Statistical models and ML searches in CheckMATE

Krzysztof Rolbiecki
University of Warsaw

HEP Theory Seminar
Bethe Center for Theoretical Physics
Universität Bonn
22.06.2026

CheckMATE is 12 years old!

CheckMATE: Confronting your Favourite New Physics Model with LHC Data

Manuel Drees,* Herbi Dreiner,† and Daniel Schmeier‡
Physikalisches Institut and Bethe Center for Theoretical Physics, University of Bonn, Bonn, Germany

Jamie Tattersall§
Institut für Theoretische Physik, University of Heidelberg, Heidelberg, Germany and
Physikalisches Institut and Bethe Center for Theoretical Physics, University of Bonn, Bonn, Germany

Jong Soo Kim¶
Instituto de Fisica Teorica UAM/CSIC, Madrid,
Spain and ARC Centre of Excellence for Particle Physics at the Terascale,
School of Chemistry and Physics, University of Adelaide, Australia
(Dated: January 29, 2014)

https://arxiv.org/abs/1312.2591

TTPU-16-38 CSIC-16-116 TTK-16-47

CheckMATE 2: From the model to the limit

Daniel Dercks[a,b,1], Nishita Desai[c,2], Jong Soo Kim[d,e,3], Krzysztof Rolbiecki[f,4], Jamie Tattersall[g,5], Torsten Weber[g,6]

[a] *II. Institute for Theoretical Physics, University of Hamburg, Luruper Chaussee 149, D-22761 Hamburg, Germany*
[b] *Bethe Center for Theoretical Physics & Physikalisches Institut der Universität Bonn, Nussallee 12, D-53115 Bonn, Germany*
[c] *Laboratoire Charles Coulomb (L2C) UMR 5221 & Laboratoire Univers et Particules de Montpellier (LUPM) UMR 5299, CNRS-Université de Montpellier, 34090 Montpellier, France*
[d] *Center for Theoretical Physics of the Universe, Institute for Basic Science (IBS), Daejeon, 34051, Korea*
[e] *Universidad Autónoma de Madrid, Instituto de Física Teórica, Calle Nicolás Cabrera 13-15, Cantoblanco, 28049 Madrid, Spain*
[f] *Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland*
[g] *Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, D-52056 Aachen, Germany*

https://arxiv.org/abs/1611.09856

A framework to create customised LHC analyses within CheckMATE

Jong Soo Kim*
Instituto de Física Teórica UAM/CSIC, Madrid, Spain

Daniel Schmeier†
Physikalisches Institut and Bethe Center for Theoretical Physics, University of Bonn, Bonn, Germany

Jamie Tattersall‡
Institut für Theoretische Physik, University of Heidelberg, Heidelberg, Germany

Krzysztof Rolbiecki§
Instituto de Física Teórica UAM/CSIC, Madrid, Spain and
Instytut Fizyki Teoretycznej, Uniwersytet Warszawski, Warsaw, Poland
(Dated: June 3, 2015)

https://arxiv.org/abs/1503.01123

Constraining electroweak and strongly charged long-lived particles with CheckMATE

Nishita Desai[a], Florian Domingo[b], Jong Soo Kim[c], Roberto Ruiz de Austri Bazan[d], Krzysztof Rolbiecki[e], Mangesh Sonawane[f], Zeren Simon Wang[gh]

https://arxiv.org/abs/2104.04542

Implementation of full and simplified likelihoods in CheckMATE

Iñaki Lara and Krzysztof Rolbiecki

https://arxiv.org/abs/2507.08565

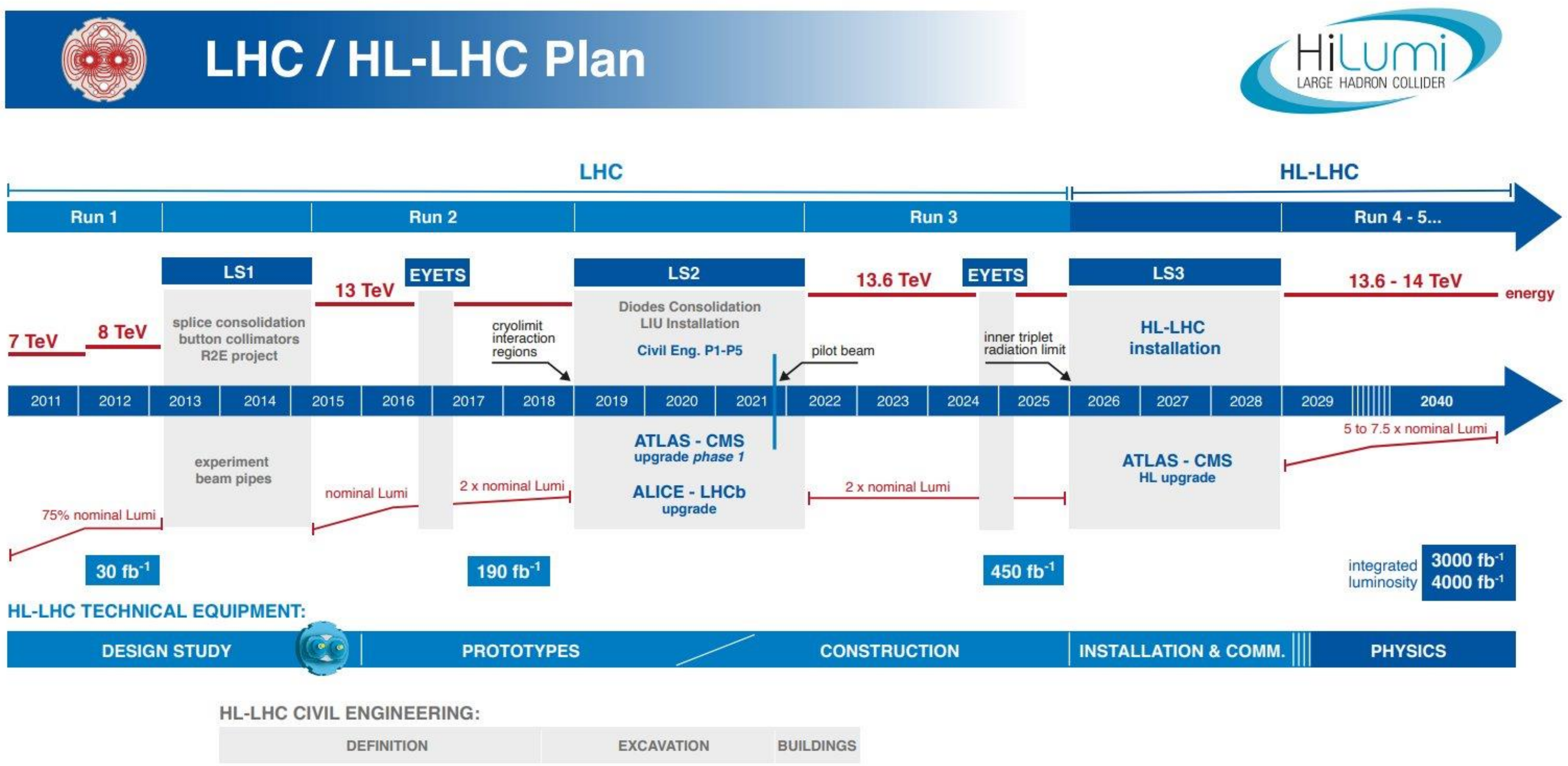
over 800 citations in total!

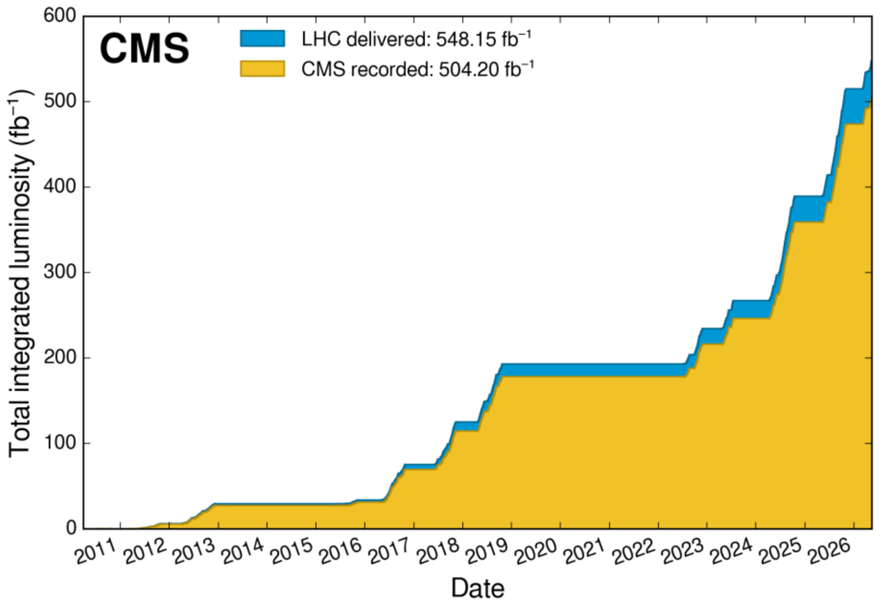
https://github.com/CheckMATE2/checkmate2

Contents

Contents

LHC timeline





Still a long road ahead!

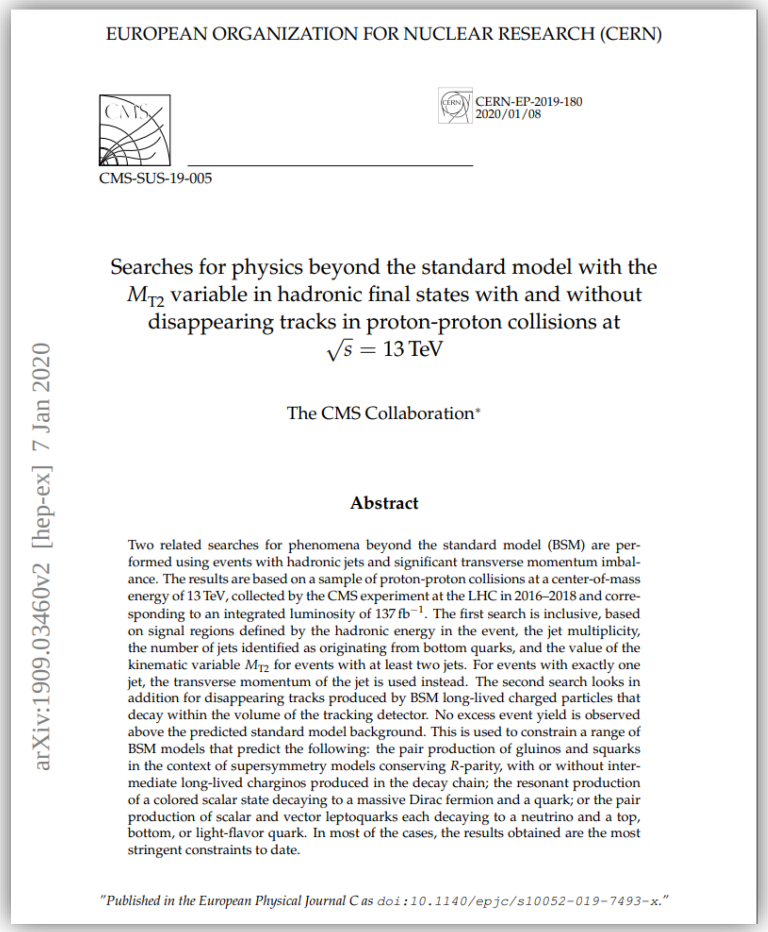
Monte Carlo tools & discoveries at the (HL)LHC

Searches for new TeV-scale physics still one of the main goals in the coming years

- Theoretical model building offers a vast number of models with particles in the LHC reach
- Experimental papers cover only a small fraction of existing models
- We need tools to cover the gap and: assess viability of models, guide future searches, looking for blind spots
- Computer tools are essential: Monte Carlo generators, fast detector simulators, cross section calculators
- We need tools to analyze MC output easily and compare it quickly and reliably with existing experimental exclusions

This is the main purpose of recasting tools

Reinterpretation/recasting in a nutshell



EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

CERN-EP-2019-180
2020/01/08

CMS-SUS-19-005

Searches for physics beyond the standard model with the M_{T2} variable in hadronic final states with and without disappearing tracks in proton-proton collisions at $\sqrt{s} = 13$ TeV

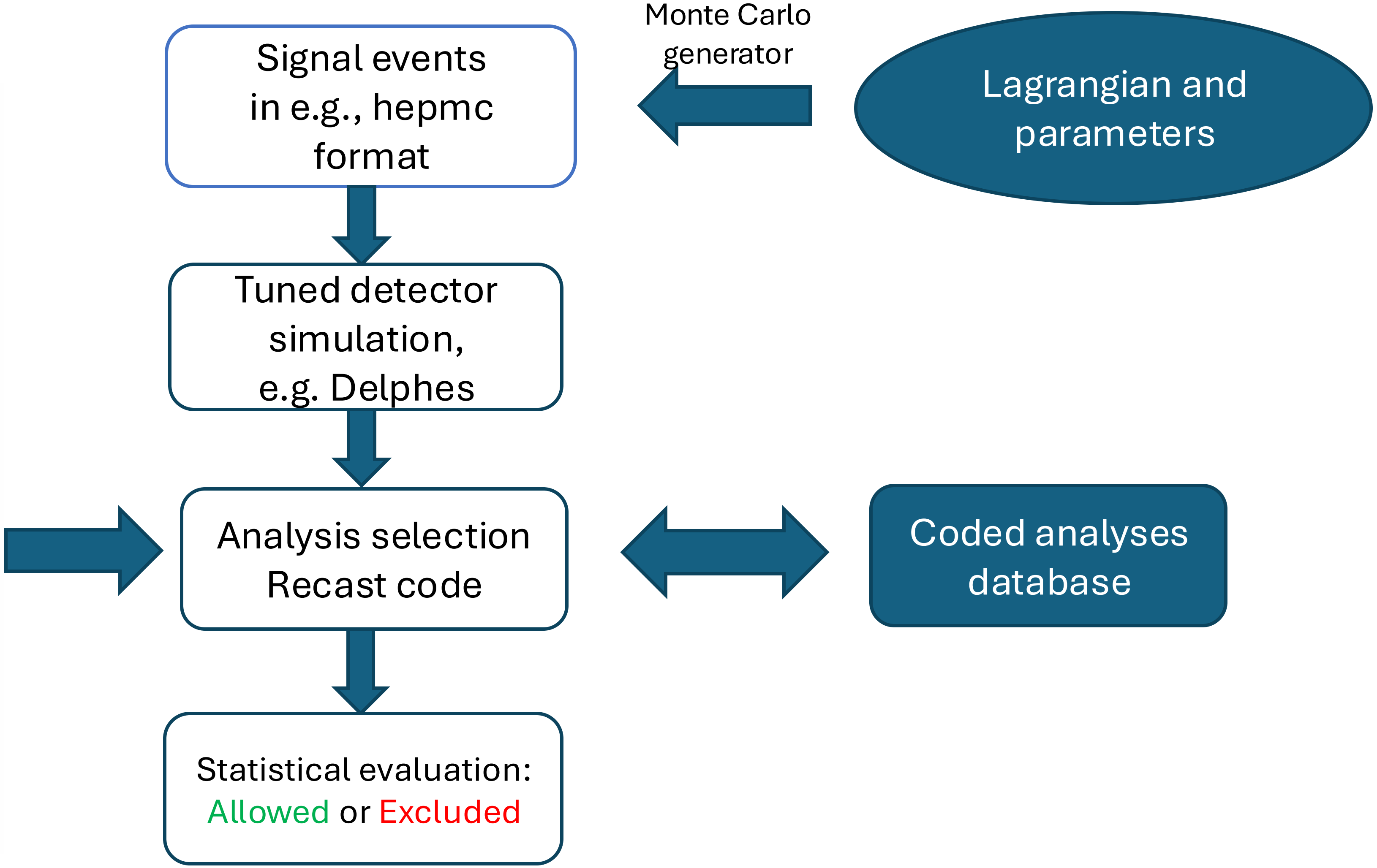
The CMS Collaboration*

Abstract

Two related searches for phenomena beyond the standard model (BSM) are performed using events with hadronic jets and significant transverse momentum imbalance. The results are based on a sample of proton-proton collisions at a center-of-mass energy of 13 TeV, collected by the CMS experiment at the LHC in 2016–2018 and corresponding to an integrated luminosity of 137 fb^{-1}. The first search is inclusive, based on signal regions defined by the hadronic energy in the event, the jet multiplicity, the number of jets identified as originating from bottom quarks, and the value of the kinematic variable M_{T2} for events with at least two jets. For events with exactly one jet, the transverse momentum of the jet is used instead. The second search looks in addition for disappearing tracks produced by BSM long-lived charged particles that decay within the volume of the tracking detector. No excess event yield is observed above the predicted standard model background. This is used to constrain a range of BSM models that predict the following: the pair production of gluinos and squarks in the context of supersymmetry models conserving R-parity, with or without intermediate long-lived charginos produced in the decay chain; the resonant production of a colored scalar state decaying to a massive Dirac fermion and a quark; or the pair production of scalar and vector leptoquarks each decaying to a neutrino and a top, bottom, or light-flavor quark. In most of the cases, the results obtained are the most stringent constraints to date.

"Published in the European Physical Journal C as doi:10.1140/epjc/s10052-019-7493-x."

arXiv:1909.03460v2 [hep-ex] 7 Jan 2020



LHC Reinterpretation Working Group

The purpose of the REI WG is to discuss topics related to the BSM (re)interpretation of LHC data, including the development of the necessary public recasting tools and related infrastructure, and to provide a platform for a continued interaction between theorists and with the experiments. The recent topics:

- the publication and reuse of statistical models
- the reinterpretation of analyses that employ machine learning
- global analyses and global fits
- preservation of data and methods for replication/reanalysis in future: for a once in a lifetime experiment, we want to make sure all the necessary information is provided and understandable for people outside of a particular analysis

(Re)interpretation of the LHC results for new physics

Feb 25 – 28, 2025
CERN
Europe/Paris timezone

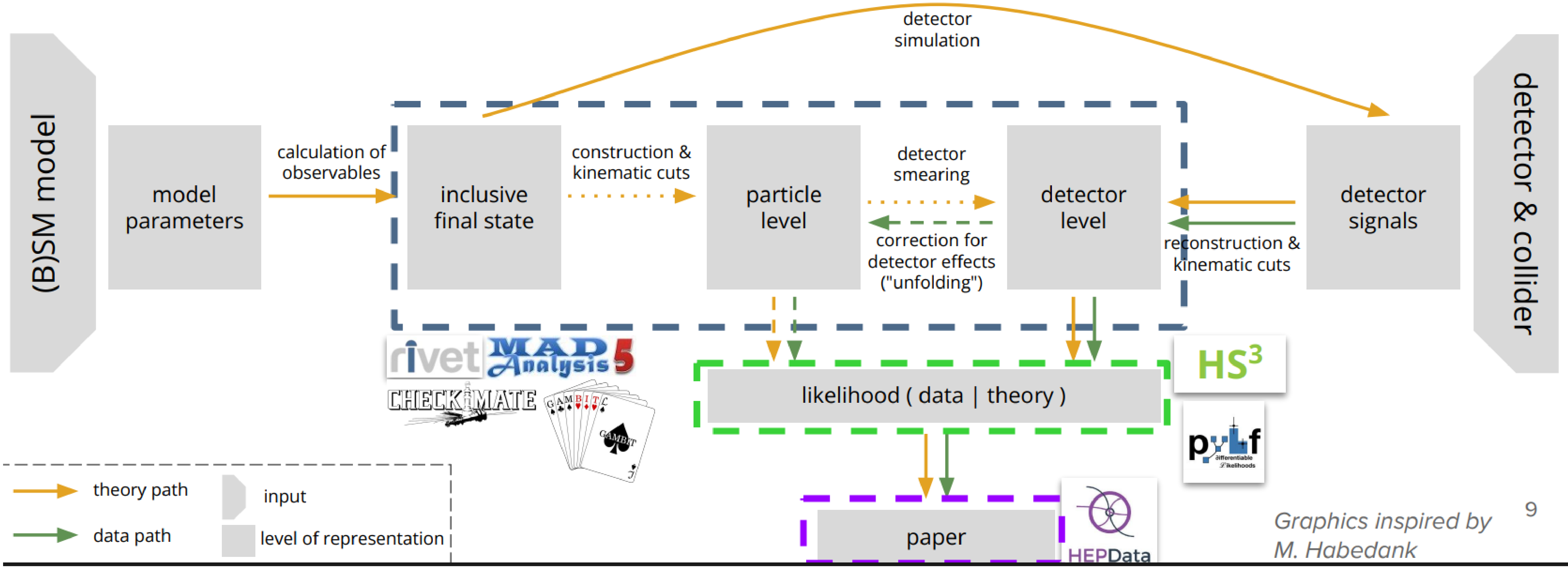
Overview
Timetable

Participant List
130 participants

https://indico.cern.ch/event/1466101/

Reinterpretation codes on the market

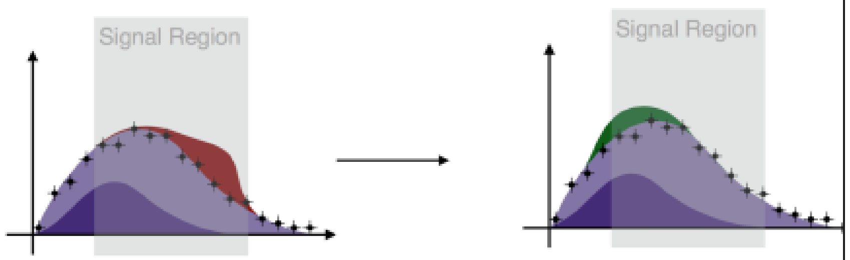
Full analysis flow

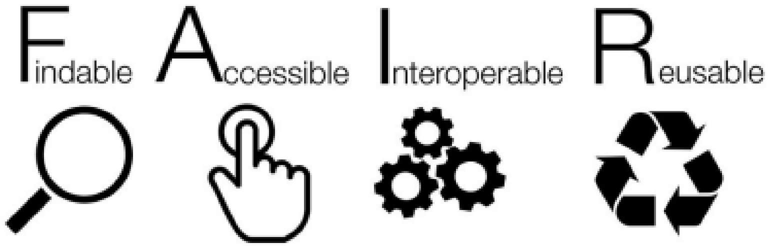


borrowed from Tomek Procter

Why bother with analysis preservation?

- All searches should be **re-exploitable:**
  - Able to constrain another model in a similar signature. **Repeating searches** because we didn't correctly preserve/present past results **is inefficient and unacceptable**.
- Serious focus on **analysis preservation** and **re-interpretability.** Results should still **be re-useable** until next hadron collider comes online in **half a century**.
- On some topics, the Run2+3 searches will be the last word. We need to get this right **NOW**.
- On a human level, knowing one is working on a long-impact publication is more motivating

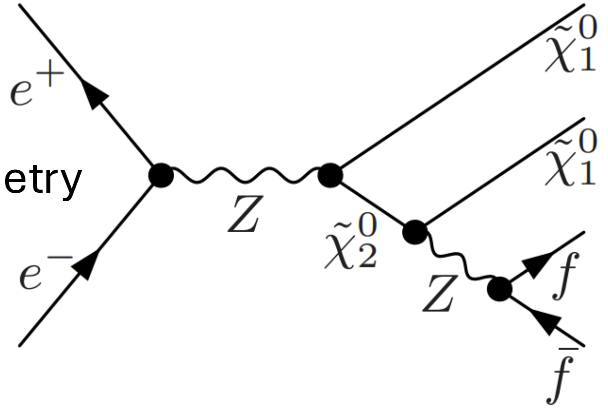




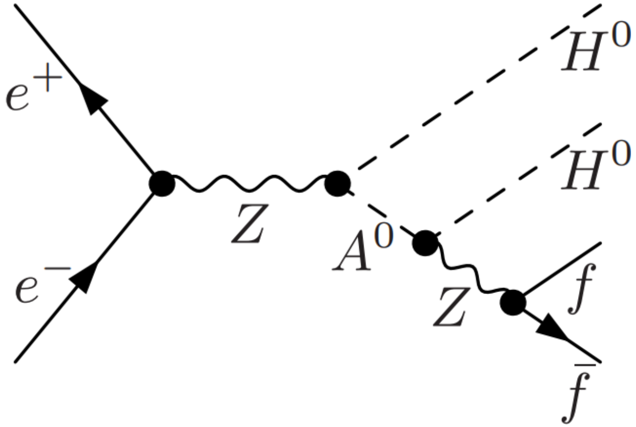
My pitch to my ATLAS colleagues !

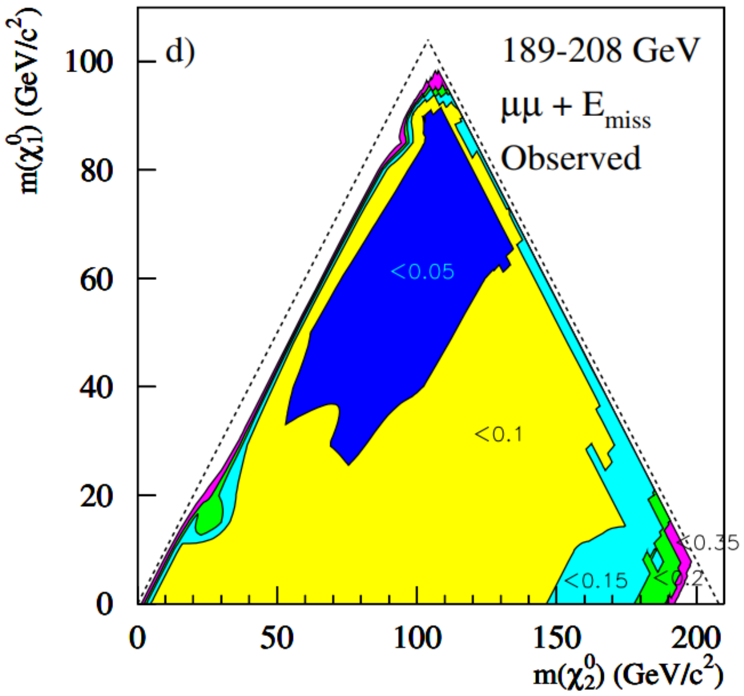
Recasting in practice from LEP to LHC

Supersymmetry

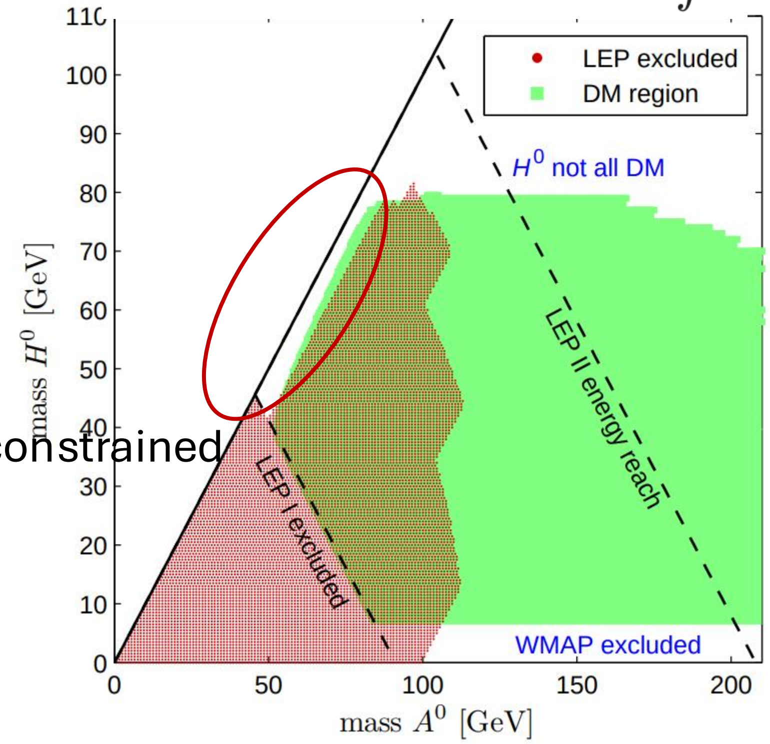


Inert doublet model:
(version of two Higgs doublets model)





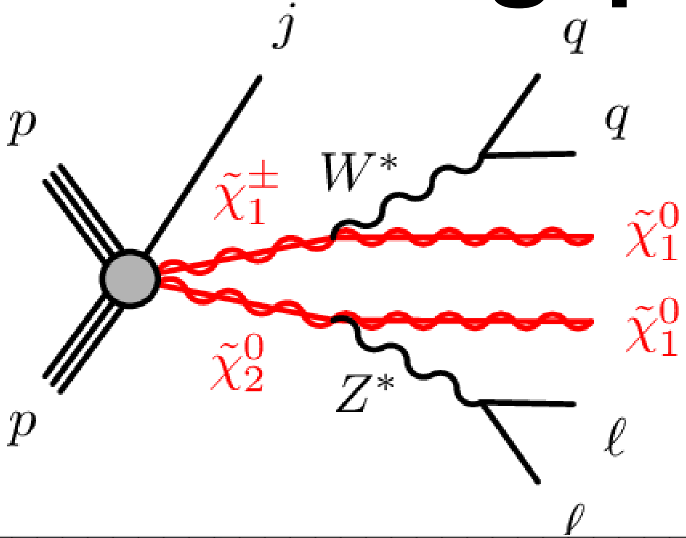
Small mass difference unconstrained

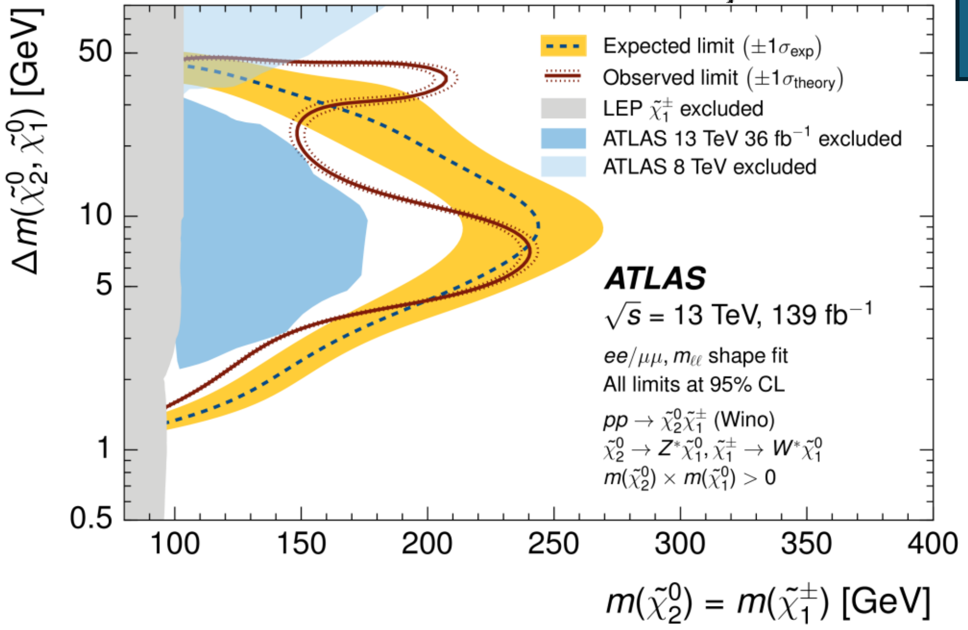


Espirito Santo ea. DELPHI PHYS 928 (2003)

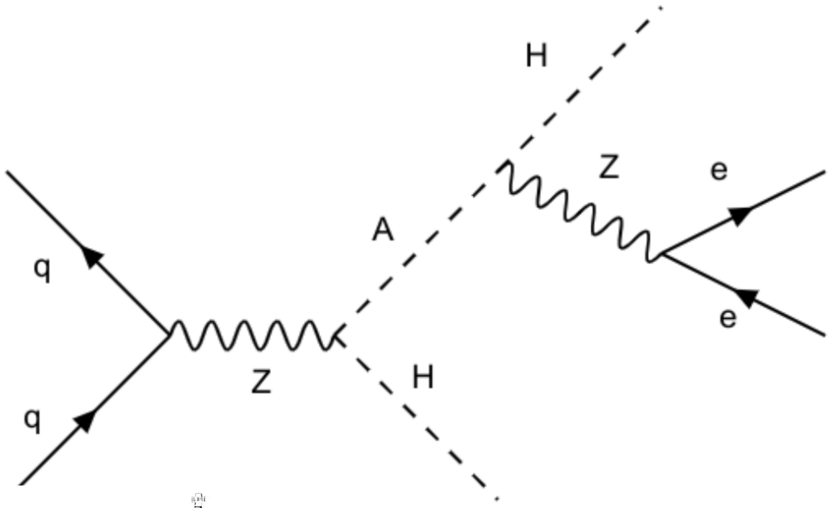
Lundstrom, Gustafsson, Edsjo, arXiv:0810.3924

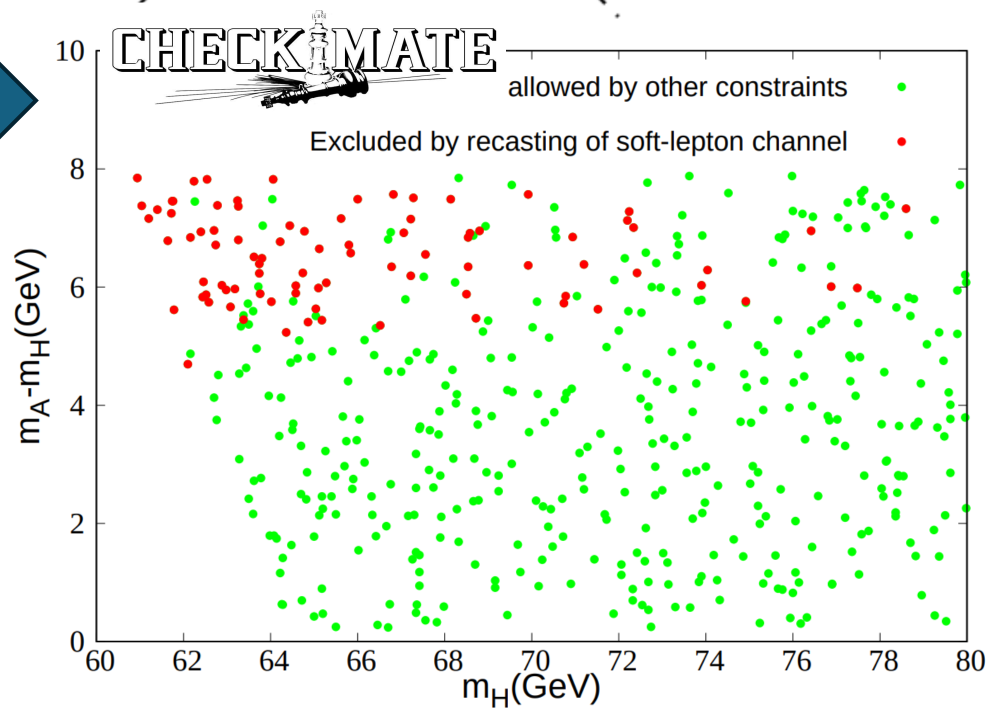
Small mass gap revisited





SUSY-2018-16



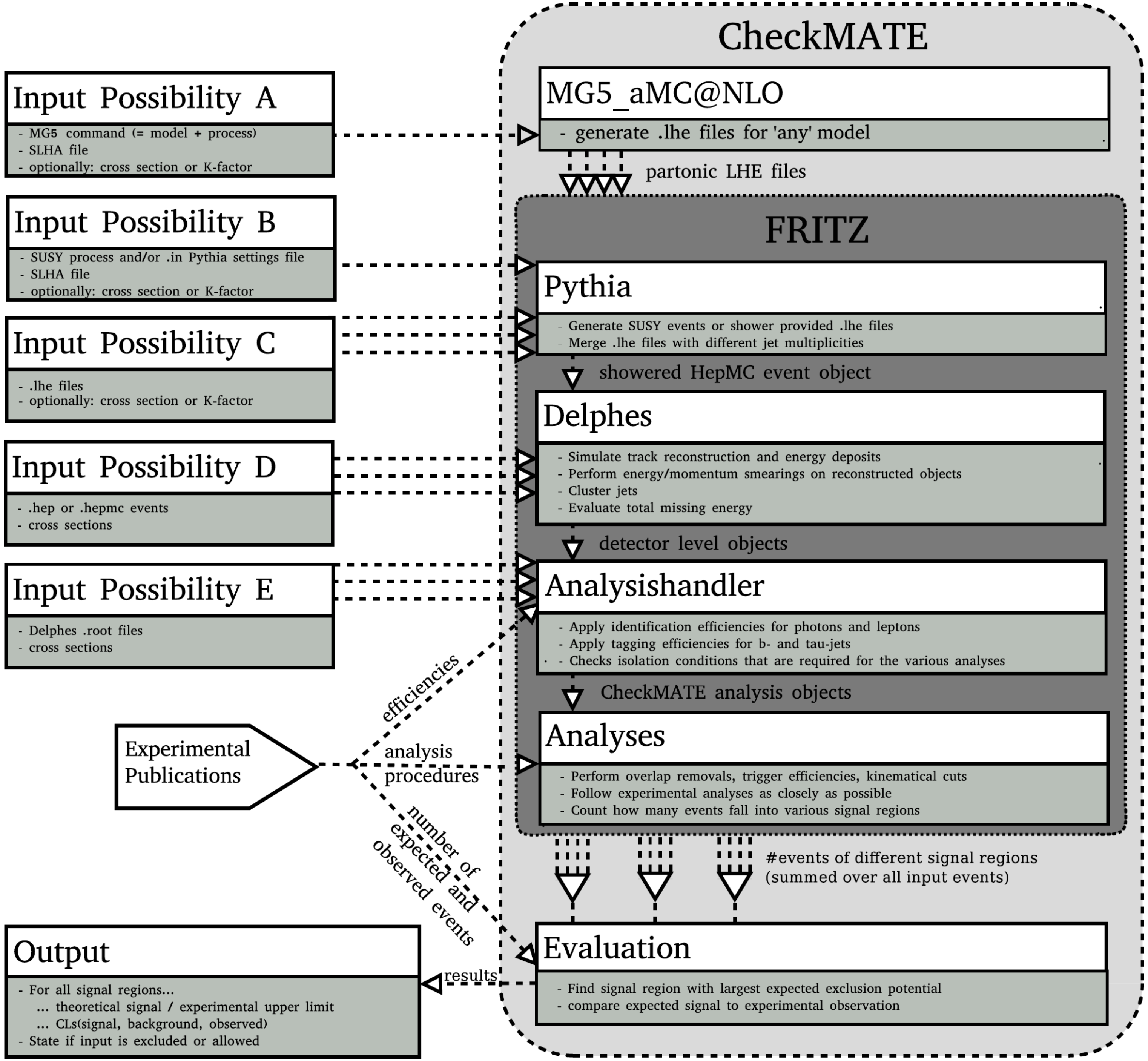


J. Lahiri, T. Robens, KR, arXiv:2511.23133

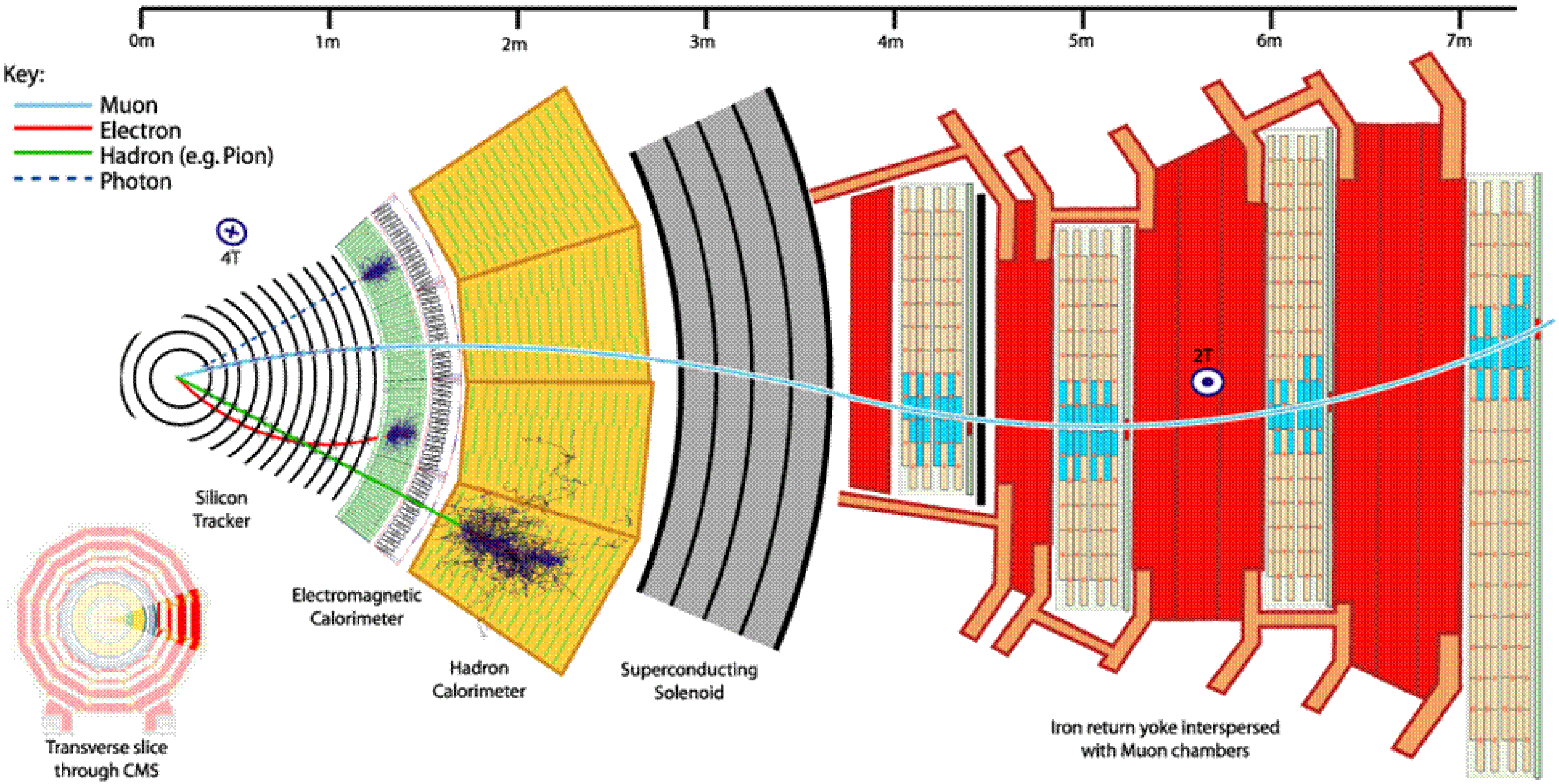
Contents

CHECKMATE

- CheckMATE is a general tool for recasting arbitrary model
- Accepts events as .hepmc, .lhe; integration with Pythia and MadGraph
- based on Delphes for detector simulation
- using existing LHC searches calculates a limit on a given parameter point
- From SLHA file to the limit in one click
- one can easily constrain models that were not covered in the original ATLAS/CMS search
- currently more than 50 searches at 13 TeV coded, including 25 with full Run 2 luminosity
- long-lived particles branch
- https://checkmate.hepforge.org/ and https://github.com/CheckMATE2/checkmate2


CheckMATE
Input Possibility A
- MG5 command (= model + process)
- SLHA file
- optionally: cross section or K-factor
Input Possibility B
- SUSY process and/or .in Pythia settings file
- SLHA file
- optionally: cross section or K-factor
Input Possibility C
- .lhe files
- optionally: cross section or K-factor
Input Possibility D
- .hep or .hepmc events
- cross sections
Input Possibility E
- Delphes .root files
- cross sections
MG5_aMC@NLO
- generate .lhe files for 'any' model
partonic LHE files
FRITZ
Pythia
- Generate SUSY events or shower provided .lhe files
- Merge .lhe files with different jet multiplicities
showered HepMC event object
Delphes
- Simulate track reconstruction and energy deposits
- Perform energy/momentum smearings on reconstructed objects
- Cluster jets
- Evaluate total missing energy
detector level objects
Analysishandler
- Apply identification efficiencies for photons and leptons
- Apply tagging efficiencies for b- and tau-jets
- Checks isolation conditions that are required for the various analyses
CheckMATE analysis objects
Analyses
- Perform overlap removals, trigger efficiencies, kinematical cuts
- Follow experimental analyses as closely as possible
- Count how many events fall into various signal regions
Experimental Publications
efficiencies
analysis procedures
number of expected and observed events
#events of different signal regions (summed over all input events)
Evaluation
- Find signal region with largest expected exclusion potential
- compare expected signal to experimental observation
results
Output
- For all signal regions...
... theoretical signal / experimental upper limit
... CLs(signal, background, observed)
- State if input is excluded or allowed

Particle detector in a nutshell



Detector simulation

Delphes 3.4 / 3.5

- Simulates tracking and energy deposition
- Applies efficiencies for photons and leptons
- Clusters jets
- Performs energy/momentum smearing of all reconstructed objects
- Evaluates total missing energy
- Checks isolation conditions for photons and leptons
- Applies b-/ tau-tag on jets

DELPHES
fast simulation

CheckMATE improvements

- Added identification and isolation flags
- Tuned to reproduce LHC detectors:
  - ATLAS for 13 TeV Run; updates in progress
  - CMS work in progress

CheckMATE: ATLAS analyses

#Name	NSR	Description	Lumi
atlas_1604_01306	1	photon + MET search at 13 TeV	3.2
atlas_1605_09318	8	>= 3 b-jets + 0-1 lepton + Etmiss	3.3
atlas_1609_01599	9	ttV cross section measurement at 13 TeV	3.2
atlas_1704_03848	5	monophoton dark matter search	36.1
atlas_conf_2015_082	1	leptonic Z + jets + Etmiss	3.2
atlas_conf_2016_013	10	4 top quark (1 lepton + jets, vector like quark search)	3.2
atlas_conf_2016_050	5	1-lepton + jets + etmiss (stop)	13.3
atlas_conf_2016_054	10	1-lepton + jets + etmiss (squarks and gluino)	14.8
atlas_conf_2016_076	6	2 leptons + jets + etmiss	13.3
atlas_conf_2016_096	8	2-3 leptons + etmiss (electroweakino)	13.3
atlas_conf_2017_060	20	monojet search	36.1
atlas_conf_2016_066	2	search for photons, jets and met	13.3
atlas_1712_08119	39	electroweakinos search with soft leptons	36.1
atlas_1712_02332	24	squarks and gluinos, 0 lepton, 2-6 jets	36.1
atlas_1709_04183	14	stop pair production, 0 leptons	36.1
atlas_1802_03158	7	search for GMSB with photons	36.1
atlas_1708_07875	2	electroweakino search with taus and MET	36.1
atlas_1706_03731	19	same-sign or 3 leptons RPC and RPV SUSY	36.1
#atlas_conf_2019_018	2	Search for direct stau production in events with two hadronic tau leptons	139
atlas_1908_08215	16	charginos/sleptons, 2 leptons + MET	139
atlas_1909_08457	5	search for squarks and gluinos with same-sign leptons	139
atlas_conf_2019_020	2	Search for chargino-neutralino production with mass splittings near the electroweak scale	139
atlas_1803_02762	20	Search for electroweakino production in final states with two or three leptons	36.1
atlas_2101_01629	32	squarks/gluinos, 1 lepton, jets, MET	139
#atlas_conf_2020_048	26	Search for dark matter with monojets	139
atlas_2004_14060	14	stops, leptoquarks, 0 lepton	139
atlas_1908_03122	10	0 leptons, 3 or more b-jets, sbottoms	139
atlas_1911_12606	87	search for sleptons and electroweakinos with soft leptons	139
atlas_1807_07447	633	general search for new phenomena	3.2
atlas_2103_11684	2	Search for SUSY in events with four or more leptons (gravitino SR)	139
atlas_2004_10894	12	EWino search in Higgs (diphoton) and met	139
atlas_2106_09609	21	Search for RPV SUSY in final states with leptons and many jets	139
atlas_1911_06660	2	search for direct stau production	139
atlas_2010_14293	78	search for squarks and gluinos in MET_jet final states	139
atlas_2211_08028	22	search for gluinos decaying via 3rd gen; multi b-jets and MET	139
atlas_2106_01676	72	electroweakinos, 3 leptons, WZ, Wh, on+off-shell	139
atlas_2006_05880	28	Search for top squarks in events whith a Higgs or a Z boson	139
atlas_2111_08372	23	dark matter and invisible Higgs with Z boson	139
atlas_2202_07953	17	invisible Higgs decays in VBF	139
atlas_2209_13935	59	search for charginos and sleptons in 2-lepton final states	139
atlas_2102_10874	26	monojet	139
atlas_2411_02040	5	triple Higgs search	126

CheckMATE: CMS analyses

#Name	NSR	Description	Lumi
cms_pas_sus_15_011	47	CMS, 13 TeV, 2 leptons + jets + MET	2.2
cms_sus_16_039	158	electrowekinos in multilepton final state	35.9
cms_sus_16_025	14	electroweakino and stop compressed spectra	12.9
cms_sus_16_048	20	two soft opposite sign leptons	35.9
cms_sus_19_005	303	hadronic final states with MT2	137.0
cms_1908_04722	186	hadronic final states with HT, post-fit and simple fitting	137.0
cms_2107_13201	88	monojet with multibin	137.0
cms_2205_09597	40	search for electroweakinos in hadronic final states	137.0

The list much shorter compared to ATLAS…

- From start CheckMATE was based on collaboration with ATLAS so the ties are still stronger
- ATLAS is by default releasing reinterpretation material for all SUSY searches: cutflows, simplified analysis code, efficiencies etc., what makes recasting much easier (big effort from CMS to change that though)
- Many searches very similar (on the other hand combinations are tempting…)

Contents

Construction of likelihoods

- Probability model with analysis-specific and analysis-independent terms

$$f(\boldsymbol{n}, \boldsymbol{a} \mid \boldsymbol{\eta}, \boldsymbol{\chi}) = \underbrace{\prod_{c \in \text{channels}} \prod_{b \in \text{bins}_c} \text{Pois}\left(n_{cb} \mid \nu_{cb}\left(\boldsymbol{\eta}, \boldsymbol{\chi}\right)\right)}_{\substack{\text{Simultaneous measurement} \\ \text{of multiple channels}}} \underbrace{\prod_{\chi \in \boldsymbol{\chi}} c_{\chi}(a_{\chi} \mid \chi)}_{\substack{\text{constraint terms} \\ \text{for ``auxiliary measurements''}}}$$

- With expected yields modelled by

$$\nu_{cb}(\boldsymbol{\phi}) = \sum_{s \in \text{samples}} \nu_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s \in \text{samples}} \underbrace{\left(\prod_{\kappa \in \boldsymbol{\kappa}} \kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right)}_{\text{multiplicative modifiers}} \left(\nu_{scb}^{0}(\boldsymbol{\eta}, \boldsymbol{\chi}) + \underbrace{\sum_{\Delta \in \boldsymbol{\Delta}} \Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})}_{\text{additive modifiers}}\right)$$

- η and χ are unconstrained and constrained parameters
- We are after μ □ unconstrained **parameter of interest**

Can have hundreds of nuisance (auxiliary) parameters!

ATL-PHYS-PUB-2019-029

HistFactory: JSON specification

- Toy model: single channel with two bins
- Uncorrelated backgrounds with 10% and 20% uncertainty
- Single measurement specified as POI

Innocent start, the files can be tens of MB in size!


```json
{
    "channels": [
        { "name": "singlechannel",
          "samples": [
            { "name": "signal",
              "data": [5.0, 10.0],
              "modifiers": [ { "name": "mu", "type": "normfactor", "data": null} ]
            },
            { "name": "background",
              "data": [50.0, 60.0],
              "modifiers": [ {"name": "uncorr_bkguncrt", "type": "shapesys", "data": [5.0,12.0]} ]
            }
          ]
        }
    ],
    "observations": [
        {"name": "singlechannel", "data": [50, 60]}
    ],
    "measurements": [
        { "name": "Measurement", "config": {"poi": "mu", "parameters": []} }
    ]
}
```

Simplified likelihoods

- ATLAS ATL-PHYS-PUB-2021-038
  - background model is approximated with a single background sample, representing the total SM background rate
  - background rate obtained in the background-only fit in the full likelihood
  - the complete set of nuisance parameters in the original full likelihood is reduced to a single constrained parameter α, representing the post-fit uncertainty on the total SM background estimate in each bin
  - constrained by a Gaussian $G(a = 0|\alpha, \sigma = 1)$ and is correlated over all bins in each channel
  - aim: computation efficiency; but lower statistical precision

Simplified likelihoods

- CMS [CMS-NOTE-2017-001]
  - background contributions correlated between bins
  - probability to observe n_i events given by

$$P(n_i|\mu \cdot s_i + b_i) = \frac{(\mu \cdot s_i + b_i)^{n_i} e^{-(\mu \cdot s_i + b_i)}}{n_i!}$$

  - full likelihood given by multivariate Gaussian distribution

$$\mathcal{L}_S(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{(\mu \cdot s_i + b_i + \theta_i)^{n_i} e^{-(\mu \cdot s_i + b_i + \theta_i)}}{n_i!} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{V}^{-1} \boldsymbol{\theta}\right)$$

  - where θ are nuisance parameters to model background uncertainty
  - V is covariance matrix for the expected background contributions

$$V_{ij} = E[\theta_i \times \theta_j]$$

Implemented searches

Name	Description	#SR	N_{bin}	Full
atlas_1908_03122	Search for bottom squarks in final states with Higgs bosons, b-jets and E_T^{miss}	2	7	✓
atlas_1908_08215	Search for electroweak production of charginos and sleptons in final states with 2 leptons and E_T^{miss}	1	52	✓
atlas_1911_06660	Search for direct stau production in events with two hadronic taus	1	2	✓
atlas_1911_12606	Search for electroweak production of supersymmetric particles with compressed mass spectra	2	76	✓
atlas_2004_14060	Search for stops in hadronic final states with E_T^{miss}	3	14	✗
atlas_2006_05880	Search for top squarks in events with a Higgs or Z boson	3	23	✗
atlas_2010_14293	Search for squarks and gluinos in final states with jets and E_T^{miss}	3	60	✓
atlas_2101_01629	Search for squarks and gluinos in final states with one isolated lepton, jets, and E_T^{miss}	1	26	✓
atlas_2111_08372	Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates	1	22	✗
atlas_2102_10874	Monojet: search for new phenomena in events with an energetic jet and E_T^{miss}	1	13	✓
atlas_2411_02040	Search for triple Higgs production in $6b$ final state	3	60	✓

Name	Description	N_{bin}
cms_1908_04722	Search for supersymmetry in final states with jets and E_T^{miss}	174
cms_1909_03460	Search for supersymmetry with M_{T2} variable in final states with jets and E_T^{miss}	282
cms_2107_13021	Search for new particles in events with energetic jets and large E_T^{miss}	66
cms_2205_09597	Search for production of charginos and neutralinos in final states containing hadronic decays of WW, WZ, or WH and E_T^{miss}	35

Options available to the user

Spey package preferred for evaluation; arxiv:2307.06996

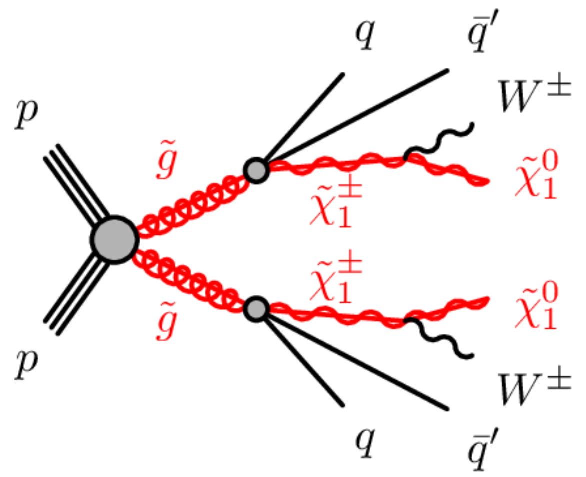
Parameter card		Terminal	X	Description and available choices
Multibin:	X	-mb X	none	No signal region combination is performed (default).
			select	Calculates user selected statistics.
			scan	Calculates observed CL_s; fast and reliable for quick assessment of exclusion.
			detailed	Calculates observed and expected upper limits and CL_s.
Expected:	False	-exp		Selects calculation of expected limits.
CLs:	False	-mbcls		Selects calculation of CL_s.
Uplim:	False	-uplim		Selects calculation of upper limits.
Statonly:	False	-so		Calculates statistical combinations without event-level analysis provided the analysis and evaluation steps were already completed.
			full	The SPEYinterface to the full likelihood model for ATLAS searches (default).
Model:	X	-mod X	simple	The simplified likelihood model for ATLAS searches.
			fullpyhf	The full likelihood model for ATLAS searches with PYHF interface.
Backend:	X		numpy	Different backends for calculations in the full model; NUMPY is always available.
			pytorch	PYTORCH backend (if available)
			tensorflow	TENSORFLOW backend (if available)
			jax	JAX backend (if available)

Table 3: Summary of options related to multibin signal regions.

Contents

Performance considerations

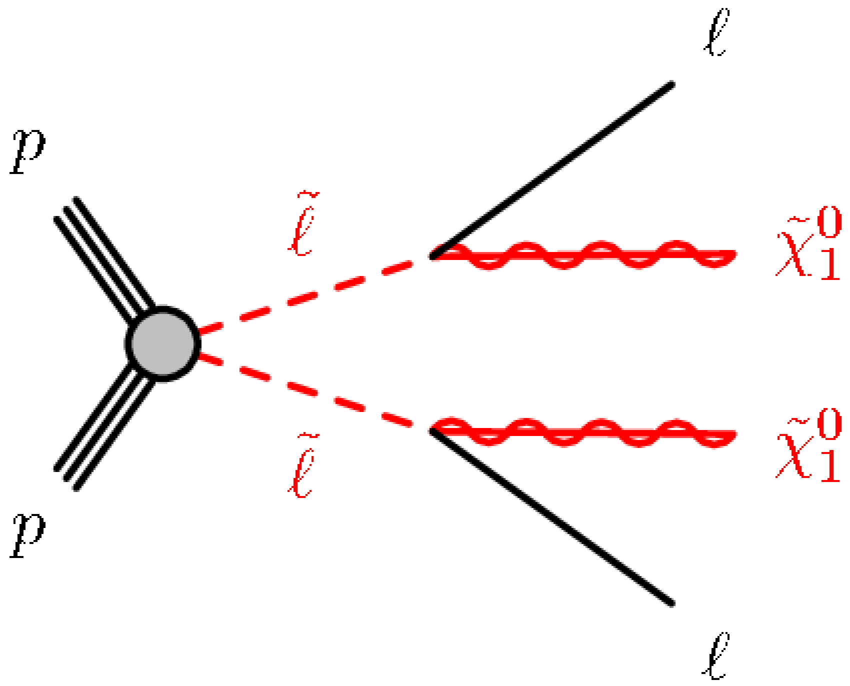
- Several backends supported for multicore calculations
- Significant improvement compared to single core numpy
- Further improvement with GPU support
- Below an example (time in seconds) for atlas_2101_01629 – a moderately complex statistical model (gluino pair production, final state qqqqWW+2LPSs)

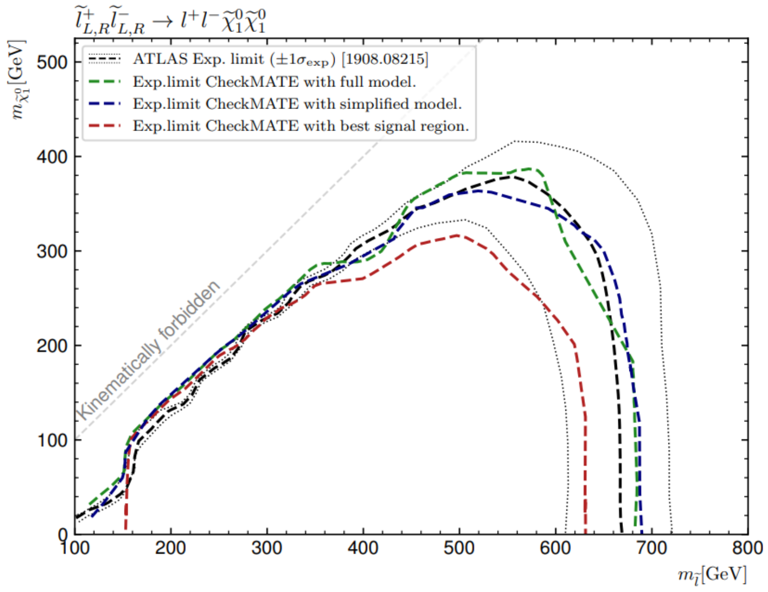


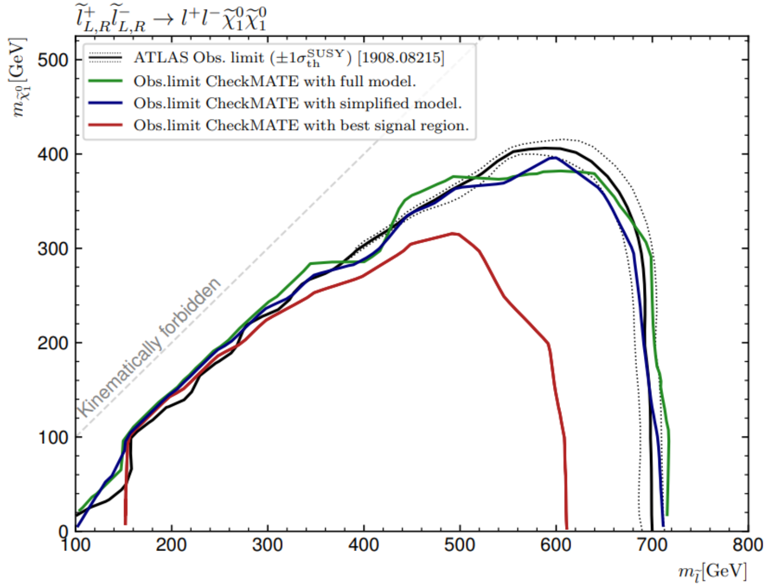
	upper limits observed and expected				CL_s
Backend	CPU A	CPU B	GPU B	CPU C	CPU A
NUMPY	11092	12621	N/A	8879	636
JAX	723	526	353	311	55
PYTORCH	583	805	N/A	404	57
TENSORFLOW	2426	2920	1906	1808	156

SUSY-2018-32 (atlas_1908_08215)

- Slepton/chargino pair production
- High-mass limit, direct decays

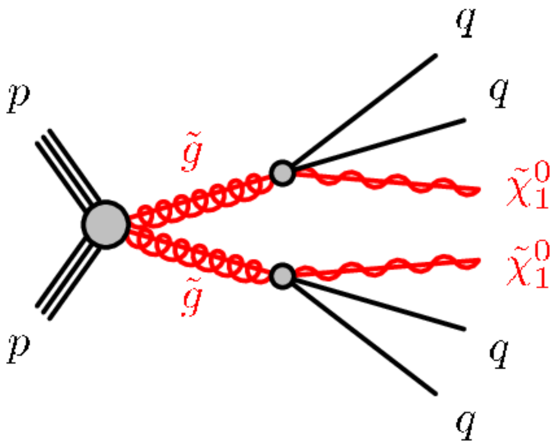


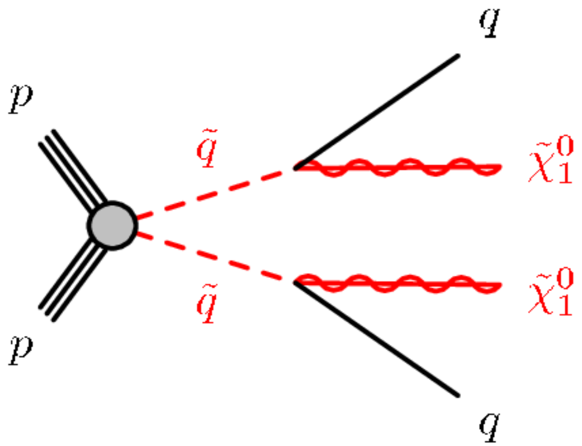


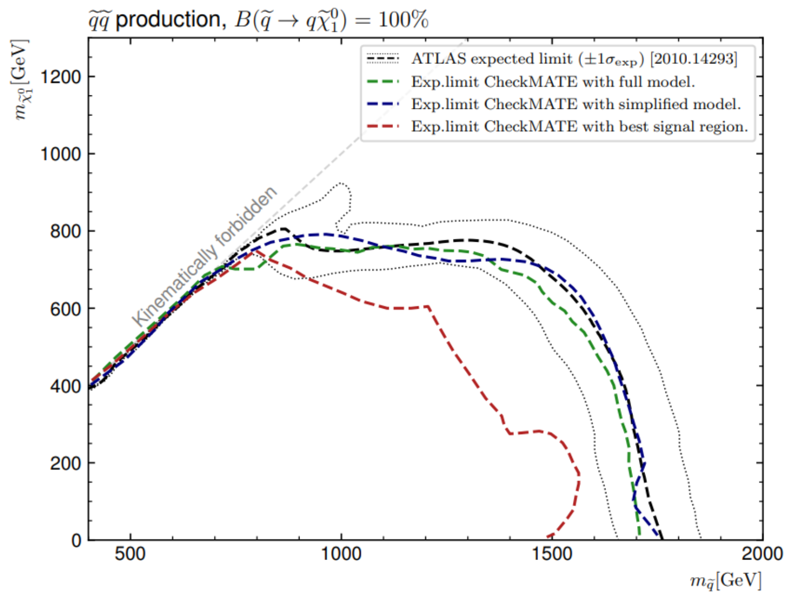


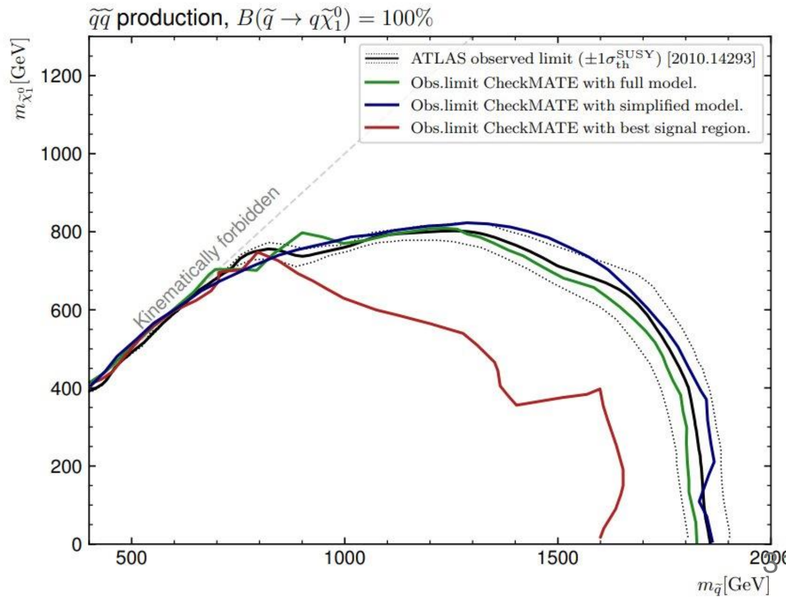
SUSY-2018-22 (atlas_2010_14293)

- Strong SUSY production, squarks and gluinos
- Direct and indirect decays
- Cut-and-count + BDT signal regions



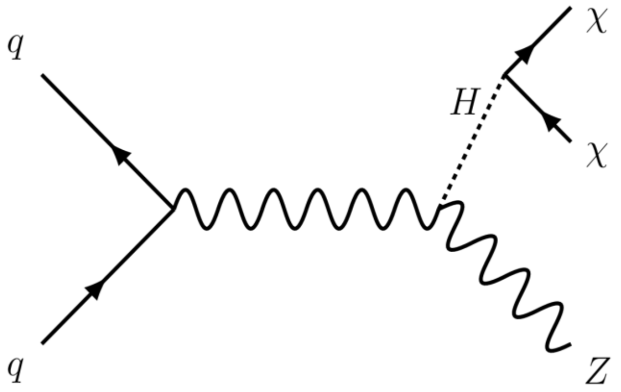


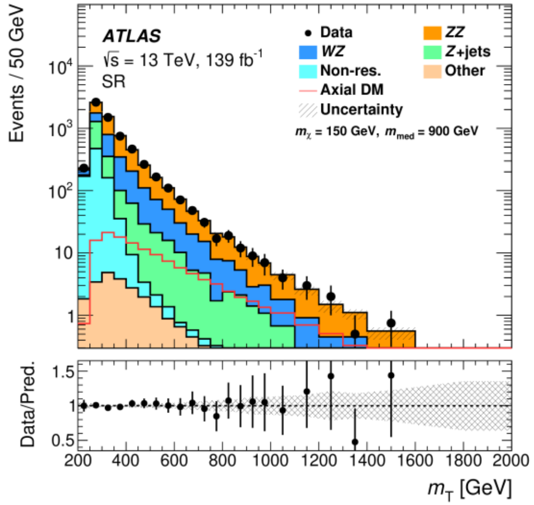


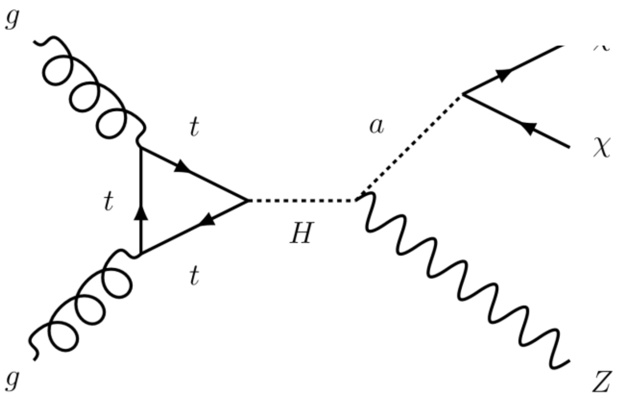


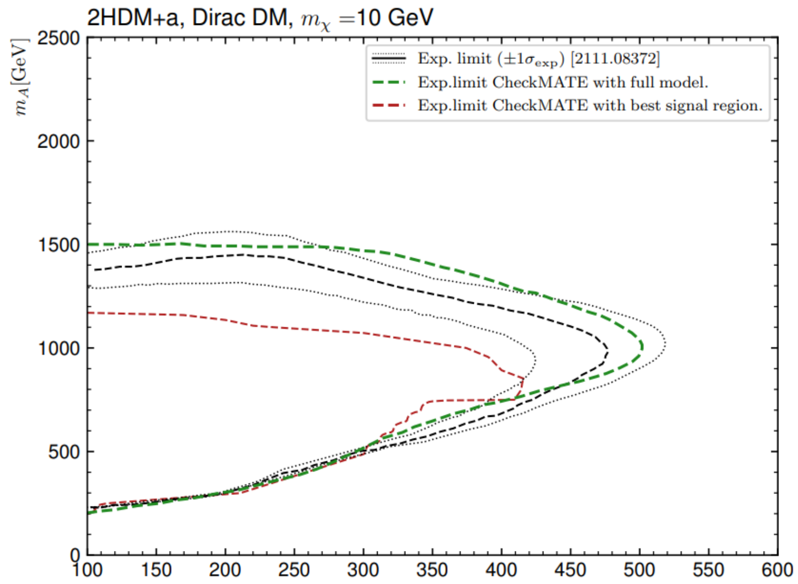
HIGG-2018-26 (atlas_2111_08372)

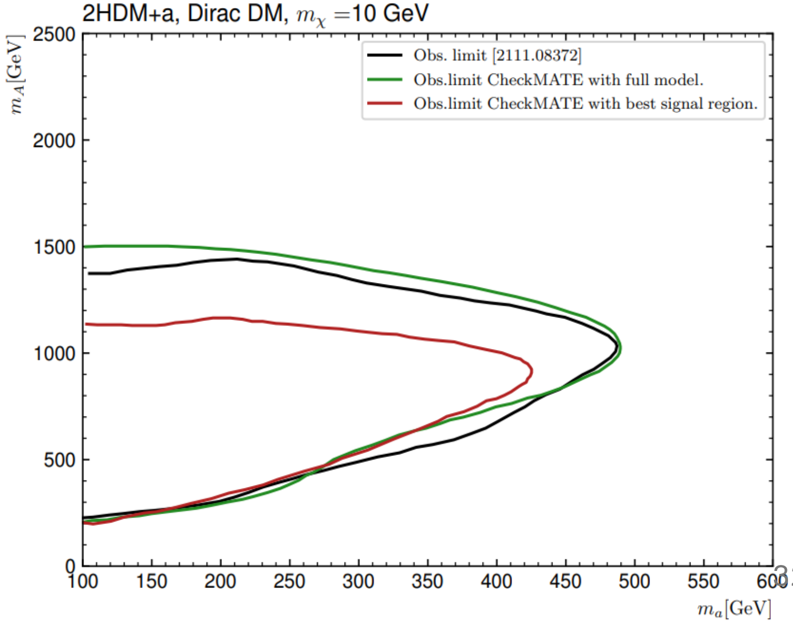
- Associated HZ production:
  - Invisible Higgs decays
  - Dark matter in 2HDMa





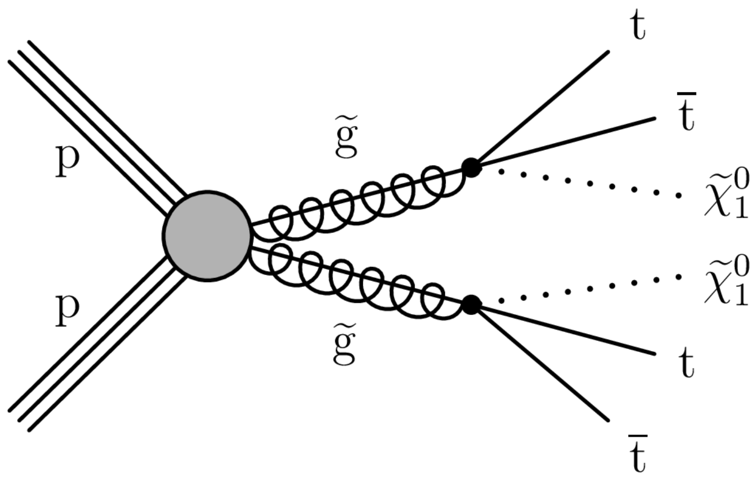


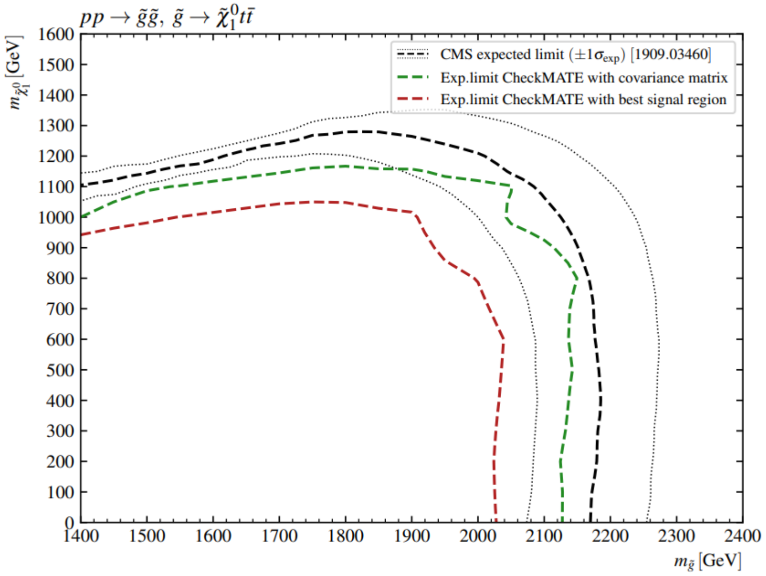


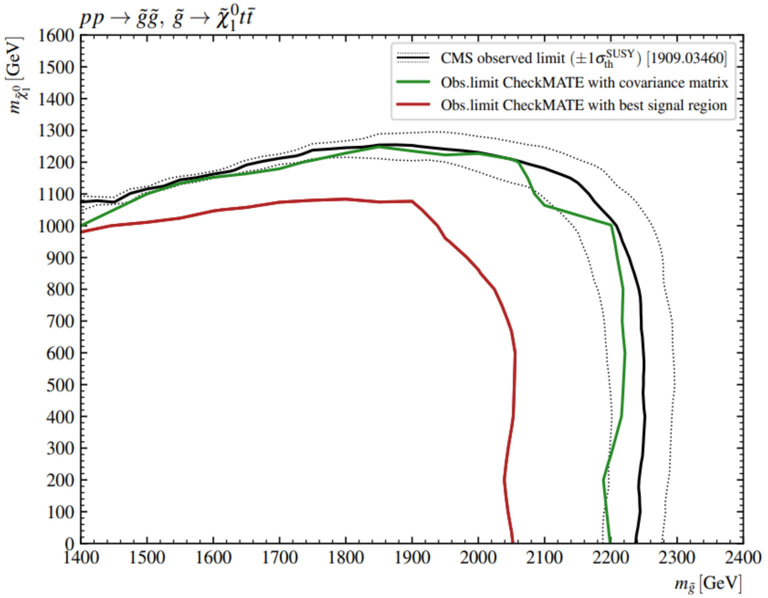


SUS-19-005 (cms_1909_03460)

- Hadronic BSM, 1-10 jets + MET
- Jets based M_{T2} variable
- Large correlation matrix: 282 signal bins

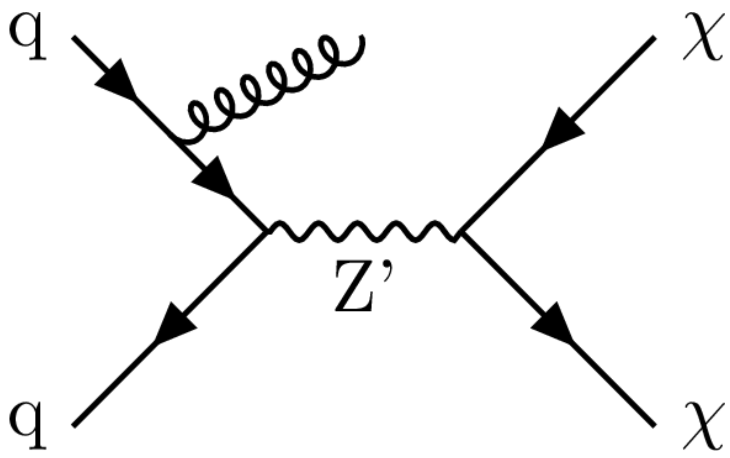


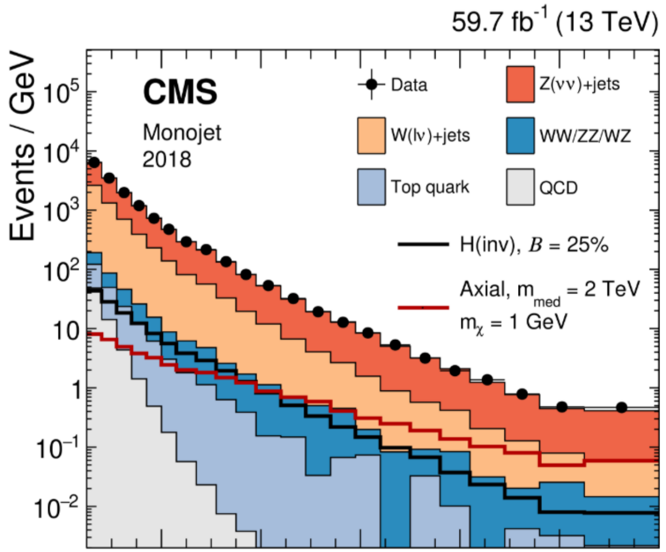


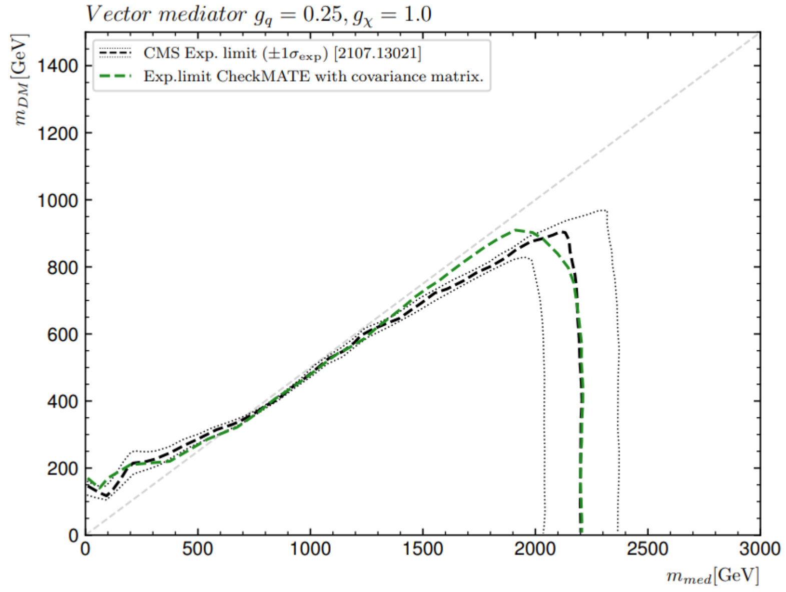


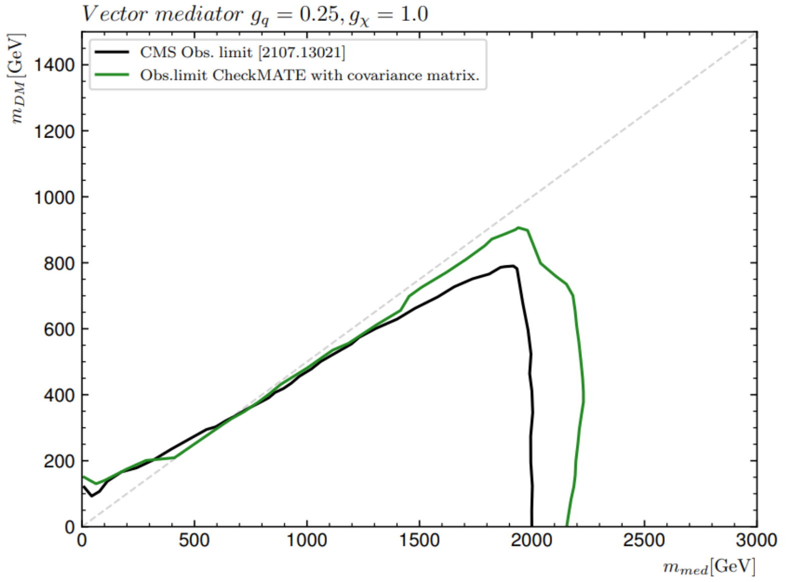
EXO-20-004 (cms_2107_13021)

- Exotics search:
  - Mono(>=1)-jet + MET
  - Dark matter, leptoquarks









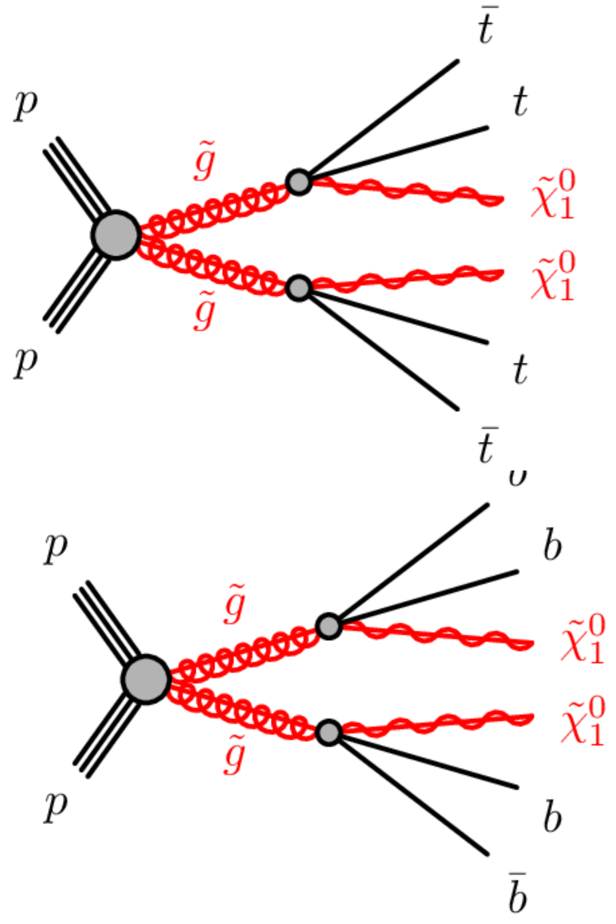
Contents

Available analyses with ML event selection

- ATLAS-SUSY-2018-30 b-jets + MET, ONNX (Rivet, CheckMATE)
- ATLAS-SUSY-2018-22 jets + MET, ROOT TMVA (CheckMATE)
- ATLAS-SUSY-2019-04 RPV jets + leptons, ONNX (CheckMATE, Rivet, MadAnalysis5, Gambit)
- ATLAS-SUSY-2019-02 leptons + MET, ROOT MVAUtils (CheckMATE)
- ATLAS-HIGP-2024-32 triple Higgs (6b), ONNX (Rivet, CheckMATE)

SUSY-2018-30 (atlas_2211_08028)

- Search for gluinos decaying to 3rd generation quarks
- Final state: at least 4 jets, at least 3 b-tagged jets, MET, 0-1 leptons (more allowed in NN analysis)
- 8 NN signal regions: 4 for gluino decaying to top pair and 4 for gluino decaying to bottom pair (still it is one net)
- The choice of the desired SR is via the last three inputs (I.e. decay type, and target masses)
- The NN has 87 input parameters: jet (small and large R) momenta, lepton momenta, MET and b-tag category (binary)
- The output gives separate background and signal probabilities



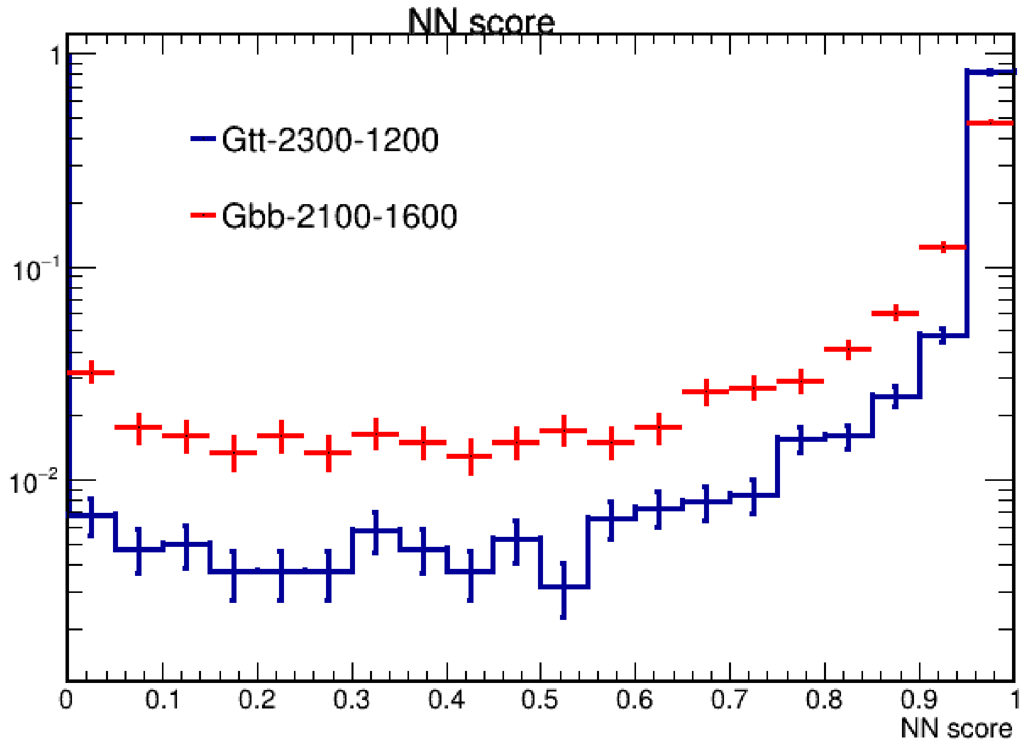
Validation

- Reasonable agreement across all channels
- The efficiency somewhat lower but similar and consistent effect also seen for cut-and-count analysis
- This is still a preprint so not too much material for comparison and testing was rather quick (nothing changed actually)

	ATLAS	CheckMATE
Gtt selection		
Common rquirem.	7.66	7.30
SR-Gtt-2100-1	2.63	1.94
SR-Gtt-1800-1	2.80	2.11
SR-Gtt-2300-1200	2.95	2.62
SR-Gtt-1900-1400	0.19	0.27
Gbb selection		
Common rquirem.	80	65
SR-Gbb-2800-1400	22	14
SR-Gbb-2300-1000	21	14
SR-Gbb-2100-1600	6.20	6.80
SR-Gbb-2000-1800	0.19	0.58

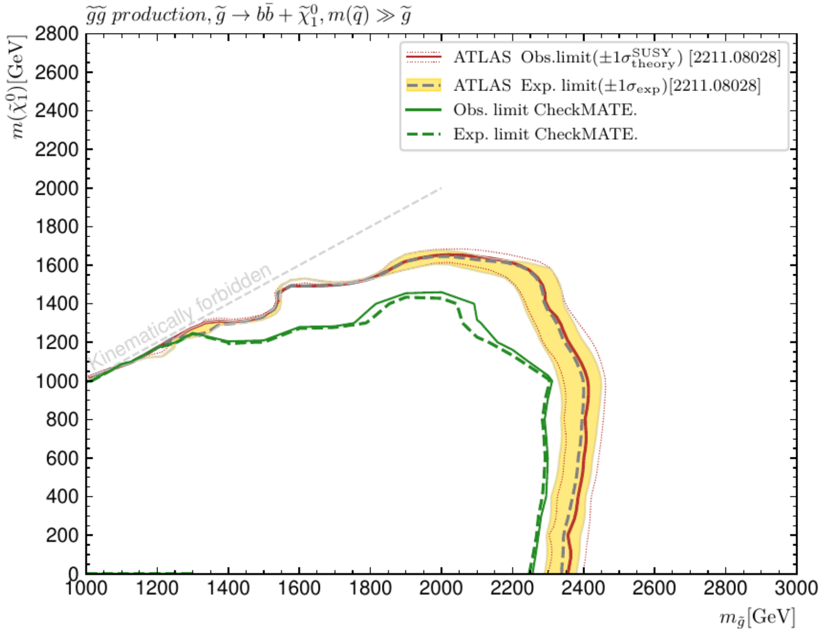
BDT Validation

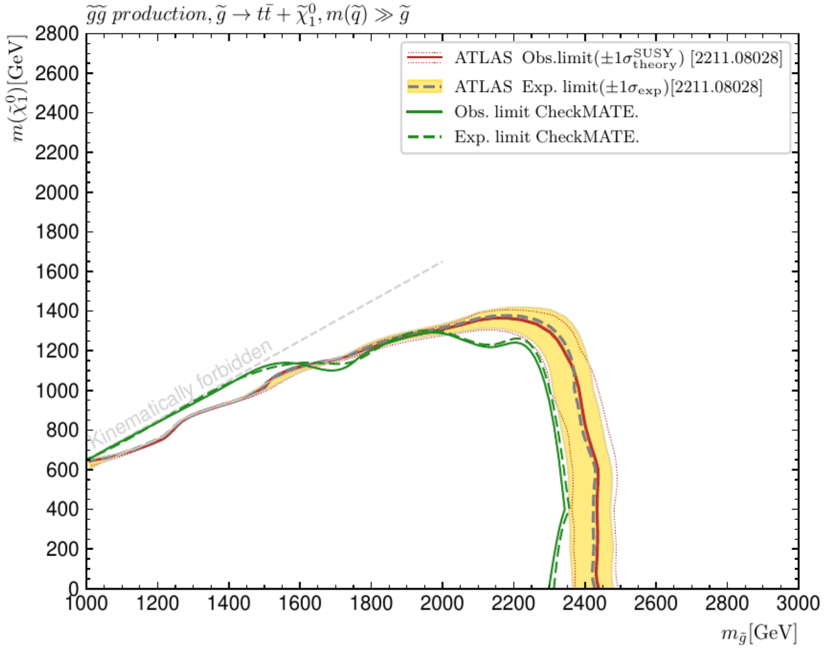
- True signal is Gtt m_{gluino} = 2400, m_{neut} = 1000
- Target signal for Gtt-2300-1200
- Less compatible with Gbb as it should be
- The last bin pronounced for Gbb but the cutoff is at 0.9993 so the acceptance is actually very low
- No figures from ATLAS for comparison



Validation

- Reasonable agreement (within 30%) for b-squarks
- Disagreement mainly due to b-tagging efficiency (not specifically tuned)
- Better yet for t-squarks (note a complicated final state)

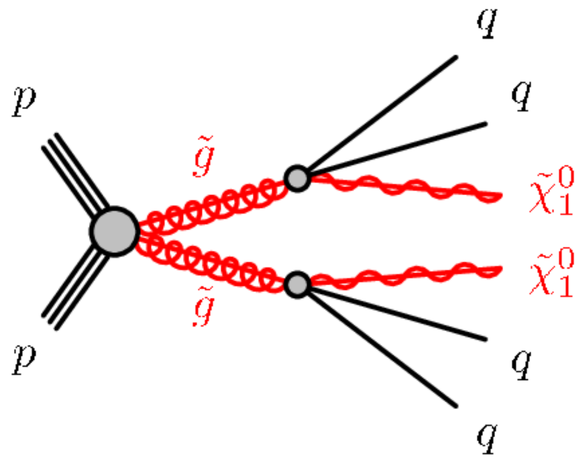


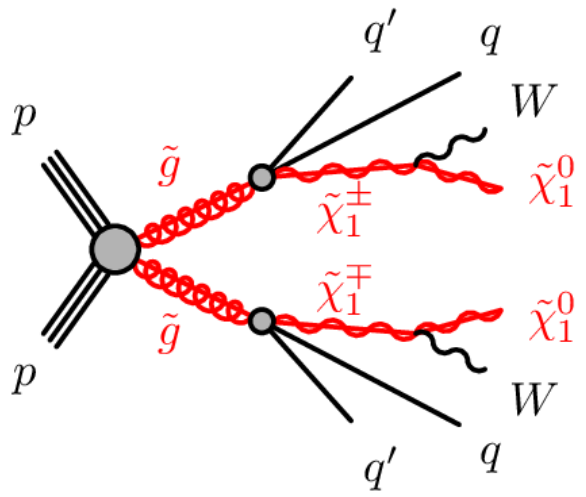


Contents

SUSY-2018-22 (atlas_2010_14293)

- Search for squarks and gluinos
- Final state: 2-6 jets + MET
- Principal variables: m_{eff} , leading jet pT and MET
- Multi-bin signal regions
- 8 boosted decision tree SRs, targeting gluino production with
  direct and indirect decays
- BDT weights released as .xml files for using with ROOT Toolkit for Multivariate Data Analysis -- TMVA





Validation

- Generally excellent agreement
- Here example for model-independent discovery channels
- See Iñaki's talk for details of multi-bin fits and validation
- Full note:

checkmate.hepforge.org

	Selection	$m_{\tilde{q}} = 1200$ GeV $m_{\tilde{\chi}_1^0} = 600$ GeV		$m_{\tilde{q}} = 1400$ GeV $m_{\tilde{\chi}_1^0} = 600$ GeV		$m_{\tilde{q}} = 1600$ GeV $m_{\tilde{\chi}_1^0} = 400$ GeV	
		ATLAS	CM	ATLAS	CM	ATLAS	CM
Generated MC events		10000	10000	6000	10000	6000	10000
Common Requirements	Preselection, $E_T^{miss} > 300$ GeV, $p_T(j_1) > 200$ GeV, $m_{eff} > 800$ GeV	1763	1780	541	546	174	176
	jet multiplicity ≥ 2	1763	1780	541	546	174	176
	Cleaning cuts	1746	–	535	–	173	–
SR-2j-1600	$\Delta\phi(j_{1,2,(3)}, E_T^{miss}) > 0.8$	1433	1434	431	433	136	139
	$\Delta\phi(j_{i>3}, E_T^{miss}) > 0.4$	1377	1353	411	410	129	130
	$p_T(j_2) > 250$ GeV	853	850	311	310	111	112
	$\lvert\eta(j_{1,2})\rvert < 2.0$	836	832	306	305	109	110
	$E_T^{miss}/\sqrt{H_T} > 16$ GeV$^{1/2}$	568	554	228	227	86.4	87.3
	m_{eff}(incl.) > 1600 GeV	366	362	202	195	83.5	84.2
SR-2j-2200	$\Delta\phi(j_{1,2,(3)}, E_T^{miss}) > 0.4$	1603	1619	483	492	156	158
	$\Delta\phi(j_{i>3}, E_T^{miss}) > 0.2$	1567	1566	470	476	151	153
	$p_T(j_1) > 600$ GeV	509	514	269	259	120	121
	$E_T^{miss}/\sqrt{H_T} > 16$ GeV$^{1/2}$	337	339	201	188	94.6	95.7
	m_{eff}(incl.) > 2200 GeV	101	96	108	101	76.1	76.4
SR-2j-2800	$\Delta\phi(j_{1,2,(3)}, E_T^{miss}) > 0.8$	1433	1434	431	433	136	138
	$\Delta\phi(j_{i>3}, E_T^{miss}) > 0.4$	1377	1352	411	410	129	130
	$p_T(j_2) > 250$ GeV	853	850	311	311	111	112
	$\lvert\eta(j_{1,2})\rvert < 1.2$	655	653	235	239	82.3	84.3
	$E_T^{miss}/\sqrt{H_T} > 16$ GeV$^{1/2}$	439	433	173	178	64.6	66.4
	m_{eff}(incl.) > 2800 GeV	15.6	10.5	18.8	15.1	29.1	27.0

BDT input

Input variables	GGd1	GGd2	GGd3	GGd4
E_T^{miss} [GeV]	•	–	•	
$p_T(j)$ [GeV]	$p_T(j_1), p_T(j_2), p_T(j_3), p_T(j_4)$			
$\eta(j)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$			
Aplanarity	•			–
m_{eff} [GeV]	•			
Total number of input variables	11	10	11	10

Input variables	GGo1	GGo2	GGo3	GGo4
E_T^{miss} [GeV]	–			•
$p_T(j)$ [GeV]	$p_T(j_1), p_T(j_2), p_T(j_3), p_T(j_4), p_T(j_5)$	$p_T(j_1), p_T(j_2), p_T(j_3), p_T(j_4)$	$p_T(j_1), p_T(j_2), p_T(j_3), p_T(j_5), p_T(j_6)$	$p_T(j_1), p_T(j_2), p_T(j_3), p_T(j_4)$
$\eta(j)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4), \eta(j_5)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_5), \eta(j_6)$	$\eta(j_1), \eta(j_2), \eta(j_3), \eta(j_4)$
Aplanarity	•			–
m_{eff} [GeV]	•			
Total number of input variables	12	10	12	10

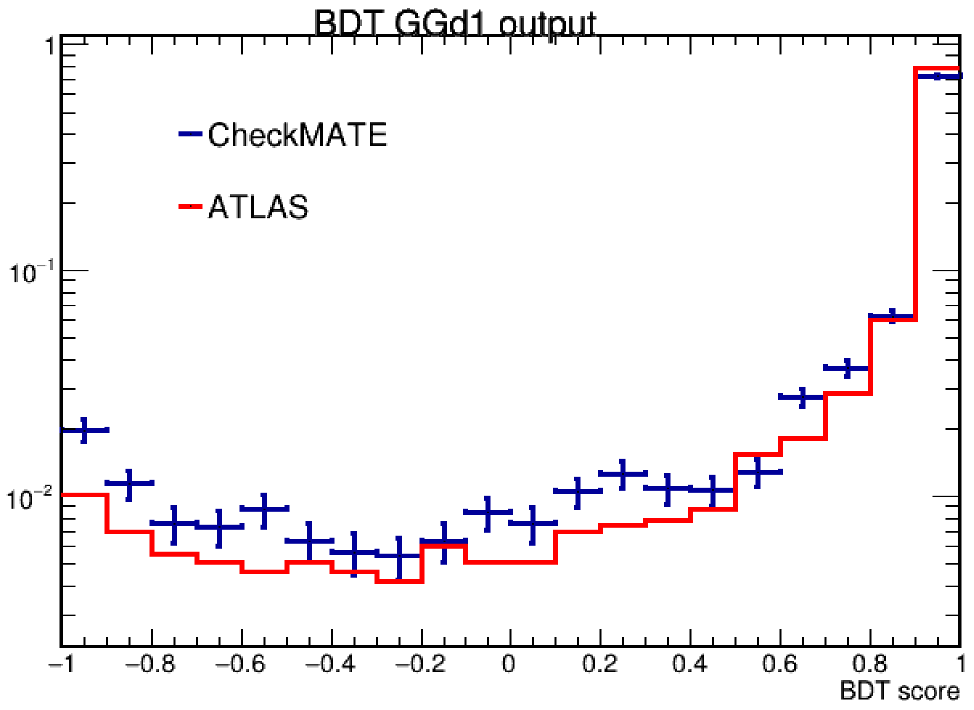
BDT validation

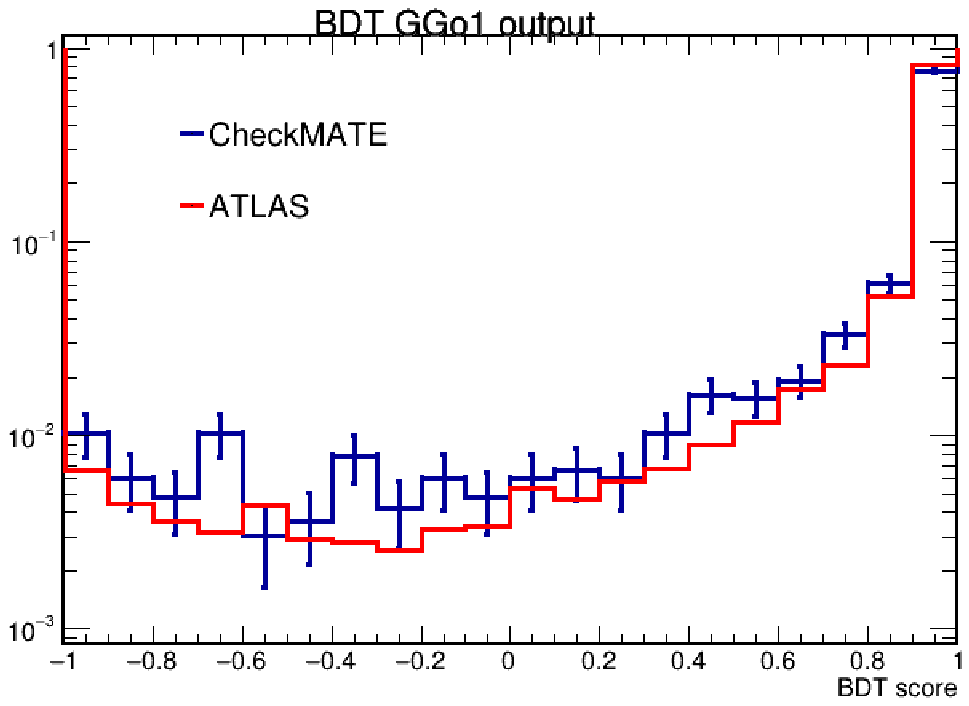
- Each SR targets direct gluino decays for specific range in $\Delta m = m_g - m_{neut}$
- GGd1: Δm = 1600-1900 GeV
- GGd2: Δm = 1000-1400 GeV
- GGd3: Δm = 600-1000 GeV
- GGd4: Δm = 200-600 GeV
- Overall, very good agreement
- No specific exclusion contour to compare with

$m_{\tilde{g}}$	2200		2200		1800		1400	
$m_{\tilde{\chi}_1^0}$	500		1000		1000		1000	
	A	C	A	C	A	C	A	C
GGd1	14.1	12.5	7.04	5.5	5.5	4.2	3.0	3.0
GGd2	14.3	13.4	11.4	10.1	19.4	14.3	8.8	9.9
GGd3	14.4	14.1	14.4	13.8	71.7	62.0	49.1	43.7
GGd3	2.9	3.4	6.0	6.1	60.5	54.0	89.6	85.8

A = ATLAS; C = CheckMATE

BDT output comparison – preferred for validation



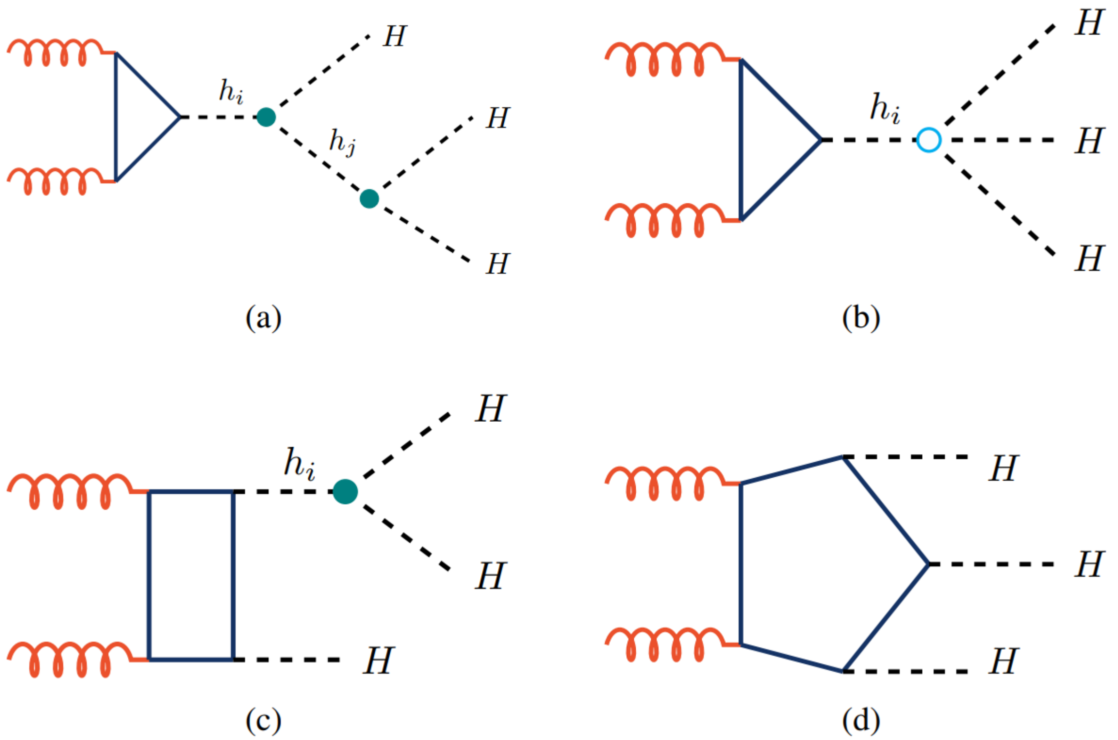


- Both signal regions show good agreement
- GGd1 = direct decay; GGo1 = one step decay

Contents

ATLAS HHH->6b

Key features:
NN classifiers motivated by 3 scenarios:
SM – non-resonant
TRSM – resonant and non-resonant
simplified – heavy resonant
statistical evaluation: fit of NN output histograms in SR and CR



Independent measurement of quartic coupling!

ATLAS provided:

- the statistical model in HS3 format (first time in BSM)
- the NN in ONNX format;
- SimpleAnalysis code
- many validation histograms (including for features)
  https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGP-2024-32/

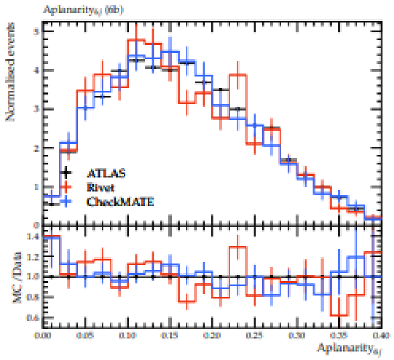
Implementation

- Parallel implementation in CheckMATE and Rivet: arXiv:2606.01966 (with Tomek Procter and Andrzej Siódmok)
- We provide additional data in the Zenodo repository
- This provided back-to-back debugging and helped to identify problems
- Largely aided by the SimpleAnalysis
- The input features very complicated
- In some cases the text (paper) and SimpleAnalysis were not enough for accurate reproduction! Many thanks to the analysis team for being extremely helpful!
- In reality, we were in contact with the analysis team before publication to provide our wishlist for reinterpretation material

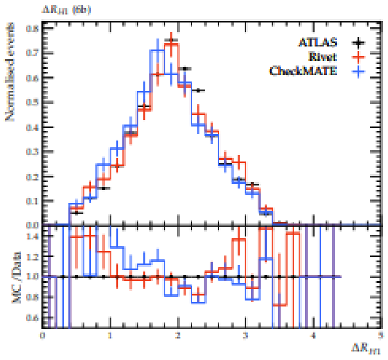
Variable	Definition	nonres	res	heavyres
m_H-radius	Distance between the event (m_{H1}, m_{H2}, m_{H3}) vector and expected vector for signal events, $(120, 115, 110)$ GeV.	✓		✓
m_{H1}	Reconstructed mass of the highest p_T Higgs boson candidate.	✓		✓
$\mathrm{RMS}(m_{jj})$	Root-mean-squared (RMS) of the invariant mass of all 15 possible jet pairs.	✓		✓
$\mathrm{RMS}(\Delta R_{jj})$	RMS of the angular separation between all 15 possible jet pairs.	✓	✓	✓
$\mathrm{RMS}(\eta)$	RMS of the pseudo-rapidity of the Higgs boson candidates.	✓		✓
Skewness ΔA_{jj}	Skewness of $\cosh(\Delta\eta_{ik}) - \cos(\Delta\phi_{ik})$, where i, k are all 15 possible jet pairs.		✓	
H_T^{6j}	Scalar sum of the p_T of the 6 jets selected to reconstruct the 3 Higgs boson candidates.		✓	
$\cos\theta$	In the (m_{H1}, m_{H2}, m_{H3}) coordinate system, θ is the angle between the vector from the origin to the event's reconstructed mass of the Higgs boson candidates, and the vector from the origin to $(120, 115, 110)$ GeV.		✓	
$\mathrm{Aplanarity}_{6j}$	The fraction of p_T from the 6 jets selected lying outside the plane formed by the 2 highest p_T jets .	✓	✓	✓
$\mathrm{Sphericity}_{6j}$	Isotropy of the momenta of the 6 jets selected to reconstruct the 3 Higgs boson candidates .		✓	
Transverse $\mathrm{Sphericity}_{6j}$	Isotropy of the p_T of the 6 jets used for Higgs reconstruction, within the $x - y$ plane .	✓		
Sphericity	Isotropy of the momenta of all jets in the event .			✓
$\eta - m_{HHH}$ fraction	$\frac{\sum_{i,k} 2p_T^i * p_T^k * (\cosh(\Delta\eta(ik)) - 1)}{m_{HHH}^2}$ where i, k are all 15 possible jet pairs, and m_{HHH} is the reconstructed tri-Higgs invariant mass.		✓	
ΔR_{H1}	Angular separation between the jets paired to form the highest p_T Higgs boson candidate.	✓	✓	✓
ΔR_{H2}	Angular separation between the jets paired to form the second-highest p_T Higgs boson candidate.	✓	✓	✓
ΔR_{H3}	Angular separation between the jets paired to form the lowest p_T Higgs boson candidate.	✓	✓	✓

Validation

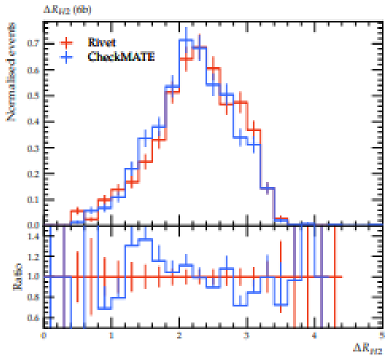
- Comparison of features for resonant NN
- Overall good agreement of ATLAS and both recasting codes
- Also, the NN output histogram is well reproduced



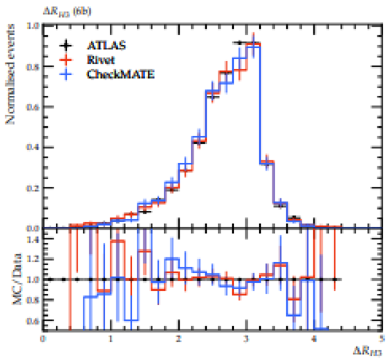
(a) Aplanarity (6 jets)



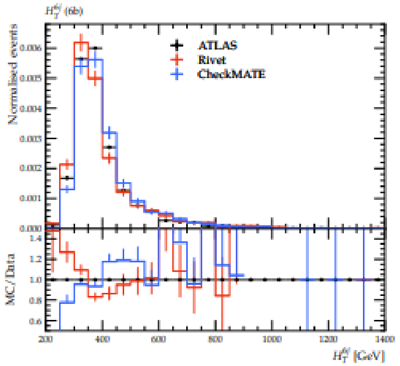
(b) ΔR_{H1}



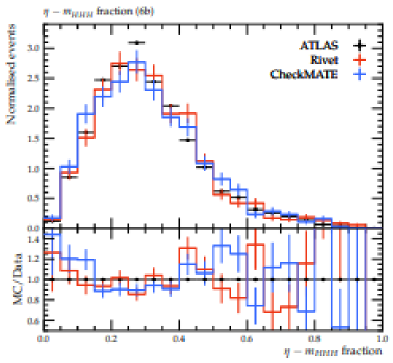
(c) ΔR_{H2}



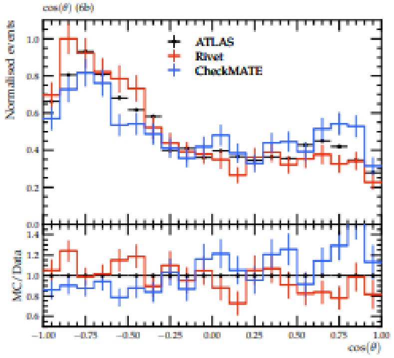
(d) ΔR_{H3}



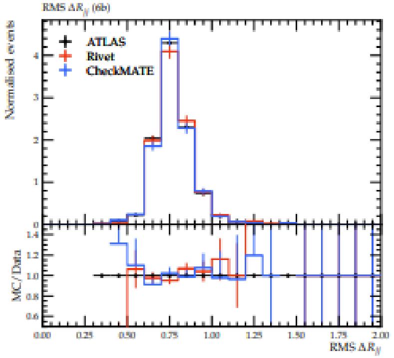
(e) H_T^{6j}



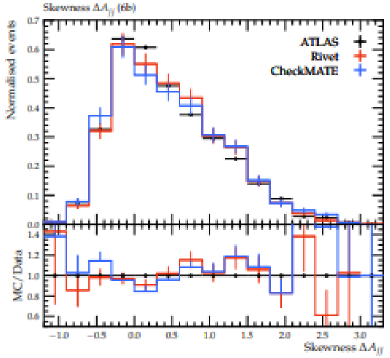
(f) $\eta - m_{HHH}$ fraction



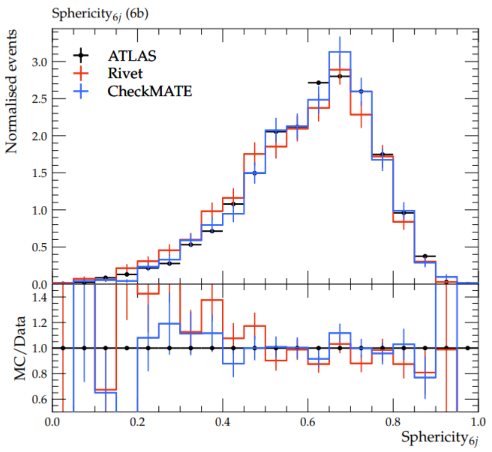
(g) $\cos(\theta)$



(h) RMS (ΔR_{jj})

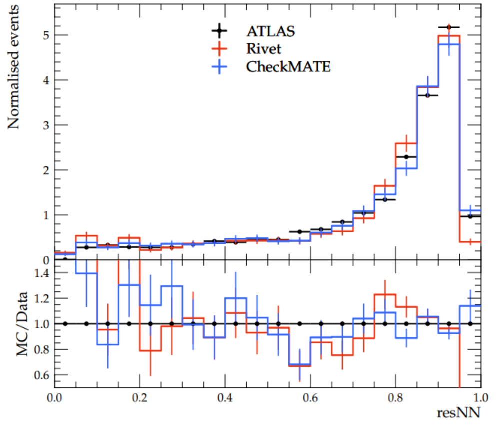


(i) Skewness (ΔA_{jj})



(j) Sphericity (6 jets)

final variable

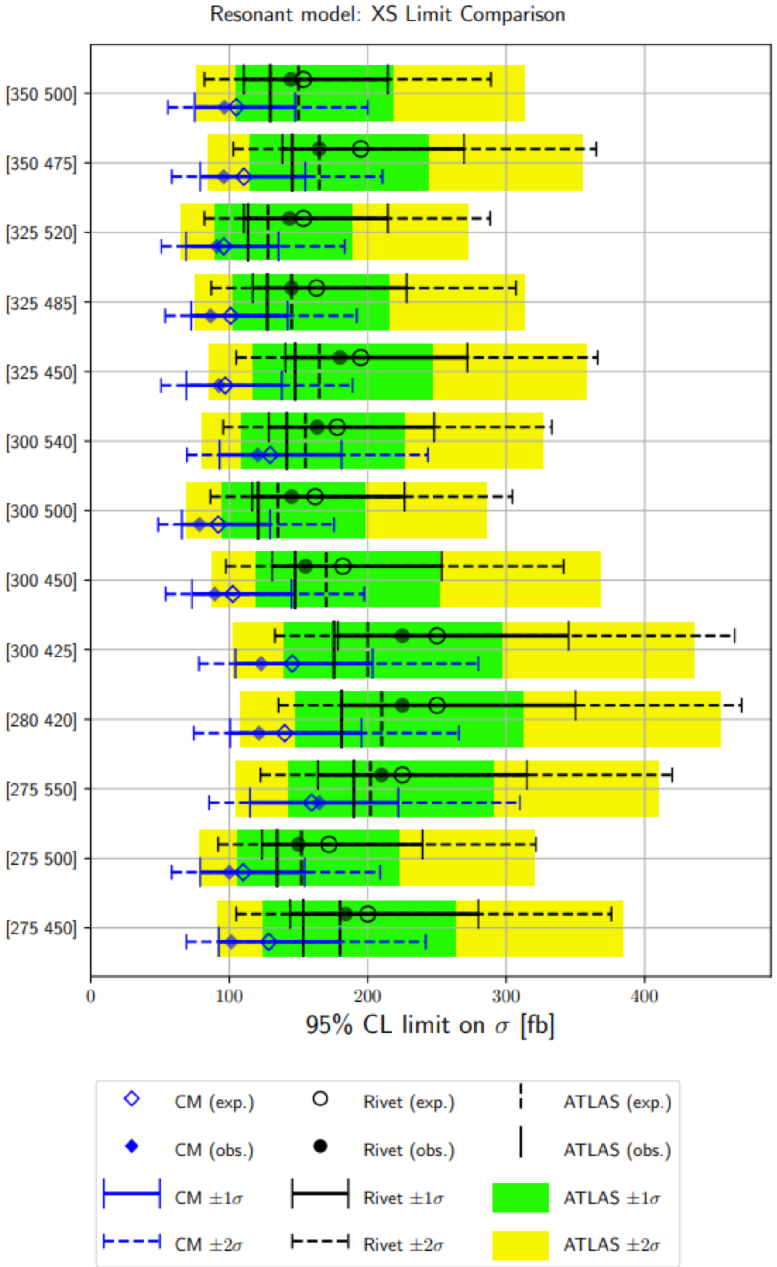


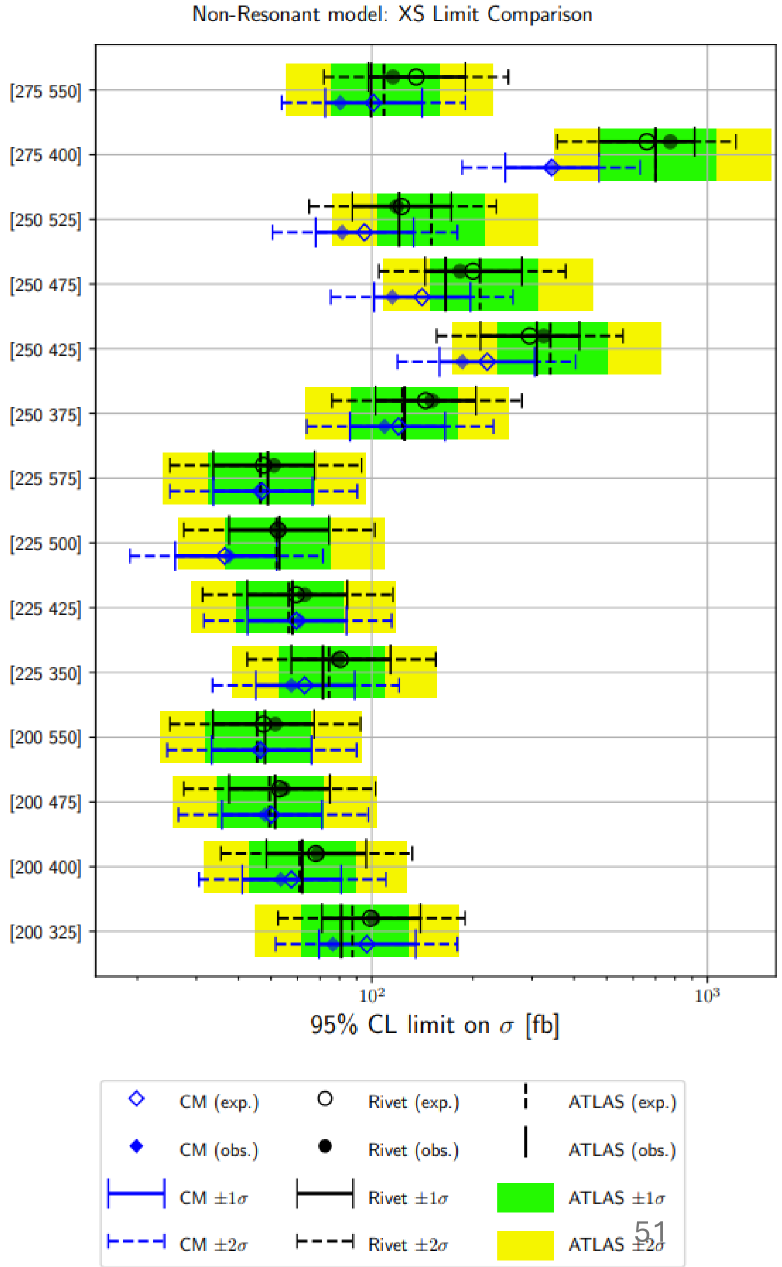
(k) Resonant NN score

arXiv:2606.01966

Validation

- Good agreement in exclusion limits
- Largest difference within 2-sigma uncertainty





arXiv:2606.01966

Moving on to HL-LHC

- Statistical model easily scalable to arbitrary luminosity
- The NN performance slightly degraded (expected)
- ATLAS made SM projection we compare to validate our rescaled statistical model

Scenario	μ (ATLAS)	μ (RIVET)	μ (CHECKMATE)
Baseline	195	185	201
Reduced bkg. uncert.	124	129	120
No systematics	99	83	83
Run 2 analysis	750	780	745

Two Real Singlet Model TRSM

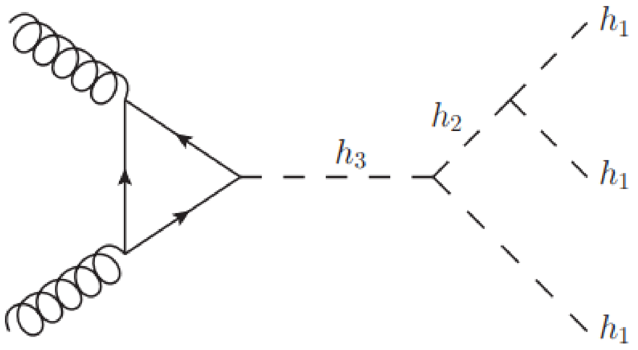
- The scalar potential with two additional scalars:

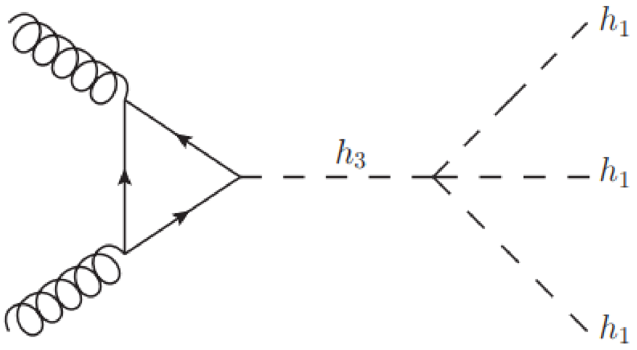
$$V = \mu_\Phi^2 \Phi^\dagger \Phi + \lambda_\Phi (\Phi^\dagger \Phi)^2 + \mu_S^2 S^2 + \lambda_S S^4 + \mu_X^2 X^2 + \lambda_X X^4 + \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2 + \lambda_{SX} S^2 X^2,$$

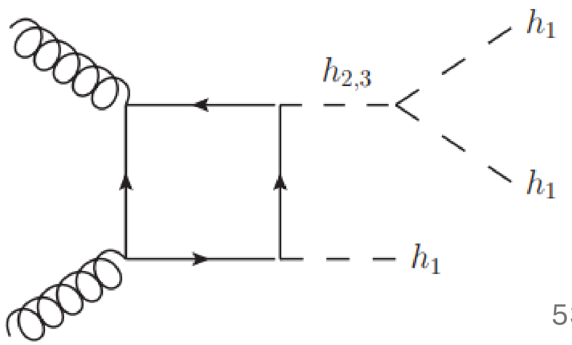
- Mixing of singlets with doublet

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = R \begin{pmatrix} \phi_h \\ \phi_S \\ \phi_X \end{pmatrix}$$

- Production at hadron colliders:

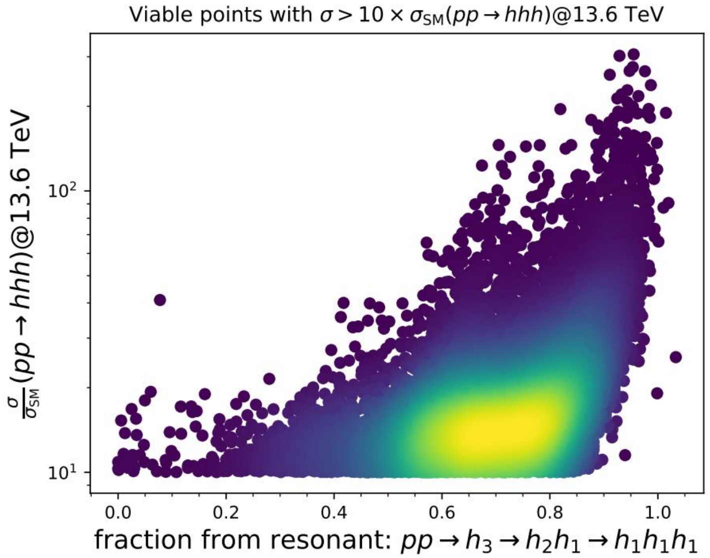


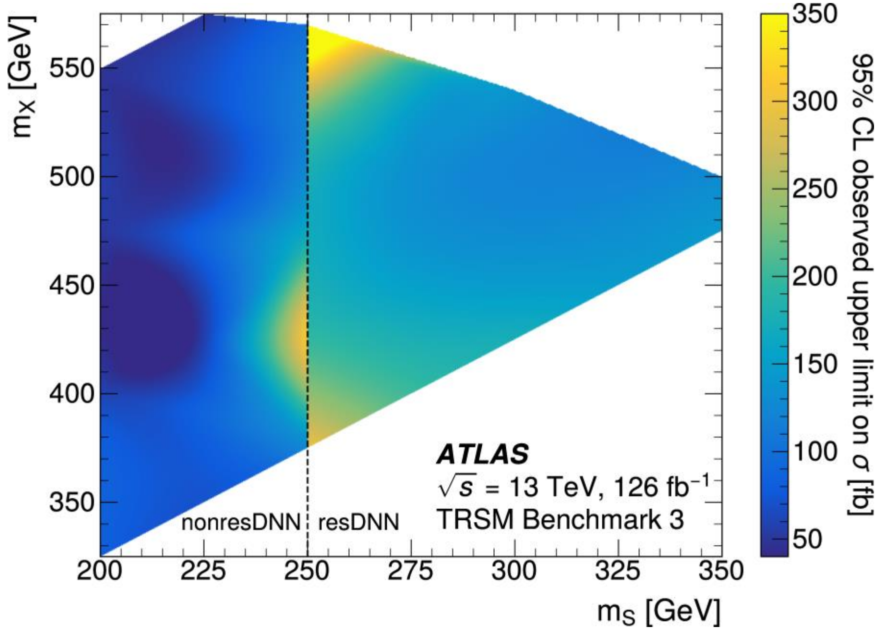




Benchmarks for HL-LHC recasting

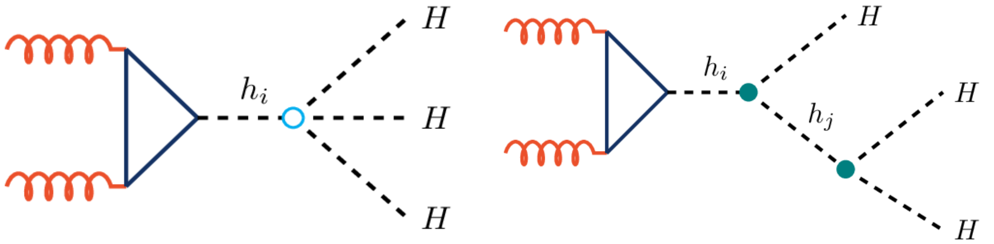
- ATLAS used TRSM benchmarks for training and validation: BP3 from T. Robens ea. arXiv:1908.08554
- New set of benchmarks proposed in O. Karkout ea. arXiv:2404.12425
- Improved perturbative behavior
- Enhanced xsec

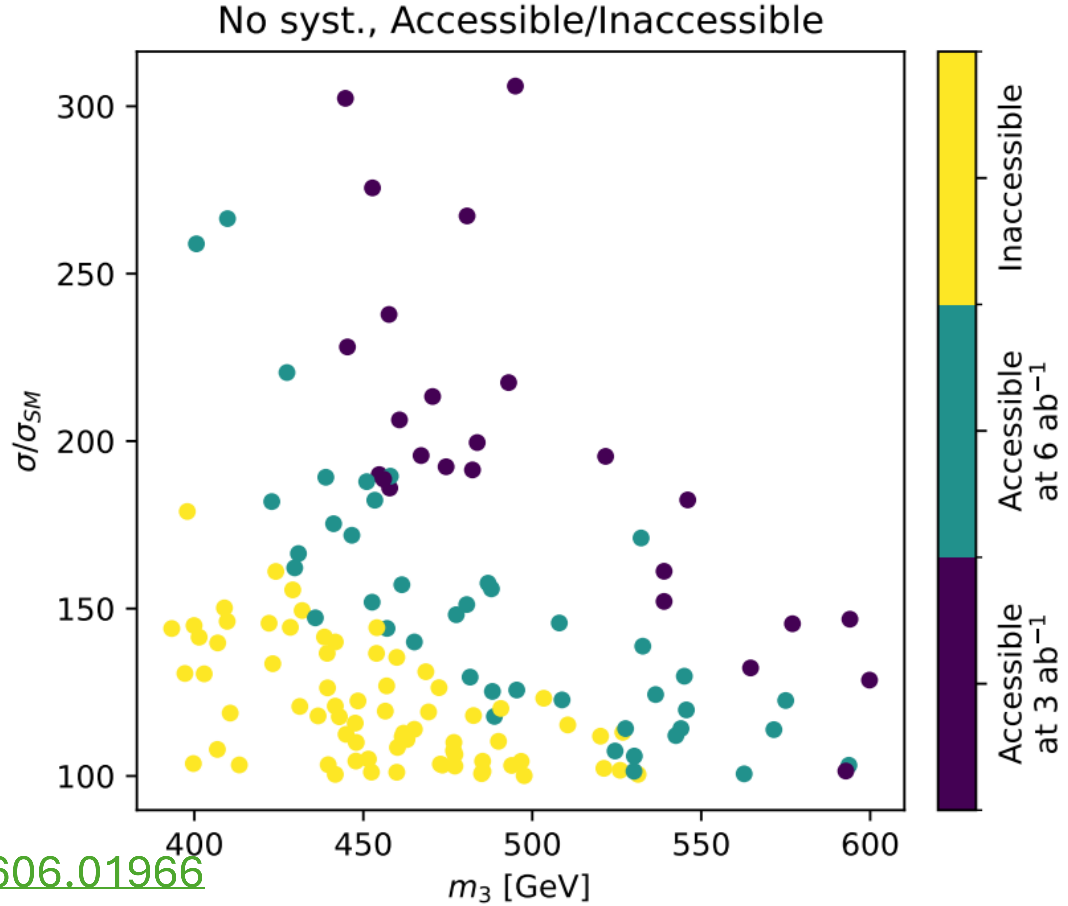


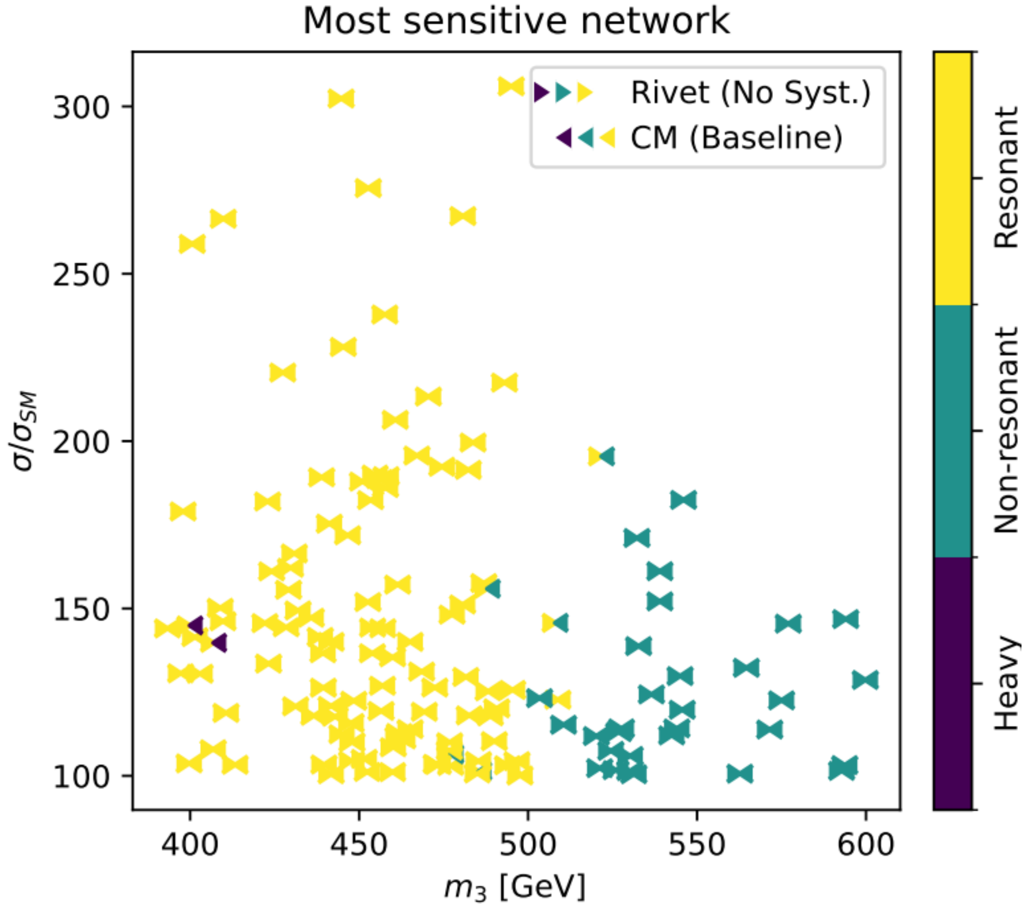


see also Osama's talk on Thursday

HL projection for two real singlets SM (TRSM)







arXiv:2606.01966

Surprisingly: significant impact from non-resonant NN, which was trained on SM only

Contents

1. Introduction
    • Reinterpretation/recasting
    • CheckMATE
2. Statistical models
    • General remarks
    • Examples
3. Machine learning in CheckMATE
    • SUSY-2018-30
    • SUSY-2018-22
4. Case study: triple Higgs search
5. Conclusions

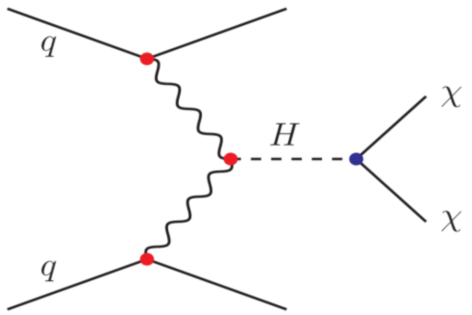
Conclusion

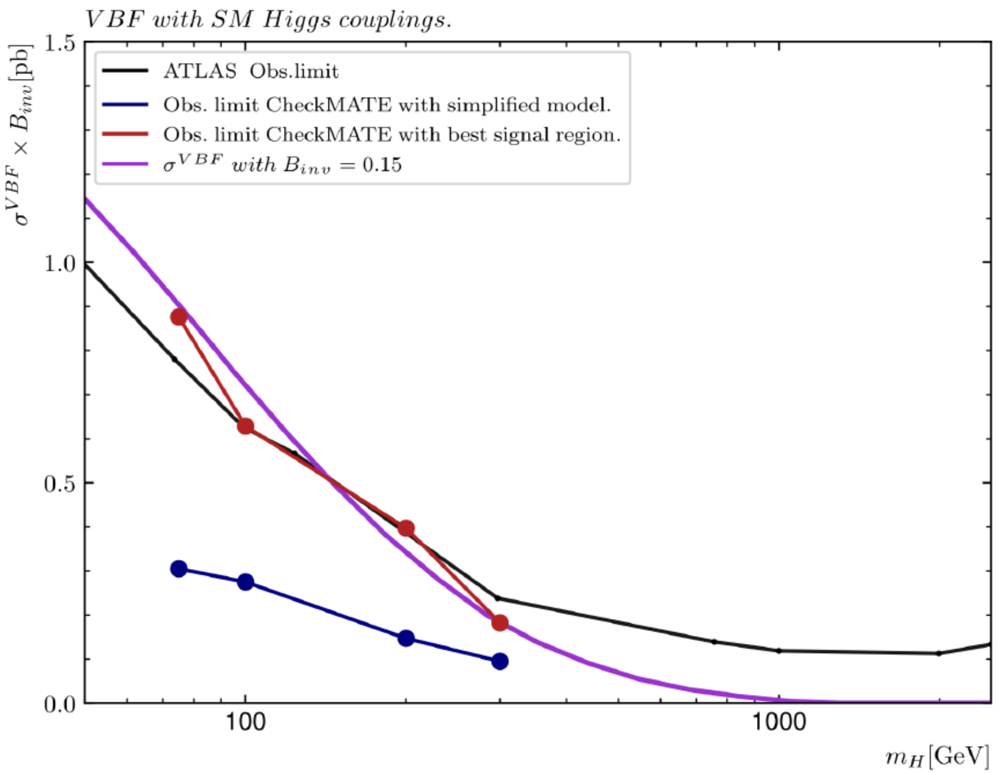
- Statistical models are standard feature now in CheckMATE
- Generally handled using Spey/pyhf or HS3/ROOT
- Several machine-learning based searches included
- Collaborations are not too enthusiastic about releasing ML models (workload is significant)
- Triple Higgs search is an excellent example combining both: advanced statistical model and machine learning
- Open to suggestions which new searches are needed

Backup

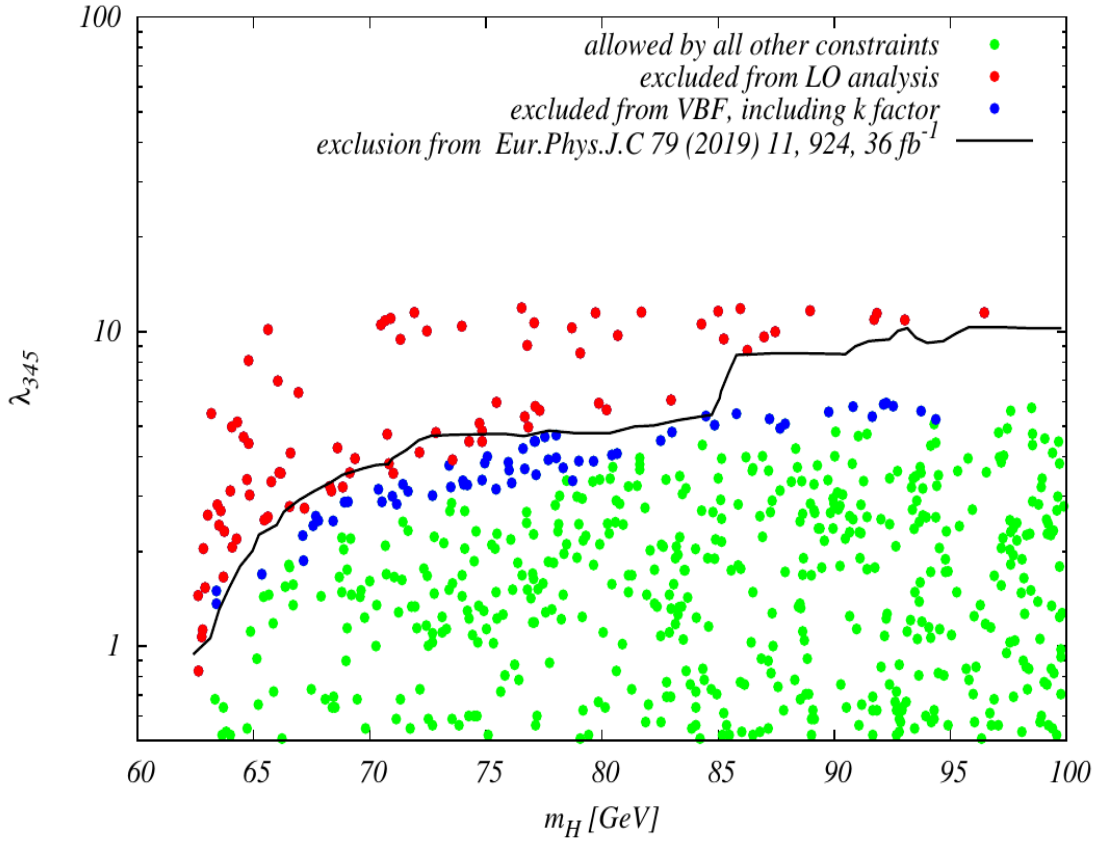
Invisible Higgs decays in VBF

ATLAS EXOT-2020-11





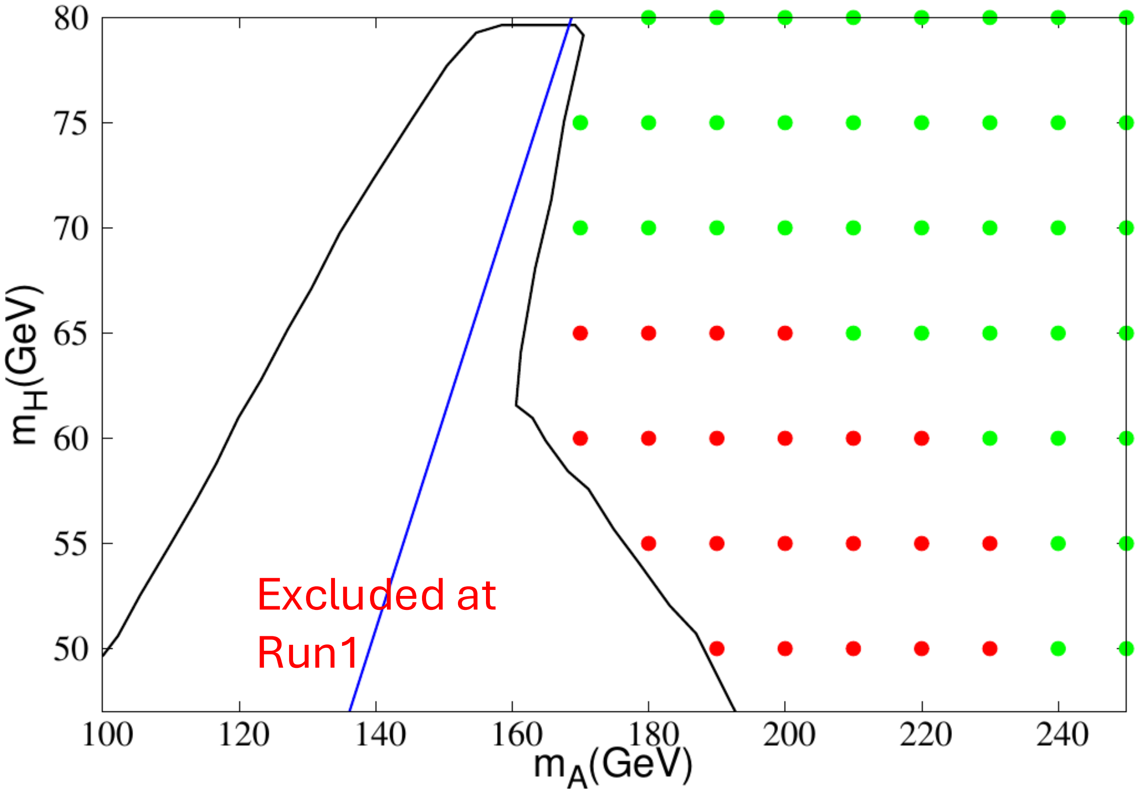
IDM constraints



J. Lahiri, T. Robens, KR, arXiv:2511.23133

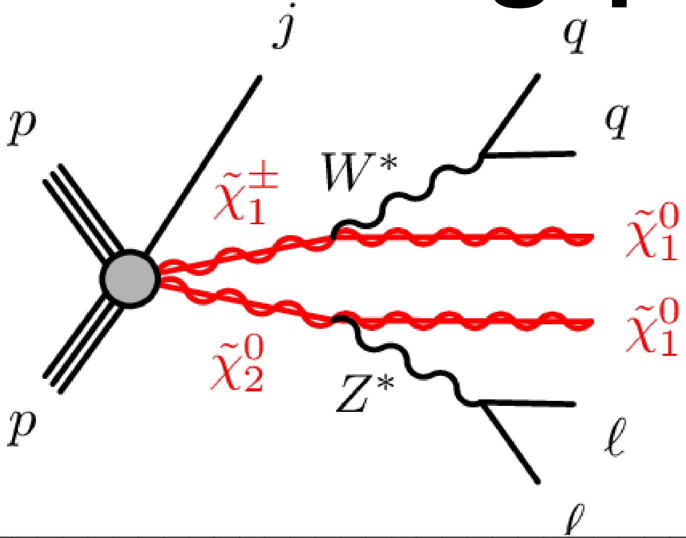
Revisiting ZH constraint after Run 2

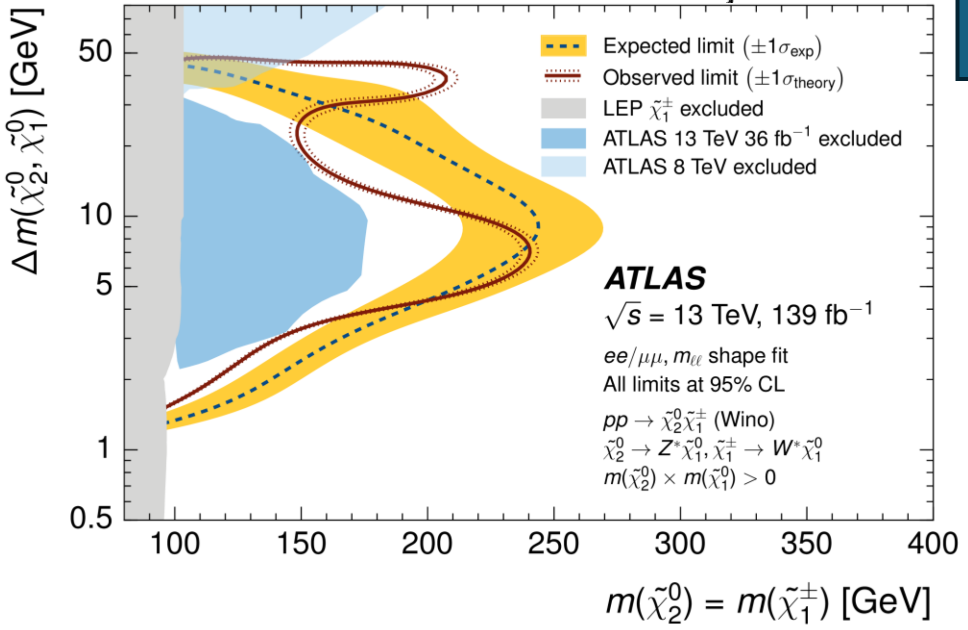
- Black line is expected Run 3 limit from the Run 1 study (extrapolated to 300/fb)
- The Run 2 constraints (red points) significantly improve
- Naïve rescaling of old results underestimates limits
- Improved statistical analysis (histogram fit instead of simple cut-and-count) gives huge advantage



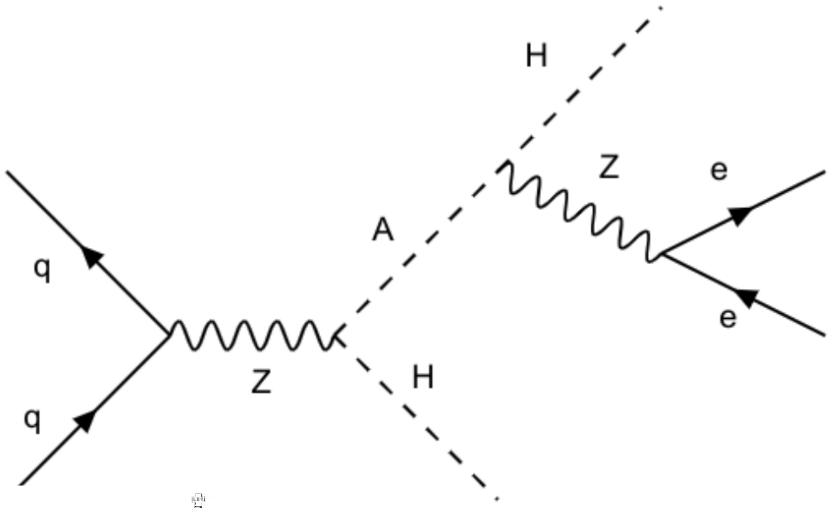
J. Lahiri, T. Robens, KR, arXiv:2511.23133

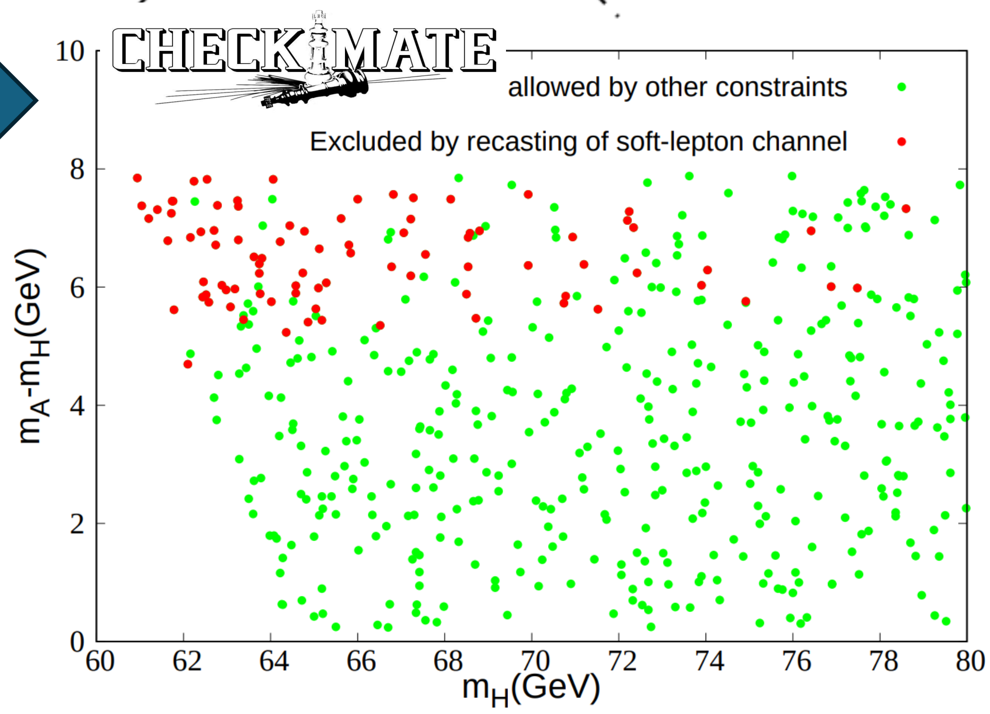
Small mass gap revisited





SUSY-2018-16





J. Lahiri, T. Robens, KR, arXiv:2511.23133

NN example
arXiv: 2106.09609

Arxiv:2106.09609

- Search for RPV-SUSY in final states with leptons and many jets (0 or 3 b-tagged)
- Signal regions count the number of jets with different pT thresholds; in general, 6-15 jets, at least 1 lepton
- Target: stops, gluino and EW higgsinos/winos
- EW signal: neutralino -> tbs; chargino -> bbs
- EW SR: 1 lepton, =6 jets, >=4 b-jets, NN discriminant
- NN released as ONNX files (in total 5, each for different jet multiplicity 4,5,6,7,8); unfortunately, very little information is provided

NN and CheckMATE implementation

- Using ONNX Runtime, https://onnxruntime.ai
- C++ library, analysis is performed on the event-by-event basis

- NN has 65 inputs: jets energy, rapidity, azimuth; MET, b-jet multiplicity; distance between jets and leading lepton, etc; some high-level combinations of jet momenta (as invariant masses);
- Problematic: b-tagging score for each jet based on DLR1 b-tagging algorithm (the pseudocode takes fixed values: 5 for b-jet; 1 for non-b-jet)

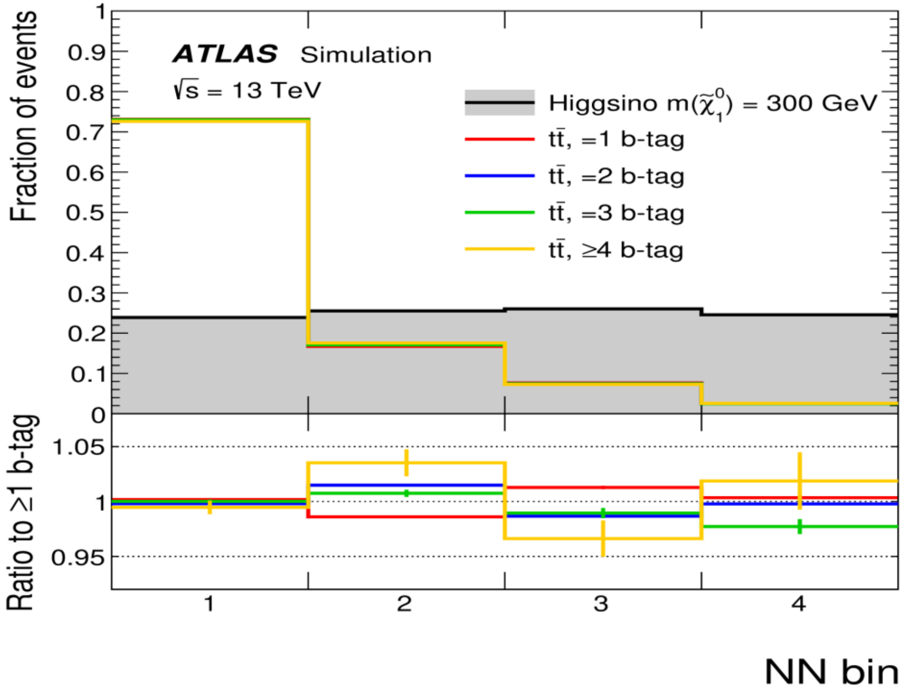
Validation - cutflow

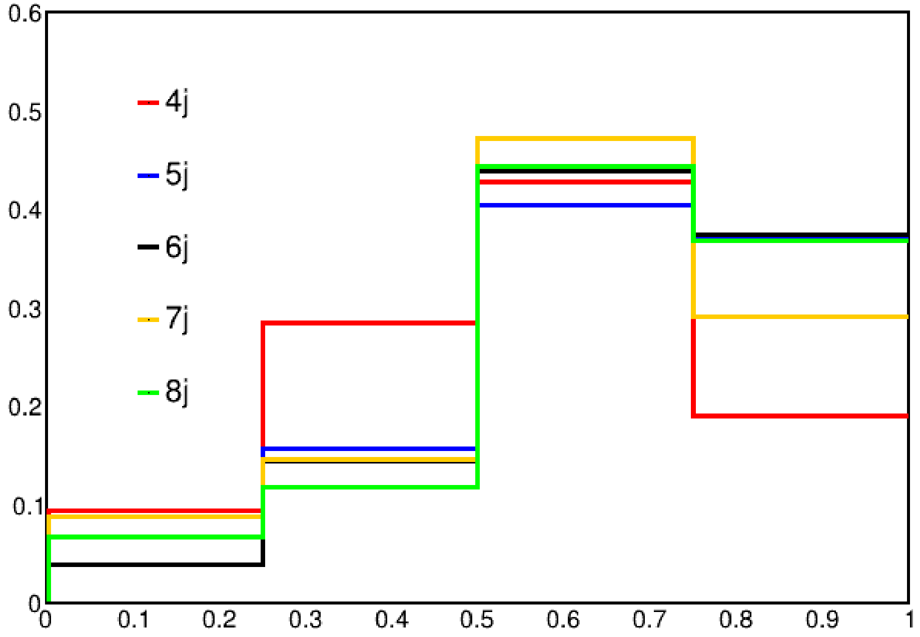
- Pretty much everything went wrong
- Clearly a problem with lepton id
- Too few events with high jet multiplicity
- After b-tagging things look better
- After NN inference the results are somewhat random
  (not surprising after looking at the histogram)
- Perhaps a general problem with signal modelling

$\tilde{\chi}_{1,2} \to tbs$	ATLAS	CheckMATE
All	14491	14491
Lead lep $p_T > 27$ GeV	5413	3304
$==$ 4 jets	681	549
$==$ 5 jets	1101	766
$==$ 6 jets	1188	734
$==$ 7 jets	840	466
$==$ 8 jets	420	186
$==$ 4 jets, $\geq$ 4b	7	5.4
$==$ 5 jets, $\geq$ 4b	29	33
$==$ 6 jets, $\geq$ 4b	57	43
$==$ 7 jets, $\geq$ 4b	61	41
$==$ 8 jets, $\geq$ 4b	39	21
NN_{4j} bin 4	2.9	0.7
NN_{5j} bin 4	8.7	6
NN_{6j} bin 4	17.4	11
NN_{7j} bin 4	18.9	19
NN_{8j} bin 4	14.5	10

Validation

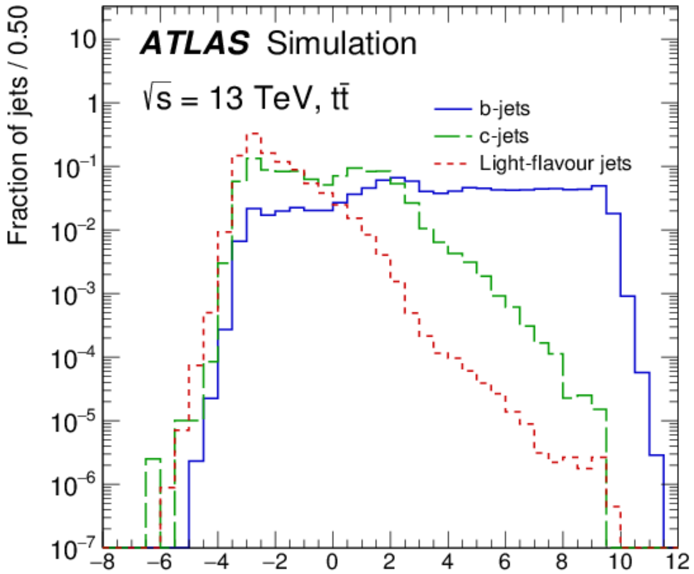
- Cutflows do not seem very useful; compare figures of NN output (unclear definitions of "bins" though)
- ATLAS distribution flat for signal
- Should be insensitive to b-tag, but it's not apparent in recast (e.g. large variation when using different ad-hoc b-tag scores)





Comparison for light jets

- Jets are assigned a score using a BDT tagger
- In recast we take ad-hoc values following SimpleAnalysis snippet
- The result seems to be very sensitive for reasonable choices of tagging score
- -1 seems to reproduce ATLAS better than 0 from SimpleAnalysi



FTAG-2018-01