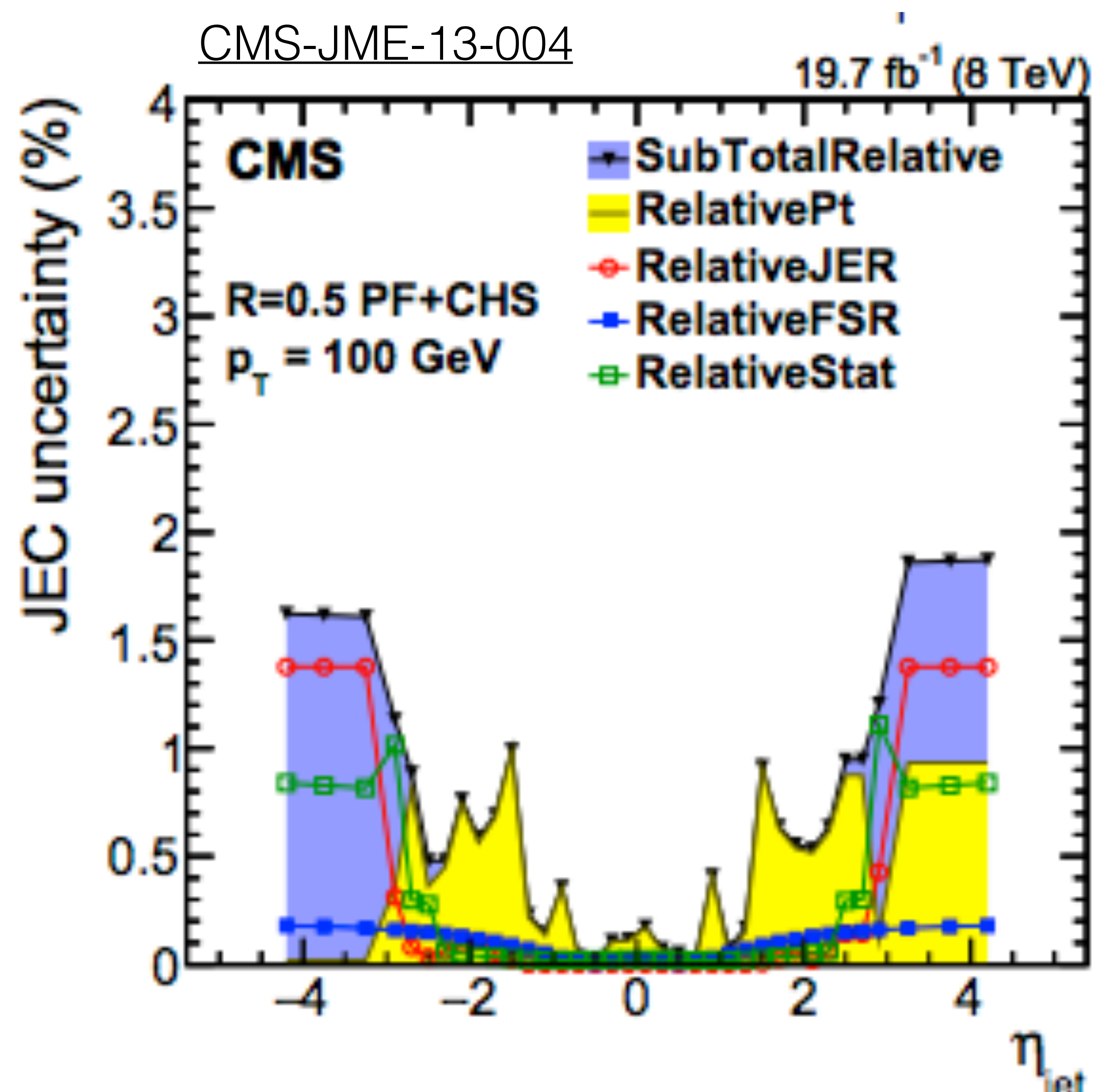




Jet energy corrections

- Jets are not measured perfectly due to pileup, non-uniform detector response
- Need to correct jet energy, resolution ~10% @ 100GeV
- Corrections have uncertainties of a few percent arising from various (~50) sources
- Uncertainties have a complex correlation structure and significantly affect the analysis

CMS-JME-13-004



e.g. 5 different major uncertainty sources.

Effect on analysis

- Mis-measured jet energies affect selection (event counts) and MVA discriminators (b-taggers, MEM)
- Typically, just re-run analysis ~100x times, recomputing affected variables
- Not feasible with MEM, which is already at ~1-2 minutes/event (mainly tt+jets MC)
- Developed an approximation in evaluating the uncertainties in MEM using vector integration

$$|\mathcal{M}(\boldsymbol{p})|^2 W(\boldsymbol{y}|\boldsymbol{p}) \to |\mathcal{M}(\boldsymbol{p})|^2 \begin{pmatrix} W(\boldsymbol{y}|\boldsymbol{p}) \\ W(\boldsymbol{y}+\delta\boldsymbol{y}_1|\boldsymbol{p}) \\ \dots \\ W(\boldsymbol{y}+\delta\boldsymbol{y}_n|\boldsymbol{p}) \end{pmatrix}$$

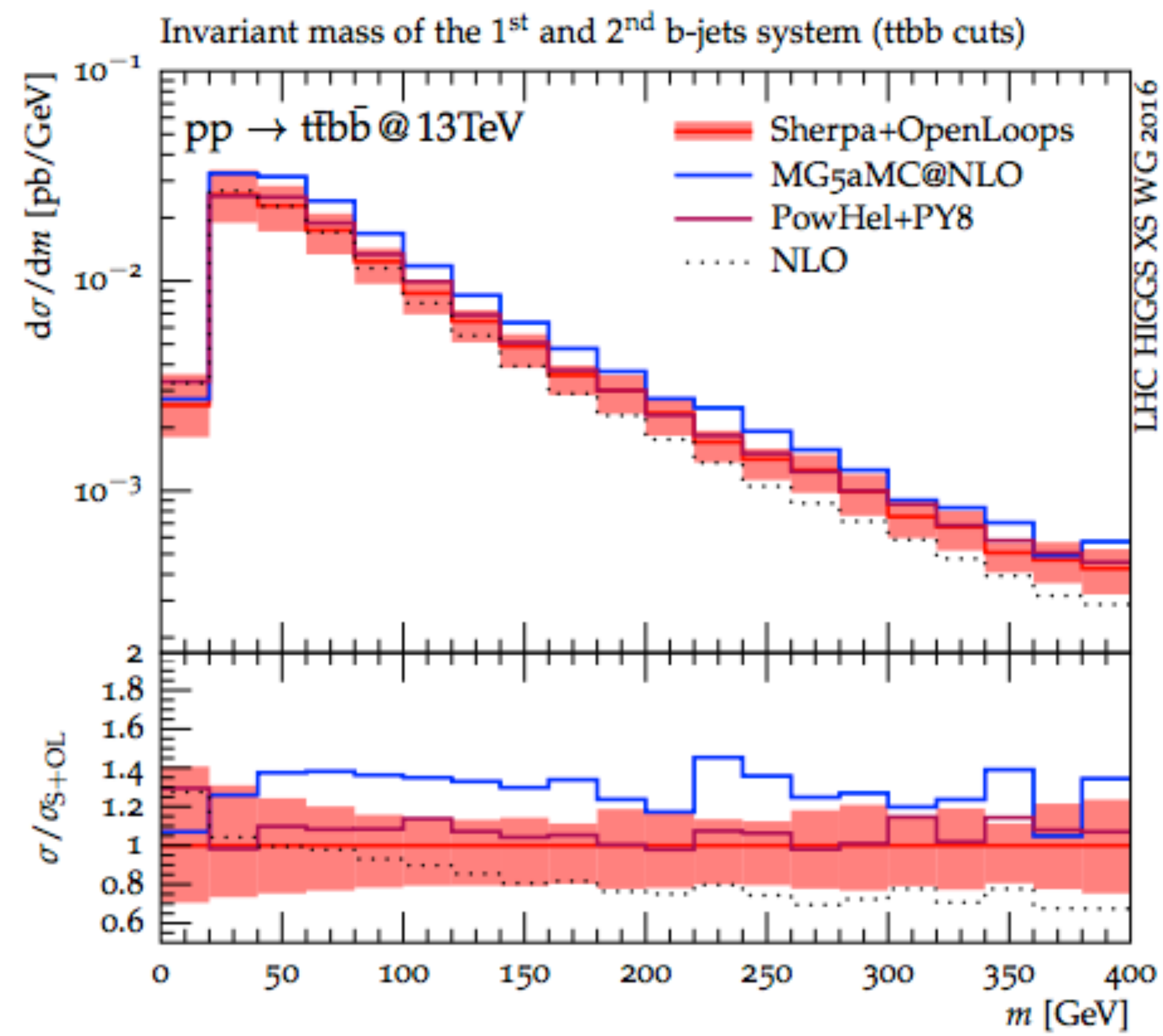
Machine learning

- Machine learning heavily used on object level: b-tagging, lepton ID, ...
- Also, ML complements MEM by using full simulation to discriminate tt+H(bb) vs tt+jets (inclusive) on event level in low-purity categories
- Relies on precise modelling of (already multivariate) b-tagging discriminators
- modeling problem: LO/PS precision for tt+HF in NLO POWHEG, assign ~50% uncertainties

tt+heavy flavour

- Currently relying on tuned POWHEG simulation (5FS)
- Major source of theory uncertainty
- Significant discrepancy between MG5@NLO and Sherpa+OpenLoops
- Looking to use data directly to constrain lack of knowledge about tt+bb

from talk by S. Pozzorini
and LHC HIGGS XS WG



Ongoing work

- Executing on our plan presented at HiggsCouplings 2016, i.e. finalize measurement with ~36/fb data
- Improving the treatment of experimental uncertainties (JEC, b-tagging)
- Measurement of tt+bb along with tt+H(bb)
- Analysis is systematically limited, need to address them with data!
- Aiming to finish thesis by end of the year 🙂

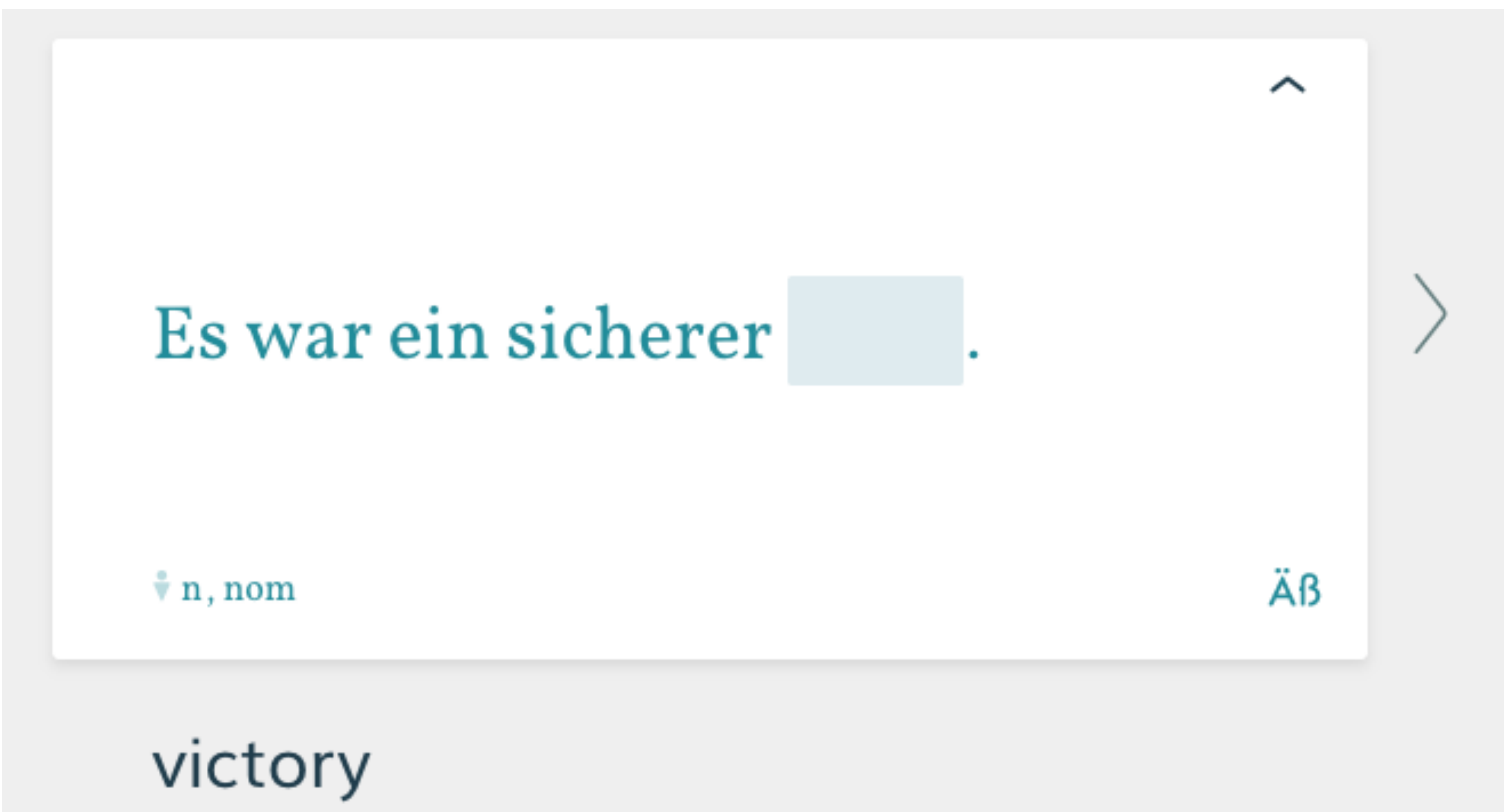
ETH Zürich

higgstools

Project at LINGVIST

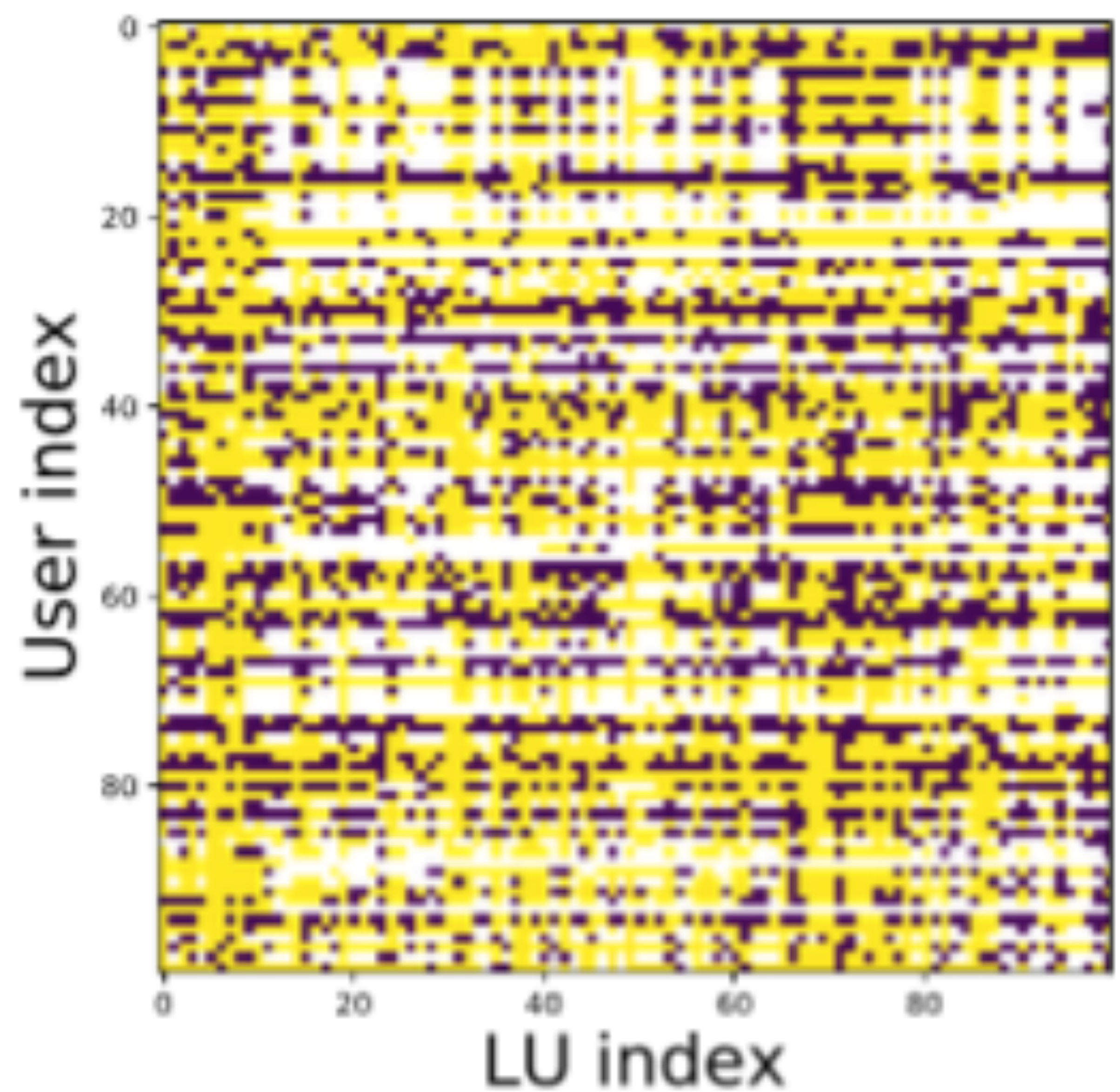
How it works

- Users try to complete the sentence in a foreign language
- We optimize the function that generates the study sequence



Background

- **Premise:** use scientifically studied methods to measure and improve learning
- Modelling of mental processes (memory, acquisition, skill)
- Specifically, given some data about users, wish to predict their future performance
- $O(10^5)$ users
- Ideal for machine learning



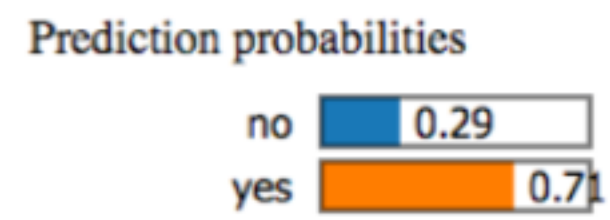
Example

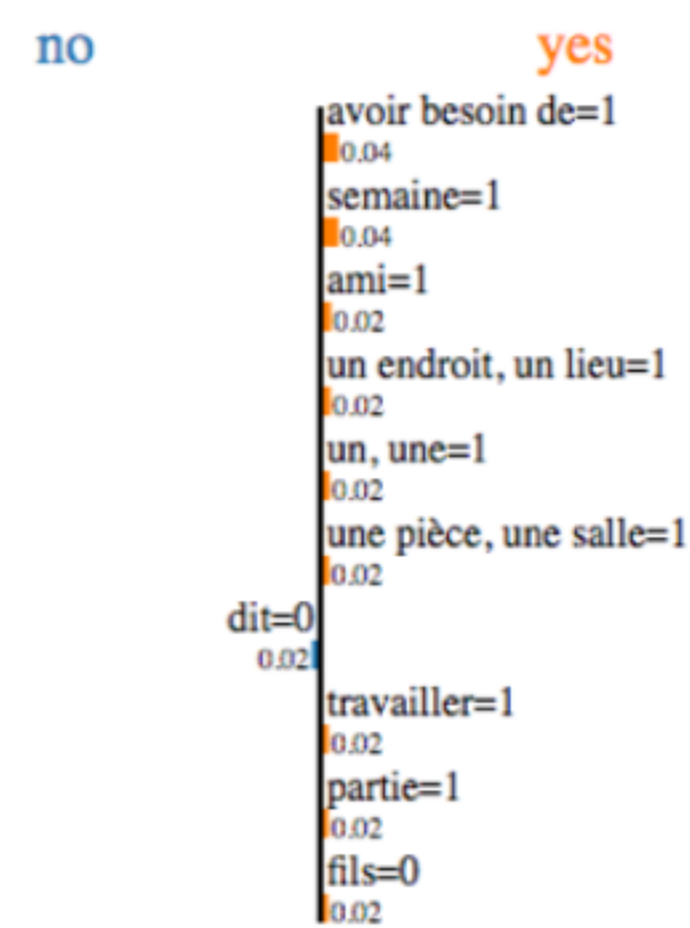
sample probe LUs

temps 0
nom 0
jour 1
vie 0
monde 1
un endroit, un lieu 1
maison 0
an 1
homme 1
nuit 1
voiture 1
école 1
nombre, chiffre 1
chose 1
le, la, les 0
fin 1
argent 1
commerce 1
partie 1
un, une 1
aujourd'hui 1
famille 1
père 0
un, une 1
un e-mail, un mél 1
je, moi 1

connaître, savoir 1
voir 1
une pièce, une salle 1
travail 1
eau 1
tu, toi 1
travailler 1
aimer 1
avoir besoin de 1
problème 1
fille 1
fils 0
il, lui 1
a besoin de 1
elle 1
dit 0
ce, il, ça 1
travaille 1
mère 1
livre 1
semaine 1
ami 1
garçon 1
bébé 1

entre (between)

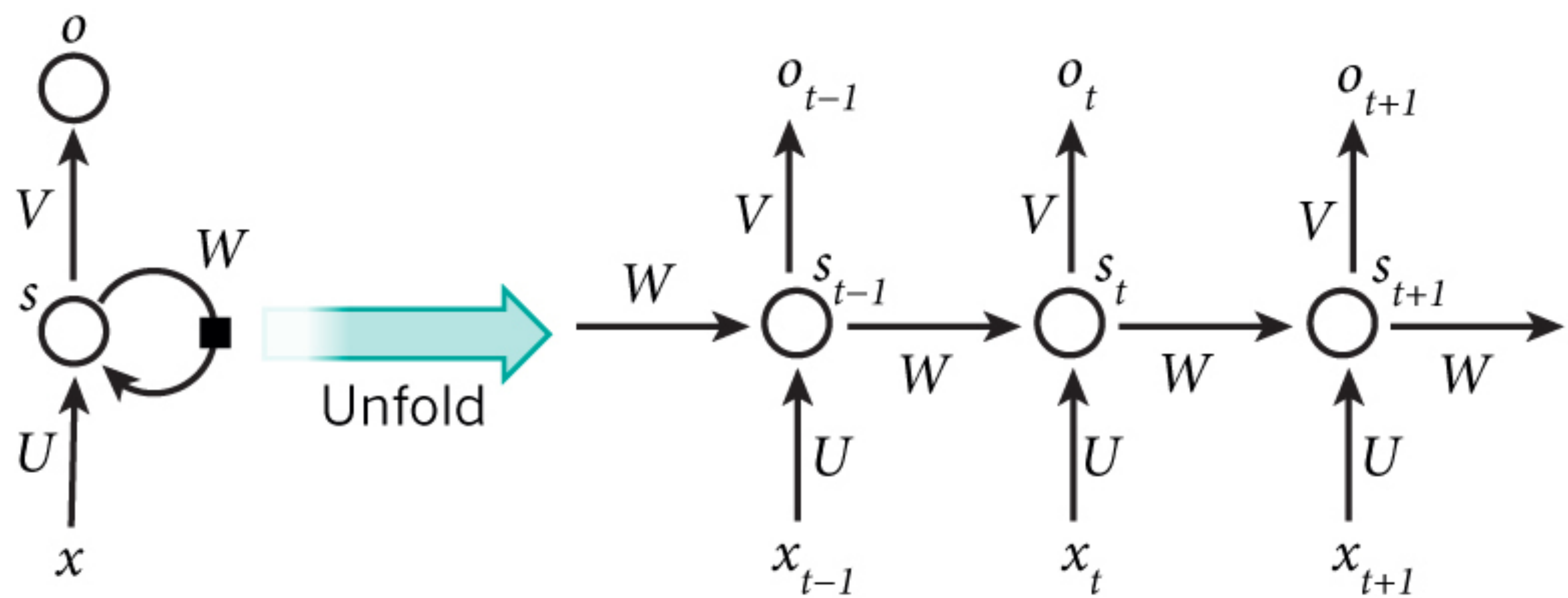




true value: yes

Implementation

- Use guess sequences to predict future performance
- based on Deep Knowledge Tracing (paper)
- Represent each word numerically using compressed sensing
- Encode guess sequences to latent vector using recurrent neural networks

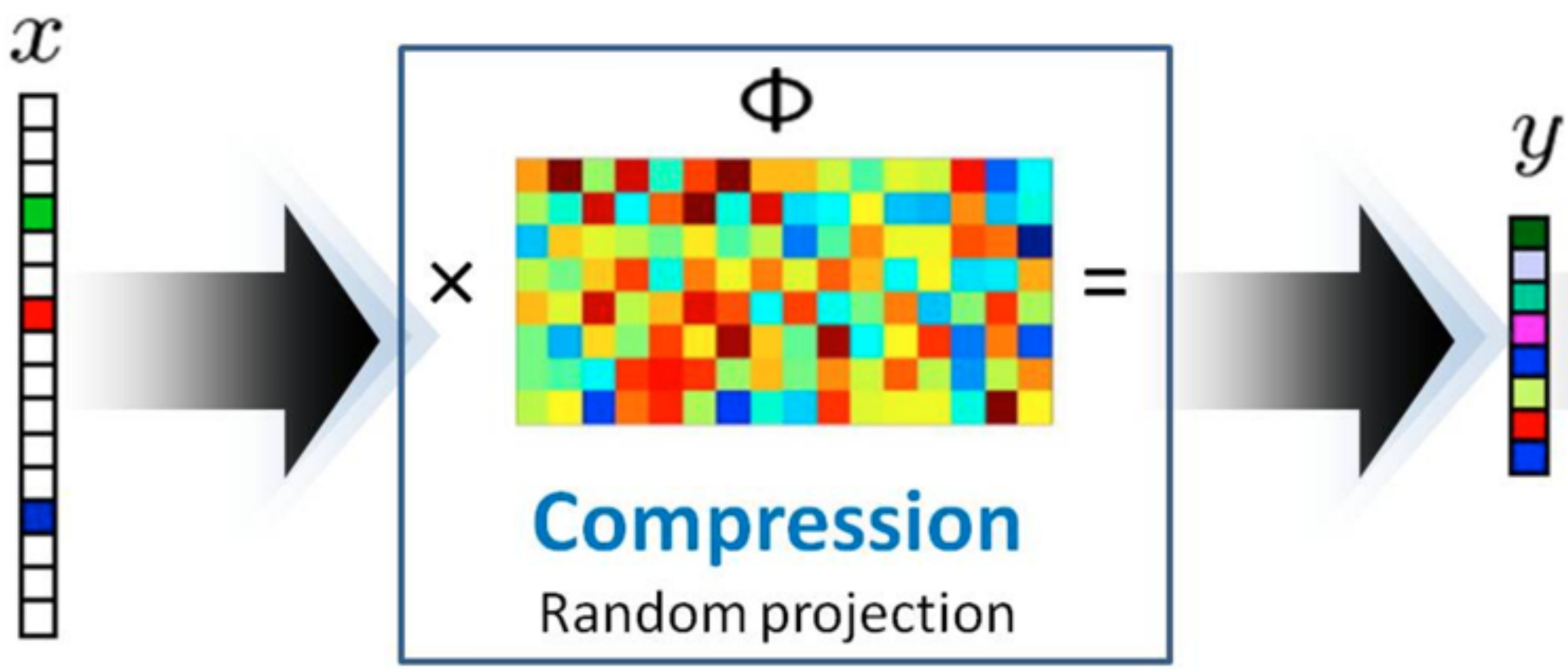


source: WildML

Compressed sensing

- How to represent sparse high-dimensional vectors, e.g. words in a large dictionary?

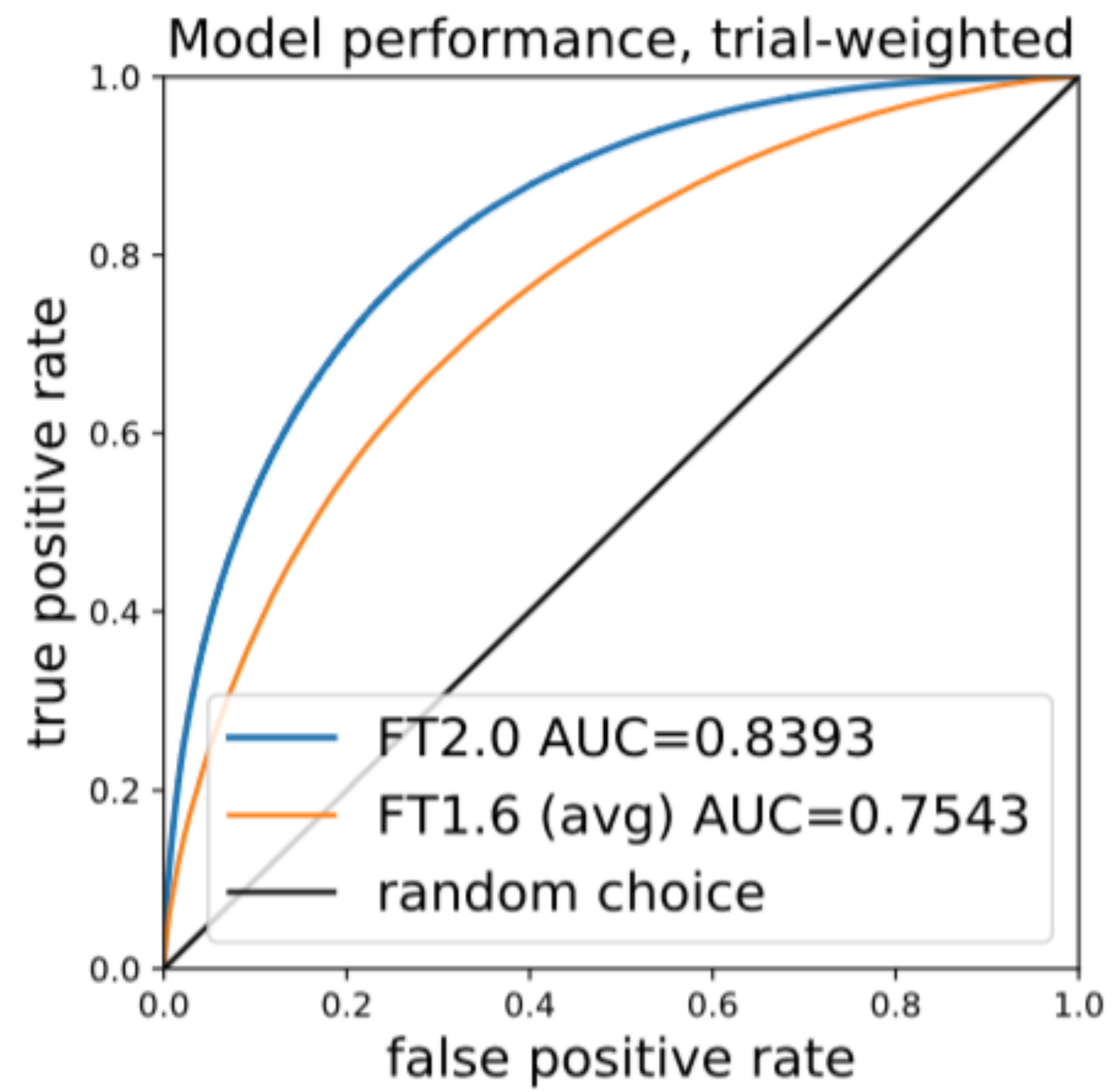
- k-sparse N dim → ~k log N dim random vectors [ref]



Performance

- Decode latent representation to knowledge vector in word space using successive dense layers

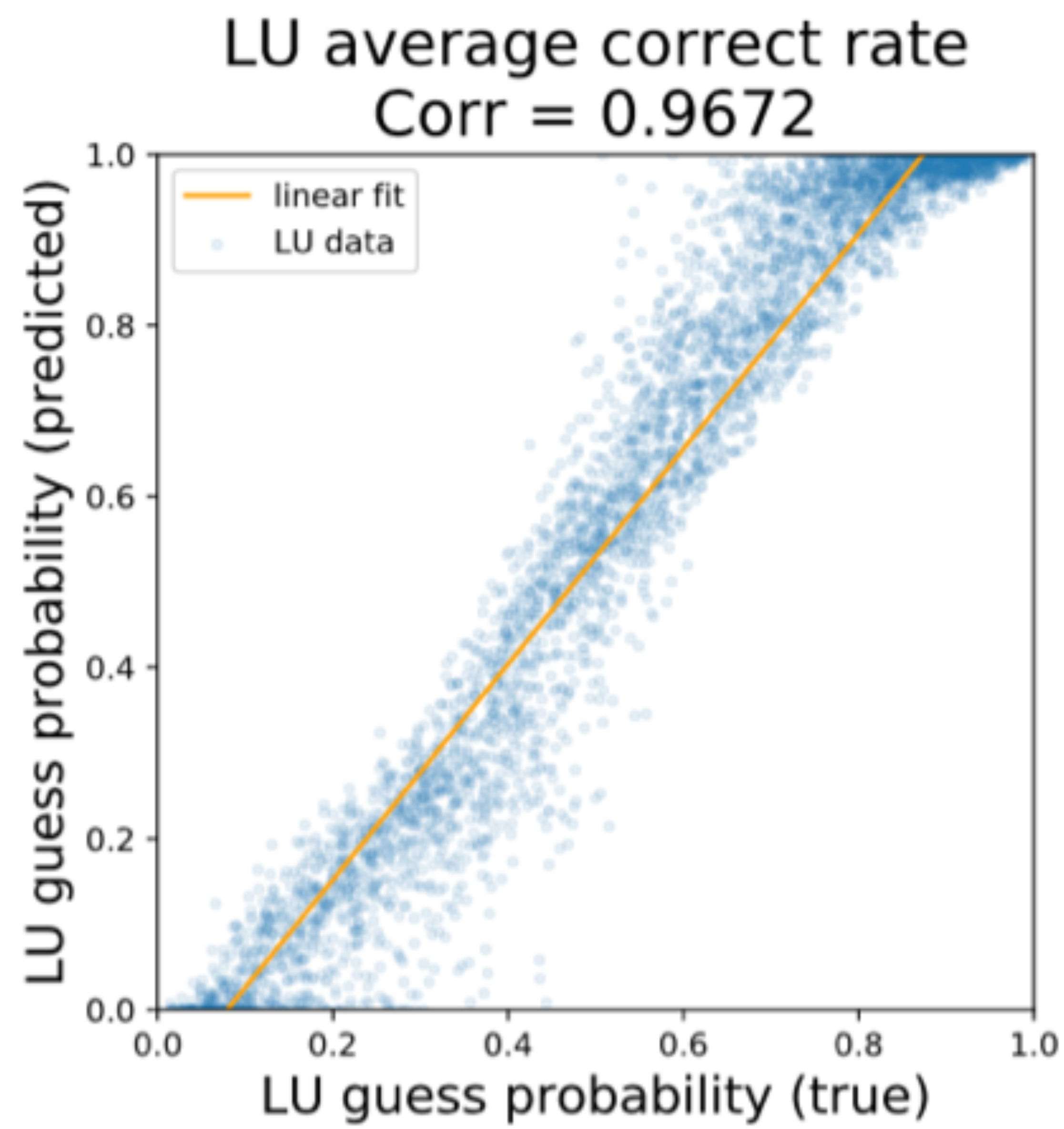
- Measure performance using correctly (wrongly) predicted words



$$\text{sequence} = [(\text{representation}, \text{guess}), (\text{representation}, \text{guess}), \ldots] \rightarrow$$
$$\rightarrow \textbf{LSTM} \rightarrow \text{encoded knowledge} \rightarrow$$
$$\rightarrow \text{dense} \rightarrow \cdots \rightarrow \text{dense} \rightarrow$$
$$\rightarrow \text{dense output with sigmoid activation} \rightarrow \vec{g}$$

Summary

- Used data analysis & mathematical modeling on a complex practical problem
- Was able to improve over existing methods by 50%
- Learned a lot about open-ended problem solving and mathematical thinking
- Machine learning complements detailed theoretical models in case you have lots of data
- Will be useful in experimental analysis: e.g. representing N-particle final states, <u>uncertainty estimation</u>



ETH Zürich

higgstools

Final words

A big thank you to everyone involved in creating and nurturing HiggsTools!

I have learned a lot and hope to keep learning.