Statistically Learning The Next Standard Model From LHC Data

Wolfgang Waltenberger
(ÖAW and Uni Wien)

Seminar for HEPHY AI group leader position
Vienna, September 2022

image courtesy of Jon Butterworth, Chris Wormell

Who Am I

Priv. Doz. Dr. Wolfgang Waltenberger

Born december 10th 1974

Father of Fabian (2000) and Yannick (2005)

2001 - now: CMS member
2004: Finished PhD at TU Wien
2008 - now: staff member at HEPHY Wien
2014/15: Guest Professor at USP São Paulo (Brazil)
since 2016: lecturing at TU Wien, Uni Wien
2018: Habilitation at Uni Wien
2021/22: Guest Professor in Grenoble (France)

Worked on: tracking, vertexing, *b*-tagging, search for SUSY, interpretation, simplified models, SModelS, ml/stats

inverse problem of particle physics

My Current Data Science Involvements

At the Austrian Academy of Sciences
Founded ÖAW AI Graduate School, directed first installment (2019), now preparing to be again director of third iteration (january 2023)

At HEPHY
(Co-)supervising three PhD students in three different groups [see next slide]
Work on learning (as opposed to postulating) the Next Standard Model from data

At TU Wien and Uni Wien
Teaching regular courses in statistics (see youtube.com/WolfgangWaltenberger)

Generally – as AI/ML is still trending: currently many invited lectures, lecture series, and seminars on AI/ML/stats related topics
Eg. WMLQ2022, September 2022, Poland, Vietnam School of Physics, July 2022, Vietnam
In the past: MCNet Summer School, Sweden, VDSP Winter School, Austria, DKPI Graduate School, Austria, ….

My Current Data Science Involvements

Co-supervising three PhD students:

	Felix Wagner	**Mark Matthewman**	**Sahana Narasimha**
group	CRESST/ COSINUS	CMS (HGCAL)	phenomenology
co-supervising with	Florian Reindl (TU Wien)	Erica Brondolin (CERN)	-
physically at	Vienna	CERN	Vienna
Data scientific highlights	Reinforcement learning of detector control	Graph neural network for HGCAL energy regression	
Financed by	FFG	CERN PhD programme	Bilateral French-Austrian ANR-FWF
Started	July 2020	March 2022	October 2022

common denominator is data science!

My Current Data Science Involvements

(Co-)supervising three PhD students:

	Felix Wagner	**Mark Matthewman**	**Sahana Narasimha**
group	CRESST/ COSINUS	CMS (HGCAL)	phenomenology
co-supervising with	Florian Reindl (TU Wien)	Erica Brondolin (CERN)	-
Data scientific highlight	Reinforcement learning of detector control	Graph neural network for HGCAL energy regression	**Topic of this talk**
Financed by	FFG	CERN PhD programme	Bilateral French-Austrian ANR-FWF
Started	July 2020	March 2022	October 2022

common denominator is data science!

Prelude:

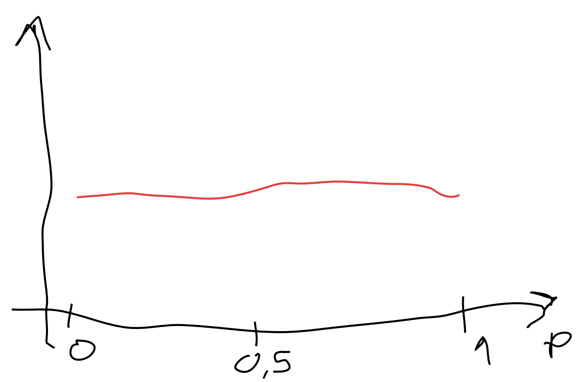
Fun With Meta-Statistics

Fun With Meta-Statistics

- Our SModelS v2.2.0 database summarizes the results of almost 1000 signal regions of about 100 CMS and ATLAS publications of searches for new physics

- For each signal region, we know the number of observed events ending up in this signal region, alongside with the number of expected Standard Model “background” events and its error. Assuming a simplified statistical model, we can compute *p*-values for the Standard Model hypothesis

- **If there is no new physics is in the data, the distribution of *p*-values should look like this:**



(p = 0 means huge excess of observed events)

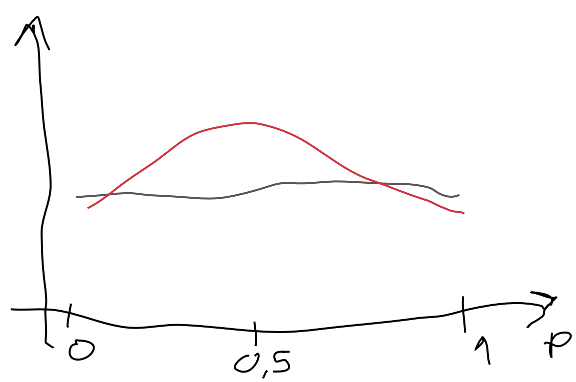
https://smodels.github.io/docs/ListOfAnalyses220

Fun With Meta-Statistics

- Our SModelS v2.2.0 database summarizes the results of almost 1000 signal regions of about 100 CMS and ATLAS publications of searches for new physics

- For each signal region, we know the number of observed events ending up in this signal region, alongside with the number of expected Standard Model “background” events and its error. Assuming a simplified statistical model, we can compute *p*-values for the Standard Model hypothesis

- **Given that the background errors are conservatively, systematically overestimated by ~ 30%, we expect the following distribution for the *p*-values:**



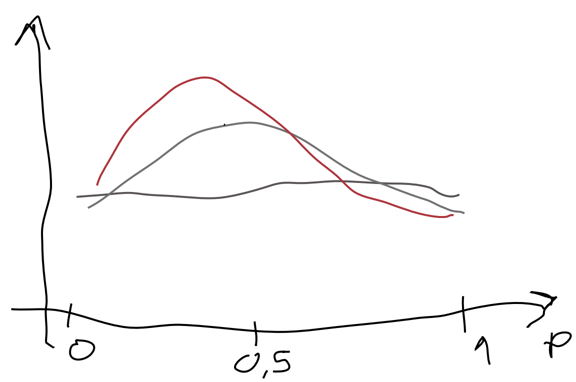
(peak at ~ 0.5)

(p = 0 means huge excess of observed events)

https://smodels.github.io/docs/ListOfAnalyses220

Fun With Meta-Statistics

- Our SModelS v2.2.0 database summarizes the results of almost 1000 signal regions of about 100 CMS and ATLAS publications of searches for new physics
- For each signal region, we know the number of observed events ending up in this signal region, alongside with the number of expected Standard Model “background” events and its error. Assuming a simplified statistical model, we can compute *p*-values for the Standard Model hypothesis
- **If dispersed new physics were slowly seeping in, it would look like this:**



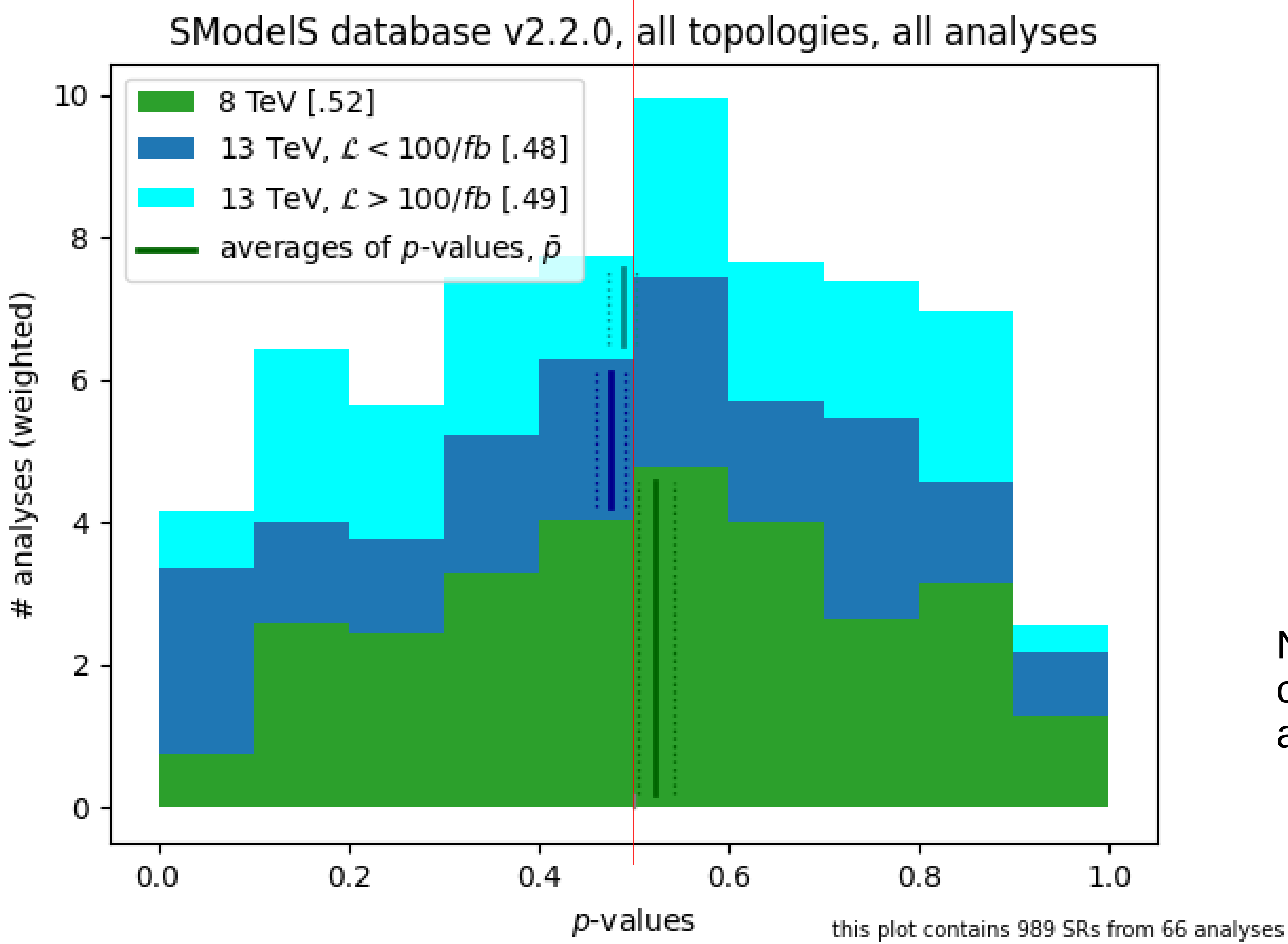
(peak moving to smaller values)

(p = 0 means huge excess of observed events)

https://smodels.github.io/docs/ListOfAnalyses220

Fun With Meta-Statistics

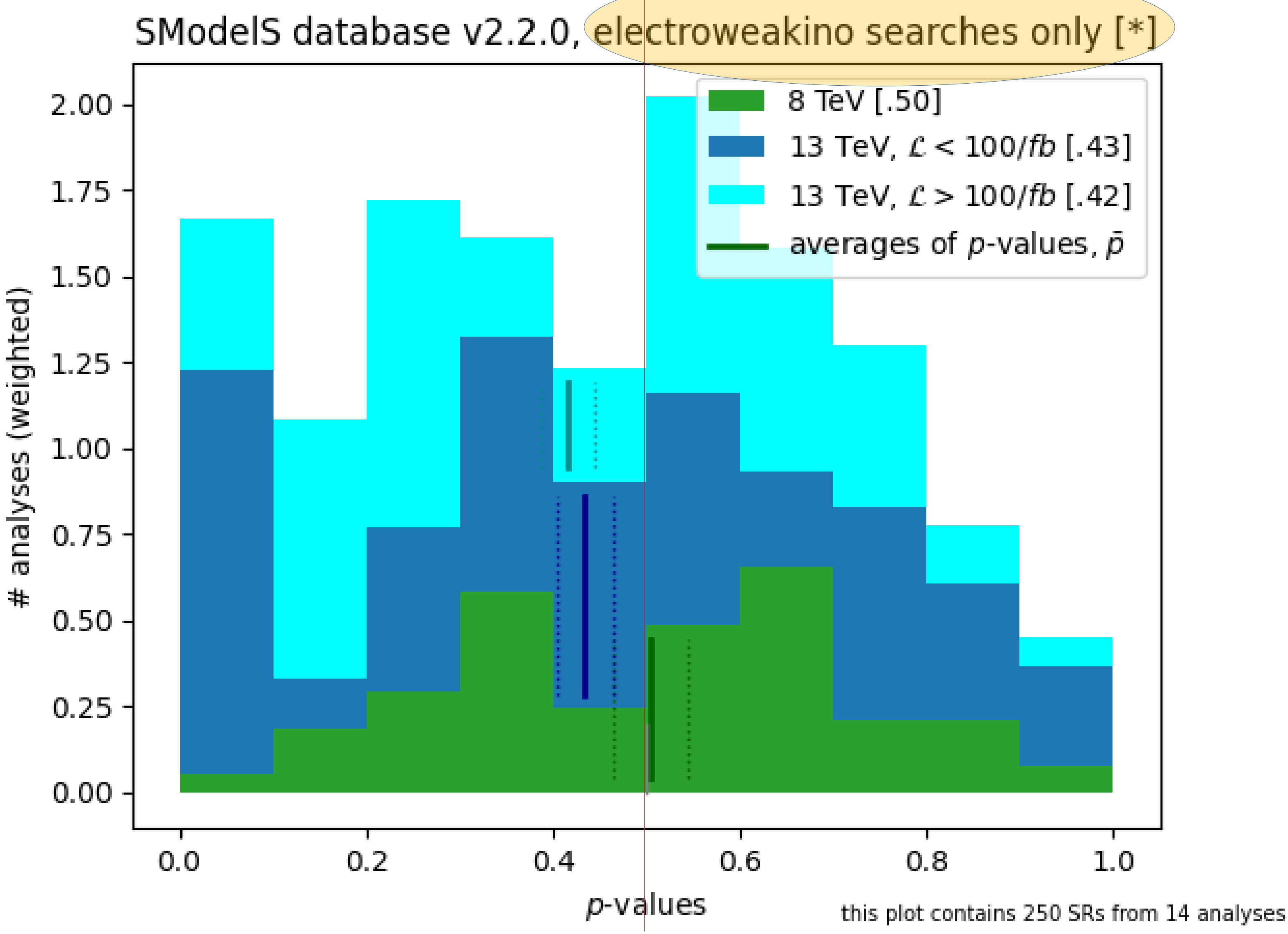
Here's with the actual, real data:



N.B: the color bars are stacked

Lookin' good! No obvious *p*-hacking in our search programme.

Fun With Meta-Statistics



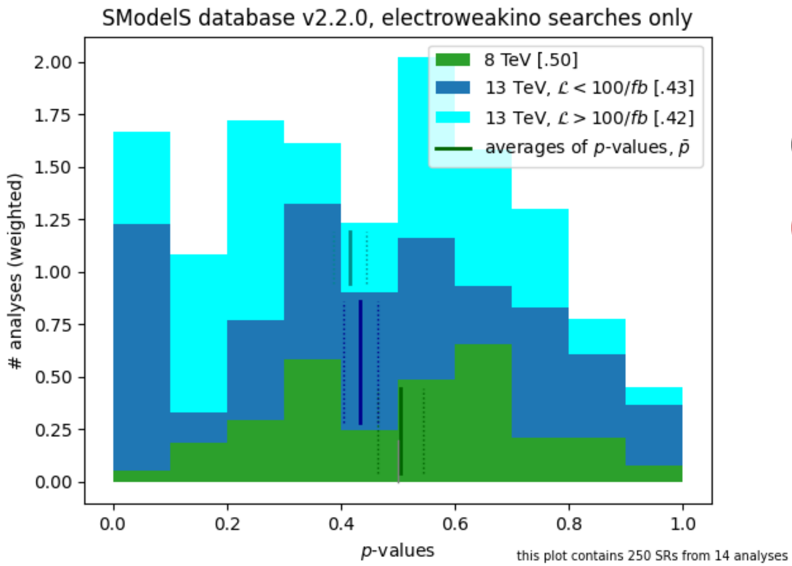
N.B: the color bars are stacked

Random fluke? Selection bias? New physics slowly seeping in?

[*] searches that target chargino/neutralino productions in RPC SUSY scenarios. Decays via W,Z,h bosons + dark matter candidate

[**] some data are used more than once in this plot. We cannot – and do not pretend to -- make too serious frequentist statements

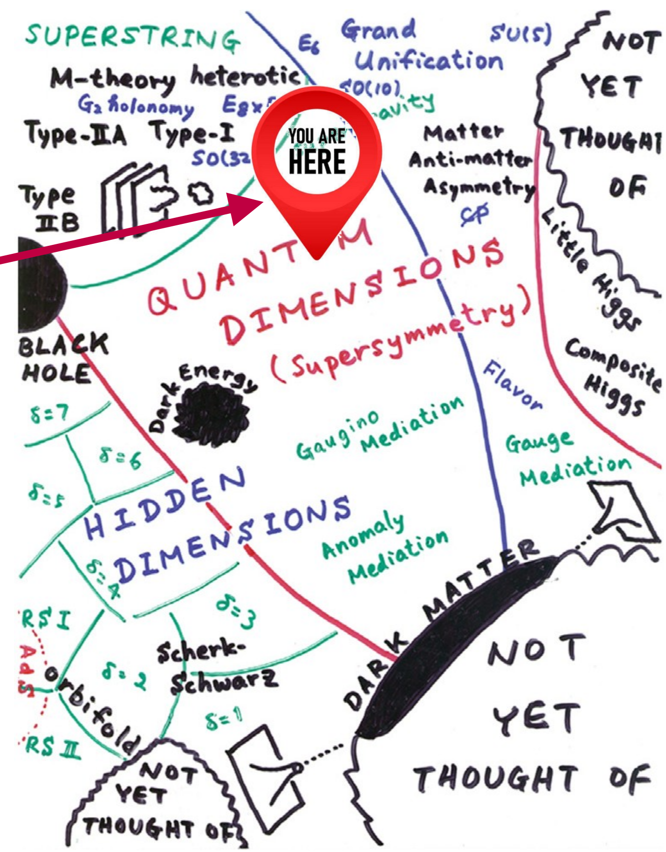
Fun With Meta-Statistics



If the trend were (conjunctive form!) due to new physics, how would we go from such dispersed signals ….

… to identifying the One True Point [*] (a.k.a. the Next Standard Model) in Hitoshi Murayama's landscape of theories?

→ **our Inverse Problem!**



[*] obviously in reality we would actually want a posterior density, not one point

Statistically Learning The Next Standard Model From LHC Data

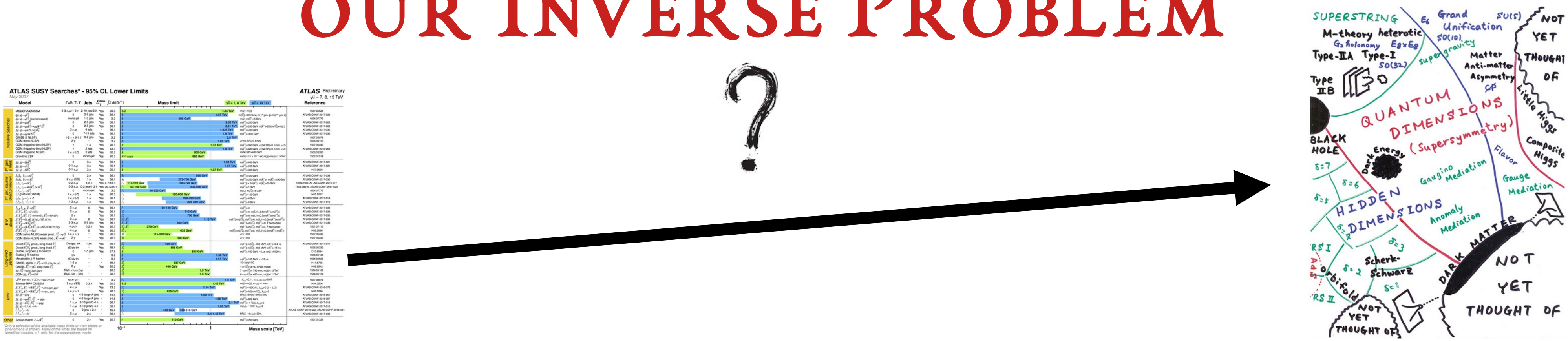
image courtesy of Jon Butterworth, Chris Wormell

Wolfgang Waltenberger (ÖAW, Uni Wien)
[presenting work in collaboration with Andre Lessa and Sabine Kraml]

https://arxiv.org/abs/2012.12246

Seminar, HEPHY AI group leader
Vienna, September 2022

OUR INVERSE PROBLEM



- too difficult a task for us humans (we are neither smart nor creative enough)
- Let the machine solve it!

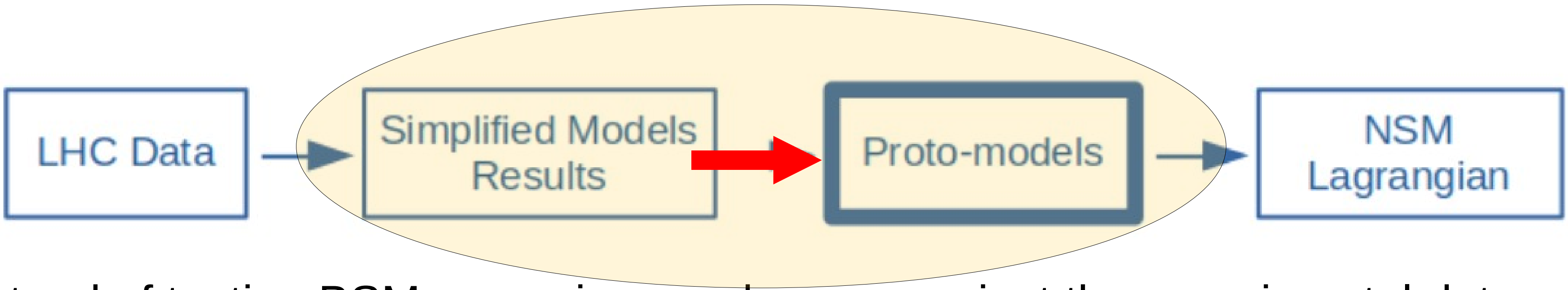
Our mission statement:

- Given the SModelS database of simplified models results, we **let a machine find the simplest possible model** that identifies **the largest possible violation of the Standard Model hypothesis** in the results, while evading all constraints from the negative search results in our database.

- Actually, we don't just want a single model, we want posteriori probabilities in these theory landscapes.

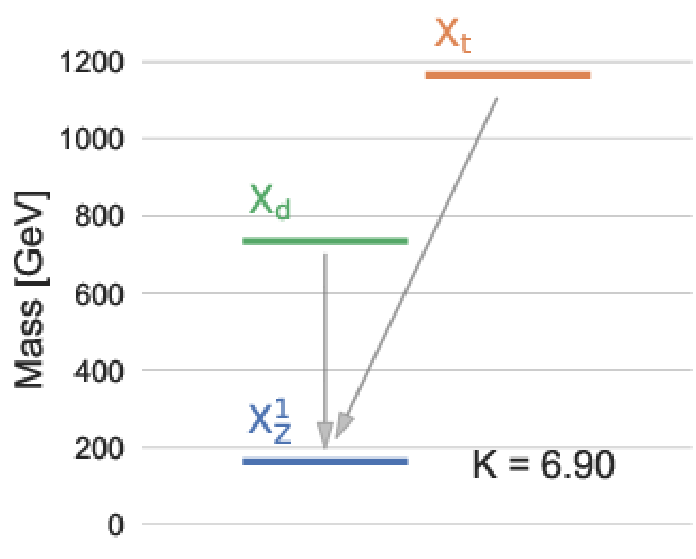
- The models are allowed to be "**incomplete**", we want precursor theories, "proto-models", whose construction is driven by data, not by abstract principles. UV-completing such models will be a separate step (of course also partly executed by machines)

Protomodels



Instead of testing BSM scenarios one-by-one against the experimental data:

- identify potential dispersed signals in the slew of published LHC analyses
- build candidate “protomodels” from them.



NSM: “Next Standard Model”, BSM: “Beyond the Standard Model”

Protomodels

- Protomodels can be thought of as **consistent sets of simplified models**.
- **Caveat:** The (variable-, or trans-dimensional) protomodels space is restricted by the SModelS software and database: currently restricted to models exhibiting a **Z_2 symmetry** (i.e. SUSY- and UED-like):

particle	decay channels	particle	decay channels
X_q	$qX_Z^j,\ q'X_W^i,\ qX_g$	X_W^1	WX_Z^j
X_t^1	$tX_Z^j,\ bX_W^i,\ WX_b^1,\ tX_g$	X_W^2	$WX_Z^j,\ ZX_W^1,\ hX_W^1$
X_b^1	$bX_Z^j,\ tX_W^i,\ WX_t^1,\ bX_g$	$X_Z^{j\neq 1}$	$WX_W^i,\ ZX_Z^k,\ hX_Z^k$
X_t^2	$tX_Z^j,\ bX_W^i,\ ZX_t^1,\ WX_b^1,\ tX_g$	X_ℓ	$\ell X_Z^j,\ \nu_\ell X_W^i$
X_b^2	$bX_Z^j,\ tX_W^i,\ ZX_b^1,\ WX_t^1,\ bX_g$	X_{ν_ℓ}	$\nu_\ell X_Z^j,\ \ell X_W^i$
X_g	$q\bar{q}X_Z^i,\ q\bar{q}'X_W^i,\ b\bar{b}X_Z^i,\ t\bar{t}X_Z^j,\ btX_W^i,\ qX_q,\ bX_b,\ tX_t$		

X ("Xeno"-) particles, X_q is squark-like, X_Z is neutralino-like, etc

Ongoing work:
Next iteration will likely not anymore require a Z_2 symmetry
[Andre Lessa, UFABC Brazil]

Building the Next Standard Model

PROTOMODEL BUILDER

Random changes to the protomodel → Test against database (SModelS) → Test statistic K (compatibility with data, model complexity)

Feedback for building new models

➡ "MCMC-type walk" over model+parameter space

➡ After many iterations/steps, the builder "learns" the best BSM model

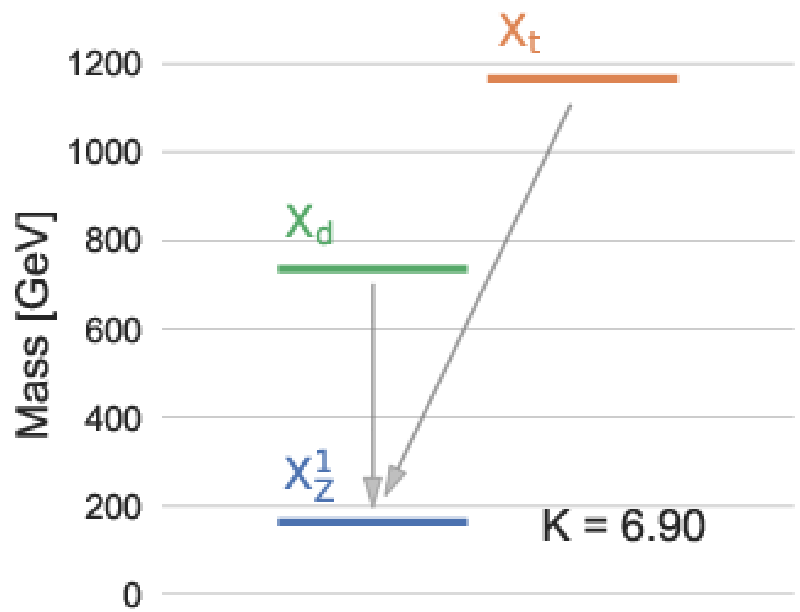
Building the Next Standard Model

- In each step of this random walk, the following changes to the existing protomodel are allowed:
  - randomly add or remove a particle
  - randomly change a branching ratio, or the mass of a particle
  - randomly change a production cross section of a particle

- after each step a test statistic K is computed that quantifies how well the protomodel describes the data.

- K got much worse? $\rightarrow$ Revert to old protomodel

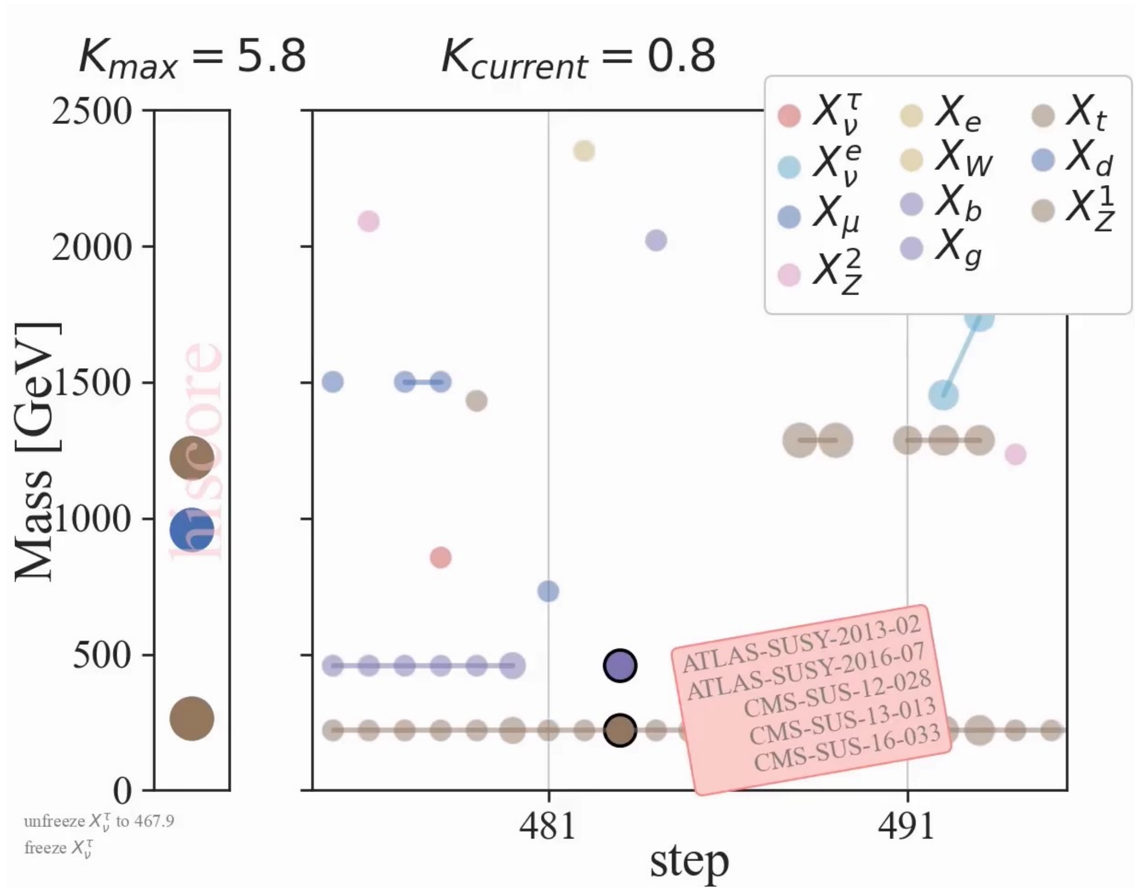
- K stayed the same or got better? $\rightarrow$ keep new protomodel



Our Approach

the test statistic

Particle spectra



A hiscore protomodel

potential dispersed signals

Random modifications

an MCMC-like walk

https://smodels.github.io/protomodels/videos

The Test Statistic

The test statistic K^c is a likelihood-ratio test that quantifies how much better the proto-model describes the data than the Standard-Model (plus a penalty for model complexity).

Quantifies violation of Standard Model hypothesis

Penalizes for model complexity

$$K^c := -2\ln \frac{\mathrm{L}^c_{\mathbf{SM}} \cdot \pi(\mathbf{SM})}{\mathrm{L}^c_{\mathbf{BSM}}(\hat{\mu}) \cdot \pi(\mathbf{BSM})}$$

We search for proto-models and combinations of results / likelihoods that maximize K^c *while remaining compatible with all negative results in our database*.

The Test Statistic

The test statistic $\mathbf{K}^c$ is a likelihood-ratio test that quantifies how much better the proto-model describes the data than the Standard-Model (plus a penalty for model complexity).

Quantifies violation of Standard Model hypothesis

Penalizes for model complexity

$$K^c := -2\ln\frac{\mathrm{L}^c_{\mathbf{SM}}\cdot\pi(\mathbf{SM})}{\mathrm{L}^c_{\mathbf{BSM}}(\hat{\mu})\cdot\pi(\mathbf{BSM})}$$

Priors of the models used to penalize for model complexity

Joint likelihoods: combining “complete” sets of results that are assumed to be approximately uncorrelated.

We search for proto-models and combinations of results / likelihoods that maximize K^c *while remaining compatible with all negative results in our database*.

Input Data

The test statistic is based on likelihoods.

- likelihood computation based on simplified models results in SModelS database
- vast number efficiency and upper limit maps from ~ 50 CMS and ~ 50 ATLAS publications.
- Assume simplified statistical models “behind” the data → simplified likelihoods

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
1	CMS-PAS-EXO-16-036	hscp search	ul, eff	12.9
2	CMS-PAS-SUS-16-052	soft l, <= 2 jets	ul, eff	35.9
3	CMS-SUS-16-009	multijets + $\not\!\!E_T$, top tagging	ul	2.3
4	CMS-SUS-16-032	Sbottom and compressed stop	ul	35.9
5	CMS-SUS-16-033	0ℓ + jets + $\not\!\!E_T$	ul, eff	35.9
6	CMS-SUS-16-034	2 OSSF l's	ul	35.9
7	CMS-SUS-16-035	2 SS l's	ul	35.9
8	CMS-SUS-16-036	0ℓ + jets + $\not\!\!E_T$	ul	35.9
9	CMS-SUS-16-037	1ℓ + jets + $\not\!\!E_T$ with MJ	ul	35.9
10	CMS-SUS-16-039	multi-l EWK searches	ul	35.9
11	CMS-SUS-16-041	multi-ls + jets + $\not\!\!E_T$	ul	35.9
12	CMS-SUS-16-042	1ℓ + jets + $\not\!\!E_T$	ul	35.9
13	CMS-SUS-16-043	EWK WH	ul	35.9
14	CMS-SUS-16-045	Sbottom to bHbH and H $\to \gamma\gamma$	ul	35.9
15	CMS-SUS-16-046	γ + $\not\!\!E_T$	ul	35.9
16	CMS-SUS-16-047	γ + HT	ul	35.9
17	CMS-SUS-16-049	All hadronic stop	ul	35.9
18	CMS-SUS-16-050	0ℓ + top tag	ul	35.9
19	CMS-SUS-16-051	1ℓ stop	ul	35.9
20	CMS-SUS-17-001	Stop search in dil + jets + $\not\!\!E_T$	ul	35.9
21	CMS-SUS-17-003	2 taus + $\not\!\!E_T$	ul	35.9
22	CMS-SUS-17-004	EW-ino combination	ul	35.9
23	CMS-SUS-17-005	1ℓ + multijets + $\not\!\!E_T$, top tagging	ul	35.9
24	CMS-SUS-17-006	jets + boosted H(bb) + $\not\!\!E_T$	ul	35.9
25	CMS-SUS-17-009	SFOS l's + $\not\!\!E_T$	ul	35.9
26	CMS-SUS-17-010	2L stop	ul	35.9
27	CMS-SUS-18-002	γ, jets, b-jets+ $\not\!\!E_T$, top tagging	ul	35.9
28	CMS-SUS-19-006	0ℓ + jets, MHT	ul	137.0

17	CMS-SUS-14-010	b-jets + 4 Ws	ul	19.5
18	CMS-SUS-14-021	soft l's, low n_{jets}, high $\not\!\!E_T$	ul	19.7

#	ID	Short Description	Type	$\mathcal{L}$ [fb^{-1}]
1	ATLAS-SUSY-2015-01	2 b-jets + $\not\!\!E_T$	ul	3.2
2	ATLAS-SUSY-2015-02	single l stop	ul, eff	3.2
3	ATLAS-SUSY-2015-06	0 l's + 2-6 jets + $\not\!\!E_T$	eff	3.2
4	ATLAS-SUSY-2015-09	jets + 2 SS l's or >=3 l's	ul	3.2
5	ATLAS-SUSY-2016-07	0ℓ + jets + $\not\!\!E_T$	ul, eff	36.1
6	ATLAS-SUSY-2016-14	2 SS or 3 l's + jets + $\not\!\!E_T$	ul	36.1
7	ATLAS-SUSY-2016-15	0ℓ stop	ul	36.1
8	ATLAS-SUSY-2016-16	1ℓ stop	ul, eff	36.1
9	ATLAS-SUSY-2016-17	2 opposite sign l's + $\not\!\!E_T$	ul	36.1
10	ATLAS-SUSY-2016-19	stops to staus	ul	36.1
11	ATLAS-SUSY-2016-24	2-3 l's + $\not\!\!E_T$, EWino	ul, eff	36.1
12	ATLAS-SUSY-2016-26	>=2 c jets + $\not\!\!E_T$	ul	36.1
13	ATLAS-SUSY-2016-27	jets + γ + $\not\!\!E_T$	ul, eff	36.1
14	ATLAS-SUSY-2016-28	2 b-jets + $\not\!\!E_T$	ul	36.1
15	ATLAS-SUSY-2016-33	2 OSSF l's + $\not\!\!E_T$	ul	36.1
16	ATLAS-SUSY-2017-01	EWK WH(bb) + $\not\!\!E_T$	ul	36.1
17	ATLAS-SUSY-2017-02	0ℓ + jets + $\not\!\!E_T$	ul	36.1
18	ATLAS-SUSY-2017-03	multi-l EWK searches	ul	36.1
19	ATLAS-SUSY-2018-04	2 hadronic taus	ul, eff	139.0
20	ATLAS-SUSY-2018-06	3 l's EW-ino	ul	139.0
21	ATLAS-SUSY-2018-31	2b + 2H(bb) + $\not\!\!E_T$	ul, eff	139.0
22	ATLAS-SUSY-2018-32	2 OS l's + $\not\!\!E_T$	ul	139.0
23	ATLAS-SUSY-2019-08	1ℓ + higgs + $\not\!\!E_T$	ul, eff	139.0

14	ATLAS-SUSY-2013-19	2 OS l's + (b-)jets + $\not\!\!E_T$	ul	20.3
15	ATLAS-SUSY-2013-21	monojet or c-jet + $\not\!\!E_T$	eff	20.3
16	ATLAS-SUSY-2013-23	1ℓ + 2 b-jets (or 2 γs) + $\not\!\!E_T$	ul	20.3
17	ATLAS-SUSY-2014-03	>= 2(c-)jets + $\not\!\!E_T$	eff	20.3

https://smodels.github.io/docs/ListOfAnalyses124

Input Data

The test statistic is based on likelihoods.

- likelihood computation based on simplified models results in SModelS database
- vast number efficiency and upper limit maps from ~ 50 CMS and ~ 50 ATLAS publications.
- Assume simplified statistical models “behind” the data → simplified likelihoods

#	ID	Short Description		
1	CMS-PAS-EXO-16-036	hscp search		
2	CMS-PAS-SUS-16-052	soft l, <= 2 jets		
3	CMS-SUS-16-009	multijets + $\not{E}_T$, top tagging		
4	CMS-SUS-16-032	Sbottom and compressed sto		
5	CMS-SUS-16-033	0ℓ + jets + $\not{E}_T$		
6	CMS-SUS-16-034	2 OSSF l's		
7	CMS-SUS-16-035	2 SS l's		
8	CMS-SUS-16-036	0ℓ + jets + $\not{E}_T$		
9	CMS-SUS-16-037	1ℓ + jets + $\not{E}_T$ with MJ		
10	CMS-SUS-16-039	multi-l EWK searches		
11	CMS-SUS-16-041	multi-ls + jets + $\not{E}_T$		
12	CMS-SUS-16-042	1ℓ + jets + $\not{E}_T$	ul	35.9
13	CMS-SUS-16-043	EWK WH	ul	35.9
14	CMS-SUS-16-045	Sbottom to bHbH and H → $\gamma\gamma$	ul	35.9
15	CMS-SUS-16-046	γ + $\not{E}_T$	ul	35.9
16	CMS-SUS-16-047	γ + HT	ul	35.9
17	CMS-SUS-16-049	All hadronic stop	ul	35.9
18	CMS-SUS-16-050	0ℓ + top tag	ul	35.9
19	CMS-SUS-16-051	1ℓ stop	ul	35.9
20	CMS-SUS-17-001	Stop search in dil + jets + $\not{E}_T$	ul	35.9
21	CMS-SUS-17-003	2 taus + $\not{E}_T$	ul	35.9
22	CMS-SUS-17-004	EW-ino combination	ul	35.9
23	CMS-SUS-17-005	1ℓ + multijets + $\not{E}_T$, top tagging	ul	35.9
24	CMS-SUS-17-006	jets + boosted H(bb) + $\not{E}_T$	ul	35.9
25	CMS-SUS-17-009	SFOS l's + $\not{E}_T$	ul	35.9
26	CMS-SUS-17-010	2L stop	ul	35.9
27	CMS-SUS-18-002	γ, jets, b-jets+ $\not{E}_T$, top tagging	ul	35.9
28	CMS-SUS-19-006	0ℓ + jets, MHT	ul	137.0

18	CMS-SUS-14-021	soft l's, low n_{jets}, high $\not{E}_T$	ul	19.7

Ongoing work for next iteration:

moving to full likelihoods, **machine-learning surrogate models** (multi-layer perceptrons, normalizing flows) as accelerators

[Humberto Gonzales, PostDoc in Genova]

12	ATLAS-SUSY-2016-26	>=2 c jets + $\not{E}_T$	ul	36.1
13	ATLAS-SUSY-2016-27	jets + γ + $\not{E}_T$	ul, eff	36.1
14	ATLAS-SUSY-2016-28	2 b-jets + $\not{E}_T$	ul	36.1
15	ATLAS-SUSY-2016-33	2 OSSF l's + $\not{E}_T$	ul	36.1
16	ATLAS-SUSY-2017-01	EWK WH(bb) + $\not{E}_T$	ul	36.1
17	ATLAS-SUSY-2017-02	0ℓ + jets + $\not{E}_T$	ul	36.1
18	ATLAS-SUSY-2017-03	multi-l EWK searches	ul	36.1
19	ATLAS-SUSY-2018-04	2 hadronic taus	ul, eff	139.0
20	ATLAS-SUSY-2018-06	3 l's EW-ino	ul	139.0
21	ATLAS-SUSY-2018-31	2b + 2H(bb) + $\not{E}_T$	ul, eff	139.0
22	ATLAS-SUSY-2018-32	2 OS l's + $\not{E}_T$	ul	139.0
23	ATLAS-SUSY-2019-08	1ℓ + higgs + $\not{E}_T$	ul, eff	139.0

14	ATLAS-SUSY-2013-19	2 OS l's + (b-)jets + $\not{E}_T$	ul	20.3
15	ATLAS-SUSY-2013-21	monojet or c-jet + $\not{E}_T$	eff	20.3
16	ATLAS-SUSY-2013-23	1ℓ + 2 b-jets (or 2 γs) + $\not{E}_T$	ul	20.3
17	ATLAS-SUSY-2014-03	>= 2(c-)jets + $\not{E}_T$	eff	20.3

https://smodels.github.io/docs/ListOfAnalyses124

Input Data

The test statistic is based on likelihoods.

- likelihood computation based on simplified models results in SModelS database
- vast number efficiency and upper limit maps from ~ 50 CMS and ~ 50 ATLAS publications.
- Assume simplified statistical models "behind" the data → simplified likelihoods

#	ID	Short Description		
1	CMS-PAS-EXO-16-036	hscp search		
2	CMS-PAS-SUS-16-052	soft l, <= 2 jets		
3	CMS-SUS-16-009	multijets + $\not{E}_T$, top tagging		
4	CMS-SUS-16-032	Sbottom and compressed sto		
5	CMS-SUS-16-033	0ℓ + jets + $\not{E}_T$		
6	CMS-SUS-16-034	2 OSSF l's		
7	CMS-SUS-16-035	2 SS l's		
8	CMS-SUS-16-036	0ℓ + jets + $\not{E}_T$		
9	CMS-SUS-16-037	1ℓ + jets + $\not{E}_T$ with MJ		
10	CMS-SUS-16-039	multi-l EWK searches		
11	CMS-SUS-16-041	multi-ls + jets + $\not{E}_T$		
12	CMS-SUS-16-042	1ℓ + jets + $\not{E}_T$	ul	35.9
13	CMS-SUS-16-043			
14	CMS-SUS-16-045			
15	CMS-SUS-16-046			
16	CMS-SUS-16-047			
17	CMS-SUS-16-049			
18	CMS-SUS-16-050			
19	CMS-SUS-16-051			
20	CMS-SUS-17-001			
21	CMS-SUS-17-003	[2 taus + $\not{E}_T$]	ul	
22	CMS-SUS-17-004	EW-ino combination	ul	35.9
23	CMS-SUS-17-005	1ℓ + multijets + $\not{E}_T$, top tagging	ul	35.9
24	CMS-SUS-17-006	jets + boosted H(bb) + $\not{E}_T$	ul	35.9
25	CMS-SUS-17-009	SFOS l's + $\not{E}_T$	ul	35.9
26	CMS-SUS-17-010	2L stop	ul	35.9
27	CMS-SUS-18-002	γ, jets, b-jets+ $\not{E}_T$, top tagging	ul	35.9
28	CMS-SUS-19-006	0ℓ + jets, MHT	ul	137.0

17	CMS-SUS-14-010	b-jets + 4 Ws	ul	19.5
18	CMS-SUS-14-021	soft l's, low n_{jets}, high $\not{E}_T$	ul	19.7

12	ATLAS-SUSY-2016-26	>=2 c jets + $\not{E}_T$	ul	36.1
			ul, eff	36.1
			ul	36.1
			ul	36.1
			ul	36.1
			ul	36.1
			ul	36.1
			ul, eff	139.0
			ul	139.0
21	ATLAS-SUSY-2018-31	2b + 2H(bb) + $\not{E}_T$	ul, eff	139.0
22	ATLAS-SUSY-2018-32	2 OS l's + $\not{E}_T$	ul	139.0
23	ATLAS-SUSY-2019-08	1ℓ + higgs + $\not{E}_T$	ul, eff	139.0

14	ATLAS-SUSY-2013-19	2 OS l's + (b-)jets + $\not{E}_T$	ul	20.3
15	ATLAS-SUSY-2013-21	monojet or c-jet + $\not{E}_T$	eff	20.3
16	ATLAS-SUSY-2013-23	1ℓ + 2 b-jets (or 2 γs) + $\not{E}_T$	ul	20.3
17	ATLAS-SUSY-2014-03	>= 2(c-)jets + $\not{E}_T$	eff	20.3

Ongoing work for next iteration:

moving to full likelihoods, **machine-learning surrogate models** (multi-layer perceptrons, normalizing flows) as accelerators

[Humberto Gonzales, PostDoc in Genova]

Also: started work to include measurements (via rivet) in addition to these searches

[Jamie Yellen, PhD student (Glasgow)]

https://smodels.github.io/docs/ListOfAnalyses124

The Combiner

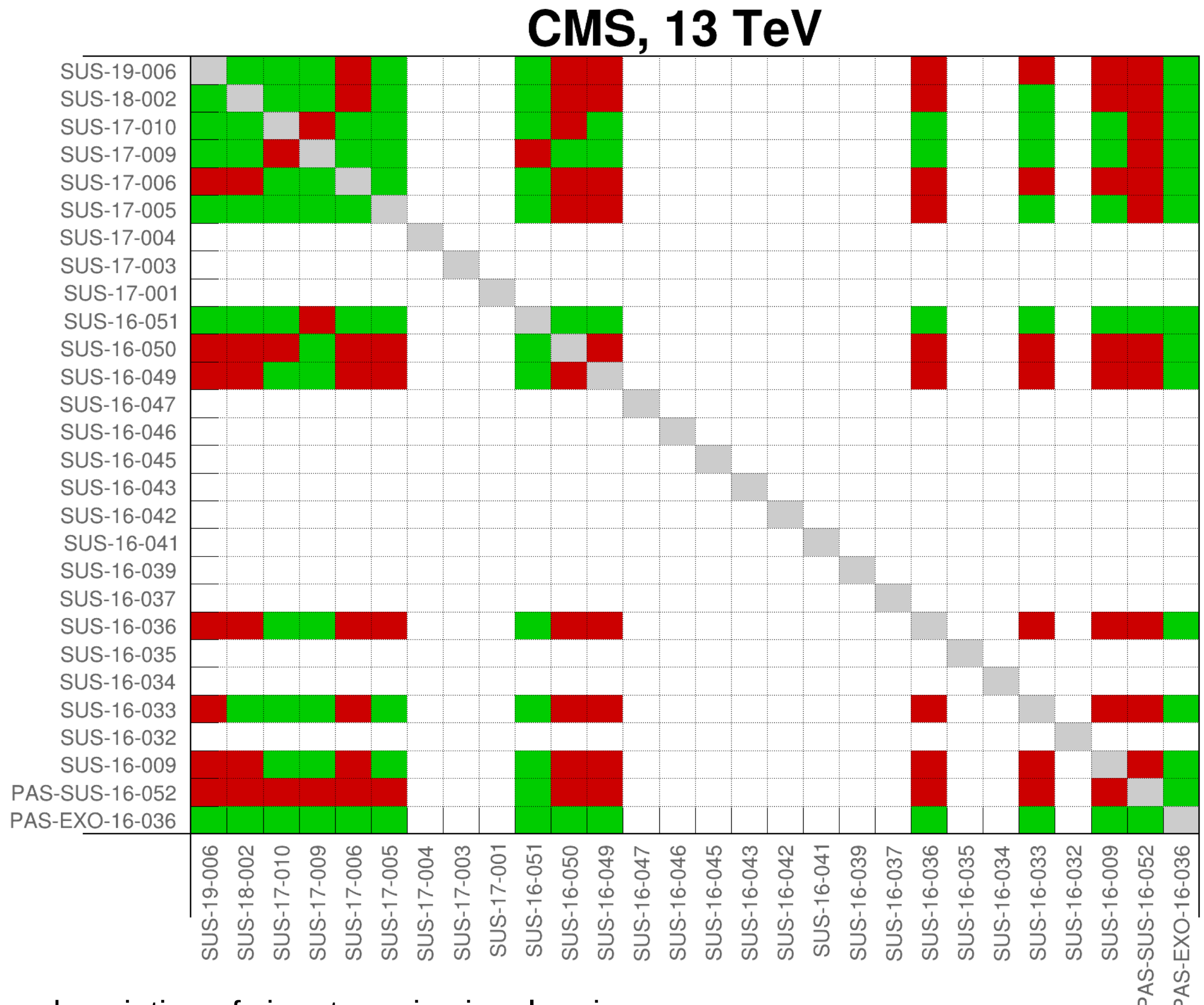
As we are chasing dispersed signals, we need to allow the machine to combine (i.e. multiply) likelihoods. Simplified, binaric “inter-analyses *exclusivity* matrix”:

green: approximately uncorrelated → combinable

red: correlated, not combinable

White: cannot construct a likelihood

Signal regions within each analysis: correlated



In this publication: “educated guesses” from description of signatures in signal regions.

The Combiner

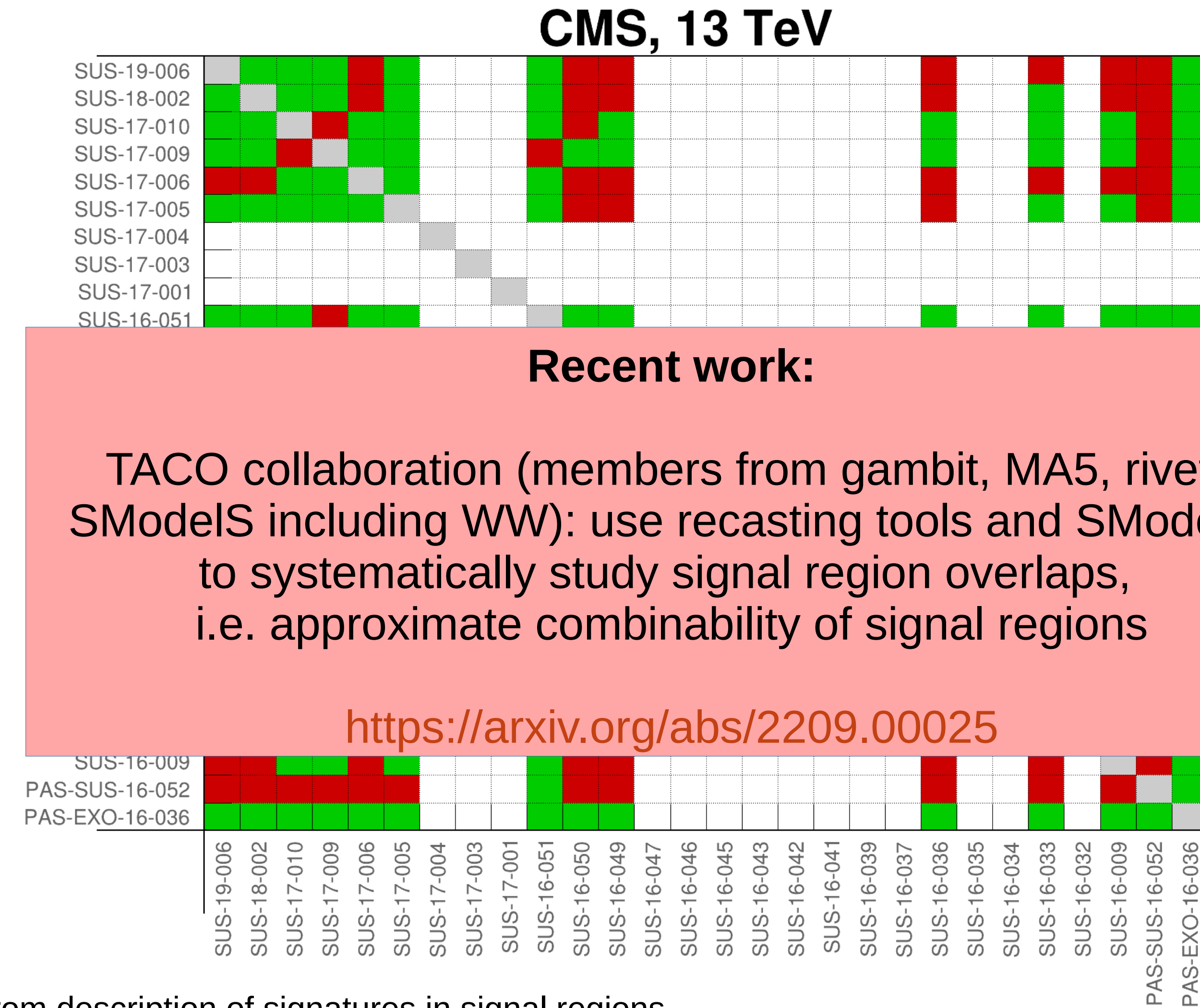
As we are chasing dispersed signals, we need to allow the machine to combine (i.e. multiply) likelihoods. Simplified, binaric “inter-analyses *exclusivity* matrix”:

green: approximately uncorrelated → combinable

red: correlated, not combinable

White: cannot construct a likelihood

Signal regions within each analysis: correlated



Recent work:

TACO collaboration (members from gambit, MA5, rivet, SModelS including WW): use recasting tools and SModelS to systematically study signal region overlaps, i.e. approximate combinability of signal regions

https://arxiv.org/abs/2209.00025

In this publication: “educated guesses” from description of signatures in signal regions.

The Penalty Term

For every legal combination, we define a test statistic K^c

$$K^c := -2 \ln \frac{\mathrm{L}^c_{\mathrm{SM}} \cdot \pi(\mathrm{SM})}{\mathrm{L}^c_{\mathrm{BSM}}(\hat{\mu}) \cdot \pi(\mathrm{BSM})}$$

π(BSM) is the prior of the BSM model. We use it to “regularize” the model, i.e. impose the *law of parsimony*:

$$\pi(\mathrm{BSM}) \approx \exp[-\mathrm{n}^{\mathrm{BSM}}_{\mathrm{particles}}]$$

Resulting in a test statistic that resembles an “Akaike Information Criterion” (AIC):

$$K^c \approx \Delta\chi^2 - 2n^{\mathrm{BSM}}_{\mathrm{particles}}$$

An additional BSM particle will have to increase the (delta-)chi-square by approximately two units.

The Penalty Term

For every legal combination, we define a test statistic K^c

$$K^c := -2 \ln \frac{\mathrm{L}^c_{\mathrm{SM}} \cdot \pi(\mathrm{SM})}{\mathrm{L}^c_{\mathrm{BSM}}(\hat{\mu}) \cdot \pi(\mathrm{BSM})}$$

π(BSM) is the prior of the BSM model. We use it to “regularize” the model, i.e. impose the *law of parsimony*:

$$\pi(\mathrm{BSM}) \approx \exp[-\mathrm{n}^{\mathrm{BSM}}_{\mathrm{particles}}]$$

Resulting in a test statistic that r

An additional BSM particle
approximately two units.

Ongoing work:

Extend penalty term of test statistic to not only correctly penalize for model complexity, but also for number of experimental results

With: Jamie Yellen (PhD student, Glasgow)

The Critic

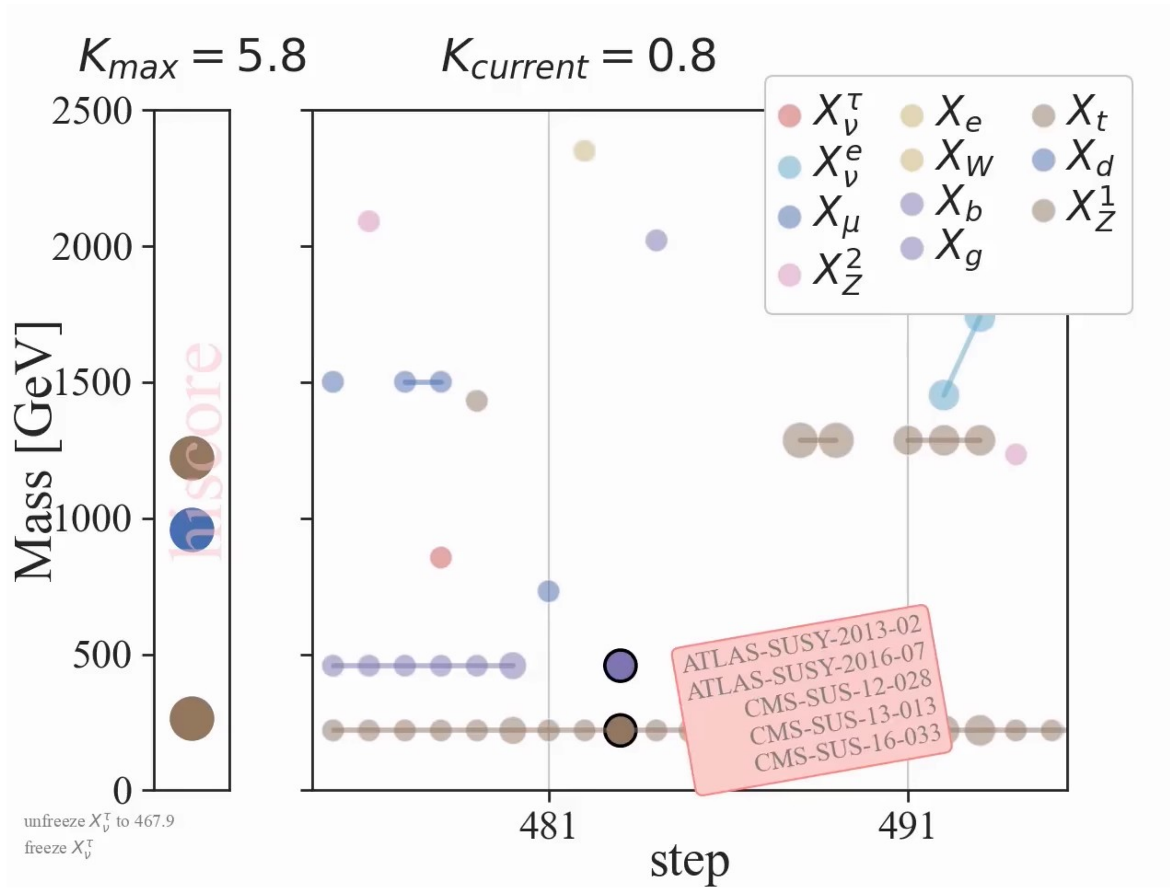
For every legal combination *c*, we define a test statistic K^c

$$K^c := -2 \ln \frac{\mathrm{L}^c_{\mathrm{SM}} \cdot \pi(\mathrm{SM})}{\mathrm{L}^c_{\mathrm{BSM}}(\hat{\mu}) \cdot \pi(\mathrm{BSM})}$$

$$\Rightarrow K = \max\{K^c \mid \forall \text{ combinations}\}$$

- $\hat{\mu}$ is the signal strength of the model that maximizes the likelihood.
- By limiting its support we guarantee compatibility with all negative results in the SModelS database.
- In allusion to adversarial setups, we also call this feature the critic

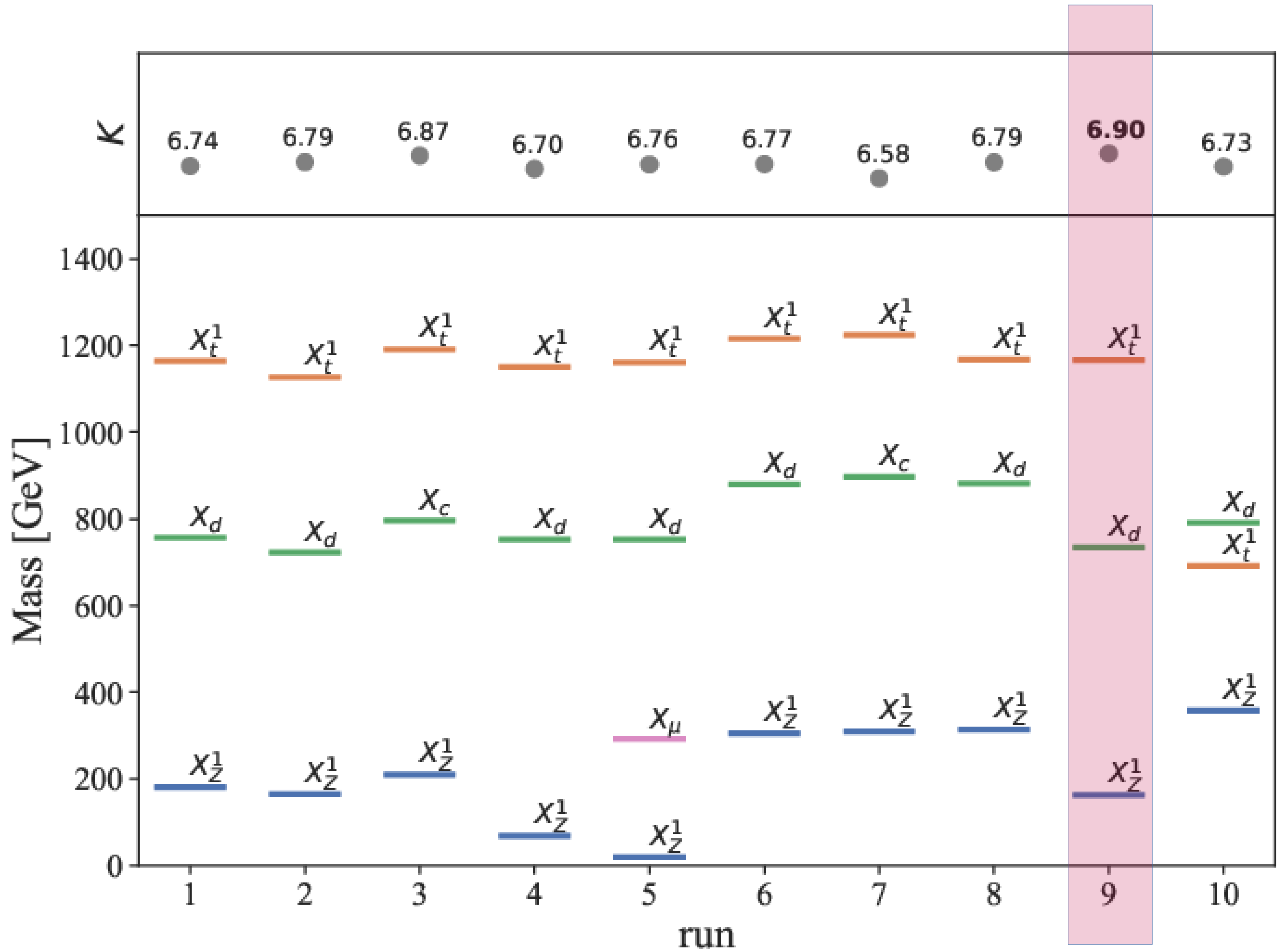
And Then We Ran the Algorithm …



We defined a “run” as 50 parallel walkers, making 1,000 steps each.
We performed 10 such runs on the SModelS database.
Total computing resources spent: ~ 1,000,000 CPU hours

… and Obtained Results

We performed 10 such runs on the SModelS database:

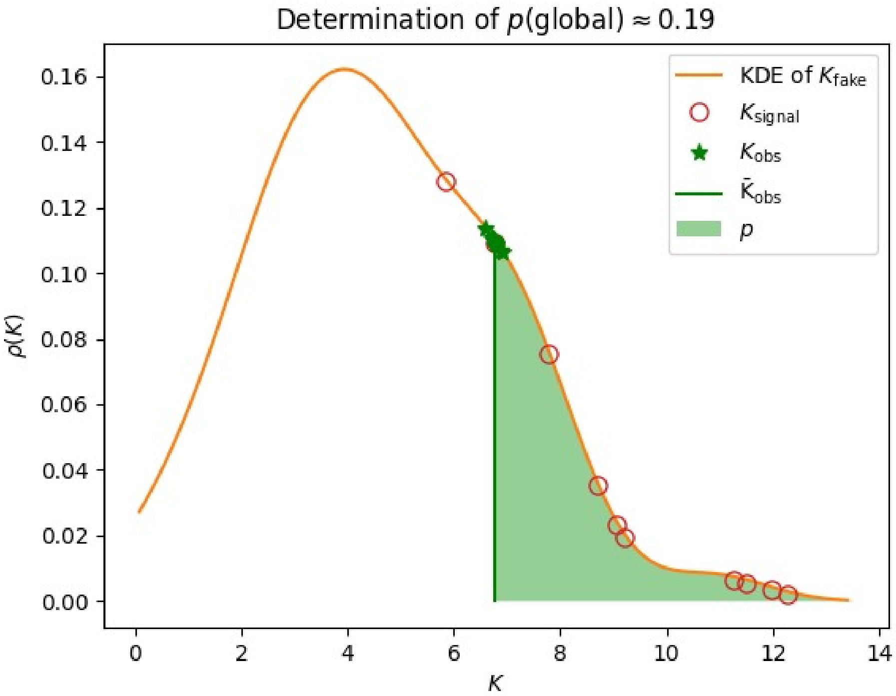


All 10 runs introduced a top partner as well as a light quark partner. The cross sections are compatible with values expected from the MSSM. The best test statistic was K=6.9.

Global p-Value

- Because we have statistical models of the search results, we can synthesize statistically correct databases of results that are “typical”, if no new physics is in the data.

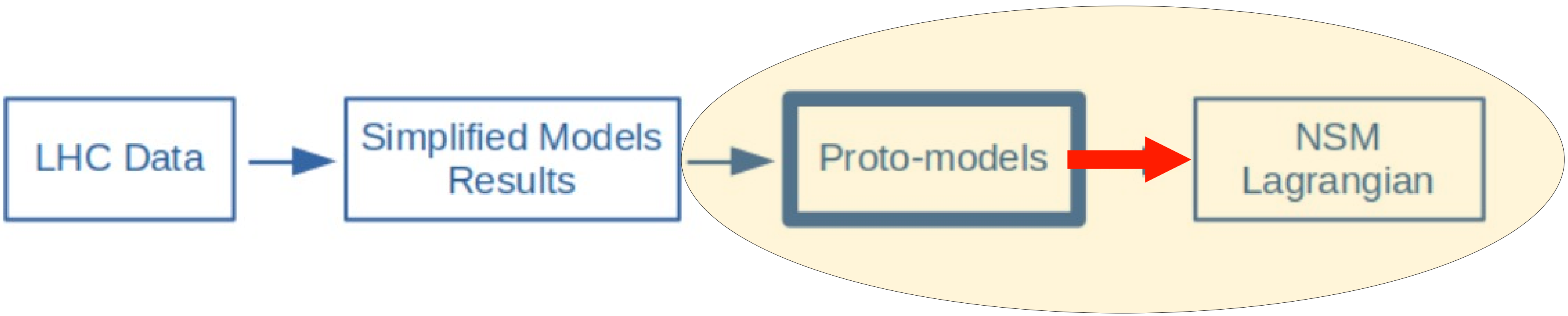
- From this we can compute a *p-value for the Standard Model hypothesis*: that is the chances that – under the SM hypothesis – we would obtain a result as extreme as ours or more extreme.



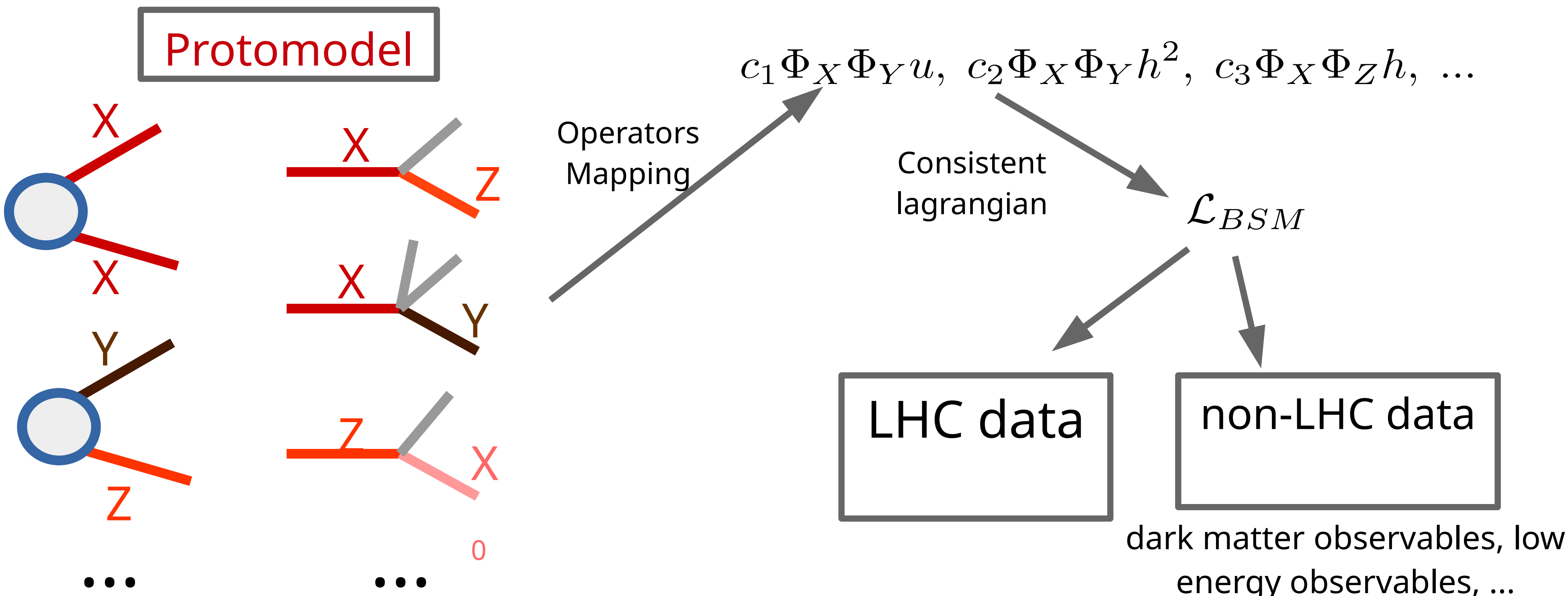
Since we did not correct for the conservativeness of the experimental results, we assume our result to also be conservative.

By construction, no Look-Elsewhere Effect applies. (within the database, the machine does look everywhere)

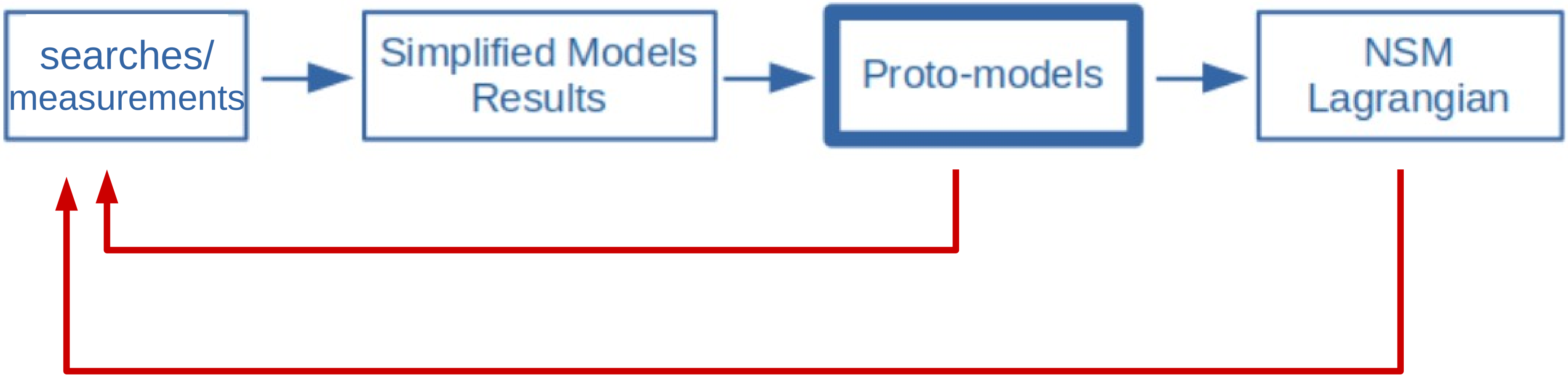
Towards UV Completion



Work from protomodels to UV complete theories has recently begun (John Gargalionis, PostDoc in Valencia) – lot's of combinatorics!



ITERATIONS



We of course envisage the procedure to be iterative:

The pheno community can point out models and combinations of interest, interpret results, make predictions about potential channels, provide feedback to the experiments

Only the experimental communities can claim potential (dispersed) discoveries.

Future: Fully Differentiable Chain of Inductive Reasoning

If we had gradients we could perform gradient descent to find the best model, and we could use e.g. the Fisher information to infer the error on its parameters (or, alternatively we can then MCMC-sample).

LHC Data → Simplified Models Results → Proto-models → NSM Lagrangian

- LHC Data: described as likelihoods L that are differentiable with respect to the yields y_i
- Simplified Models Results: SModelS is – in principle -- differentiable
- Proto-models: that's just a sum of simplified models → differentiable!
- NSM Lagrangian: for individual candidates we can make this differentiable w.r.t fundamental parameters Θ_l, via neural networks

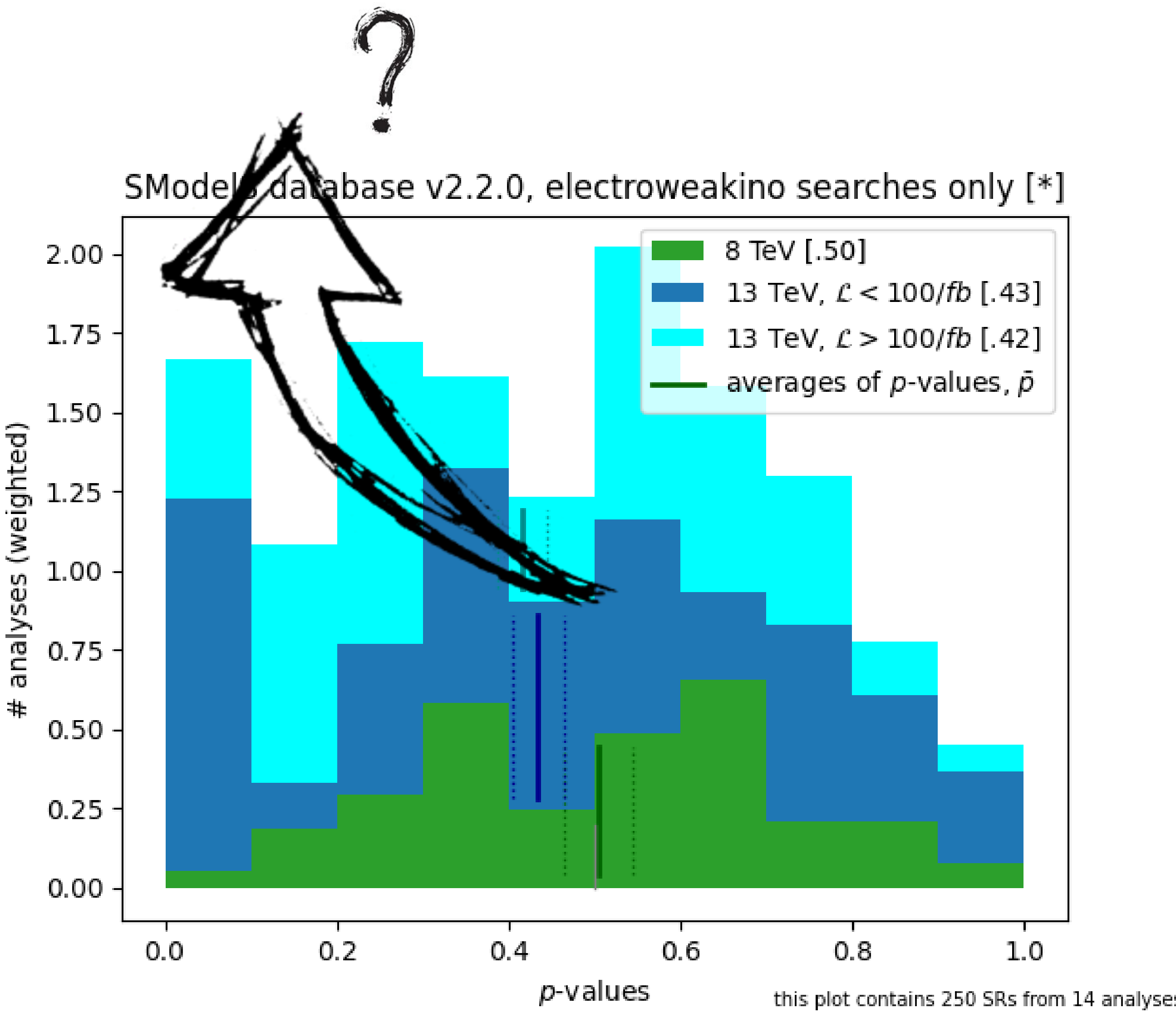
$$\frac{\partial L}{\partial \theta_l} = \frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial p_j} \cdot \frac{\partial p_j}{\partial (m_k, \Gamma_k, \sigma_k)} \cdot \frac{\partial (m_k, \Gamma_k, \sigma_k)}{\partial \theta_l}$$

Not yet required (theory space as well as space of measurements are still low-dimensional enough).

Summary, Outlook

- In light of no clear evidence for new physics in the individual channels/results, a more global attempt at finding new physics seems appropriate
- First prototype run of a machine that builds protomodels with results from ~ 100 analyses resulted in p-value of SM hypothesis of ~ 0.2: a very small tension with the Standard Model hypothesis (but also some tension between some results)
- Working on next iteration with more results, better likelihoods, surrogate models as accelerators, covering more signatures, larger protomodels space, UV completion
- Extend to measurements, not only searches
- Can we make the entire chain differentiable?

Run-3 will continue the trends or refute them

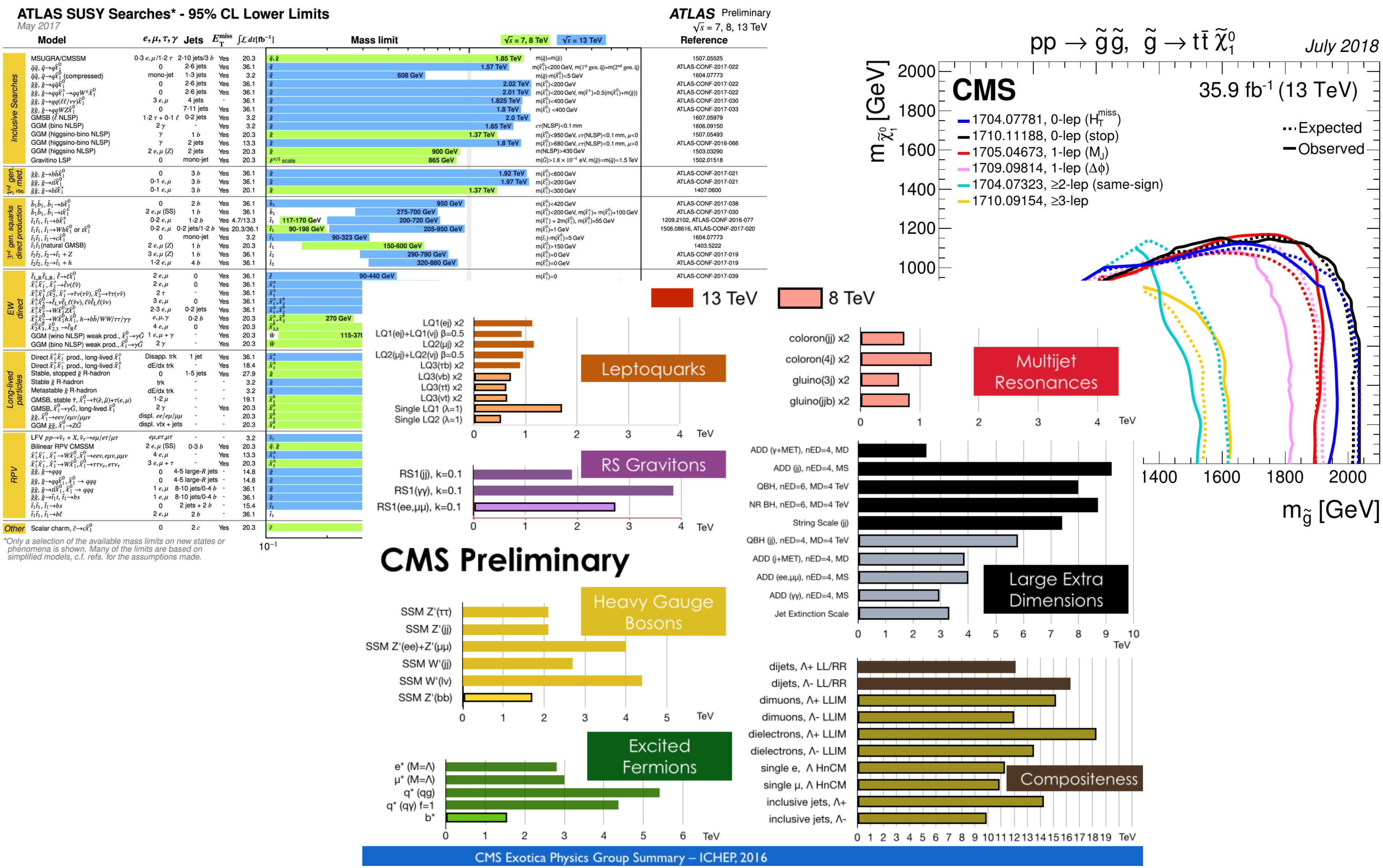


The Jury Is Out [*]

[*] no pun intended

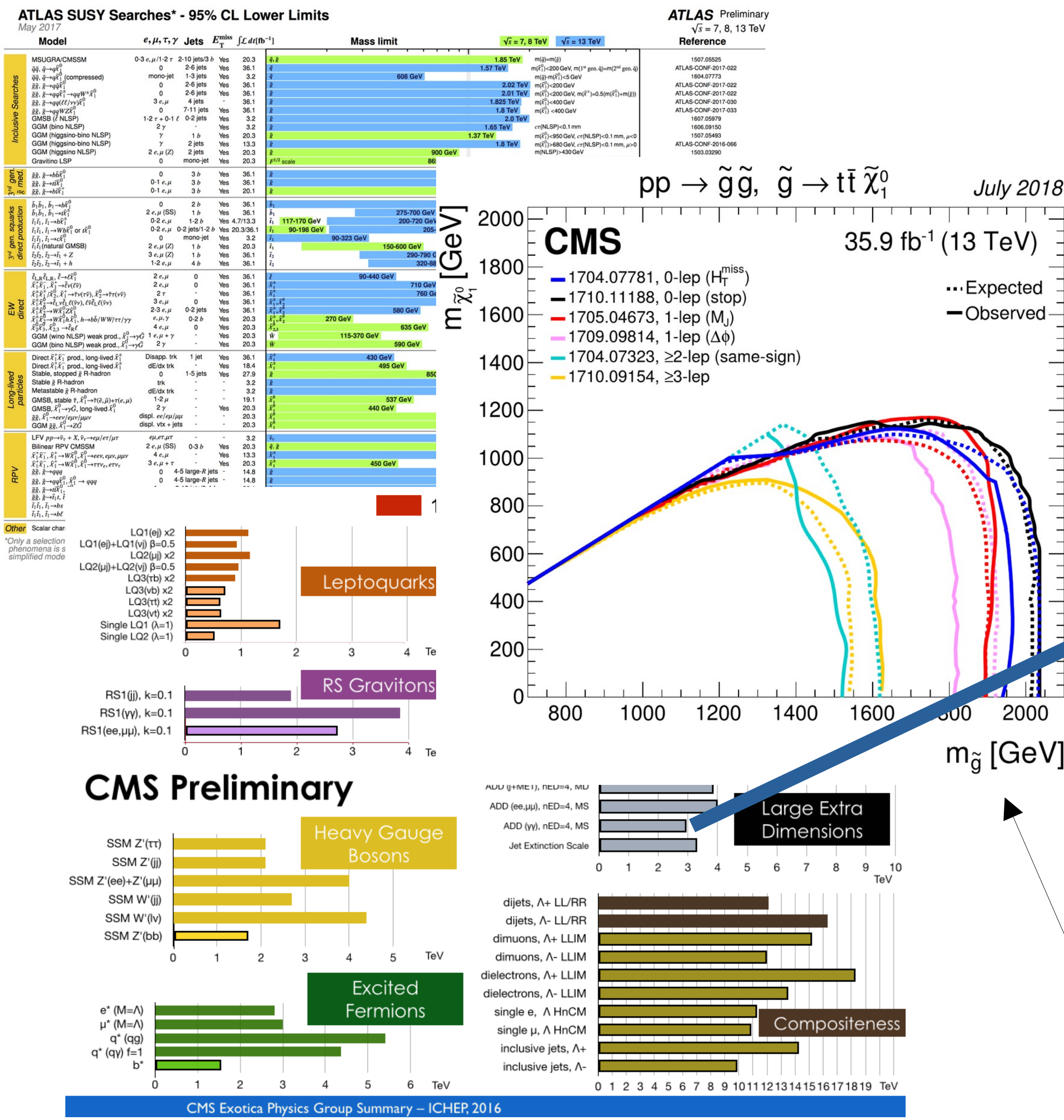
Backup

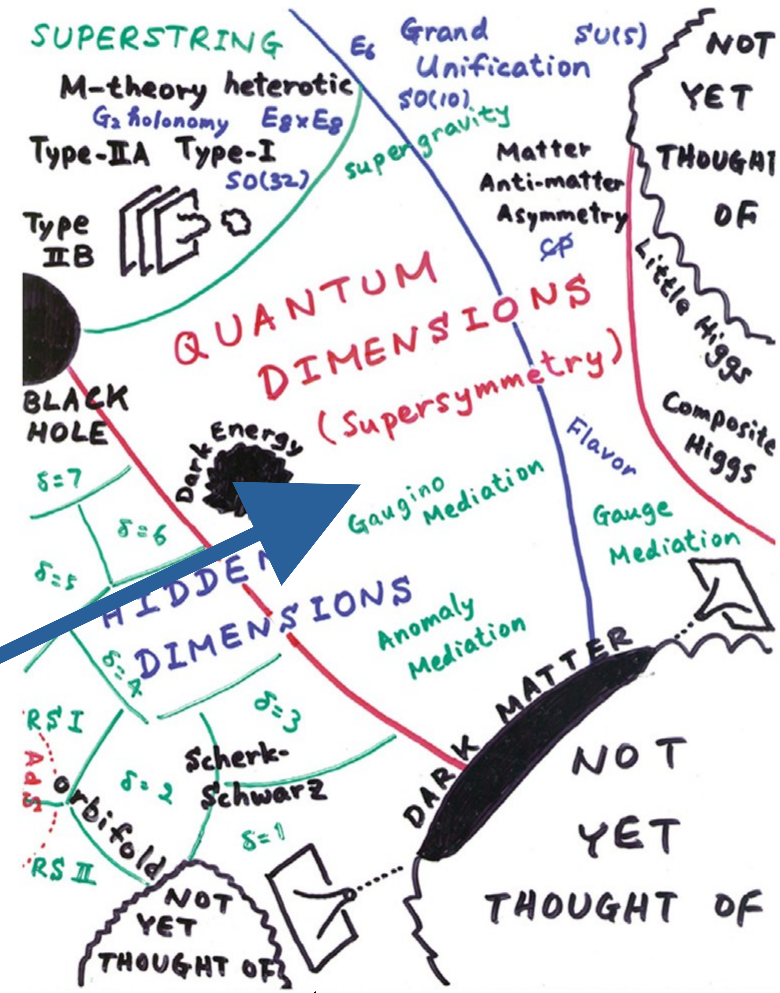
A situation unlike in the Past



Not a handful of experimental signatures. Hundreds of publications with a wide range of signatures!

The Inverse Problem

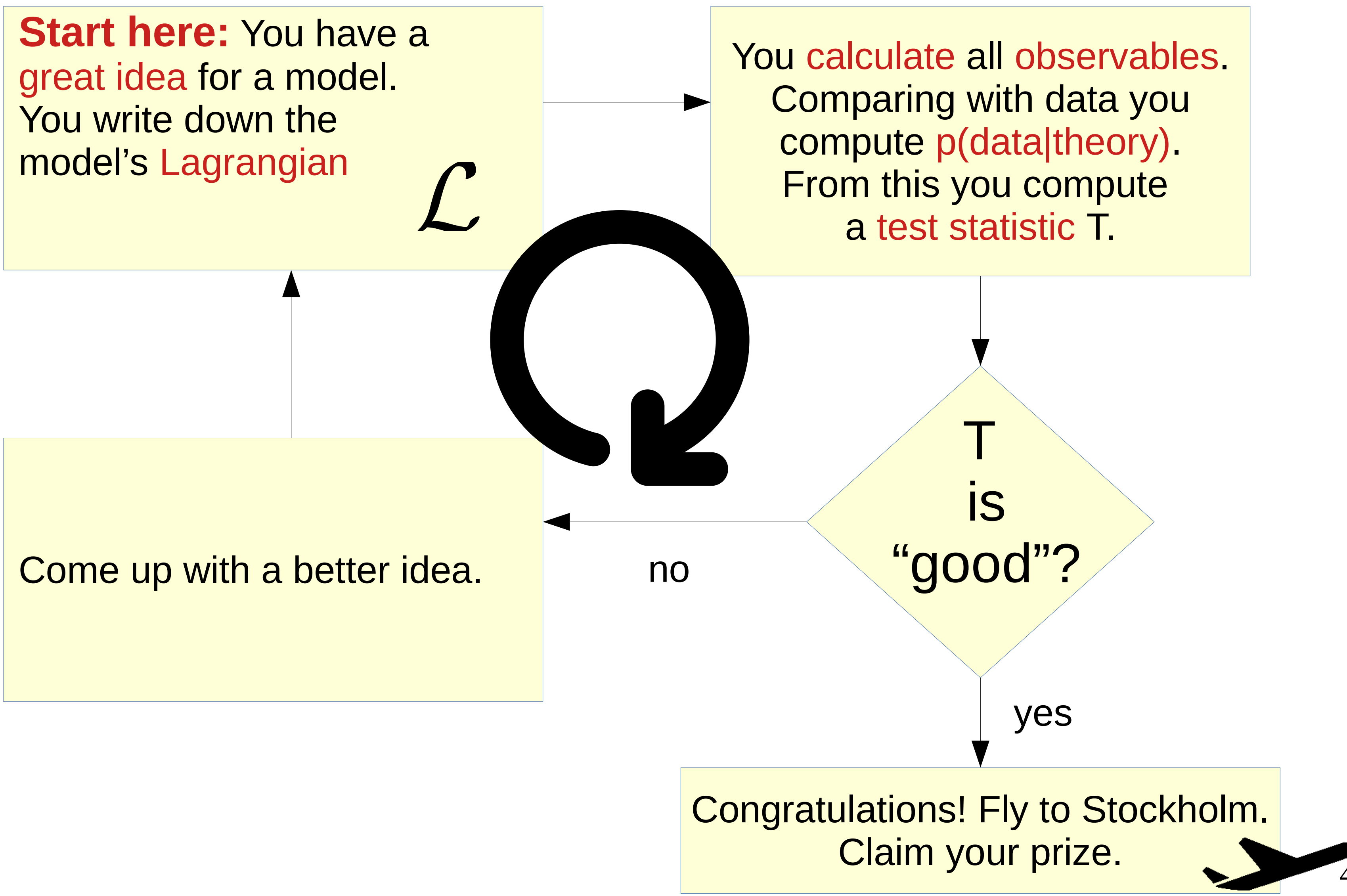




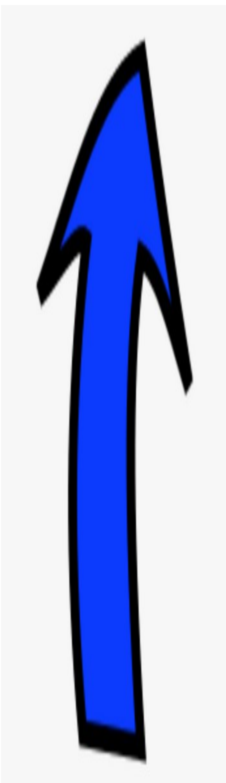
So how do we get from here to here?

Top-down versus bottom-up

Top-Down:



Top-down versus bottom-up



Bottom-Up:

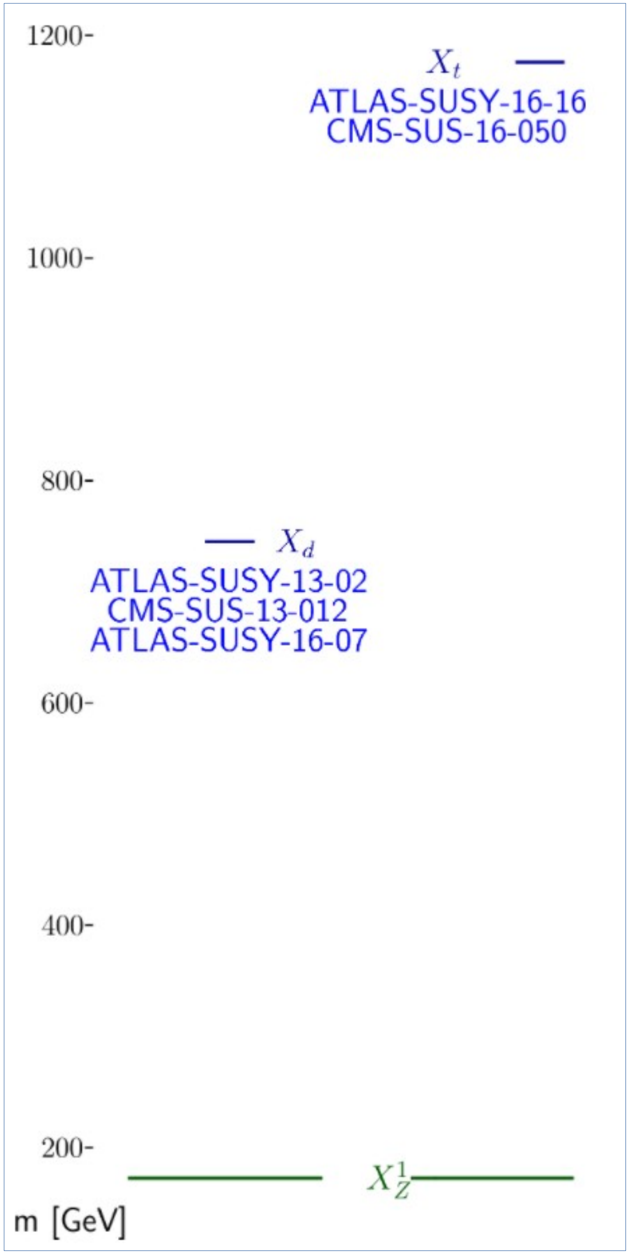
Only now do you think about symmetries, gauge groups, etc that may underlie all observations. Construct your Lagrangian.

$\mathcal{L}$

From the descriptions you try and construct precursor theories to the NSM that describe everything you really know about TeV-scale (and below) physics

Start here: You describe your experimental findings in a language amenable to theoretical physics, e.g. simplified models for on-shell effects ("searches"), effective field theories for off-shell effects ("measurements").

The Hiscore Proto-Model



Analysis	Dataset	Obs	Exp	Z	P	Signal
ATL multijet, 8 TeV [54]	SR6jtp	6	4.9 ± 1.6	$0.4\ \sigma$	X_d	0.25
ATL multijet, 13 TeV [55]	2j_Me ...	611	526 ± 31	$2.2\ \sigma$	X_d	44.18
ATL 1ℓ stop, 13 TeV [48]	tN_high	8	3.8 ± 1	$1.9\ \sigma$	X_t	3.93
CMS multijet, 8 TeV [56]		30.8 fb	19.6 fb	$1.1\ \sigma$	X_d	2.66 fb
CMS 0ℓ stop, 13 TeV [49]		4.5 fb	2.5 fb	$1.6\ \sigma$	X_t	2.62 fb

Table 3: Analyses contributing to the K value of the highest score proto-model

Tension!

the dispersed excess

Analysis (all CMS 13 TeV)	Prod	σ_{XX} (fb)	$\sigma^{\mathrm{UL}}_{\mathrm{obs}}$ (fb)	$\sigma^{\mathrm{UL}}_{\mathrm{exp}}$ (fb)	r_{obs}
CMS multijet, M_{H_T}, 137 fb^{-1} [15]	$(\bar{X}_d, X_d)$	23.96	18.45	21.57	1.30
CMS multijet, M_{H_T}, 137 fb^{-1} [15]	$(\bar{X}_t, X_t)$	2.62	2.04	2.08	1.28
CMS multijet, M_{H_T}, 36 fb^{-1} [57]	$(\bar{X}_d, X_d)$	23.96	19.26	28.31	1.24
CMS multijet, M_{T2}, 36 fb^{-1} [58]	$(\bar{X}_d, X_d)$	23.96	26.02	31.79	0.92
CMS 1ℓ stop, 36 fb^{-1} [59]	$(\bar{X}_t, X_t)$	2.62	2.91	4.44	0.90

Table 4: List of the most constraining results for the highest score proto-model. The

what is driving the "critic"

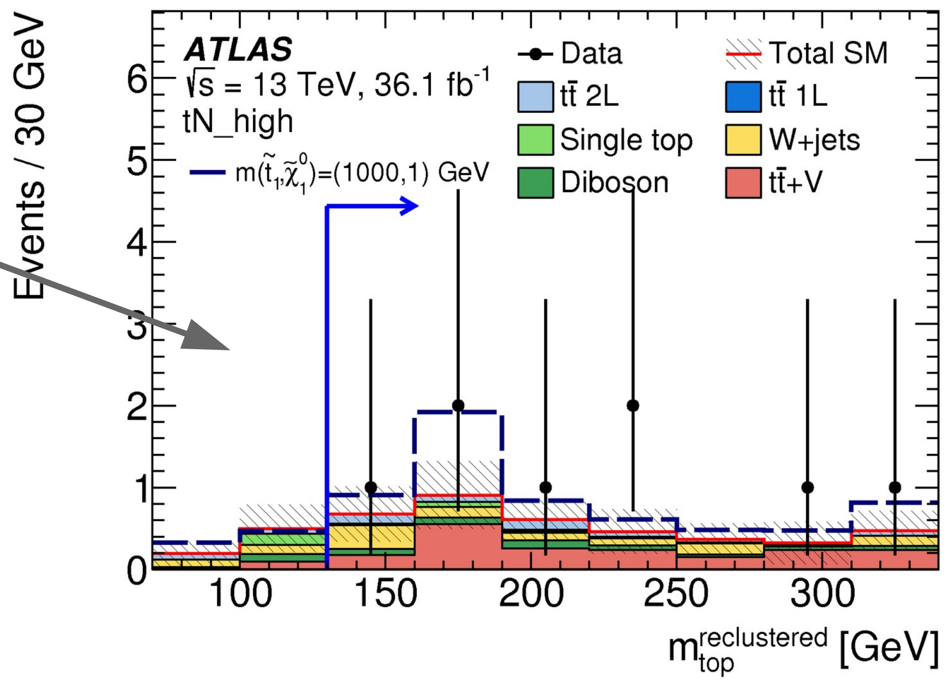
Signal strength multipliers: $(\bar{X}_t, X_t) = 1.2; (\bar{X}_d, X_d), (X_d, X_Z^1), (\bar{X}_d, X_Z^1) = 0.49$

Contributions by particles: $X_t : K_{\mathrm{without}} = 2.59(59\%), X_d : K_{\mathrm{without}} = 3.90(41\%)$
Last updated: Mon Dec 14 20:08:06 2020

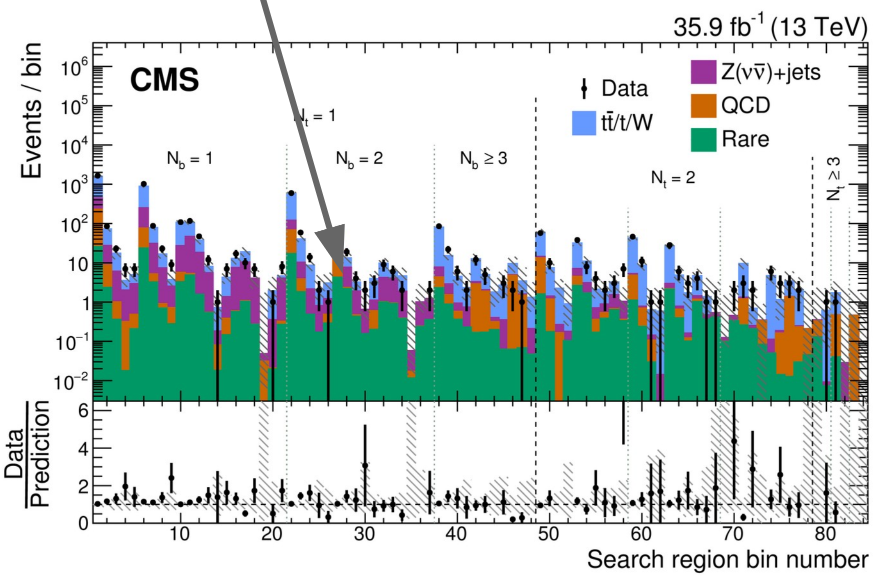
https://smodels.github.io/protomodels/2020_PioneerStudy/real9/index.html

Data driving the protomodel

Nothing unusual is visible in the individual results



ATLAS 1l stop (ATLAS-SUSY-2016-16)



CMS 0l stop (CMS-SUS-16-050)

But when interpreted as a correlated signal originating from the same protomodel, things may seem different!

Likelihoods

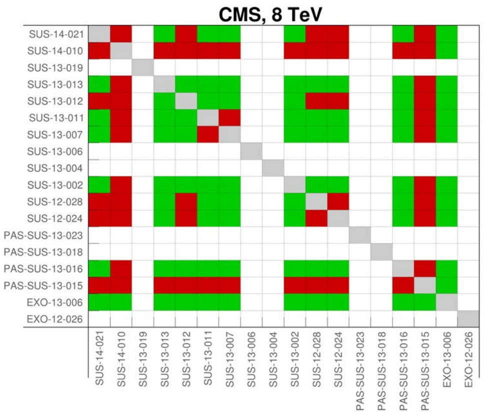
- **Only exclusion lines**
  If only exclusion lines are given, without upper limits, we can do nothing
- **Observed 95% CL upper limits only:**
  cannot construct likelihood, binary decision “excluded” / “not-excluded” only (“critic”)
- **Expected and observed 95% CL upper limits**
  can construct an approximate likelihood with truncated Gaussian, cannot combine topologies, very crude approximation
- **Efficiency maps**
  can construct a likelihood as Gaussian (for the nuisances) * Poissonian (for yields), can work per SR, and combine topologies in each SR [*]
- **Efficiency maps + correlation matrices**
  can combine signal regions via multivariate Gaussian * Poissonians
- **Efficiency maps + full likelihoods**
  full realism, correct statistical model

Combos

Likelihoods

BETTER

[*] if efficiency maps are not supplied, we can try to produce them with recasting frameworks



The Combiner

we allow the machine
to combine likelihooods.

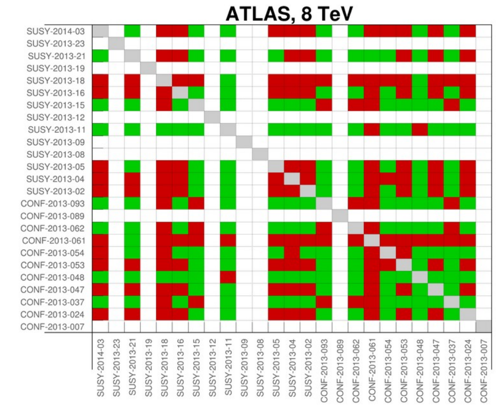
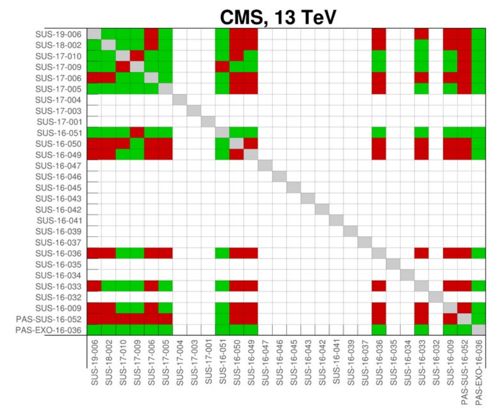


Fig. 2

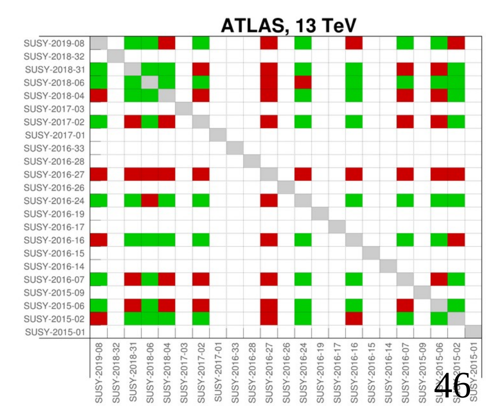
Approximately uncorrelated are analyses that are:

- from different runs, and/or
- from different experiments, and/or
- looking for (clearly) different signatures



A combination “c” of analyses is “legal” if the following conditions are met:

- all results are mutually uncorrelated (= ”combinable”)
- if a result can be added, it has to be added (any subset of a legal combination is not itself legal)
- combined likelihood: $L_c = \prod_{i \in c} L_i$



The Test Statistic

For every legal combination, we define a test statistic *K*

$$K^c := -2\ln\frac{\mathrm{L}^c_{\mathbf{SM}}\cdot\pi(\mathbf{SM})}{\mathrm{L}^c_{\mathbf{BSM}}(\hat{\mu})\cdot\pi(\mathbf{BSM})}$$

Eq. 6

(Remember, we have a database of results from ~ 100 CMS+ATLAS searches. We want to find the most interesting combinations of these results, i.e. the ones that maximally violate the SM hypothesis)

Of all “legal” combinations of experimental results, the builder chooses the one combination “c” that maximizes *K*:

$$K := \max_{\forall c\in C} K^c$$

Eq. 7

μ denotes an global signal strength multiplier – the production cross sections are free parameters

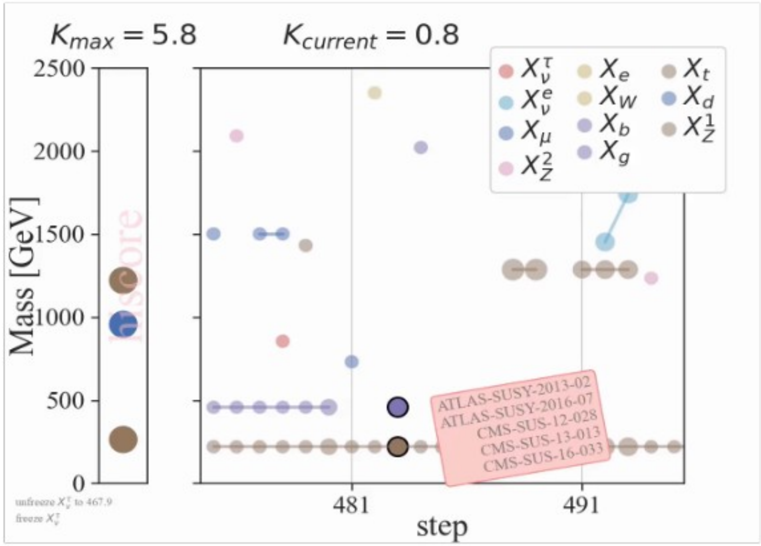
$$\forall i,j : \sigma\left(pp\to X_iX_j\right) = \mu\bar{\sigma}\left(pp\to X_iX_j\right)$$

It is maximized in the denominator, but its support is confined such that no limits in the SModelS database are violated (the “critic”),

$$\hat{\mu}\in[0,\mu_{\mathrm{max}}]$$

The Walker

The Walker takes care of moving in the protomodel space with varying dimensionality by performing the following types of modifications to the protomodel:



- add or remove particles from the protomodel
- change the masses of particles
- change the signal strengths of production modes
- change decay channels and branching ratios

At each step the test statistic *K* is computed. An MCMC-like procedure[*] is then applied in the sense that the step is reverted with a probability of

$$\exp\left[\frac{1}{2}(K_i - K_{i-1})\right]$$

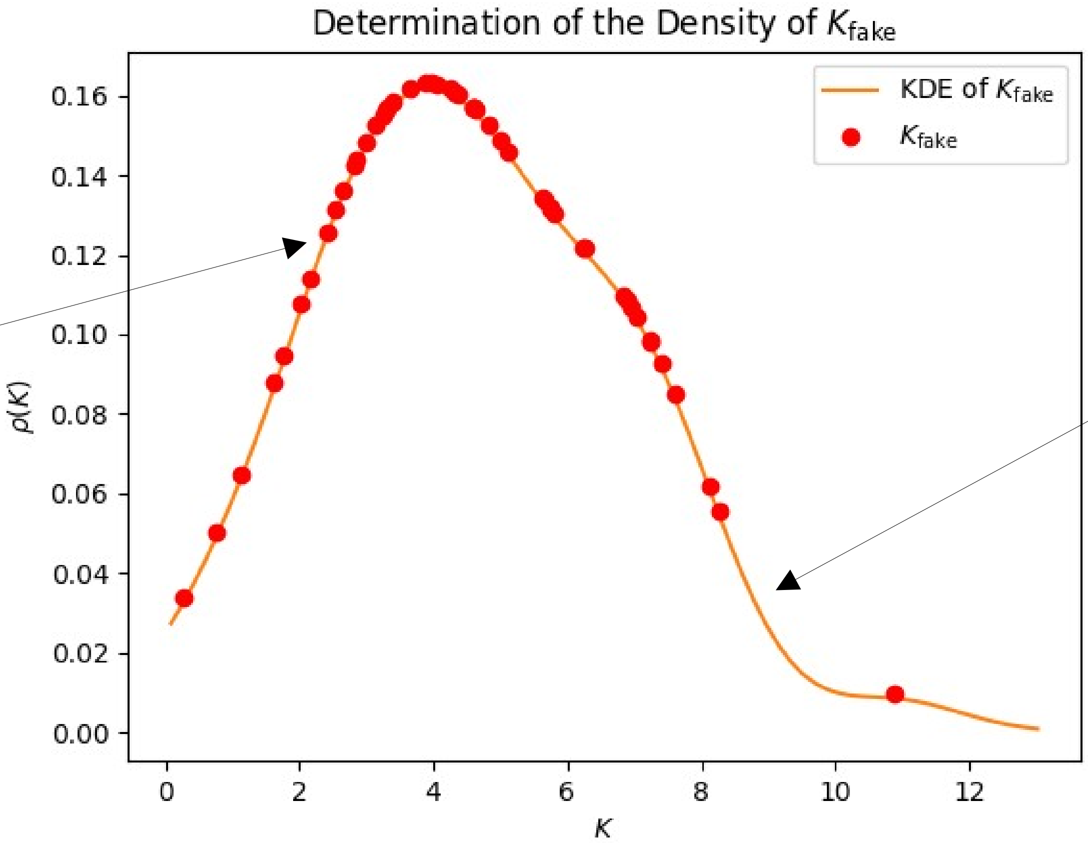
if and only if K_i is smaller than K_{i-1}

* (note however, instead of ratios of unnormalized posteriors we have ratios of ratios of unnormalized posteriors)

Walking Over Fake Standard Model Databases

- Produced 50 “fake” SModelS databases by sampling background models
- Corresponds to typical LHC results if no new physics is in data
- Determine 50 “fake” K values by running 50 walkers on each of the 50 databases (50 x 50 walkers in total) → density of K under null SM-only hypothesis

K for one “fake” background-only database.



Density of K estimated via a simple Kernel density estimator.

The Walks

We define a “run” as 50 parallel walks, each taking 1000 steps.

We performed

- 10 runs on the SModelS database (Sec. 5.2)
- 50 runs on fake “Standard Model-like” databases (Sec 5.1)
  to be able to determine a global *p*-value under the SM hypothesis
- 2x10 runs on fake “Signal-like” databases (Sec 5.3)
  to show closure of the method

Walking Over Databases With Fake Signals

To show closure of our method, we inject the winning protomodel as a signal in fake databases, and see if the algorithm can reconstruct the injected signal.

Sec 5.3

Technical closure test

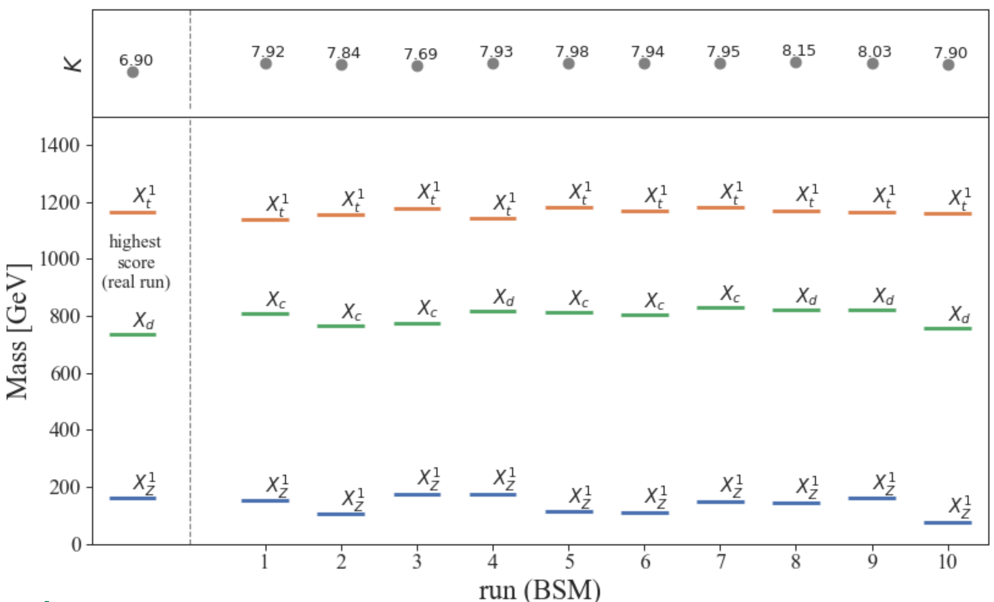


Fig. 11

No sampling of the models for the SRs, i.e. observed events := expected SM + expected signal events

Physics closure test

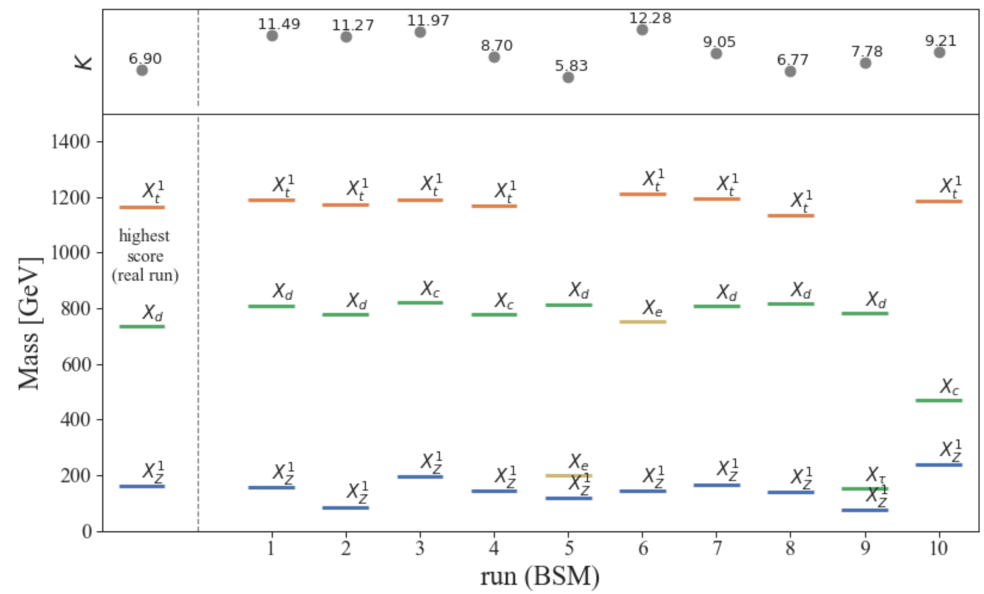


Fig. 10

Sampling turned on

Future Improvements

Improvements of the SModelS database:

- add latest full run-2 CMS and ATLAS publications (Moriond!)
- produce efficiency maps for existing results
- enlarge mass range of older efficiency maps

Improvements in speed:

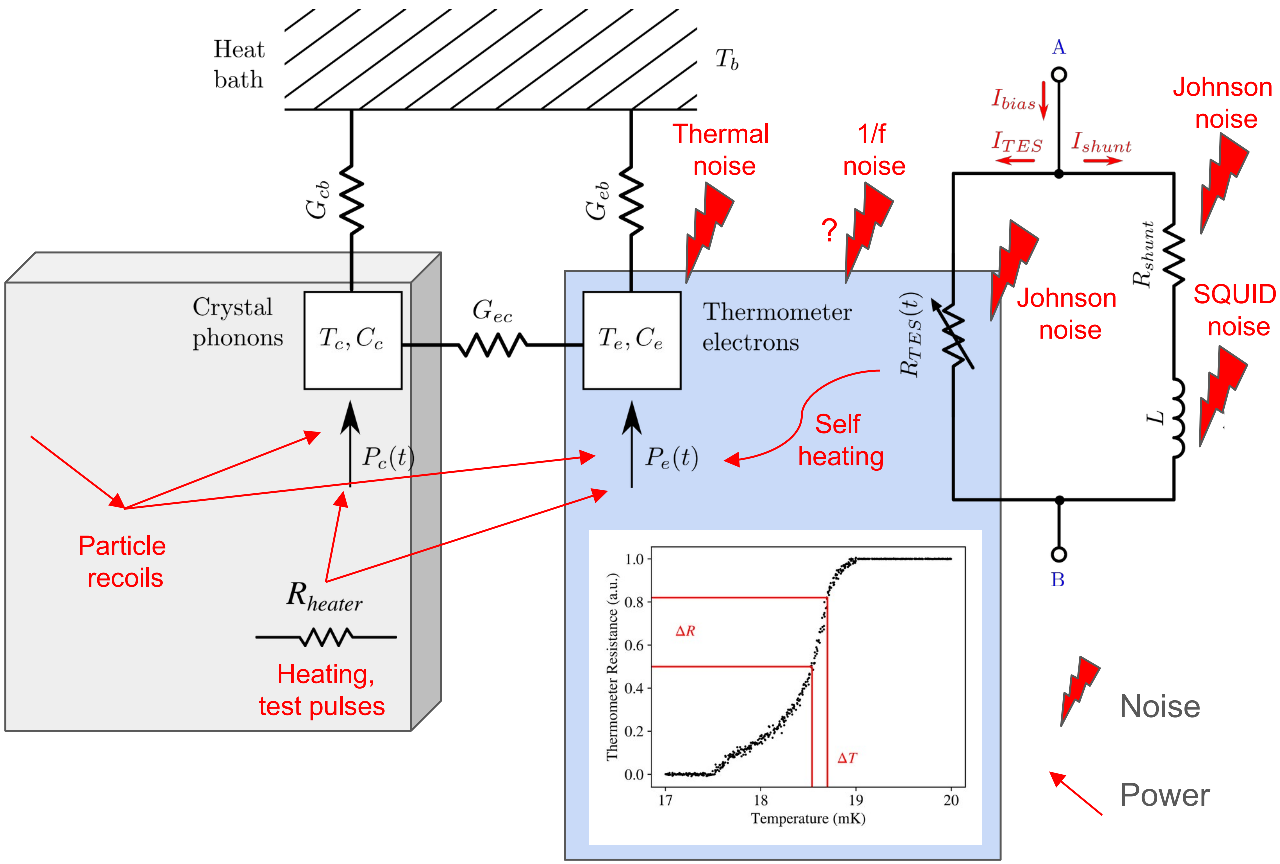
- learn the SModelS database
- make everything differentiable

Improvements in procedure:

- improve the “analyses correlation matrix”, automate the determination
- ponder relationship between proto-models and effective field theories
- connect proto-models with complete theories

Reinforcement Learning, Cryogenic Detectors

The situation with cryogenic detectors

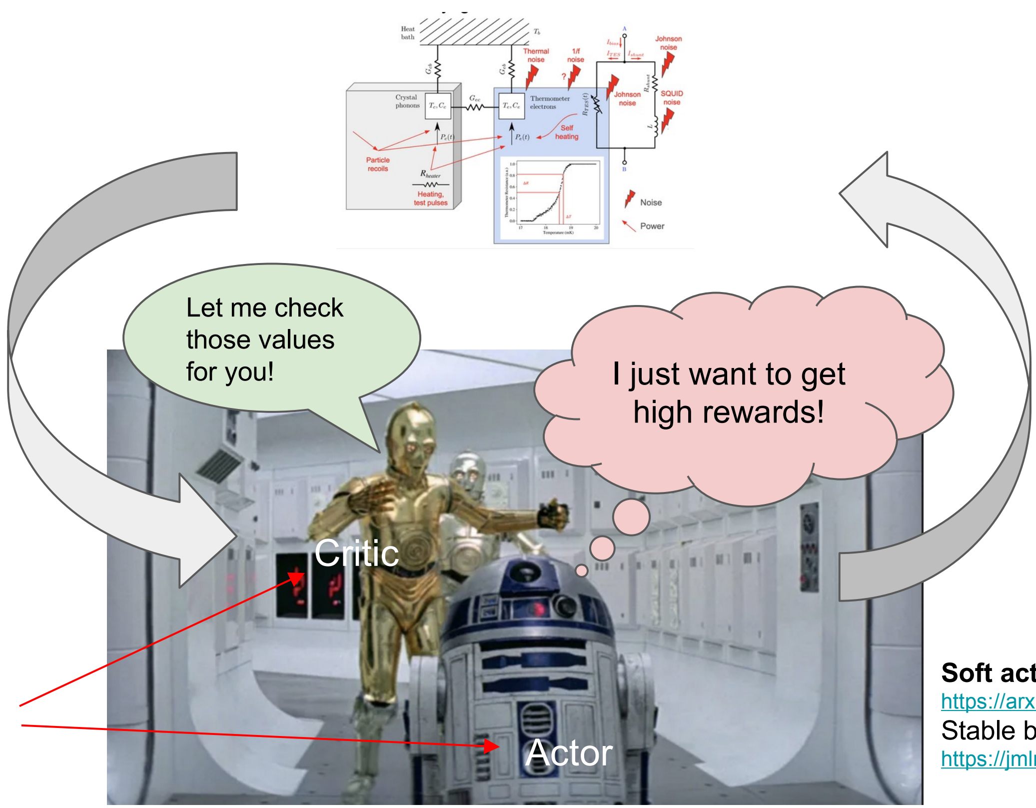


by Felix Wagner

Reinforcement Learning, Cryogenic Detectors

A framework for policy optimization: *reinforcement learning*

Observation (PH, RMS) and Reward (LT objective)



Action (setting DAC and Ib)

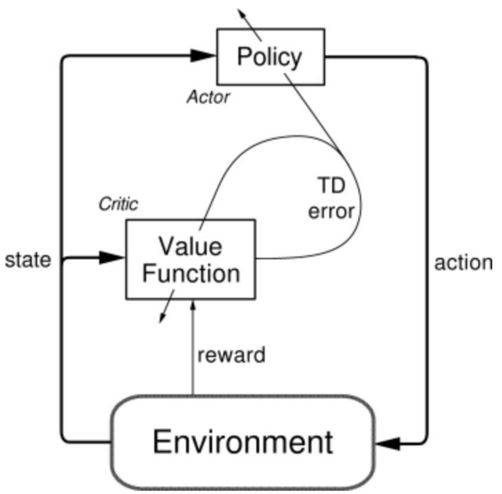
2 neural networks

Soft actor critic (SAC)
https://arxiv.org/abs/1801.01290v2
Stable baselines 3 implementation
https://jmlr.org/papers/volume22/20-1364/20-1364.pdf

by Felix Wagner

Reinforcement Learning, Cryogenic Detectors

Actor critic



$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t),$$

Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, Sergey Levine

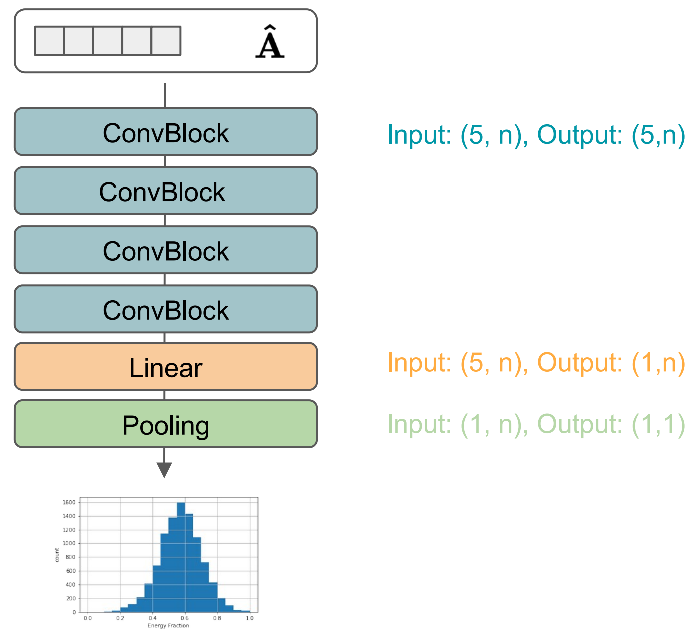
Model-free deep reinforcement learning (RL) algorithms have been demonstrated on a range of challenging decision making and control tasks. However, these methods typically suffer from two major challenges: very high sample complexity and brittle convergence properties, which necessitate meticulous hyperparameter tuning. Both of these challenges severely limit the applicability of such methods to complex, real-world domains. In this paper, we propose soft actor-critic, an off-policy actor-critic deep RL algorithm based on the maximum entropy reinforcement learning framework. In this framework, the actor aims to maximize expected reward while also maximizing entropy. That is, to succeed at the task while acting as randomly as possible. Prior deep RL methods based on this framework have been formulated as Q-learning methods. By combining off-policy updates with a stable stochastic actor-critic formulation, our method achieves state-of-the-art performance on a range of continuous control benchmark tasks, outperforming prior on-policy and off-policy methods. Furthermore, we demonstrate that, in contrast to other off-policy algorithms, our approach is very stable, achieving very similar performance across different random seeds.

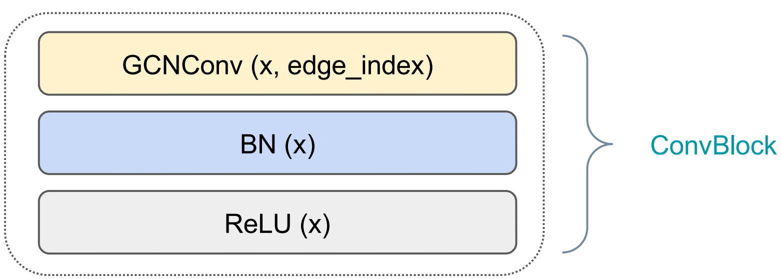
$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=1}^{T} \underbrace{R(s_t, a_t)}_{\text{reward}} + \alpha \underbrace{\mathrm{H}(\pi(\cdot\,|\,s_t))}_{\text{entropy}}\right]$$

by Felix Wagner

Graph Neural Networks

- using pytorch geometric
- consists of ConvBlocks, a linear model and a pooling layer
- GCNConv
  - updates nodes according to neighbors via adjacency matrix
  - symmetric normalization





$$\mathbf{X}' = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}\mathbf{X}\mathbf{\Theta}$$

https://arxiv.org/abs/1609.02907

by Mark Matthewman

Graph Neural Networks

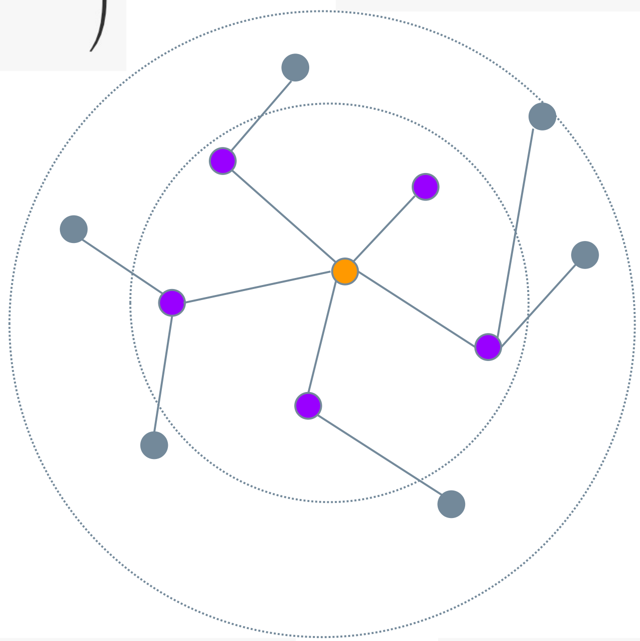
Hackathon - Convolution Operations

$$h_v^{(k)} = f^{(k)}\left(W^{(k)} \cdot \frac{\sum_{u \in \mathcal{N}(v)} h_u^{(k-1)}}{|\mathcal{N}(v)|} + B^{(k)} \cdot h_v^{(k-1)}\right)$$

GCNConv

$$h_v^{(k)} = f^{(k)}\left(W^{(k)} \cdot \left[\underset{u \in \mathcal{N}(v)}{\mathrm{AGG}}(\{h_u^{(k-1)}\}),\ h_v^{(k-1)}\right]\right)$$

SAGEConv



GATConv

$$h_v^{(k)} = f^{(k)}\left(W^{(k)} \cdot \left[\sum_{u \in \mathcal{N}(v)} \alpha_{vu}^{(k-1)} h_u^{(k-1)} + \alpha_{vv}^{(k-1)} h_v^{(k-1)}\right]\right)$$

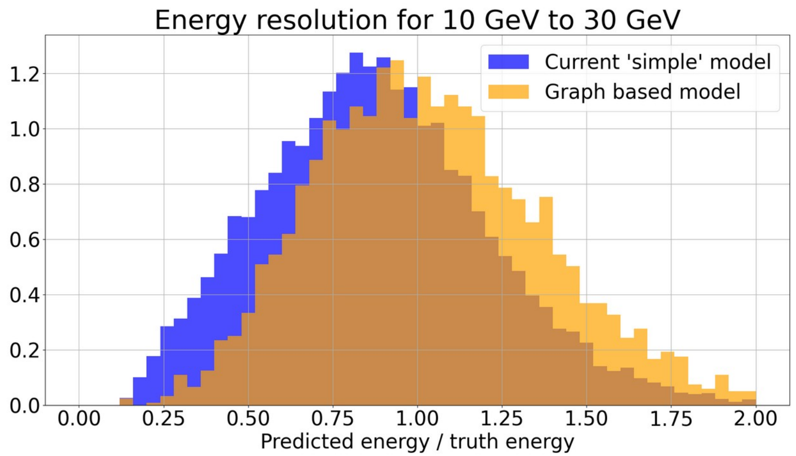
GINConv

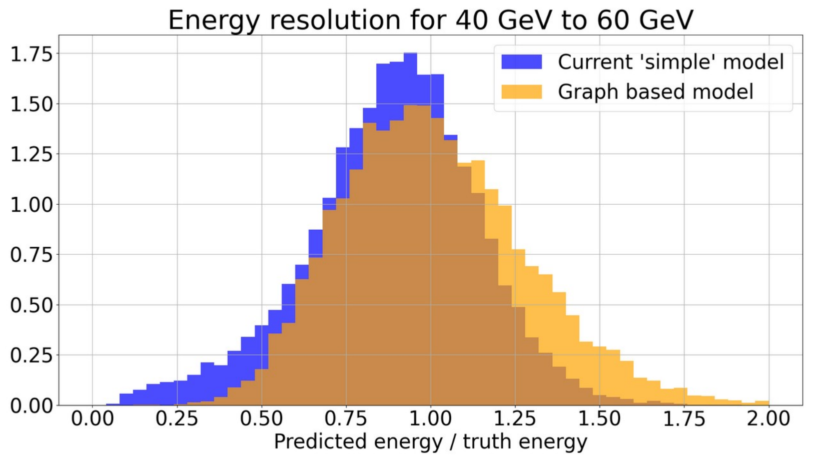
$$h_v^{(k)} = f^{(k)}\left(\sum_{u \in \mathcal{N}(v)} h_u^{(k-1)} + (1 + \epsilon^{(k)}) \cdot h_v^{(k-1)}\right)$$

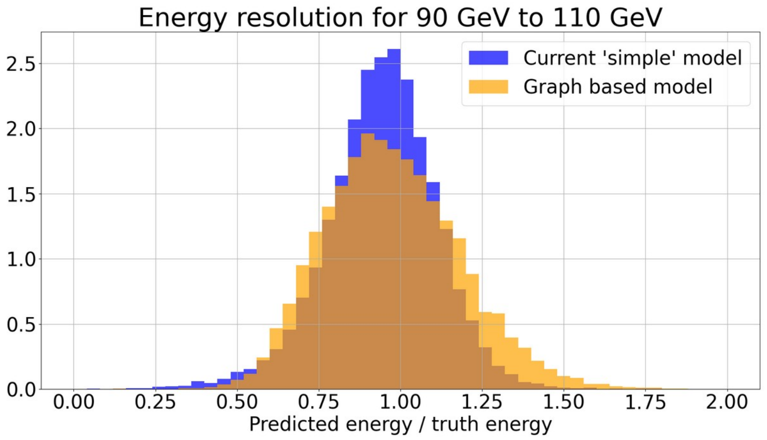
by Mark Matthewman

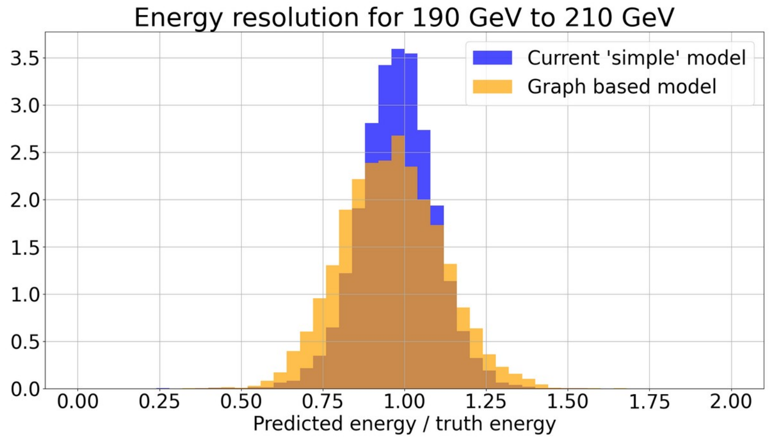
Graph Neural Networks

Hackathon - Results









by Mark Matthewman