Can we "machine-learn" the Next Standard Model?

image courtesy of Jon Butterworth, Chris Wormell

Wolfgang Waltenberger (ÖAW and Uni Vienna)

HEPHY
INSTITUT FÜR HOCHENERGIEPHYSIK

MCnet graduate summer school,
Lund, June 2020

universität wien

Can we "machine-learn" the Next Standard Model?

image courtesy of Jon Butterworth, Chris Wormell

Which of these fantastic beasts (if any) are real?
And how can we find out?

Wolfgang Waltenberger

HEPHY
INSTITUT FÜR HOCHENERGIEPHYSIK

universität wien

MCnet graduate summer school,
Lund, June 2020

Statistics Versus Machine Learning

Historically we tend to differentiate between statistics and machine learning as related but separate discplines.

In this talk, I wish to deemphasize this distinction, and more think in terms of:

- **Model-free versus model-based**

  Can I construct a low-dimensional statistical model that describes my data, founded in my domain-specific knowledge? Can I construct a likelihood? Do I have domain-specific knowledge about how the data come about? Or do I need to resort to a high-dimensional empirical parametrization of the dependency of a label w.r.t. the features?

- **Gradient-free versus gradient-based**

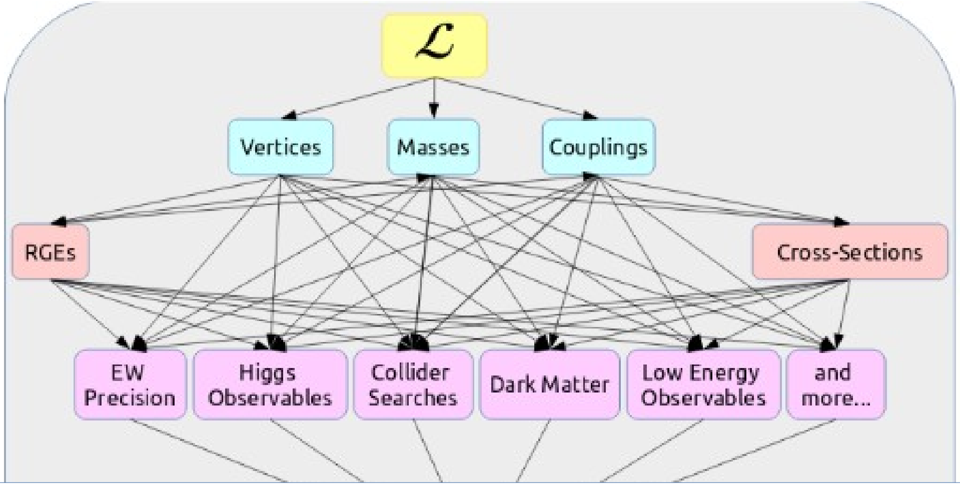
  Many of our data science problems are ultimately optimization problems. Do I have an analytical gradient for my objective function? Can I perform gradient descent to find the extremum of my objective function?

In this talk I will argue that with novel hardware and data science tools, we can aim at higher dimensional BSM models, merge the notion of “model building” with statistical learning, and employ gradient-free and ultimately gradient-based learning techniques to infer the NSM.

An inverse problem

Say you have a favorite BSM model.

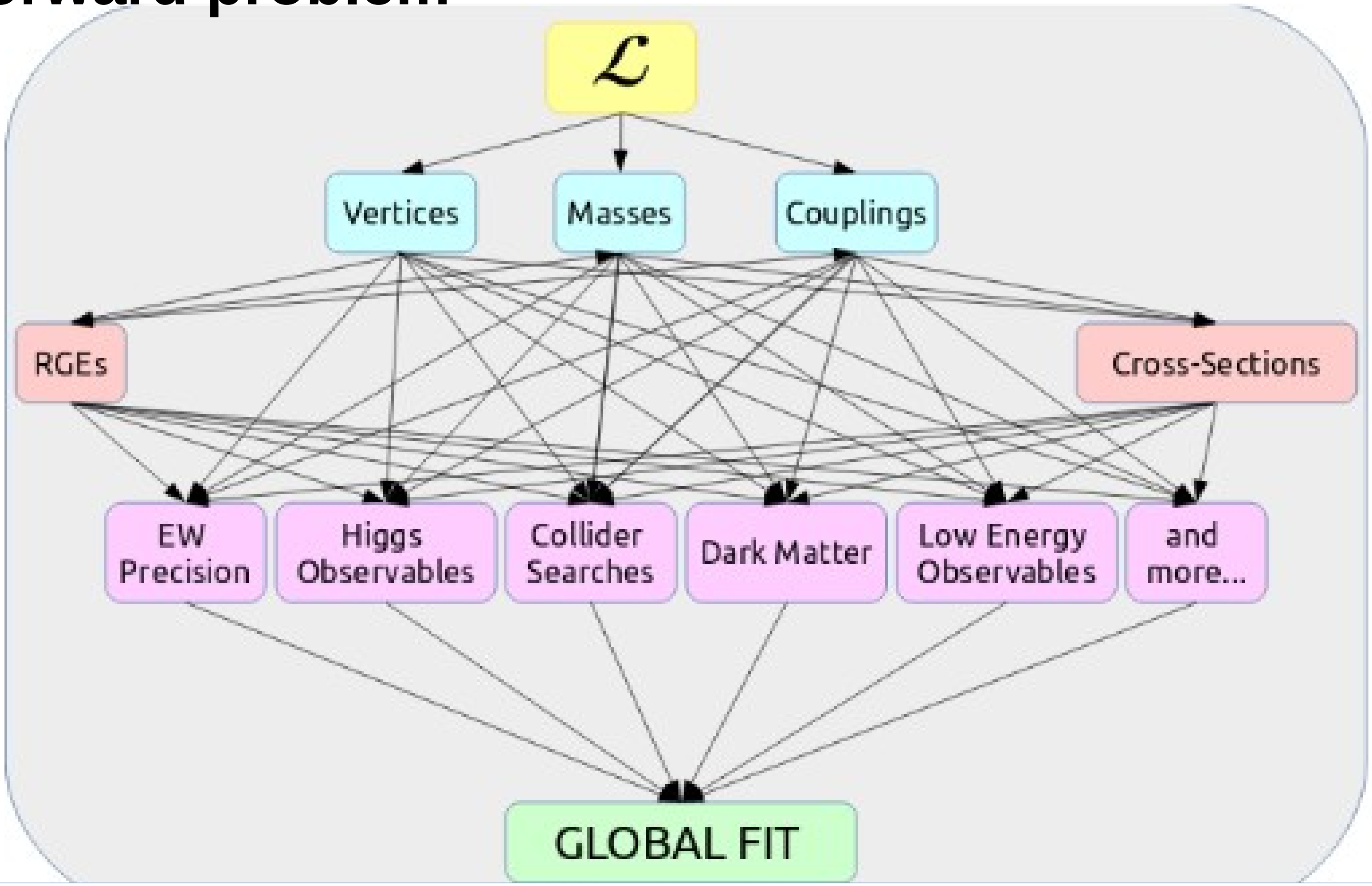
Computing the LHC observables for your favorite model is often technically challenging – as this audience knows very well – but it is (usually) a clearly defined, deductive task. If your BSM model is not too exotic, the tools are alreay in place – again, thanks also to major intellectual efforts like the ones at Mcnet. I will refer to this as the "forward problem".



Plot stolen from Jamie Tattersall's slides

An inverse problem

Computing the LHC observables for your favorite BSM model is a very difficult but clearly defined, deductive task: the **“forward problem”**



Plot stolen from Jamie Tattersall’s slides

An inverse problem

But how about the other way round?

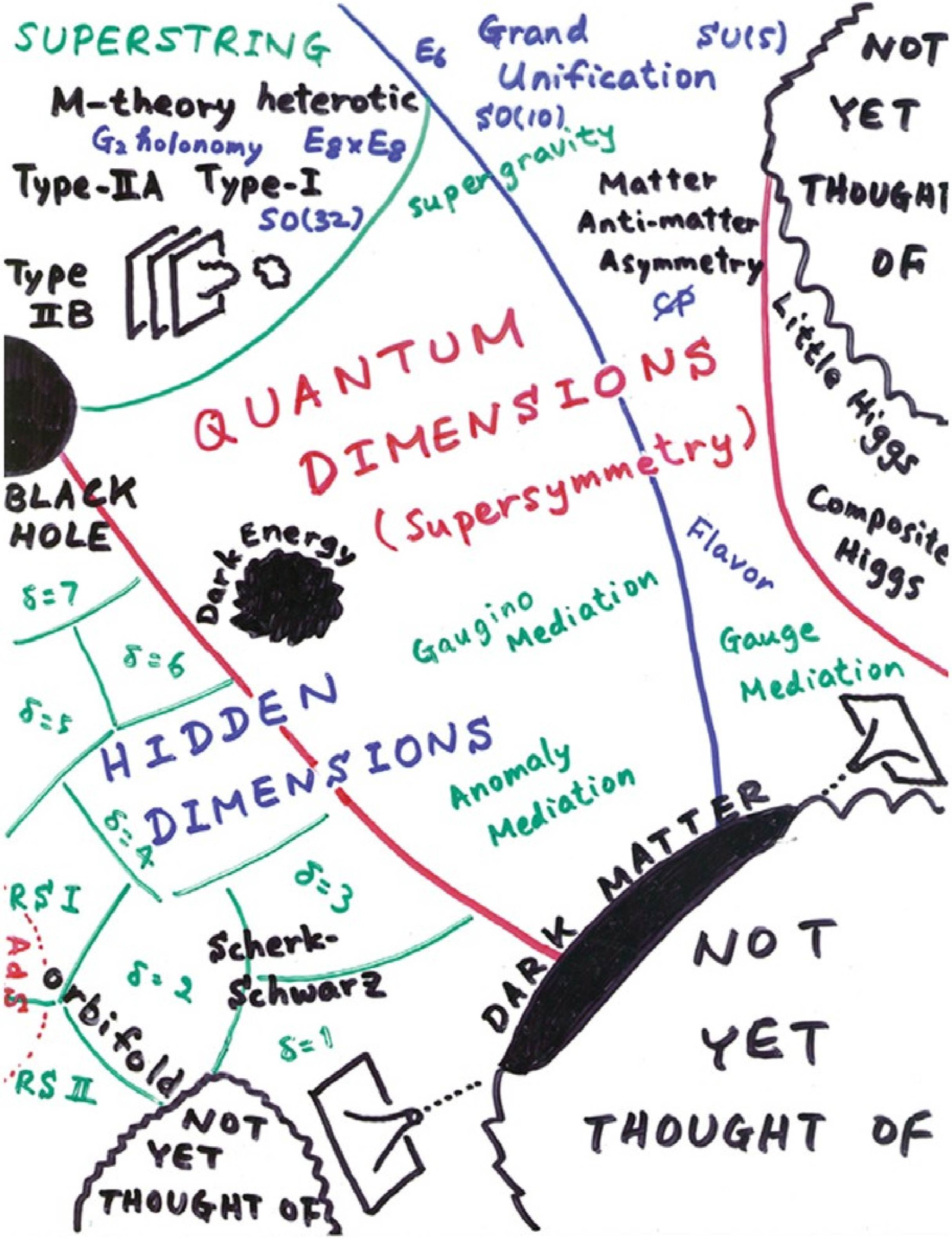
Building the prospective Next Standard Model (**NSM**) from all our wonderful LHC results is inductive reasoning – one tries to infer the general rules behind one's concrete observations.

It is by construction ill-defined, and there is no guaranteed recipe for success.

And yet, constructing the NSM is our ultimate goal as we search for new signs for physics, is it not?

I shall refer to this challenge as our "Inverse Problem" (PI)*.

*PI because it is inverse



Hitoshi Murayama's impression of *The* Theory Landscape –

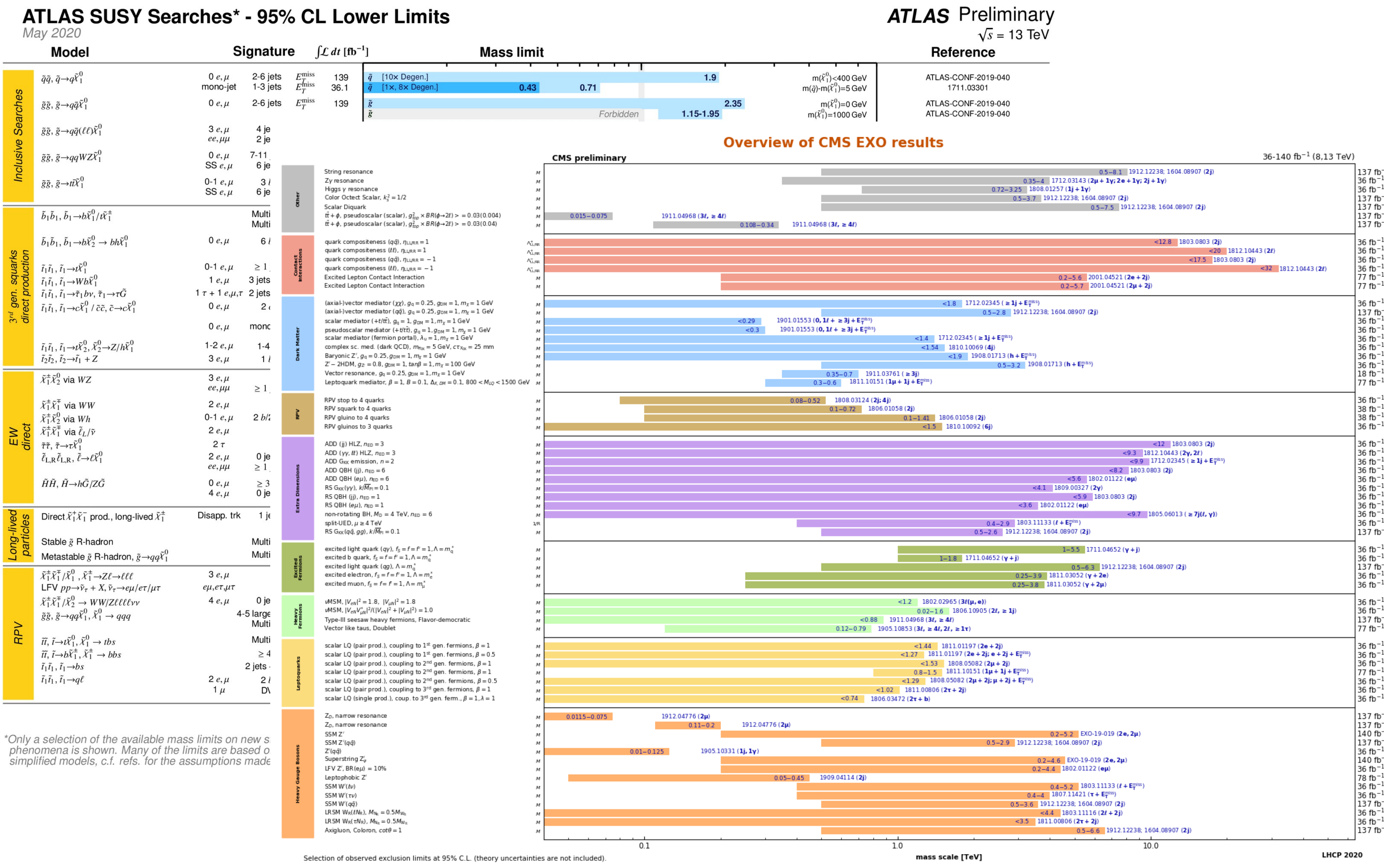
It shows a large number of ideas.

Most ideas come with a large number of free parameters!

Presumably still our best NSM candidate is the minimal supersymmetric standard model, the MSSM.

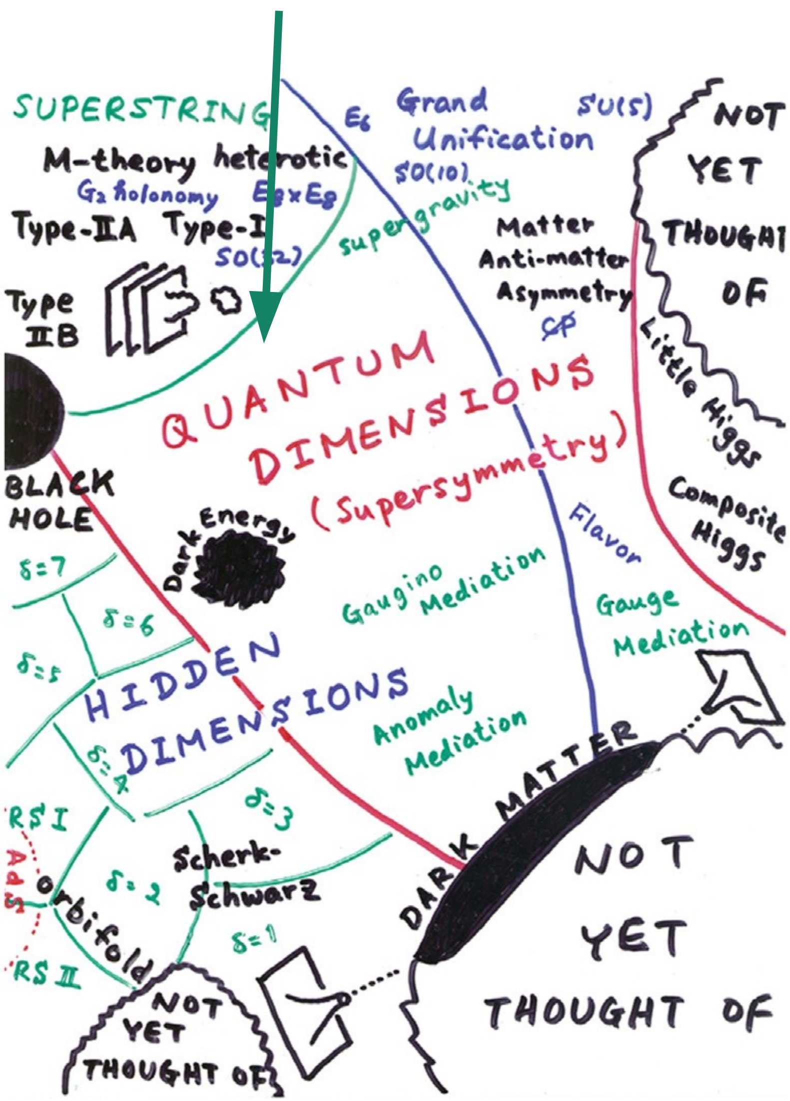
It has 100+ parameters.

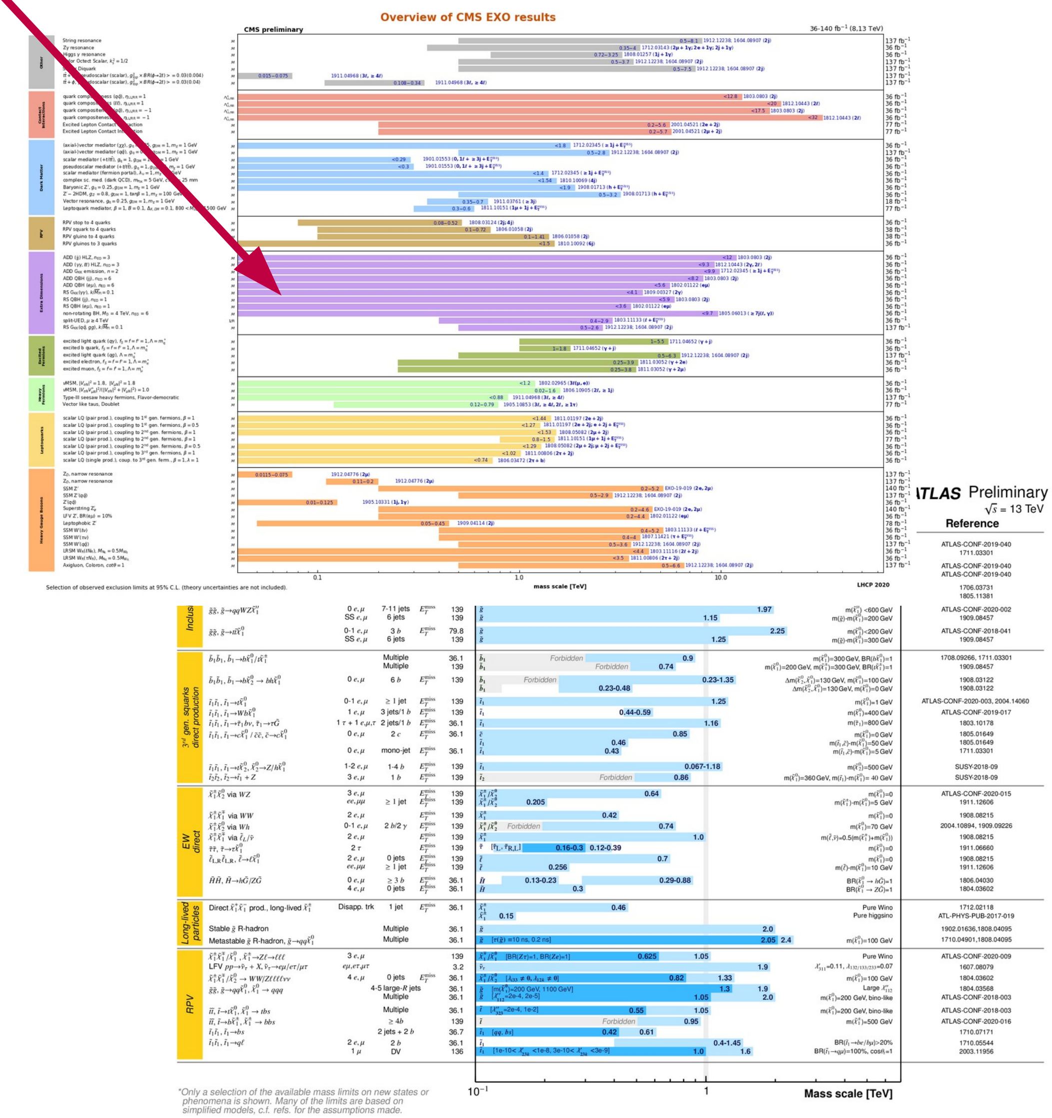
These ideas need to be systematically confronted with LHC and non-LHC results.
The number of LHC physics publications alone is O(1000) and counting!



So what do all these results

tell us about
all these ideas?





Now you might wonder:

Didn’t we have similar problems in the past?

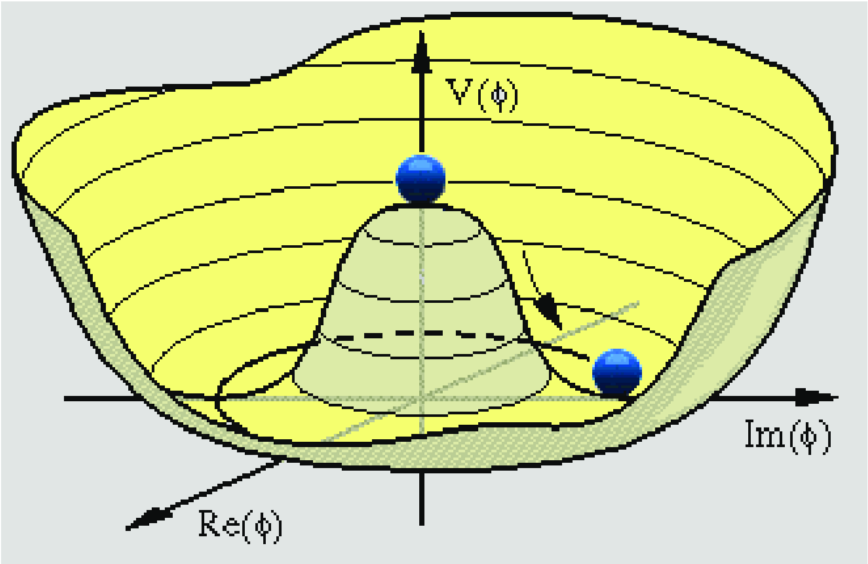
If yes, how did we solve them?

If no, what’s different?

A situation unlike in the past

How did we solve such problems in the past?

Previous successful endeavours like the Higgs or the top discoveries were driven by very clearly defined models. E.g. the Higgs mechanism had only one free parameter – the Higgs mass. Classical hypothesis testing was all that was needed.



In other words ...

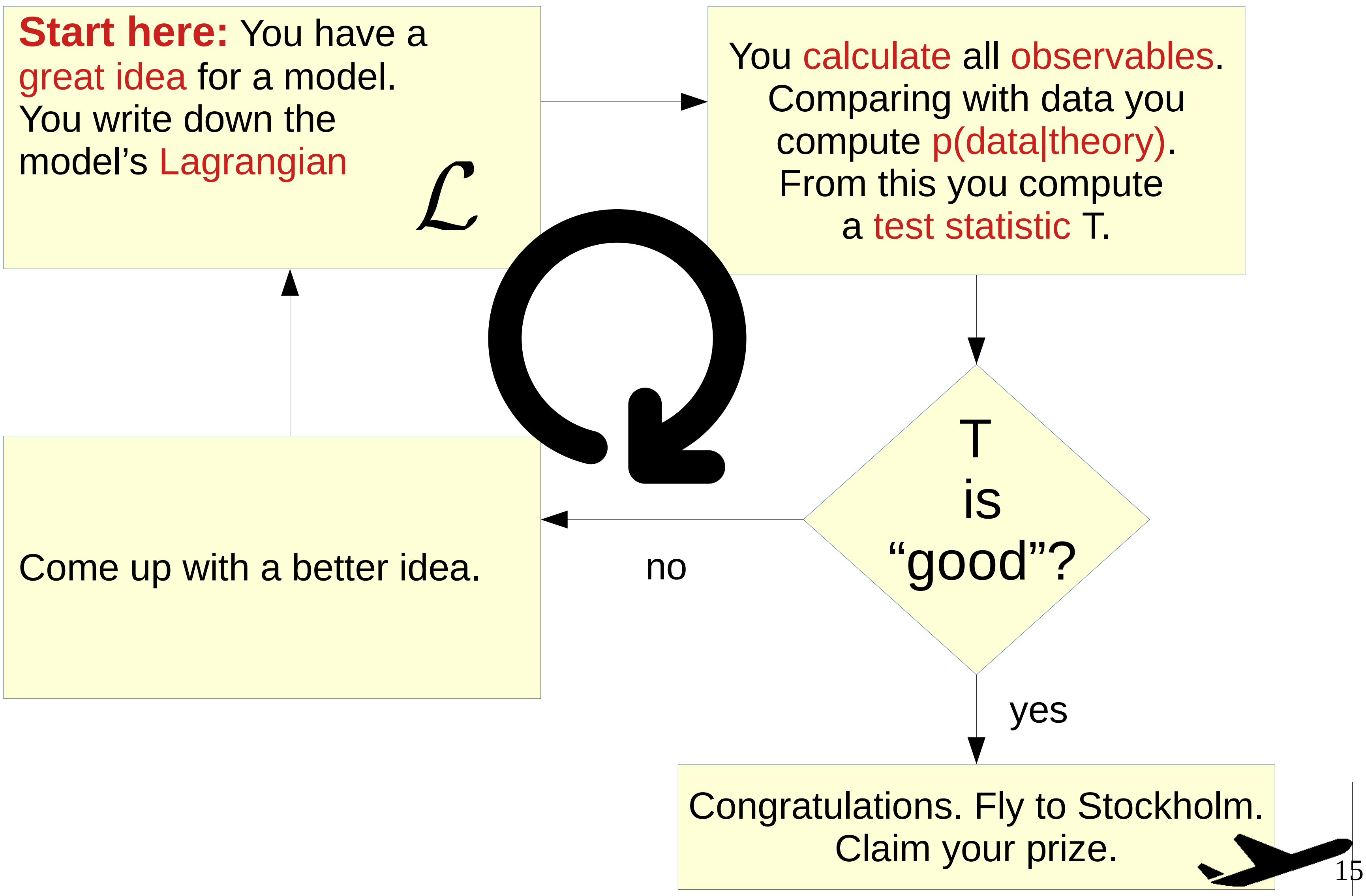
Let's say we see a few mild excesses in a few channels/analyses, hints of a "dispersed signal". How would we proceed?

How would be build, establish, and endorse a Next Standard Model?

Top-down versus bottom-up

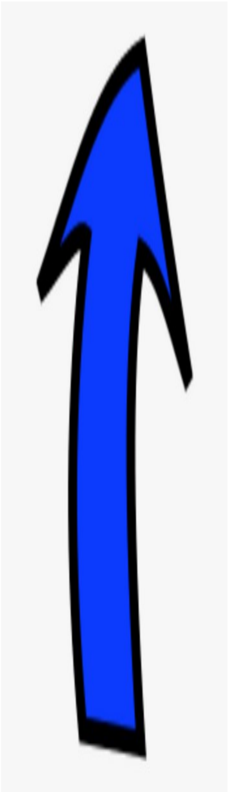
Throughout my talk I shall distinguish between “top-down” and “bottom-up” approaches.

Top-Down:



Top-down versus bottom-up

Throughout my talk I shall distinguish between “top-down” and “bottom-up” approaches.



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

↑

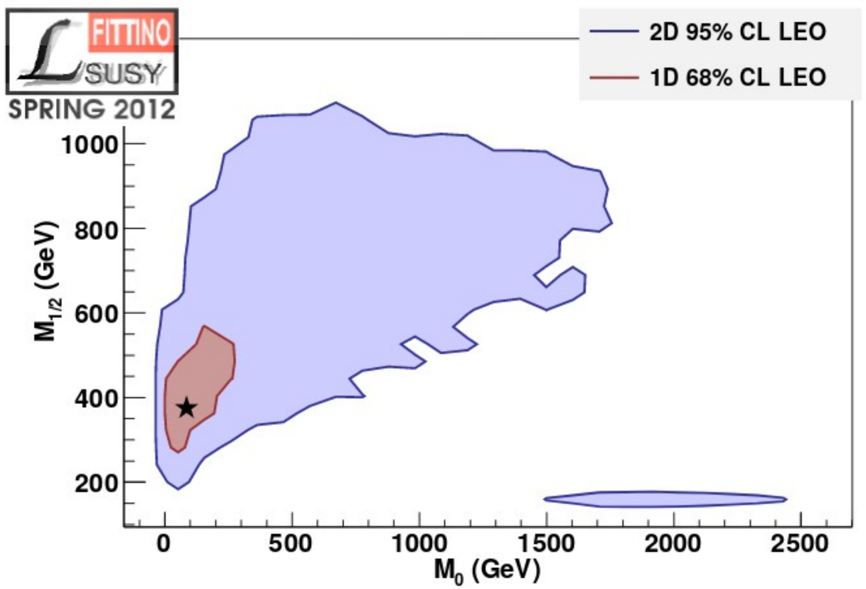
Start here: You describe your experimental findings in a language amenable to theoretical physics, e.g. simplified models for on-shell effects (“searches”), effective field theories for off-shell effects (“measurements”).

Can we "machine-learn" the Next Standard Model?

Part 2 – Top-Down

The early days: (frequentist) hypothesis testing in low-dimensional model spaces

Example: fits by the Fittino collaboration of the parameters of possibly the simplest supersymmetric theory: the "constrained Minimal Supersymmetric Model" (cMSSM), spring 2012 (right before the discovery of the Higgs boson)



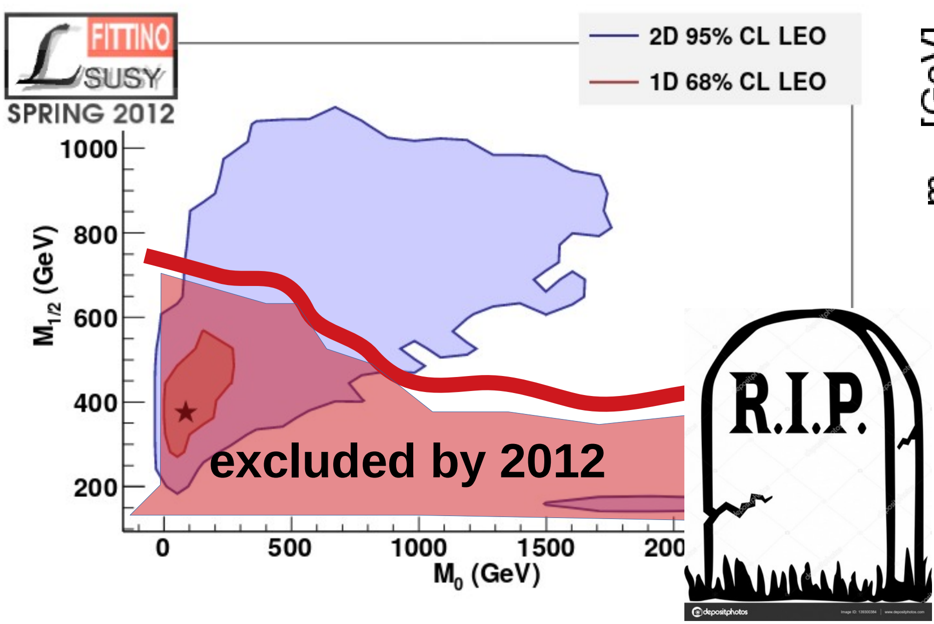
The model (CMSSM) was a 4 (5) - dimensional model, with only two "essential" parameters: the model was just systematically "scanned".

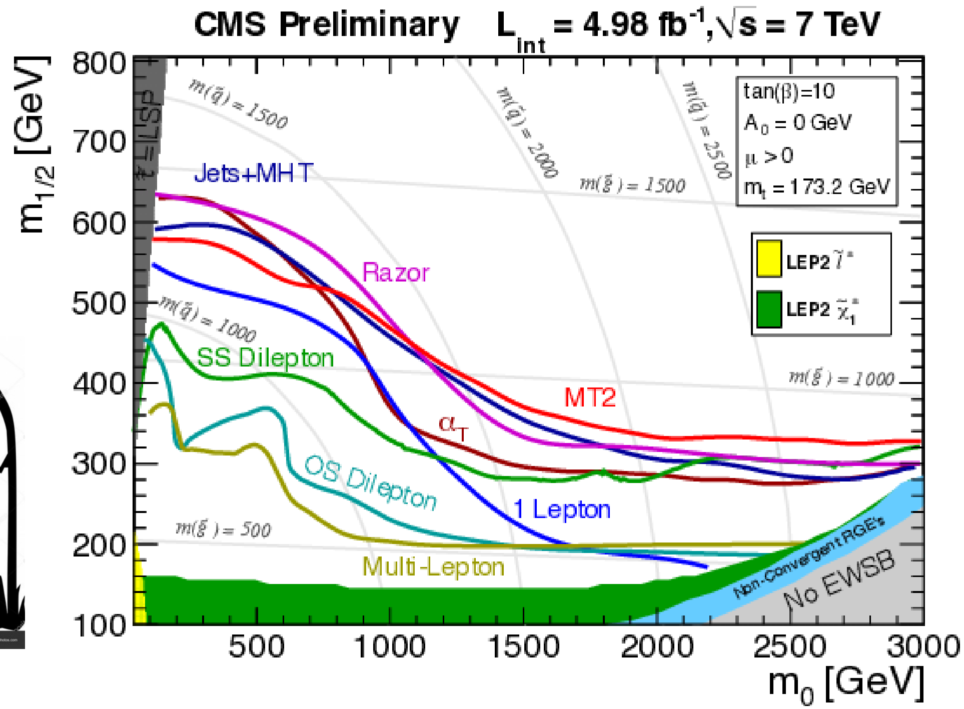
The plot on the left shows the frequentist CL quantiles of low-energy observables (LEO).

"LEOs" comprised excesses in B-meson measurements, as well as e.g. $(g-2)_\mu$

arxiv:1204.4199

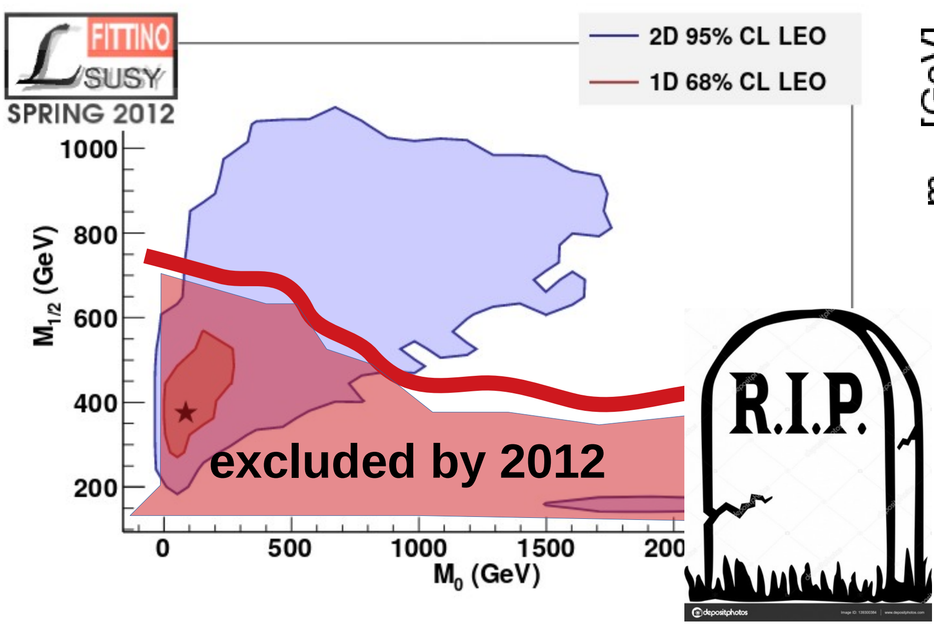
The early days: (frequentist) hypothesis testing in low-dimensional model spaces

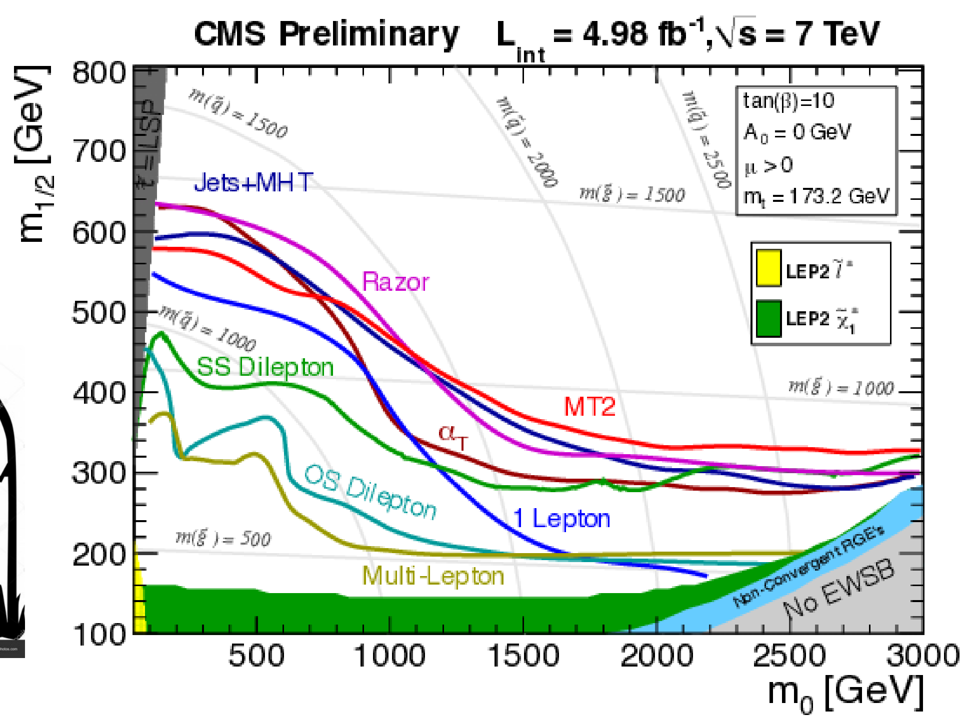




Needless to say, soon after these models and model points were essentially killed by CMS and ATLAS searches.

The early days: (frequentist) hypothesis testing in low-dimensional model spaces





Take-away messages: negative statements, exclusion lines are simple, well defined, and epistemologically correct. Positive statements about BSM are very relative statements only ("this region is favored over that region"). The theory is not disseminated. The model selection problem (why cMSSM?) is not addressed.

Raising the stakes: higher dimensional models and Bayesian statistics

If we allow for Bayesian statistics, we can ask a slightly different question:

What does the LHC teach us about NSM physics, that we didn't know before?

$$p(\mathrm{NSM}|\mathrm{LHC}) \propto L(\mathrm{LHC}|\mathrm{NSM})\pi(\mathrm{NSM})$$

Our knowledge of the NSM in light of LHC data ...

… is proportional to ...

… the likelihood of our data, **given** the theory ...

… times our prior knowledge of NSM physics

JHEP 1610 (2016) 129

Bayesians have more Fun

In their publication, CMS answered a smaller question: what did the CMS searches for new physics teach us about the phenomenological Minimal Supersymmetric Model (pMSSM)?

$$p(\text{pMSSM}|\text{CMS}) \propto L(\text{CMS}|\text{pMSSM})\pi(\text{pMSSM})$$

Our knowledge of the pMSSM in light of the CMS search results ...

… is proportional to ...

… the likelihood of our data, given the pMSSM ...

… times our prior knowledge of the pMSSM

The pMSSM is a phenomenological, "stripped-down" version of the Minimal Supersymmetric Model (MSSM), with constraints put on all model parameters that have no big effect on LHC "phenomenology". It has 18 or 19 free parameters → a major "upgrade" from the 4 or 5 free parameters of the cMSSM!

JHEP 1610 (2016) 129

BAYESIANS HAVE MORE FUN

$\pi(\text{pMSSM})$ → what's our information on the pMSSM **prior** to looking at CMS'es search results?

i	Observable $\mu_i(\theta)$	Constraint $D_i^{\text{non-DCS}}$	Likelihood function $L[D_i^{\text{non-DCS}} \mid \mu_i(\theta)]$	Comment
1	$\mathcal{B}(\text{b} \to \text{s}\gamma)$ [45]	$(3.43 \pm 0.21^{\text{stat}} \pm 0.24^{\text{th}} \pm 0.07^{\text{sys}}) \times 10^{-4}$	Gaussian	reweight
2	$\mathcal{B}(\text{B}_\text{s} \to \mu\mu)$ [46]	$(2.9 \pm 0.7 \pm 0.29^{\text{th}}) \times 10^{-9}$	Gaussian	reweight
3	$R(\text{B} \to \tau\nu)$ [45, 47]	1.04 ± 0.34	Gaussian	reweight
4	Δa_μ [48]	$(26.1 \pm 6.3^{\text{exp}}$ [illegible] $) \times 10^{-10}$	Gaussian	
5	$\alpha_\text{s}(m_\text{Z})$ [49]	0.1184 ± 0.0007	Gaussian	
6	m_t [50]	173.20 [illegible] GeV	Gaussian	reweight
7	$m_\text{b}(m_\text{b})$ [49]	$4.19^{+0.18}_{-0.06}$ GeV	Two-sided Gaussian	
8	m_h	LHC: $m_\text{h}^{\text{low}} = 120\,\text{GeV}$, $m_\text{h}^{\text{high}} = 130\,\text{GeV}$	1 if $m_\text{h}^{\text{low}} \leq m_\text{h} \leq m_\text{h}^{\text{high}}$ 0 if $m_\text{h} < m_\text{h}^{\text{low}}$ or $m_\text{h} > m_\text{h}^{\text{high}}$	reweight
9	μ_h	CMS and ATLAS in LHC Run 1, Tevatron	LILITH 1.01 [51, 52]	post-MCMC
10	sparticle masses	LEP [53] (via MICROMEGAS [54–56])	1 if allowed 0 if excluded	

"Previous measurements"

Bayesians have more Fun

$L(\mathrm{CMS}|\mathrm{pMSSM})$ → what's the **likelihood** of CMS'es search results, given the pMSSM?

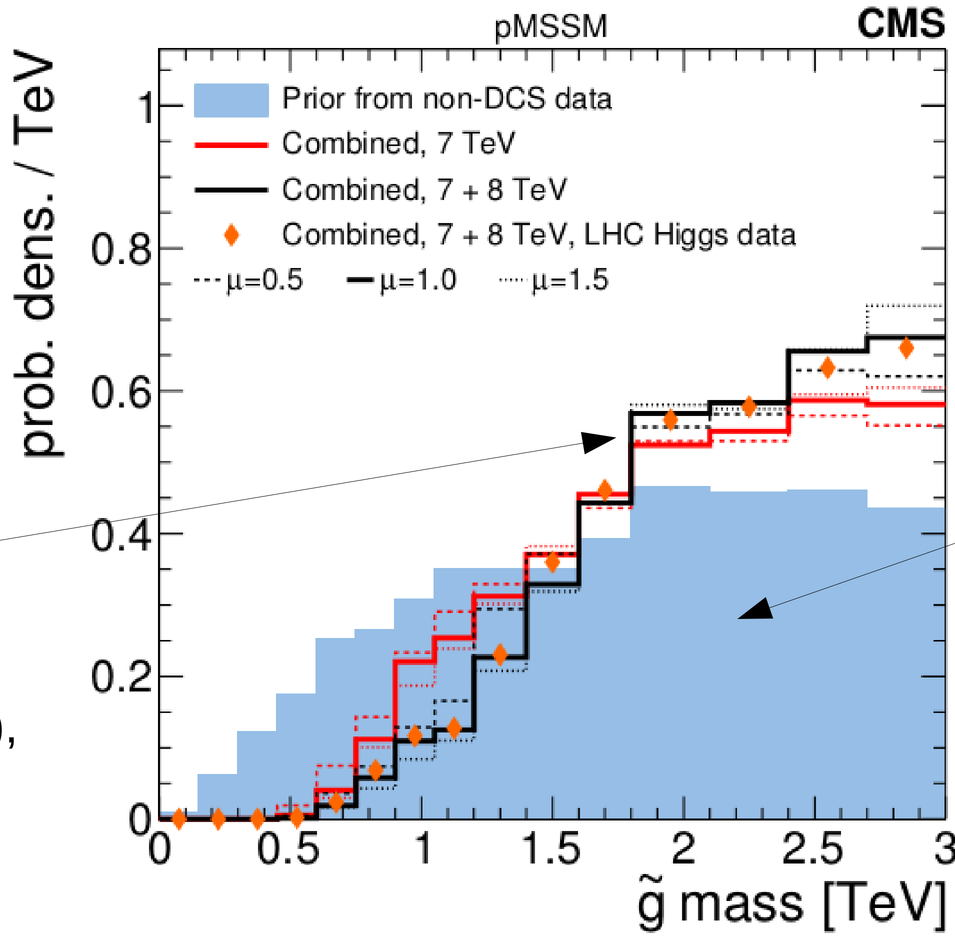
Analysis	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb^{-1}]	Likelihood
Hadronic H_T + H_T^{miss} search [8]	7	4.98	counts
Hadronic H_T + E_T^{miss} + b-jets search [9]	7	4.98	counts
Leptonic search for EW prod. of $\tilde{\chi}^0$, $\tilde{\chi}^{\pm}$, $\tilde{l}$ [10]	7	4.98	counts
Hadronic H_T + H_T^{miss} search [11]	8	19.5	counts
Hadronic M_{T2} search [12]	8	19.5	counts
Hadronic H_T + E_T^{miss} + b-jets search [13]	8	19.4	χ^2
Monojet searches [14]	8	19.7	binary
Hadronic third generation squark search [15]	8	19.4	counts
OS dilepton (OS ll) search [16] (counting experiment only)	8	19.4	counts
LS dilepton (LS ll) search [17] (only channels w/o third lepton veto)	8	19.5	counts
Leptonic search for EW prod. of $\tilde{\chi}^0$, $\tilde{\chi}^{\pm}$, $\tilde{l}$ [18] (only LS, 3 lepton, and 4 lepton channels)	8	19.5	counts
Combination of 7 TeV searches	7	—	binary
Combination of 7 and 8 TeV searches	7, 8	—	binary

About 10 new CMS searches for new physics!

Bayesians have more Fun

$L(\mathrm{pMSSM}|\mathrm{CMS})$

→ what did CMS'es searches teach us about the pMSSM? Prior versus posterior!



Our *a posteriori* knowledge of the mass of the gluino, after the first LHC "run" (red line), and the second run (black line).

Blue area: Our *prior* knowledge about the mass of the partner particle of the gluon, the *gluino*

Bayesian statistics can help us describe what we learned. Higher dimensional phenomenological models are better suited to describe what we are actually seeing. In this publication, however, the model selection problem was also not addressed.

Intermission: Model Selection

Question: In statistics, what are the default approaches to solve model selection problems?

Answer 1: A widely-used frequentist method is the Akaike Information Criterion (AIC).

$$\mathrm{AIC} = -2\ln\hat{L} + 2k$$

$\hat{L}$ is the maximum likelihood of the model, k are the free parameters. Models get punished for adding parameters. Compute AIC for all your models. Choose model with lowest AIC.

Answer 2: In Bayesian statistics, Bayes factors and the closely connected Bayes Information Criterion (BIC) are much used.

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\int \Pr(\theta_1|M_1)\Pr(D|\theta_1,M_1)\,d\theta_1}{\int \Pr(\theta_2|M_2)\Pr(D|\theta_2,M_2)\,d\theta_2} = \frac{\frac{\Pr(M_1|D)\Pr(D)}{\Pr(M_1)}}{\frac{\Pr(M_2|D)\Pr(D)}{\Pr(M_2)}} = \frac{\Pr(M_1|D)}{\Pr(M_2|D)}\frac{\Pr(M_2)}{\Pr(M_1)}.$$

(Bayes factors are likelihood ratios with marginalized (=integrated) theory parameters.)

https://pubmed.ncbi.nlm.nih.gov/22309957/

Intermission: Model Selection

Question: So why can't we just employ such a model selection algorithm and be done with the inverse problem?

Answer: these algorithms work well with a finite set of models, with not too many (<< 100) parameters. The true model ideally should be an element of the set of models being tested (e.g. the proof of the AIC being "consistent" depends on it).

So again, our situation is too vague to naively apply the standard procedures.

We may make use of these algorithms, but we may need to be smarter still.

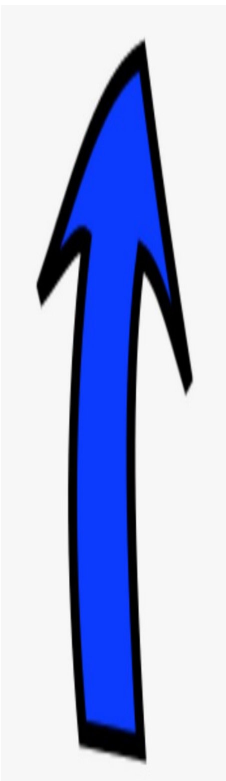
My proposal in this lecture will be an algorithm that "merges" model building with model selection.

Can we “machine-learn” the Next Standard Model?

Part 3 – Bottom-up

In lack of a clear idea of what theory are looking for ....

… why not start with data?



Bottom-Up:

Only now 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").

Abstraction layers

Q: How can we describe our experimental findings in a language that plays to theory?

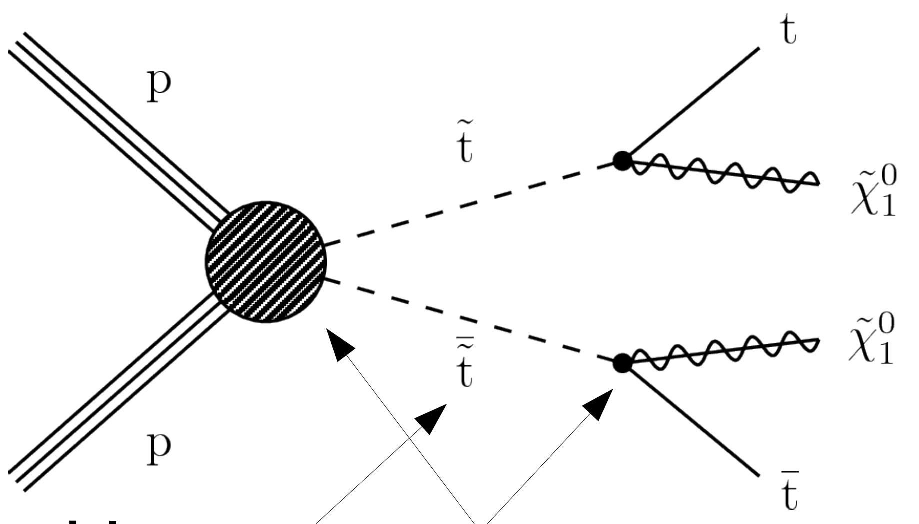
A: Simplified models (SMS) were developed to summarize the results of searches for new physics.

Likewise, effective field theories allow for a simple description of measurements – in the form of likelihoods on Wilson coefficients.

As I myself have been working on simplified models but not on effective field theories I shall focus on SMS.

Recap: What is a Simplified Model?

A visual representation of one specific simplified model:



1. The **BSM "mother" particles** are **fixed**.
In this case, they are top-partner quarks ("stops").

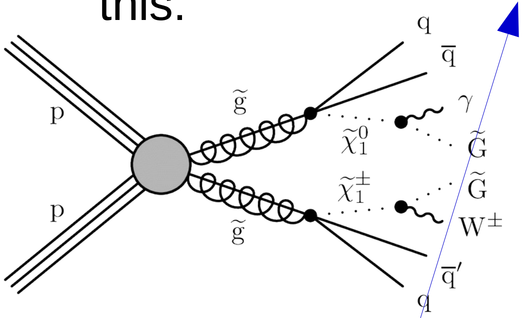
2. The **decay** is also **fixed**: in this simplified model, the stops **always** decay to a top and the dark matter candidate ("neutralino")

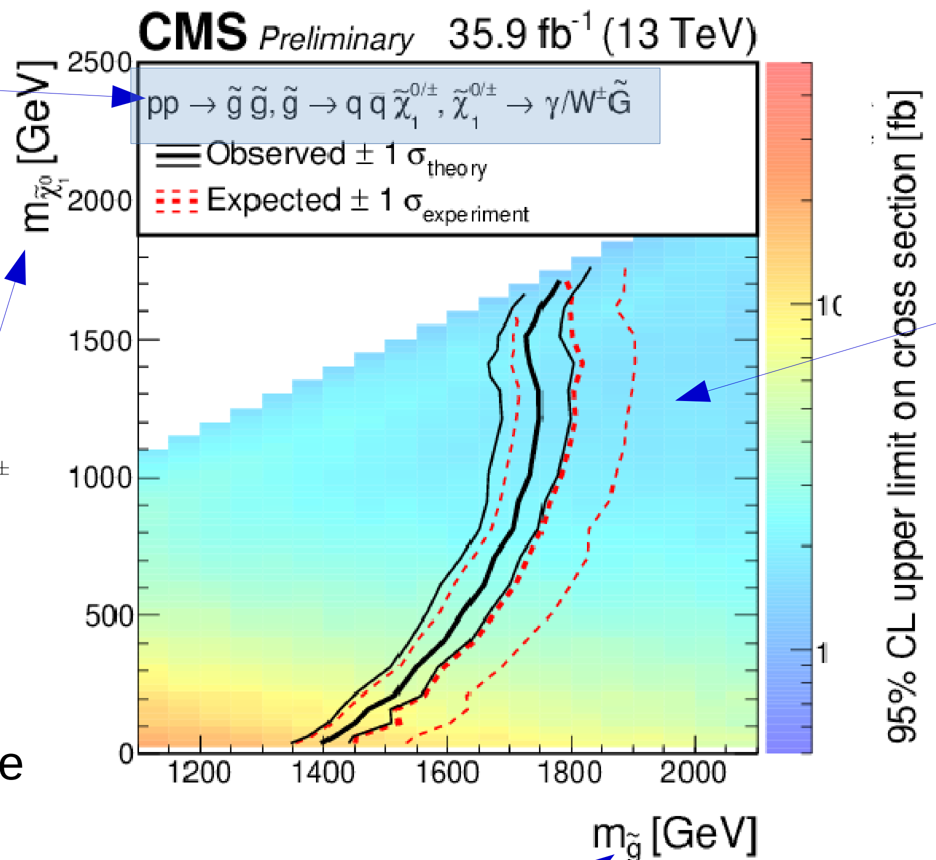
3. The production cross section is **a free parameter**. "Full" theories will predict production cross sections for simplified models; the experimental results will be formulated as limits on these production cross sections.

Recap: What is a Simplified Model?

We were able to convince the CMS and ATLAS collaborations to present the experimental findings as **upper limits on the production cross sections of simplified models.**

specifies the simplified model; it corresponds to this:





the mass of the "daughter" particle

The color encodes the 95% confidence level limit on the production cross section, for that model with given masses.

the mass of the "mother" particle

CMS-PAS-SUS-17-012

Disclaimer

In the remainder of this talk I shall present work that is being done within our SModelS collaboration.

Sorry about the apparent self-promotion, but I promise I will focus on the big concepts. I shall also argue that the building blocks I will present –

- proto-models,
- MCMC-based automated model building,
- gradient-accelerated model building,
- differentiable inductive reasoning

– can be repurposed, and reused in other contexts (e.g. with effective field theories)

SModelS – a decomposer and a database

SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry

Sabine Kraml[1*], Suchita Kulkarni[1†], Ursula Laa[2‡], Andre Lessa[3§], Wolfgang Magerl[2¶], Doris Proschofsky-Spindler[2‖], Wolfgang Waltenberger[2**]

[1] *Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 Avenue des Martyrs, F-38026 Grenoble, France*

[2] *Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfer Gasse 18, 1050 Wien, Austria*

[3] *Instituto de Física, Universidade de São Paulo, São Paulo - SP, Brazil*

← Our first publication

https://arxiv.org/pdf/1312.4175.pdf

Abstract

Letters in High Energy Physics LHEP xx, xxx, 2020

SModelS database update v1.2.3

Charanjit K. Khosa,[1] Sabine Kraml,[2] Andre Lessa,[3] Philipp Neuhuber,[4] and Wolfgang Waltenberger[4]

[1] *Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*

[2] *Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble INP, 53 Avenue des Martyrs, 38000 Grenoble, France*

[3] *Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, 09210-580 SP, Brazil*

[4] *Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfer Gasse 18, 1050 Wien, Austria*

Our most recent publication →

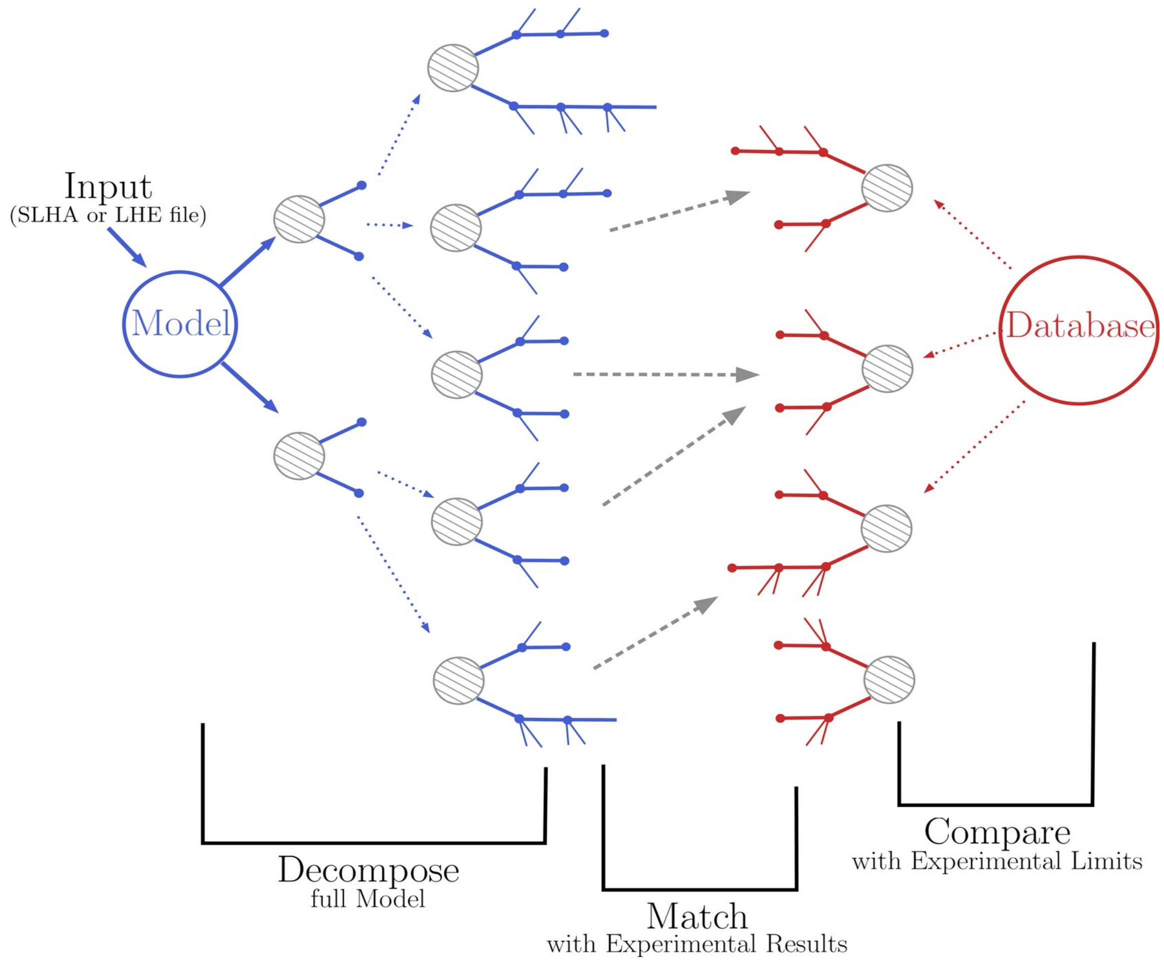
Abstract

We present an update of the SModelS database with simplified model results from 13 ATLAS and 10 CMS searches for supersymmetry at Run 2. This includes 5 ATLAS and 1 CMS analyses for full Run 2 luminosity, i.e. close to 140 fb^{-1} of data. In total, 76 official upper limit and efficiency map results have been added. Moreover, 21 efficiency map results have been produced by us using MadAnalysis5, to improve the coverage of gluino-squark production. The constraining power of the new database, v1.2.3, is compared to that of the previous release, v1.2.2. SModelS v1.2.3 is publicly available and can readily be employed for physics studies.

https://arxiv.org/pdf/2005.00555.pdf

SModelS – a decomposer and a database

We decompose full theories into simplified models, and match them against our database.



SModelS – a decomposer and a database

We laboriously collected CMS' and ATLAS' SMS results and are now proud owners of a database of results from approx. 50 (CMS) + 50 (ATLAS) analyses.

#	ID	Pretty Name	Type	$\mathcal{L}$ [fb^{-1}]
1	CMS-PAS-EXO-16-036	hscp search	ul, eff	12.9
2	CMS-PAS-SUS-16-022	>= 3 l's + $\not{E}_T$	ul	12.9
3	CMS-PAS-SUS-16-052	soft l, <= 2 jets	ul, eff	35.9
4	CMS-PAS-SUS-17-004	multi-l EWK searches	ul	35.9
5	CMS-SUS-16-009	multijets + $\not{E}_T$, top tagging	ul	2.3
6	CMS-SUS-16-032	Sbottom and compressed stop	ul	35.9
7	CMS-SUS-16-033	0L + jets + $\not{E}_T$	ul, eff	35.9
8	CMS-SUS-16-034	2 OSSF l's	ul	35.9
9	CMS-SUS-16-035	2 SS l's	ul	35.9
10	CMS-SUS-16-036	0L + jets + $\not{E}_T$	ul	35.9
11	CMS-SUS-16-037	1L + jets + $\not{E}_T$ with MJ	ul	35.9
12	CMS-SUS-16-039	multi-l EWK searches	ul	35.9
13	CMS-SUS-16-041	multi-ls + jets + $\not{E}_T$	ul	35.9
14	CMS-SUS-16-042	1L + jets + $\not{E}_T$	ul	35.9
15	CMS-SUS-16-043	EWK WH	ul	35.9
16	CMS-SUS-16-045	Sbottom to bHbH and H $\to \gamma\gamma$	ul	35.9
17	CMS-SUS-16-046	γ + $\not{E}_T$	ul	35.9
18	CMS-SUS-16-047	γ + HT	ul	35.9
19	CMS-SUS-16-049	All hadronic stop	ul	35.9
20	CMS-SUS-16-050	0L + top tag	ul	35.9
21	CMS-SUS-16-051	1L stop	ul	35.9
22	CMS-SUS-17-001	Stop search in dil + jets + $\not{E}_T$	ul	35.9
23	CMS-SUS-17-003	2 taus + $\not{E}_T$	ul	35.9
24	CMS-SUS-17-004	EW-ino combination	ul	35.9
25	CMS-SUS-17-005	1L + multijets + $\not{E}_T$, top tagging	ul	35.9
26	CMS-SUS-17-006	jets + boosted H(bb) + $\not{E}_T$	ul	35.9
27	CMS-SUS-17-009	SFOS l's + $\not{E}_T$	ul	35.9
28	CMS-SUS-17-010	2L stop	ul	35.9
29	CMS-SUS-18-002	γ, jets, b-jets+ $\not{E}_T$, top tagging	ul	35.9
30	CMS-SUS-19-006	0L + jets, MHT	ul	137.0

#	ID	Pretty Name	Type	$\mathcal{L}$ [fb^{-1}]
				18.8
				18.8
				19.4
			f	19.7
				19.4
				18.9
			f	19.4
				11.7
				19.5
				19.3
				19.5
12			ul, eff	19.3
13	CMS-SUS-13-011	1 l + >= 4 (1b-)jets + $\not{E}_T$	ul, eff	19.5
14	CMS-SUS-13-012	n_{jets} + HTmiss	ul, eff	19.5
15	CMS-SUS-13-013	2 SS l's + (b-)jets + $\not{E}_T$	ul, eff	19.5
16	CMS-SUS-13-019	>= 2 jets + $\not{E}_T$, MT2	ul	19.5
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	Pretty Name	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, eff	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	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	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	139.0

#	ID	Pretty Name	Type	$\mathcal{L}$ [fb^{-1}]
				20.7
				20.5
				20.7
				20.3
				20.3
				20.1
				20.3
				20.1
				20.3
				20.3
				20.3
				20.3
				20.3
				20.1
				20.3
				20.3
				20.3
				20.3
				20.3
				20.1
				20.1
22	ATLAS-SUSY-2013-19	2 OS l's + (b-)jets + $\not{E}_T$	ul	20.3
23	ATLAS-SUSY-2013-21	monojet or c-jet + $\not{E}_T$	eff	20.3
24	ATLAS-SUSY-2013-23	1 l + 2 b-jets (or 2 γs) + $\not{E}_T$	ul	20.3
25	ATLAS-SUSY-2014-03	>= 2(c-)jets + $\not{E}_T$	eff	20.3

SModelS – a decomposer and a database

For many (but not all) of the results in our database we can construct approximate likelihoods. Now if we know which of these likelihoods are approximately uncorrelated, we can perform combinations: searching for hints of potential dispersed signals in published results becomes a combinatorial problem!

Proto-models

In order to identify potential dispersed signals, we need to build theoretical contexts for them: we need to build precursor models for the NSM, with a particle content that is typically larger than that of individual SMSes, but much smaller than the particle content of SUSY. We call these precursor theories “proto-models”. Think of them as “stacked up” simplified models.

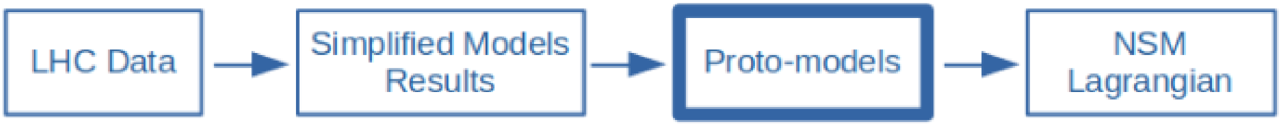


Figure 2: The overall strategy at how we envisage to construct the NSM from LHC Data: the raw data are described via Simplified Models results. From these, we shall construct proto-models. These proto-models are intended to serve as the input to constructing the NSM. The construction of proto-models is subject of this publication.

Artificial Proto-Modelling: Building Precursors of a Next Standard Model from Simplified Models Results

to appear on arXiv hopefully soonish

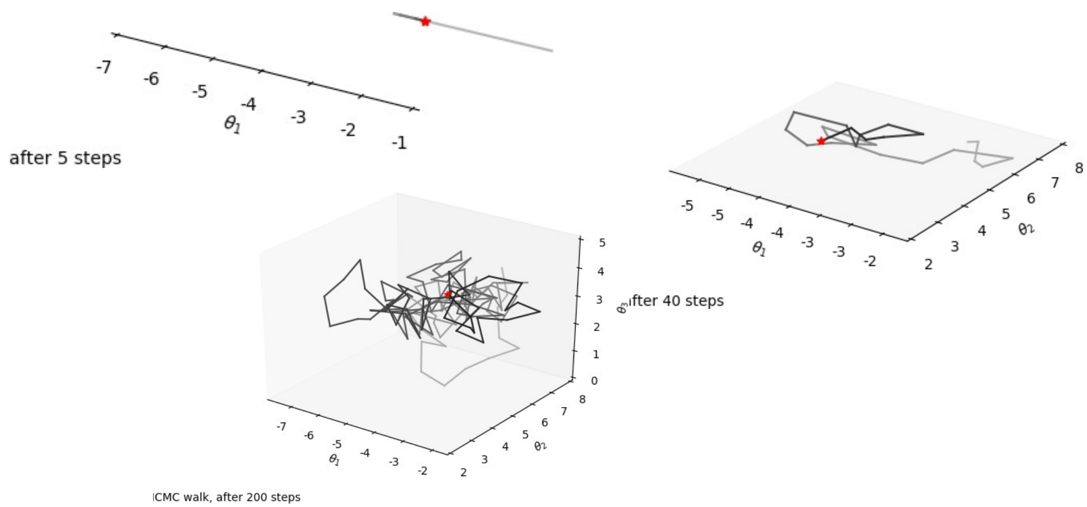
http://www.hephy.at/user/wwaltenberger/models/mcmc.webm

Proto-models

How do we construct such protomodels? In the publication we are working on right now, we propose that we construct them in a random walk.
Instead of walking the parameter space of e.g. the pMSSM, we walk in the space of all possible protomodels.

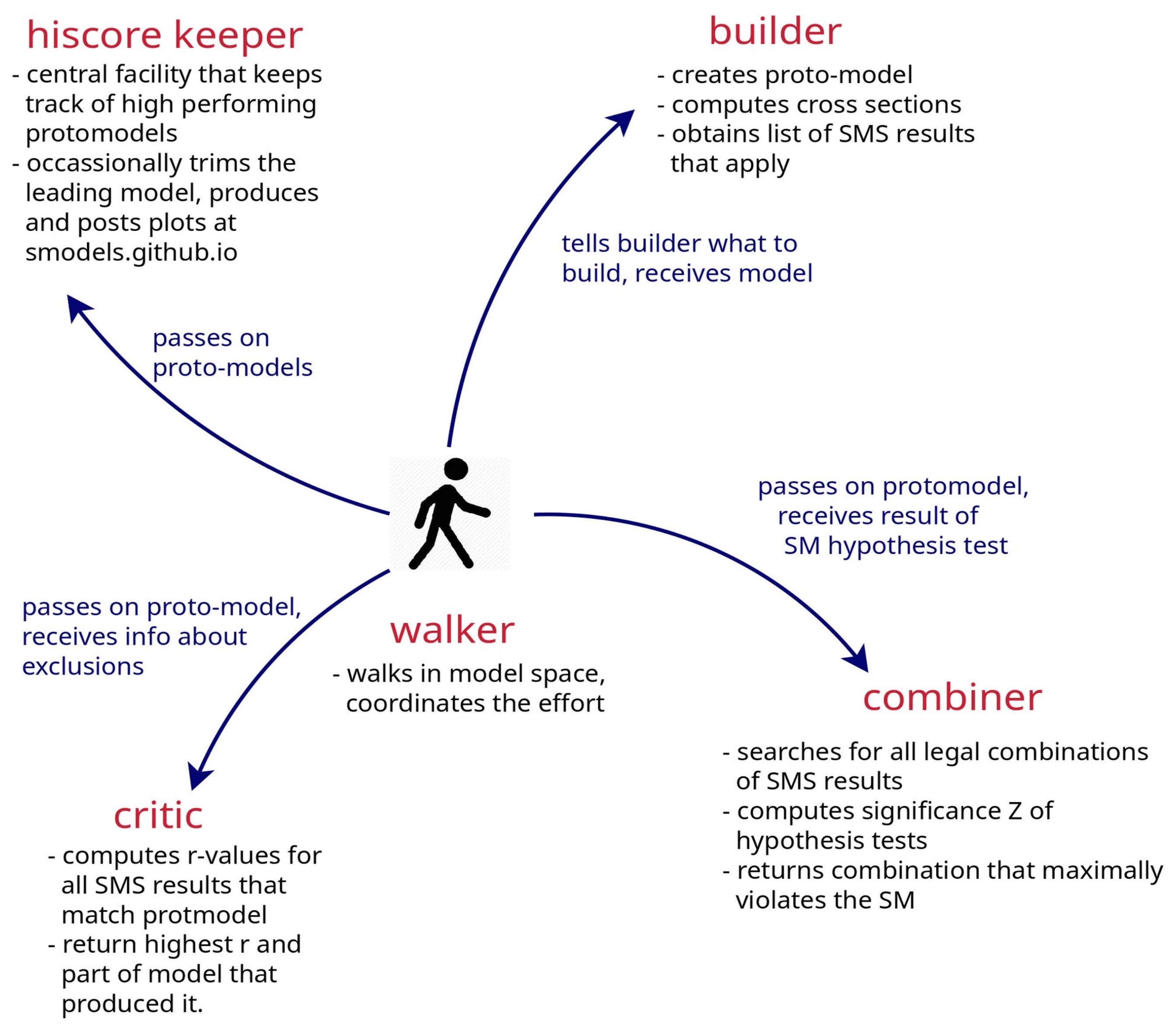
Possible actions being taken within the MCMC walk:

- randomly add a BSM particle
- randomly remove a BSM particle
- randomly change a particle’s mass
- randomly change the decay of a particle (channels and ratios)
- randomly change a signal strength multiplier
- ….



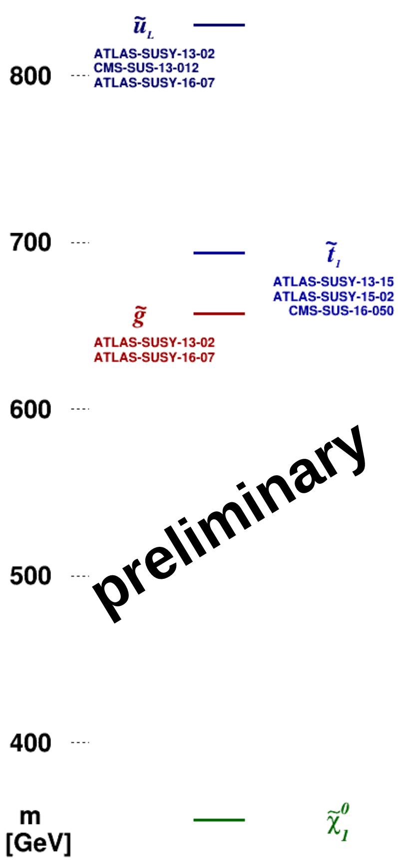
An AIC-like criterion penalizes for newly introduced degrees of freedom.
This is very similar to “weight decay” in neural networks, or “regularization” in classical regression.

Proto-models



Proto-models

The overall vision of this being that instead of postulating NSM candidates and then falsifying them (or failing to do so), we put the model building into the statistical procedure itself. A slow, bottom-up procedure, starting from data.



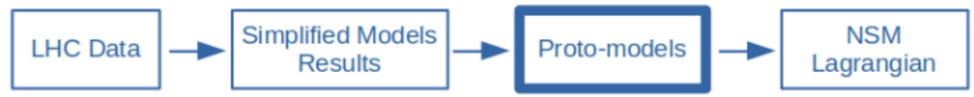


Figure 2: The overall strategy at how we envisage to construct the NSM from LHC Data: the raw data are described via Simplified Models results. From these, we shall construct proto-models. These proto-models are intended to serve as the input to constructing the NSM. The construction of proto-models is subject of this publication.

preliminary

Analysis Name	Type	Dataset	Observed	Expected	Approx σ	Particles
ATLAS-SUSY-2015-02	em	SR1	12	5.5 +/- 0.72	2.6 σ	$\sim t_1$
ATLAS-SUSY-2016-07	em	2j_Meff_1200	611	526 +/- 31	2.2 σ	$\sim u_L$,$\sim g$
CMS-SUS-16-050	ul	-	96.2 fb	51.9 fb	1.7 σ	$\sim t_1$
ATLAS-SUSY-2013-02	em	SR6jtp	6	4.9 +/- 1.6	0.4 σ	$\sim u_L$,$\sim g$
ATLAS-SUSY-2013-15	em	tNboost	5	3.3 +/- 0.7	0.9 σ	$\sim t_1$
CMS-SUS-13-012	ul	-	42.6 fb	25.8 fb	1.3 σ	$\sim u_L$

A handful of "mild" excesses. Irrelevant, if taken individually.

Proto-models

Our MCMC walks are but a crutch, a burden we must carry because we do not have derivatives, i.e. gradients and Hessians.

If we had gradients we could instead perform gradient descent to find the best model, and we could use the Fisher information to infer the errors on its parameters.

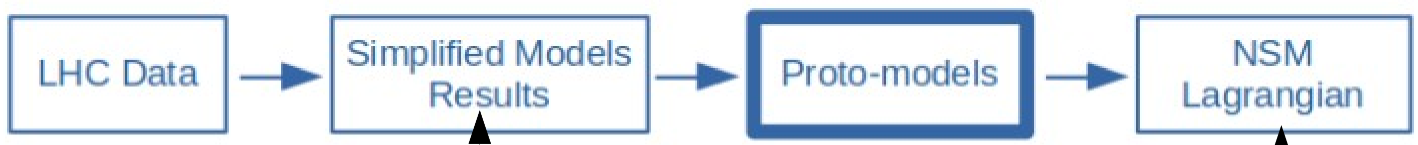
So, how about we make the whole chain differentiable?

One Chain To Rule Them All

Our MCMC walks are but a crutch, a burden we must carry because we do not have derivatives, i.e. gradients and Hessians.

If we had gradients we could instead perform gradient descent to find the best model, and we could use the Fisher information to infer the error on its prameters (if you want non-Gaussian posteriors you can still MCMC-sample if you wish).

So, how about we make the whole chain differentiable?



described as likelihoods L that are differentiable with respect to the yields y_i

we have started an effort to make SModelS differentiable w.r.t SMS parameters p_j, by learning our entire database:

thats just a sum of simplified models → differentiable!

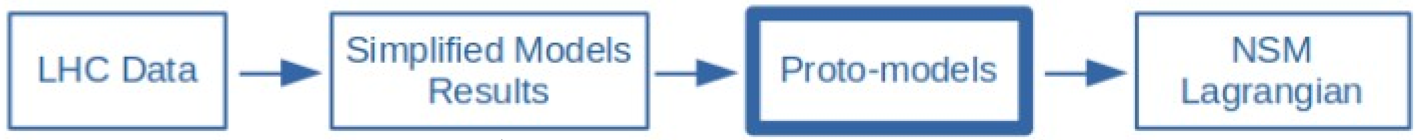
for individual candidates we can make this differentiable w.r.t fundamental parameters Θ_l, via neural networks, with efforts similar to DeepXS, or “TheoryGANs” [*]:

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

Needless to say, the data pipeline sketched above is not the only feasible one. Differentiability however would be a helpful tool for all possible data pipelines. A similar rationale would apply also to EFTs, Wilson coefficients and data from measurements.

https://arxiv.org/abs/1810.08312

Differentiable induction



described as likelihoods L that are differentiable with respect to the yields y_i

we have started an effort to make SModelS differentiable w.r.t SMS parameters p_j, by learning our entire database:

thats just a sum of simplified models → differentiable!

for individual candidates we can make this differentiable w.r.t fundamental parameters Θ_l, via neural networks, with efforts similar to DeepXS, or “TheoryGANs” [*]:

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

“statistics”, statistical models

“machine learning”, surrogate models

differentiable programming

Differentiable induction

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

All of this is to say, that we realistically can try to “learn” the fundamental laws of the universe from data, as opposed to postulating them. Gradient-free for starters, adding gradients in the long run.

“differentiable inductive reasoning”, if you wish.

Summary

Inferring fundamental laws of physics from observations is an inductive step. We may be lucky and "guess" correctly, but in general there is no guarantee for success.

We can think of the approaches pursued as being either "top-down" (starting from an idea in theoretical physics) or "bottom-up" (starting from data).

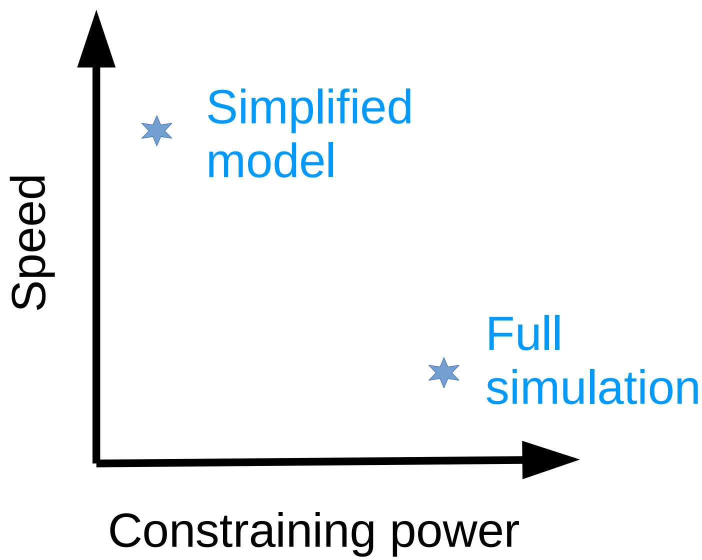
Algorithmically we can distinguish between gradient-free approaches, (e.g. MCMC walks) or gradient-based methods (e.g. using surrogate NN theory models). So far, however approaches have been gradient-free.

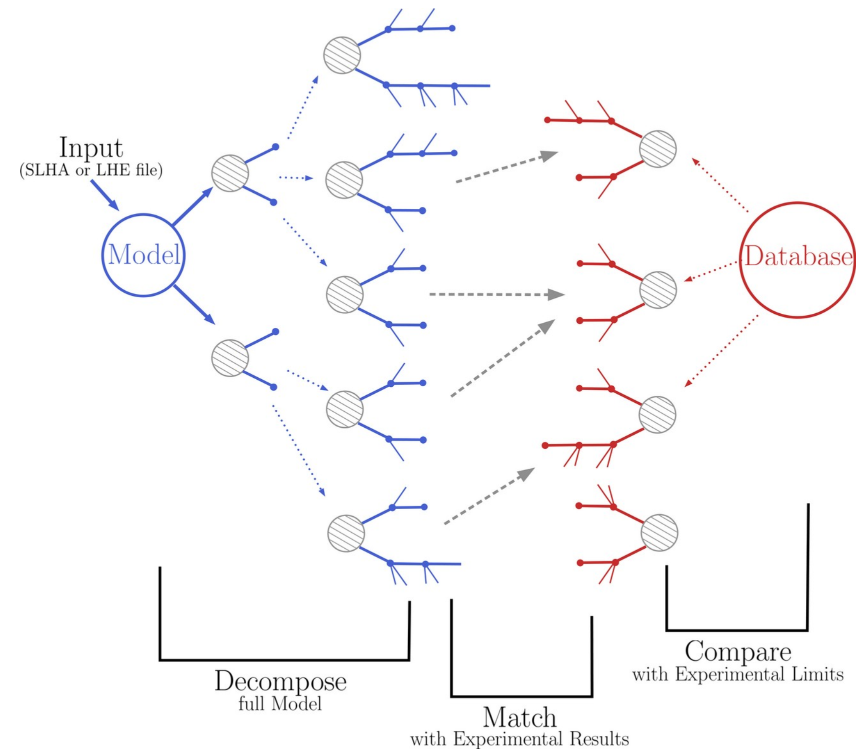
Bayesian MCMC walks can naturally be combined with the notion of phenomenological "bottom-up" model building.

Gradients can majorly speed up inference and allow for higher dimensional models. Surrogate (neural network) models automatically deliver gradients! **Differentiable programming for the win!**

Recap: the Idea behind SModelS

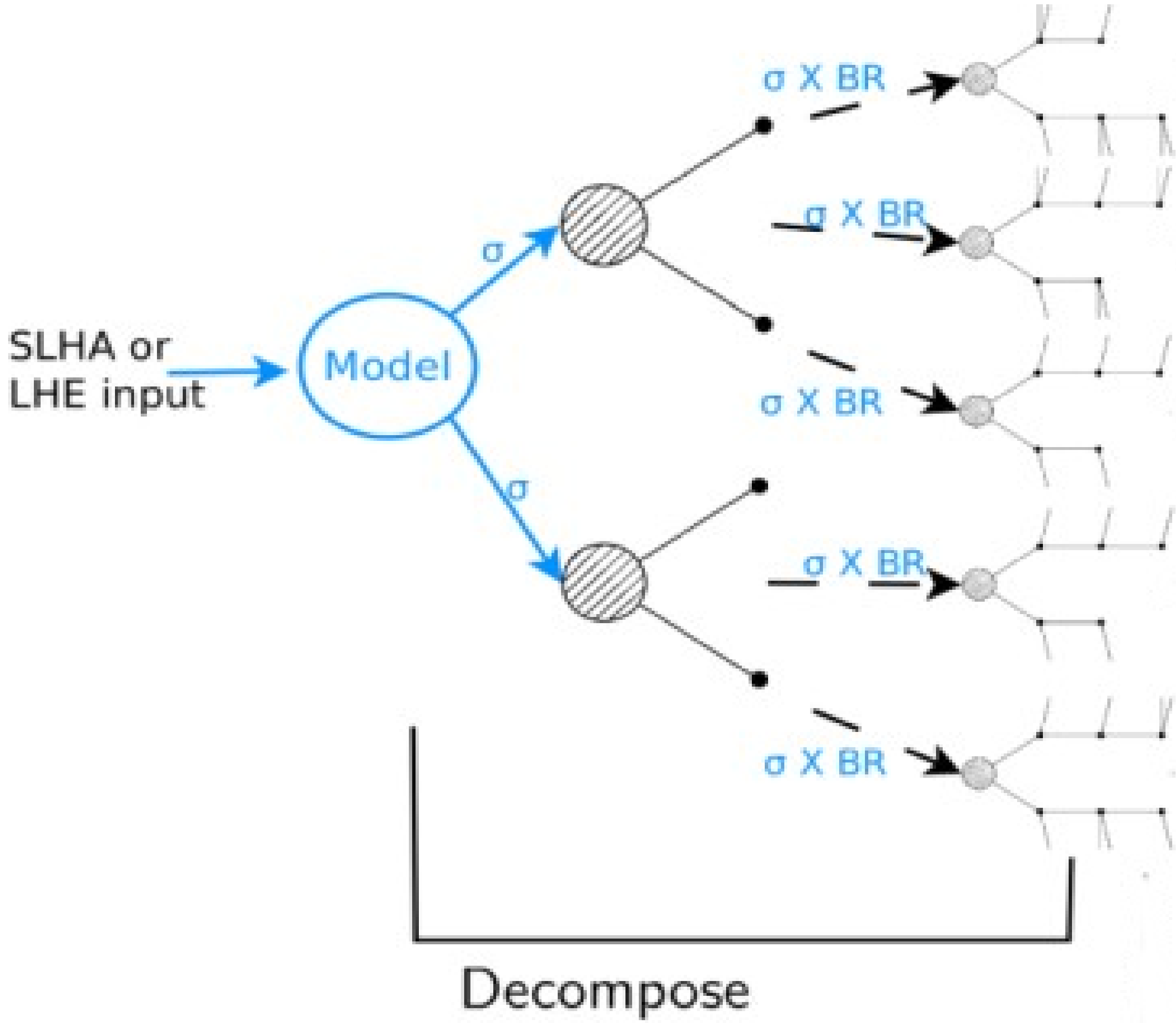
SModelS confronts theories beyond the Standard Model (BSM) with LHC search results by decomposing full models into their simplified models topologies, and comparing the cross section predictions of these individual topologies with a database of SMS results.





Recap: How SModelS works

1) Decomposition of a fundamental model



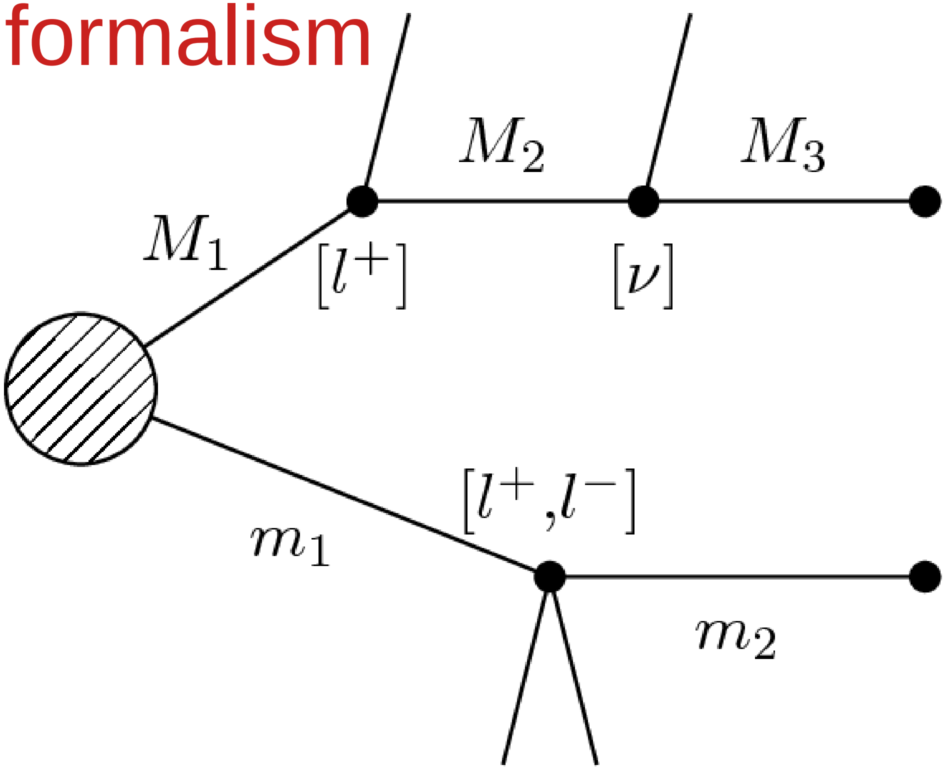
Input: SLHA file (mass spectrum, BRs) or LHE file (parton level)

Currently the model must have a $\mathbf{Z}_2$ symmetry

The decomposition produces a set of simplified model topologies (dubbed “elements”)

Recap: How SModelS works

2) Description of the topology in the SModelS formalism



$$= [[l^+],[\nu]]$$

$$= [\ [[l^+],[\nu]]\ ,\ [[l^+,l^-]]\ ]$$
$$([[M_1,M_2,M_3],[m_1,m_2]])$$

$$= [[l^+,l^-]]$$

Each topology is described by:

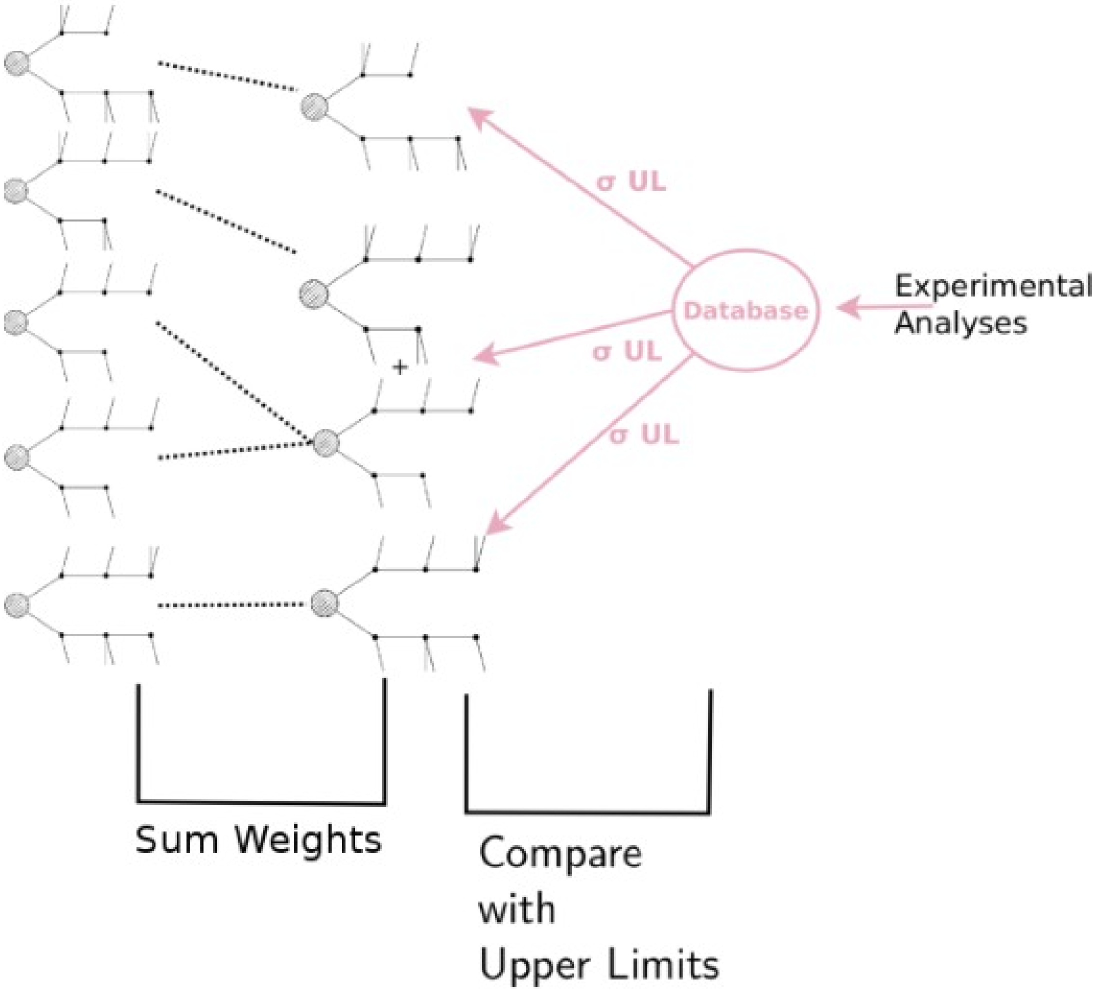
- Topology shape + final states
- BSM masses
- σ x BR

We (currently) ignore spin, color, etc of the BSM particles

It is model independent, there is no reference to the original model

Recap: How SModelS works

3) Comparison of predicted signal strengths with experimental result:



- **Upper Limit Results:**
  Predicted signal strength = σ x BR
  Experimental result: σ_{UL}

- **Efficiency Map Results:**
  Predicted signal strength = ∑ σ x BR x ε
  Experimental result: $\sigma_{UL} = N_{UL} / L$ from $N_{observed}$, expected(BG), error(BG)

- r = predicted / σ_{UL}

- Model is excluded if most constraining analysis has r > 1