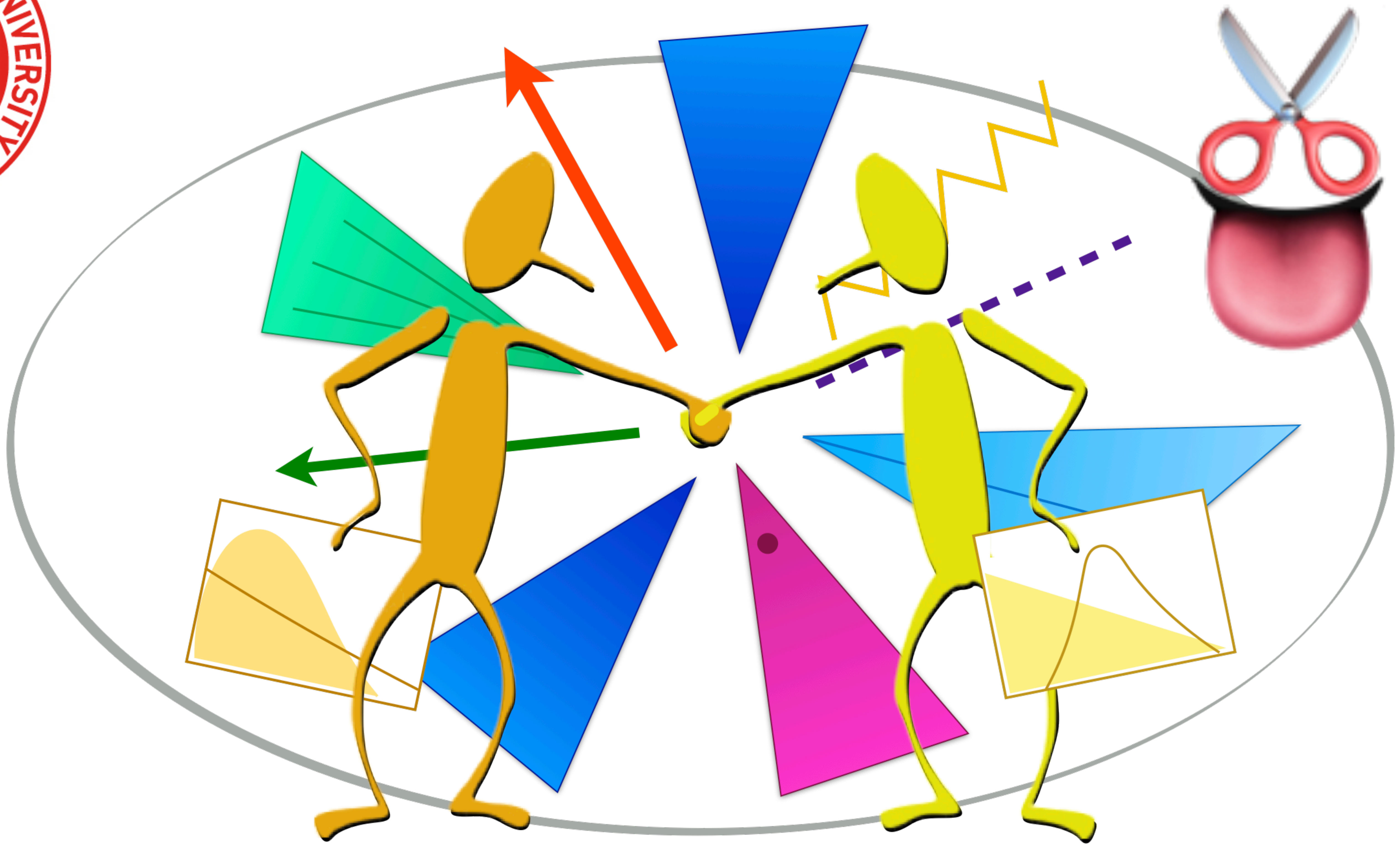


ADL/CutLang for (re)interpretation of experimental search results

Sezen Sekmen (KNU), Gökhan Ünel (UC Irvine)

Analysis Description Language Tutorial & Hackathon

21-22 Nov 2022, Kyungpook National University, Center for HEP

What is an experimental result?

An experimental result is the empirical outcome of the experiment, the measurement of some physical quantity, such as:

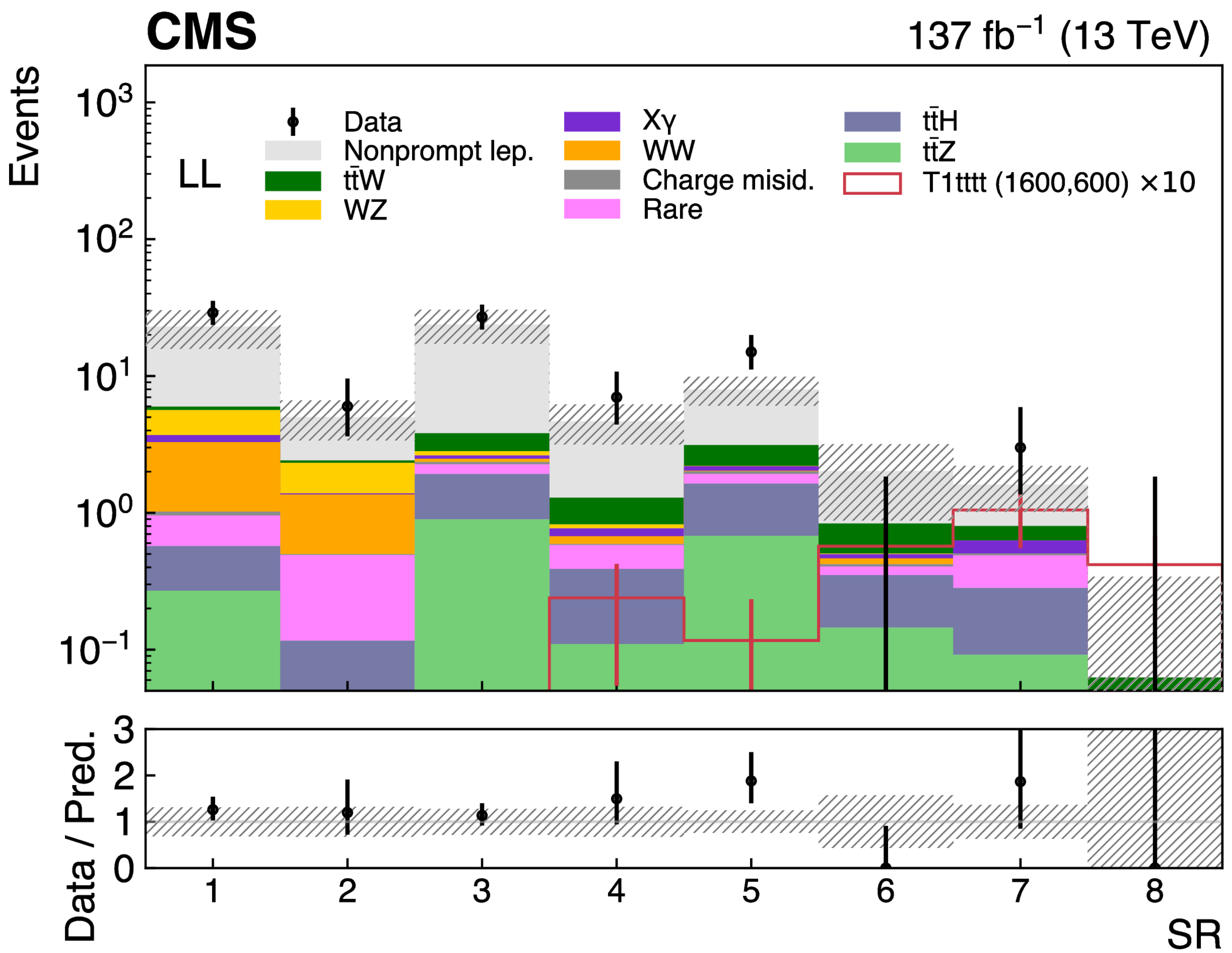
- Event counts <— mainly in new physics search analyses.
- Masses
- Cross sections/branching ratios
- Signal strengths
- Couplings
- Kinematic shapes, peaks, edges, endpoints
  (usually to derive masses or mass differences)
- Decay widths
- Charge asymmetry
- Spin correlations
- etc.

What is interpretation?

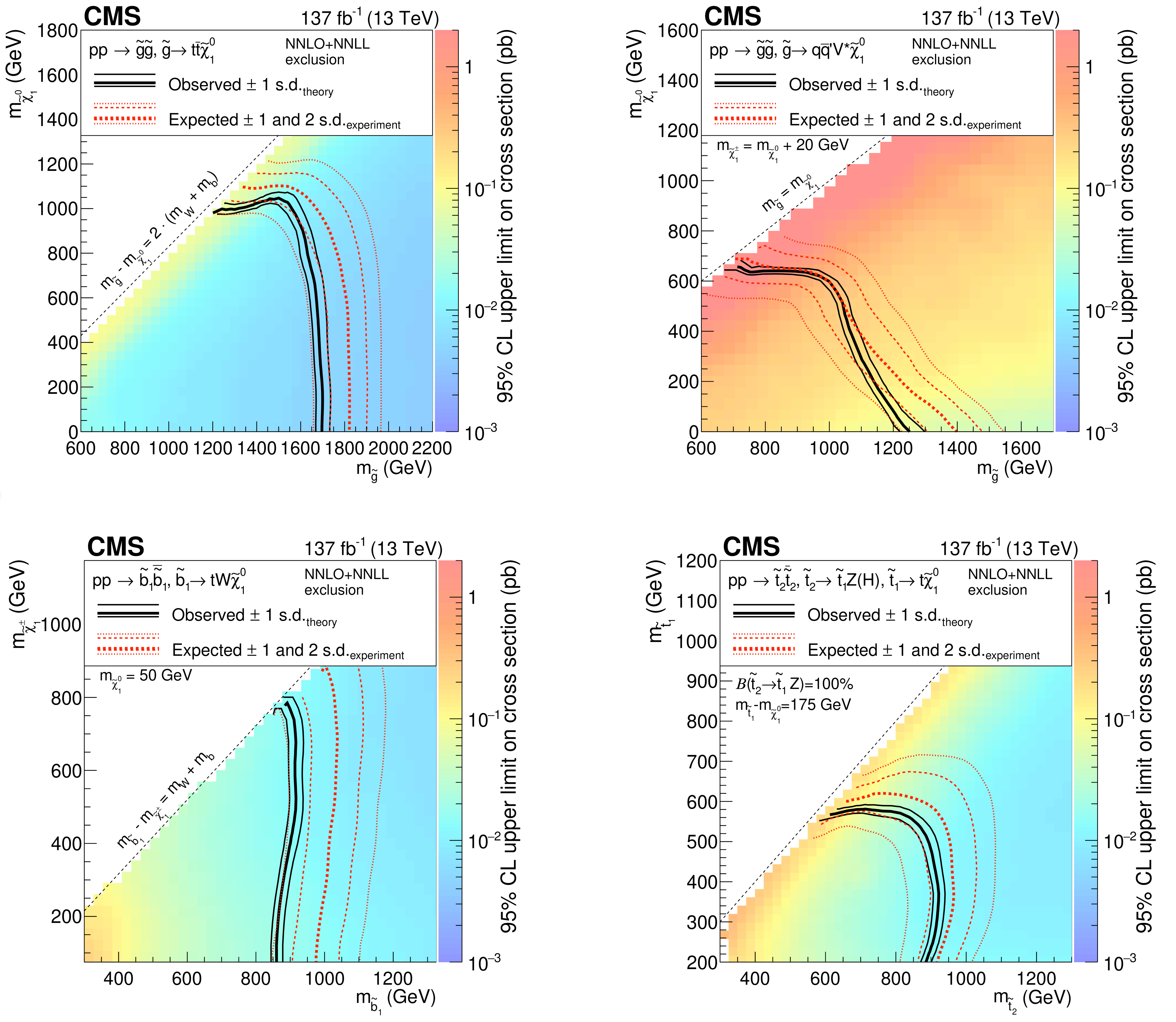
Interpretation is the comparison of experimental results with the expectations of a given a theoretical model.
CAUTION! Interpretation is NOT the experimental result.

- We use a statistical model and likelihood to interpret the experimental result.
  - The statistical model of an analysis provides the complete mathematical description of that analysis.
  - It relates the observed quantities x to the theoretical model's parameters θ_i through the probability density $p(x|\theta_i)$. Parameters θ_i include signal model parameters θ_s and background model parameters θ_b.
  - The likelihood $L(\theta_i) = p(X_0|\theta_i)$ is the probability density $p(x|\theta_i)$ evaluated at the observed values X_0 of the observables x.
  - The likelihood is the starting point of any interpretation.
- We estimate parameters θ using statistical procedures, and test the validity of the model.
- An experimental result can be interpreted with multiple theoretical models.

Results to interpretation



Experimental result including data counts, background estimation and uncertainties.



Interpretation in 4 different models.

SUS-19-006: Search for SUSY with jets and MHT

What is reinterpretation?

Reinterpretation consists of recoding a published analysis from scratch for the purpose of interpreting it in terms of other physics models not interpreted in the original publication.

Reinterpretation is usually done by phenomenologists to see the impact of experimental results on their favorite physics model.

- An analysis code is rewritten outside the experimental frameworks.
  - Analysis recoding is done less accurately compared to the original experimental code, since detector information does not exist in full detail in public simulation tools.
  - Public tools exist for rewriting analyses: CheckMate, MadAnalysis, Rivet, … now ADL/CutLang.
- Monte Carlo simulated events for the signal models are produced.
- Analysis code is run on the events to obtain predicted signal counts / efficiencies.
- Predicted signal counts are used together with observed data and background estimation results from the experimental publication to calculate limits.

What is validation?

Validation is to ensure that an analysis reimplementation is done correctly.

A rewritten analysis can reproduce the results of the original analysis to a degree of accuracy.

Ways to validate an analysis:

- If counts or cutflows are available for a certain (signal) process:
  - Produce signal events for the process, only for a few signal points.
  - Run your own analysis code on these events and obtain your counts and cutflows.
  - Compare with counts / cutflows published by the experiment.
- If counts / cutflows are not available, but only limits are available:
  - Mass-produce signal events for a lot of signal points.
  - Run your analysis code, obtain counts and calculate limits.
  - Compare with limits published by the experiment.

ADL for reinterpretation

ADL allows practical exchange of experimental analysis information with the pheno community.

- Unobscured description of the complete analysis algorithm.
- Enables straightforward adaptation from experiments to public input event formats.
- Use the same ADL file with a few changes:
  - Biggest difficulty is in reproducing an analysis is adapting object definitions:
    In ADL, e.g. just swap experimental object IDs with numeric efficiency maps.
  - Event selections stay ~the same (can swap trigger selections with efficiencies)

Rewriting simpler objects

Experimental analyses could be reusable after simplifying adaptations in their descriptions:

b-tagging for UL NanoAODv9 → **b-tagging for public use, e.g. with Delphes**

b-tagging for UL NanoAODv9

```
# AK4 jets
object JetAK4
  take Jet
  select jetID(Jet) > 0
  select pt(Jet) > 30
  select abs(eta(Jet)) < 2.4

# b-tagged jets - medium
object MediumBTag
  take JetAK4
  select btagDeepB(JetAK4) >= 0.2598
```

b-tagging for public use, e.g. with Delphes

```
# AK4 jets
object JetAK4
  take Jet
  # select jetID > 0
  select pt(Jet) > 30
  select abs(eta(Jet)) < 2.4

# b-tagged jets - medium
object MediumBTag
  take JetAK4
  select applyhitandmiss( btagdeepBmediumeff( pt, abs(eta) ) ) == 1
```

- applyhitandmiss: A generic function reading efficiencies, object attributes and applying the hit & miss method.
- btagdeepBmediumeff: Efficiencies provided by CMS. ADL table blocks can host numerical efficiencies.

ADL for reinterpretation

ADL allows practical exchange of experimental analysis information with the pheno community.

- Generic structure available for expressing analysis output in the ADL file:
  Data counts, BG estimates, signal predictions —> counts, uncertainties, cutflows.
  - Running CutLang puts preexisting results in histograms with the same format as the run output. —> Direct comparison of cutflows, limit calculations.

- ADL could facilitate providing information on analysis results to HEPDATA or similar platforms.

Counts simple example: ex12_counts.adl

Real life example - CMS-SUS-19-006:

- ADL with data counts and BG estimates
- ADL with signal cutflows

ADL database for reinterpretation

Our aim is to provide a large ADL database with written and validated analyses for use in reinterpretation studies.

- We have written various analyses in ADL.
- We have exercised validating some of these analyses with CutLang.
- We are setting up a large system for ADL analysis validation at KNU (see Junghyun's talk)

You are very welcome to join the effort!