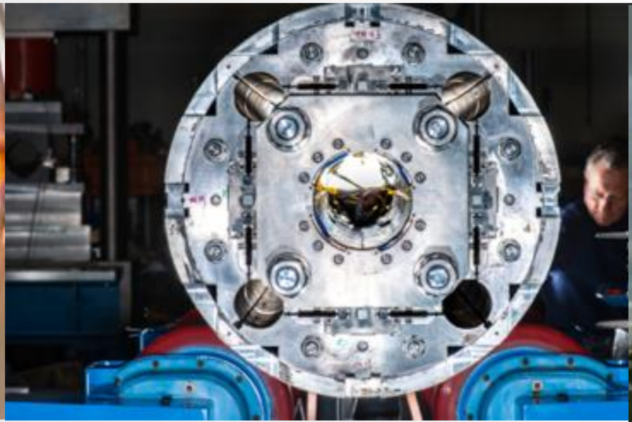
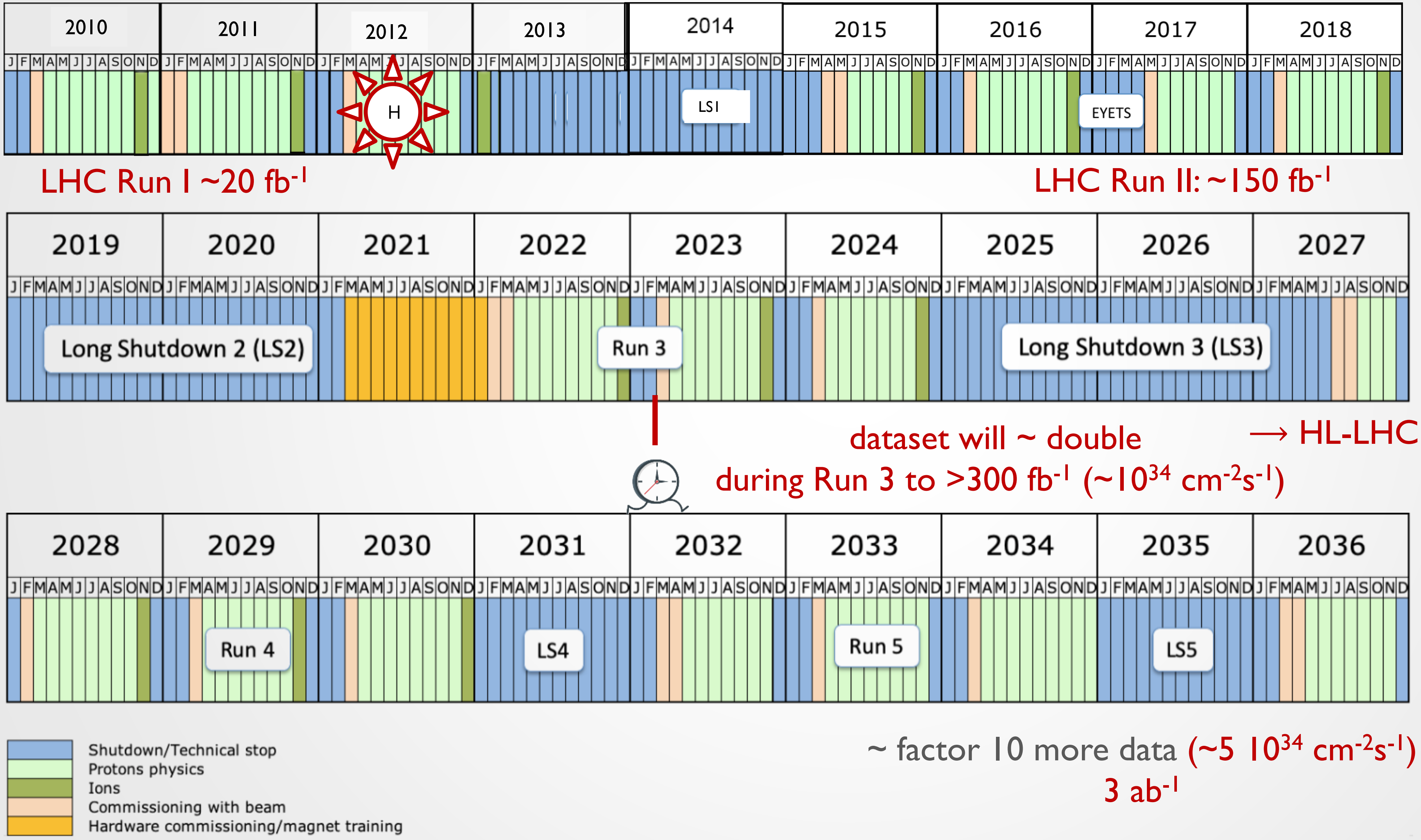
HEPHY CMS ANALYSIS GROUP SEMINAR

R. Schöfbeck (HEPHY Vienna), Feb. 10th, 2023 : W. Adam, S. Chatterjee, A. Escalante, C. Giordano, P. S. Hussain, M. Jeitler, D. Liko,
M. Matthewman, I. Mikulec, M. Shooshtari, D. Schwarz, M. Sonawane , W. Waltenberger, C. Wulz
Master's students: Rosmarie Schöfbeck, M. Kettner, O. Rothbacher, L. Wild

LHC LONG TERM SCHEDULE



<u>2022 data taking</u>

40.9/fb delivered
37.53/fb recorded
34.30 /fb certified

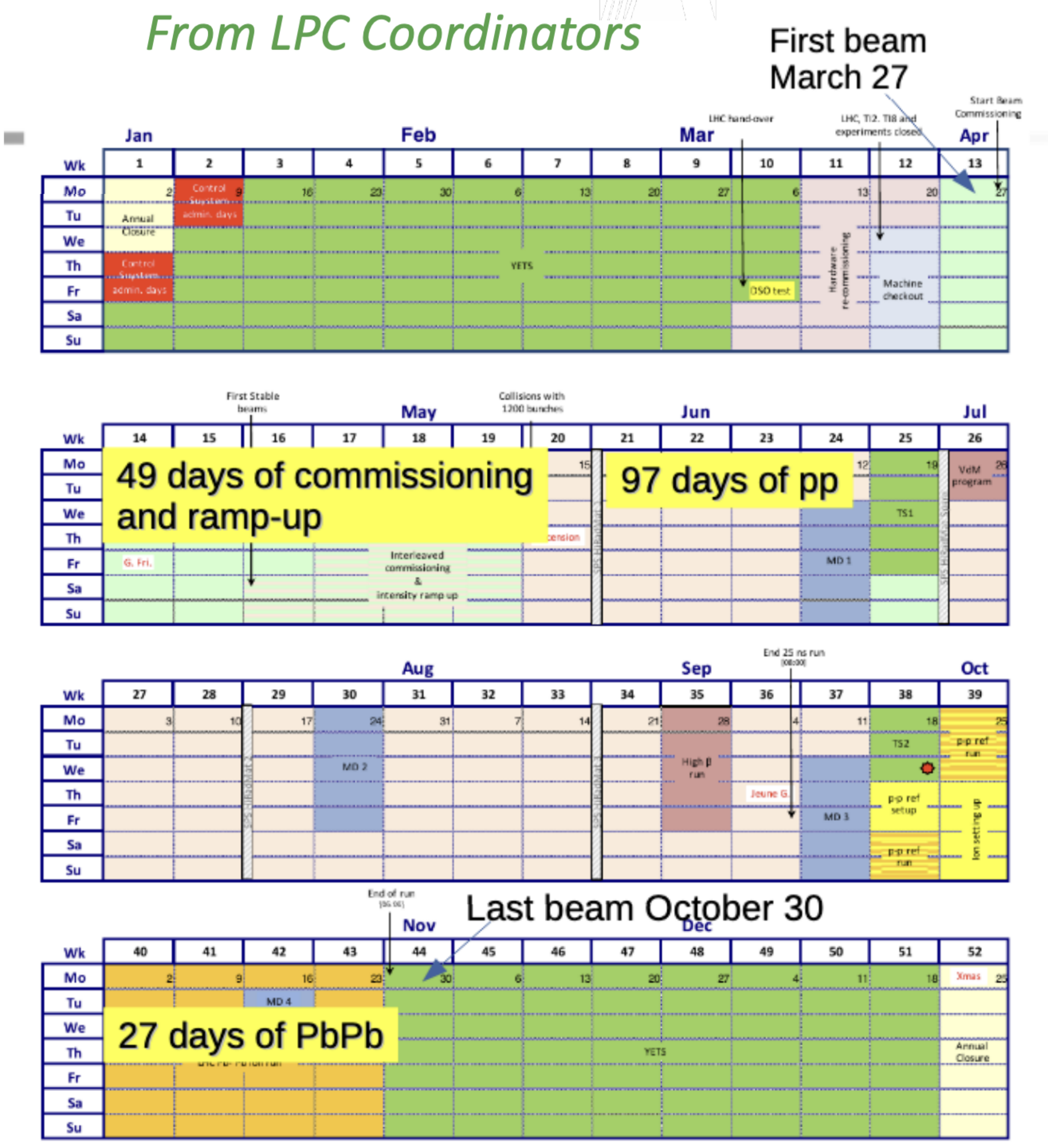
- 91% certified

OUTLOOK: 2023 DATA TAKING

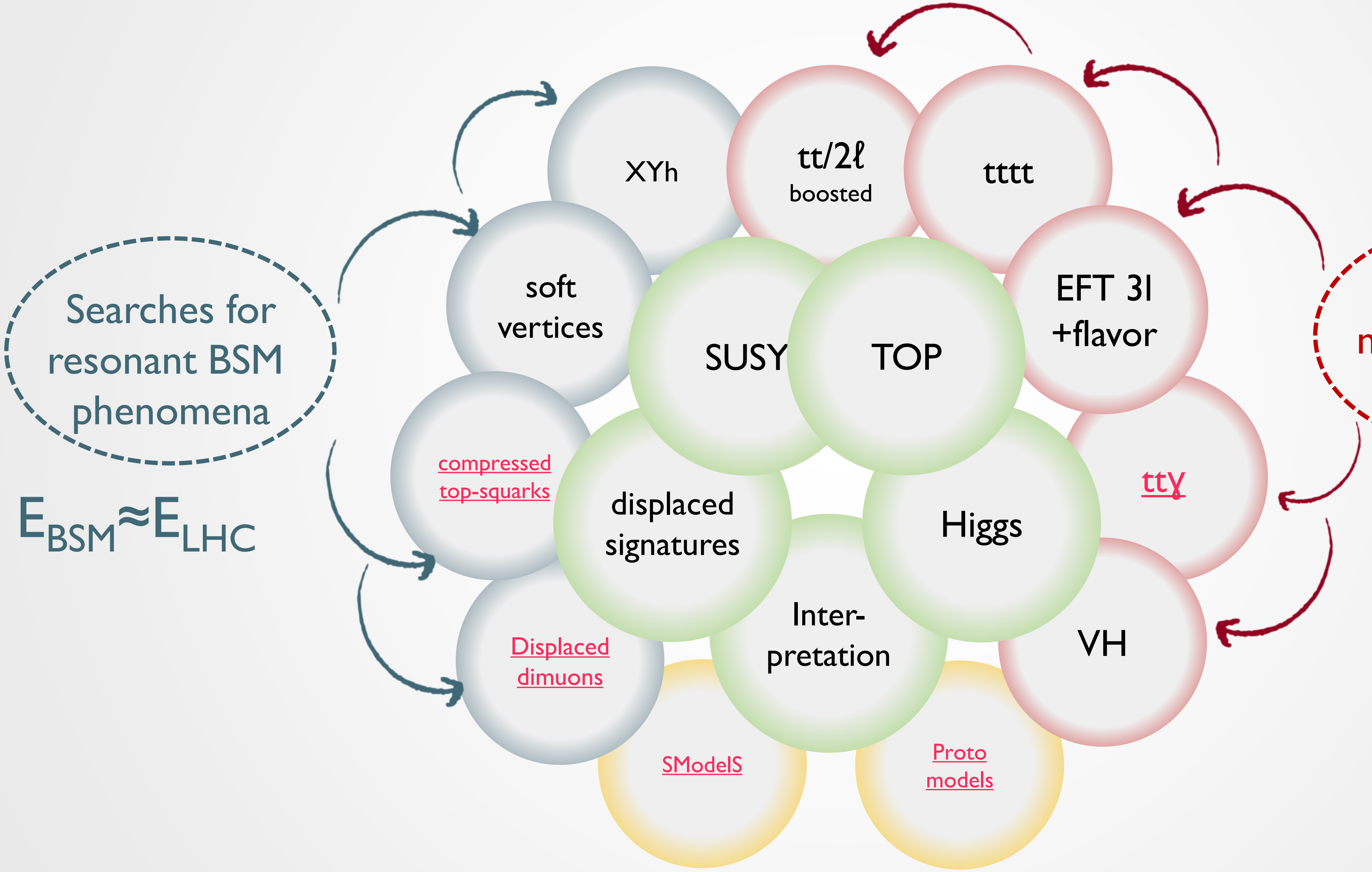
- 2023 is a short year for pp physics due to YETS and extended ion run:
  - ~13.5 weeks of pp physics
  - ~4 weeks of ion physics
  - If 75/fb in 2023 expect 300/fb for Run 3
    [Patricia McBride CMS week, internal]
- PU 2023 targeted in the range 60-65!!
  - LHC design: PU25; 2022: PU54; HL-LHC: PU140-200

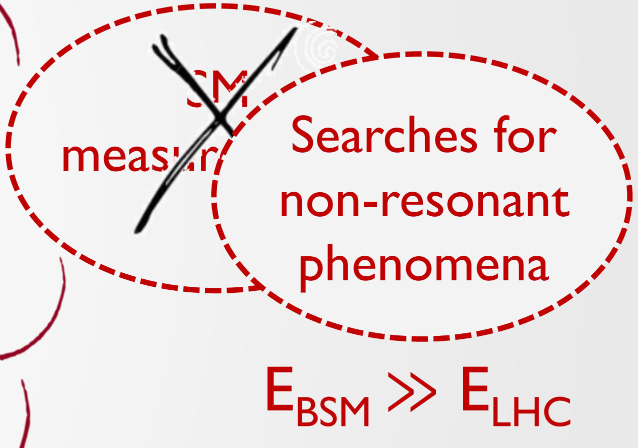
Important dates to keep in mind:

- First beam on March 27th
- First stable beams on April 22nd
- Start of physics run on May 15th
- Last day of pp physics: September 12th
- Start of pp reference run on September 21st
- Start of ion physics on October 2nd
- Last day of beam in 2023: October 30th



ACTIVITIES @ HEPHY (CMS DATA ANALYSIS)





<u>transition period</u>

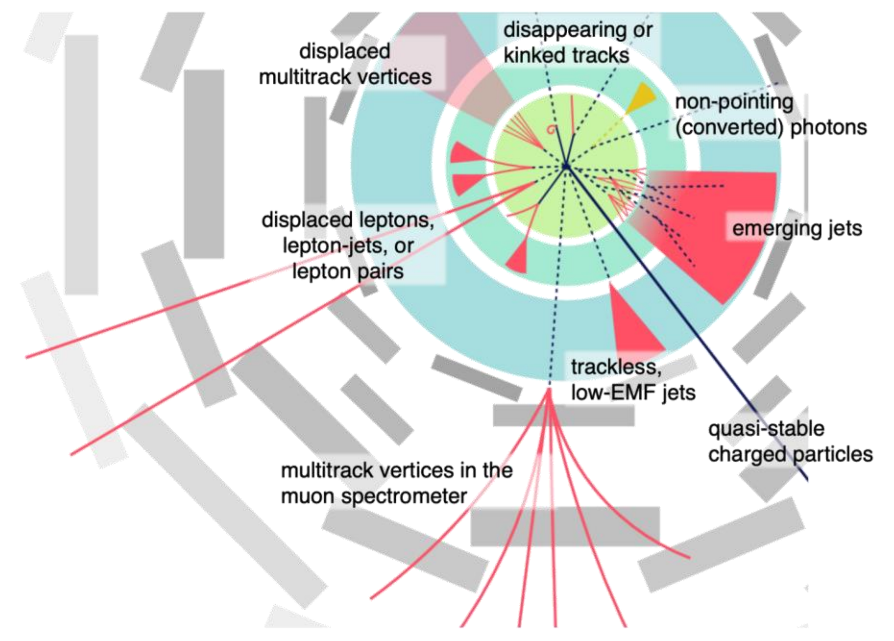
- ttγ done
- displaced 2μ accepted
- VH/EFT full swing
- tttt started
- soft vertices next
- many smaller items

SEARCHES

LONG-LIVED PARTICLES

A. Escalante, M. Sonawane, I. Mikulec,
W. Adam, S. Templ, C. Wulz, S. Kulkarni,

- Atypical experimental signatures are a wide field
- LLPs are predicted in many BSM physics scenarios [ref]
  - Decays mediated by heavy virtual mediators (e.g. HNL)
  - Nearly mass degenerate states (e.g. compressed SUSY)
  - Small couplings to SM particles (e.g. dark mediators)

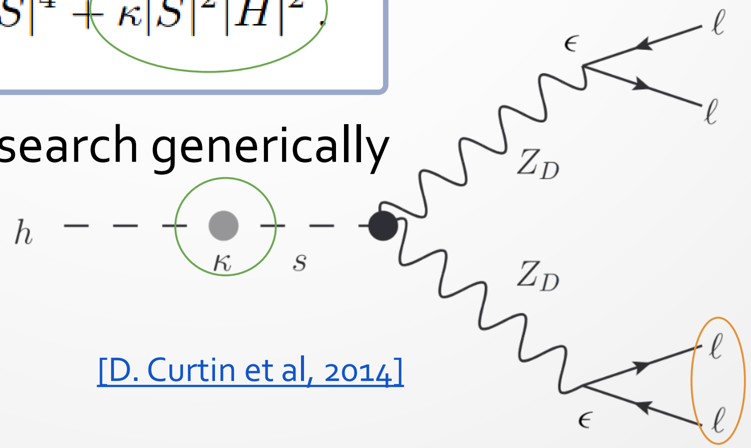


- E.g. Dark sector portal

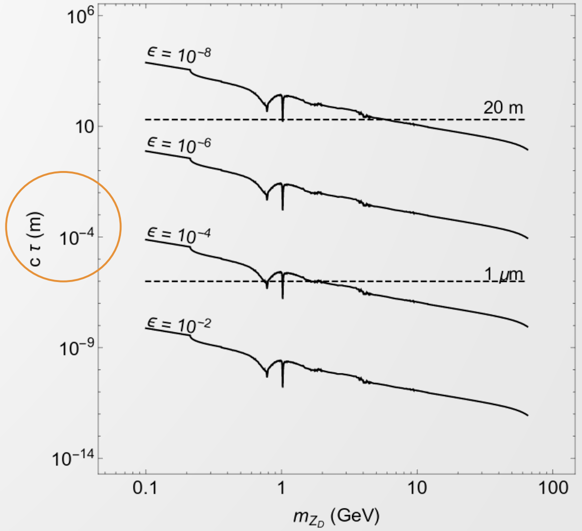
$$\mathcal{L} \subset -\frac{1}{4}\hat{B}_{\mu\nu}\hat{B}^{\mu\nu} - \frac{1}{4}\hat{Z}_{D\mu\nu}\hat{Z}_D^{\mu\nu} + \frac{1}{2}\frac{\epsilon}{\cos\theta}\hat{Z}_{D\mu\nu}\hat{B}^{\mu\nu} + \frac{1}{2}m_{D,0}^2\hat{Z}_D^{\mu}\hat{Z}_{D\mu}$$

$$V_0(H,S) = -\mu^2|H|^2 + \lambda|H|^4 - \mu_S^2|S|^2 + \lambda_S|S|^4 + \kappa|S|^2|H|^2$$

- BR, $c\tau$, L_{xy} strongly model-dependent → search generically
  - With UCLA
  - Accepted by JHEP



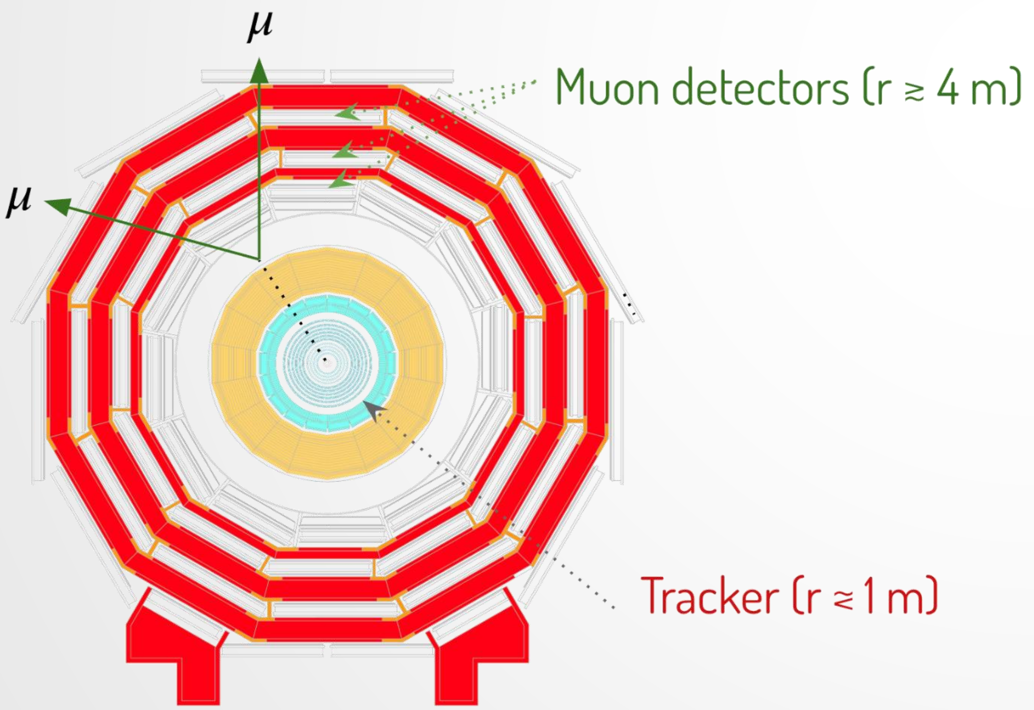
[D. Curtin et al, 2014]



DISPLACED DIMUONS

EXO-21-006, accepted by JHEP

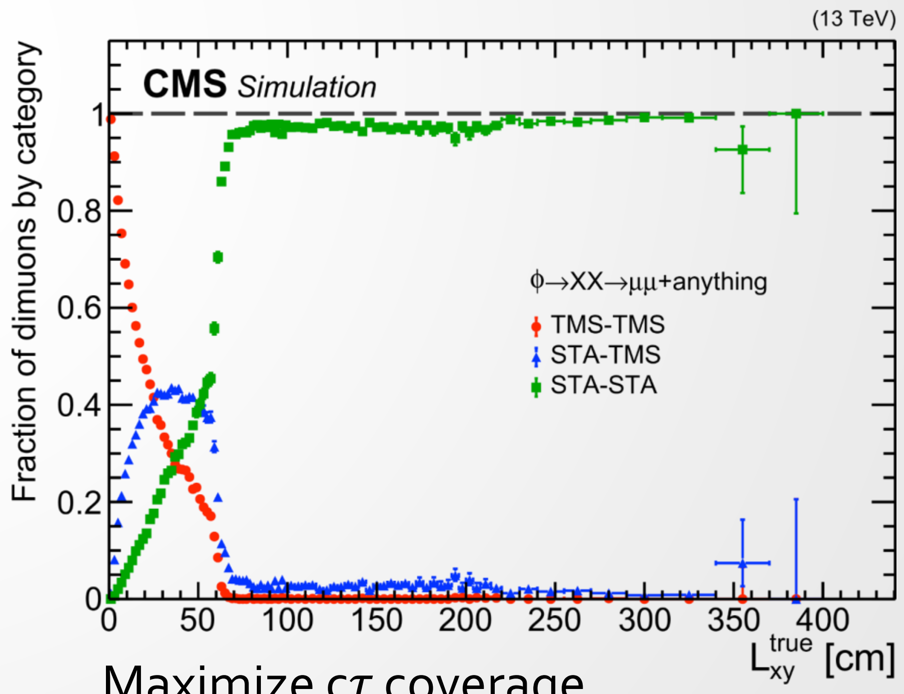
- Search generically for **displaced dimuons within and beyond the tracker**
  - Double muon triggers relying on muon system information alone



- Search uses 3 exclusive dimuon categories, STA-STA, STA-TMS, TMS-TMS, defined by two types of reconstructed muons:

  STA: muon system only

  TMS: STA + tracker information

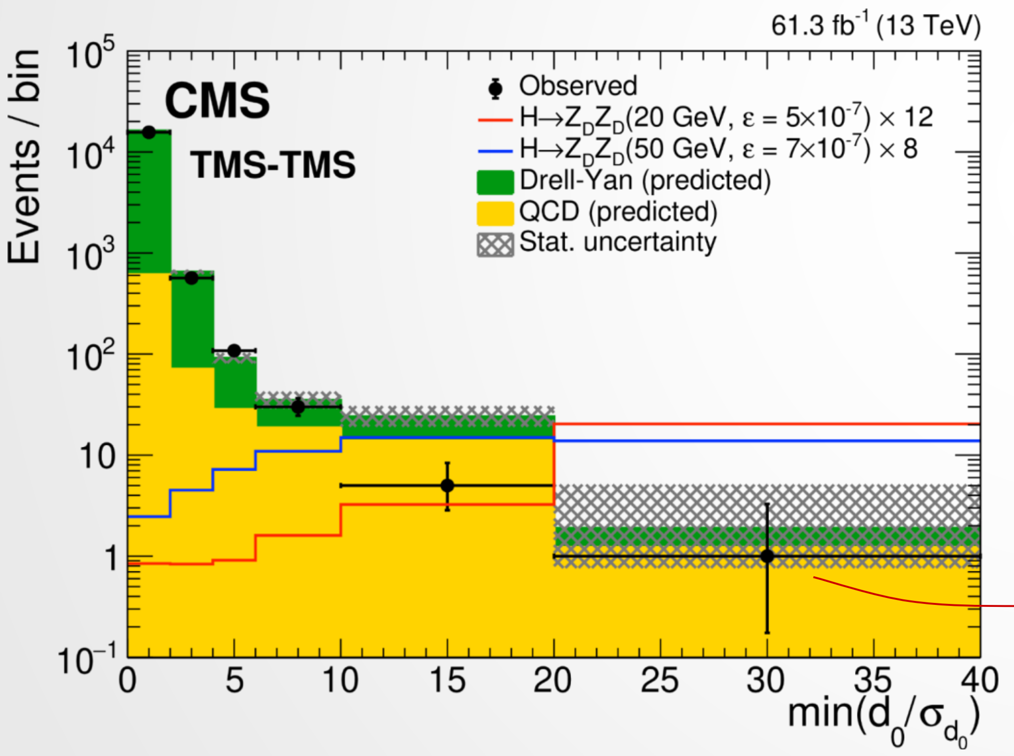


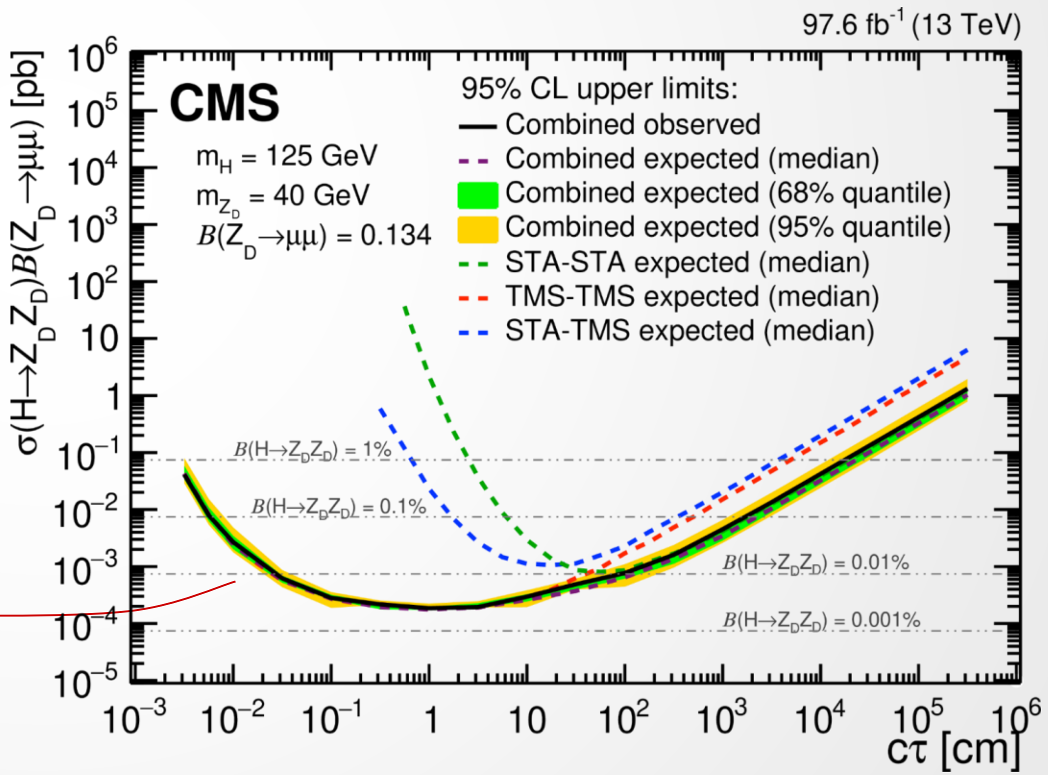
Maximize $c\tau$ coverage

DISPLACED DIMUONS: TRACKER

EXO-21-006, accepted by JHEP

- Rely on transverse impact parameter (d_o) resolution (~ tens of μm)
  - Isolated muons with **large d_o and L_{xy} significance**
  - Signal events cluster in $m_{\mu\mu}$



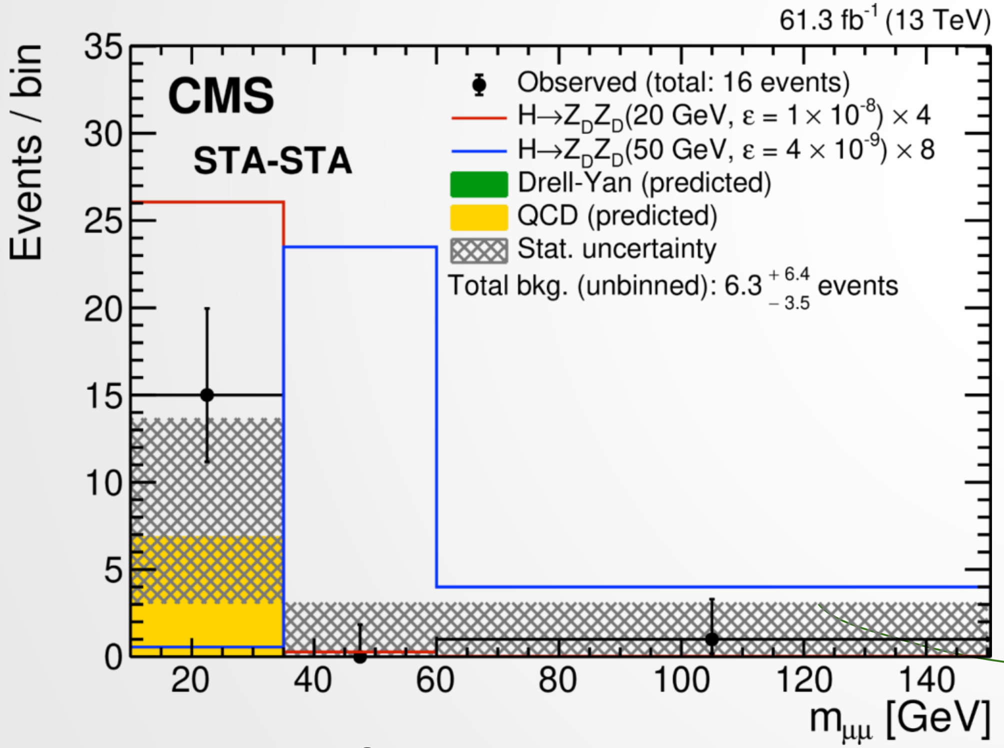


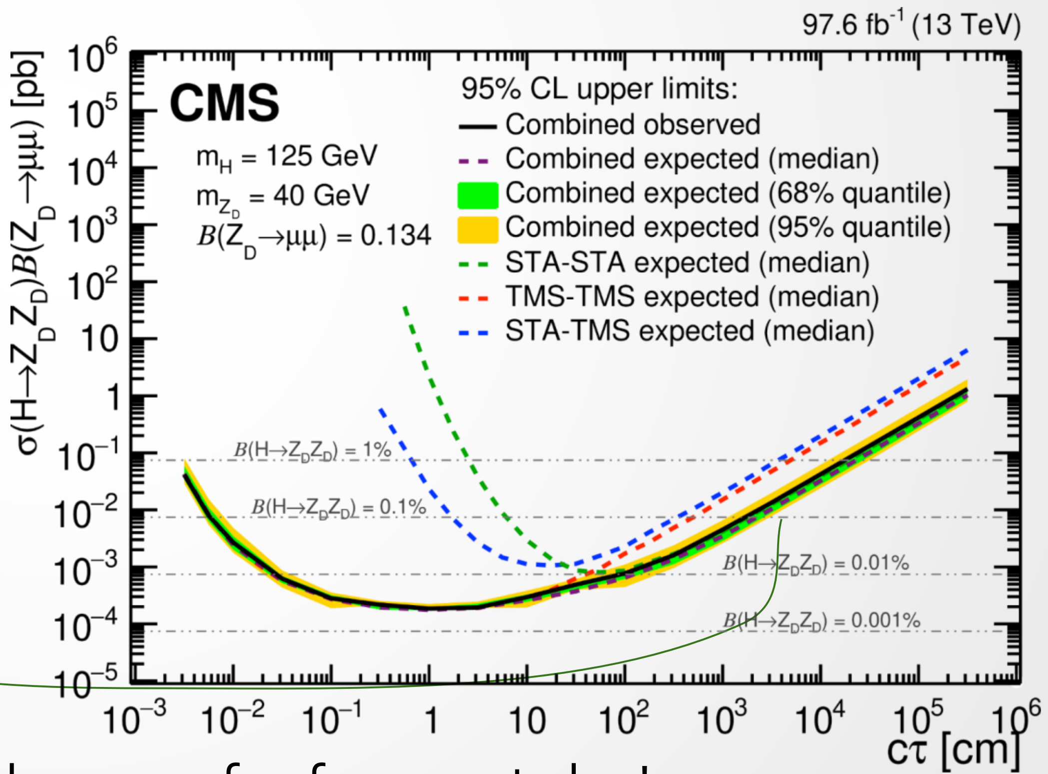
- Backgrounds evaluated using data in dedicated control regions

DISPLACED DIMUONS: MUON SYSTEM

EXO-21-006, accepted by JHEP

- STA muons not associated to TMS muons
  - Sensitivity relies on veto of prompt activity (~100% muon reconstruction efficiency in tracker)
  - Signal characteristic: Large L_{xy} significance and cluster in $m_{\mu\mu}$





- Combination of categories → sensitivity to a wide range of $c\tau$ from μm to km!
  - Excluded $B(H \rightarrow Z_D Z_D) > 10^{-4}$-$10^{-5}$, depending on $(m_{ZD}, c\tau_{ZD})$

[PRD 99, 012001]

DISPLACED DIMUONS: RUN 2 "LEGACY" RESULT

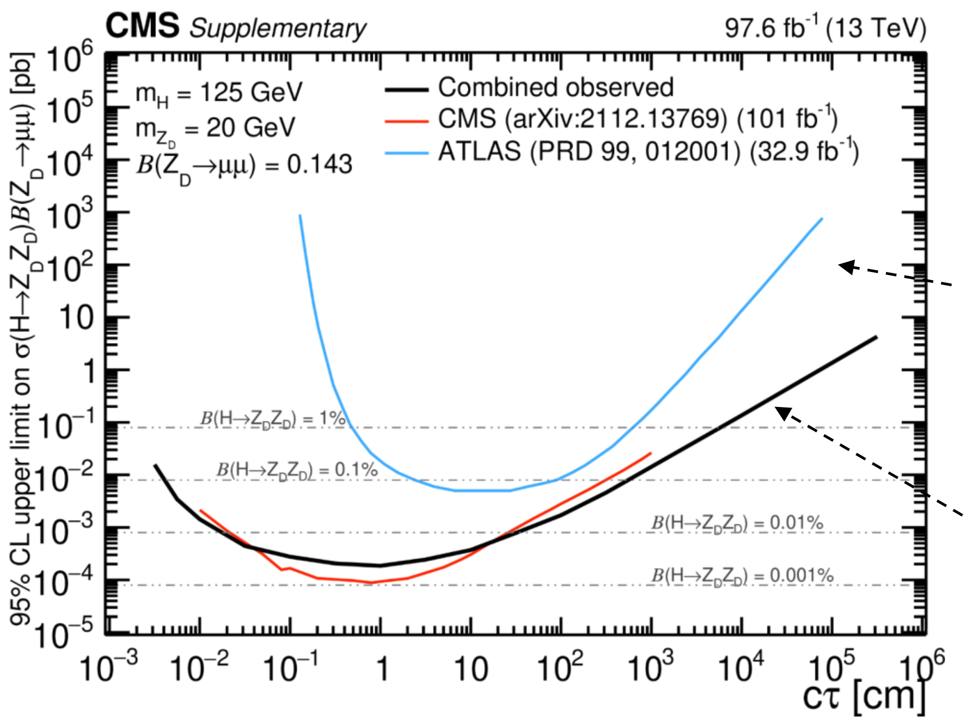
EXO-21-006, accepted by JHEP

Research line initiated '**from scratch**' in 2017.

Today:

- **State-of-the-art** reference for displaced muons.
- pioneered triggering, lepton ID, and background suppression methods in CMS.
- World-leading constraints for long-lived dark photons in most $c\tau$ (far superior wrt ATLAS).
- Early Run2 competitors (IFCA, Oviedo) now joined

- Alberto will join Madrid on a tenure-track position " Atracción de Talento Investigador de la Comunidad de Madrid Modalidad-1"
    - Will continue to be involved



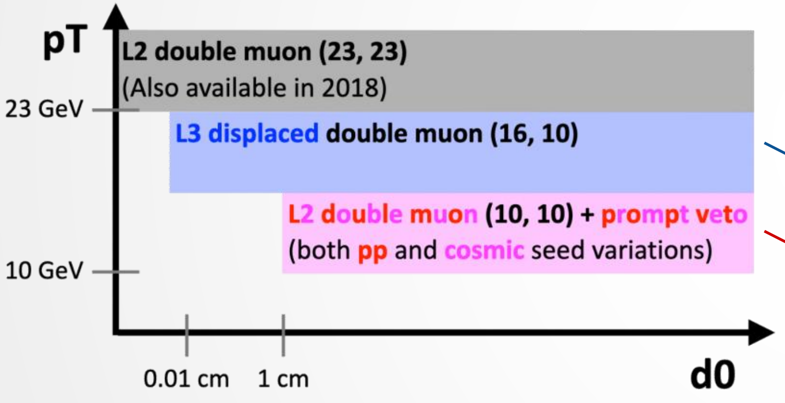
EXO-21-006

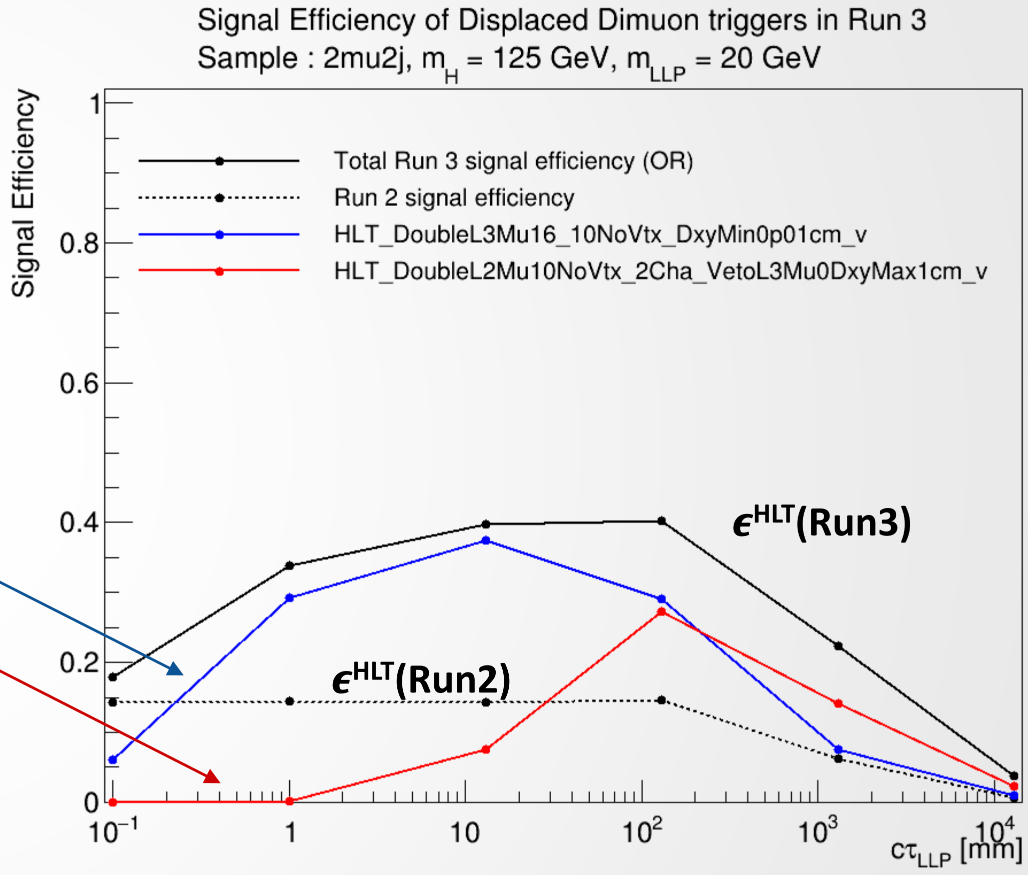
- [Physics briefing]

DISPLACED DIMUONS: RUN 3 IMPROVEMENTS

M. Sanowane

- Used Run 2 L1-experience for new HLT paths for Run 3
  1. Remove beamspot constraint in p_T measurement at L1
  2. Lower p_T thresholds at HLT (p_T> 23 GeV in Run 2)
  3. Prompt veto

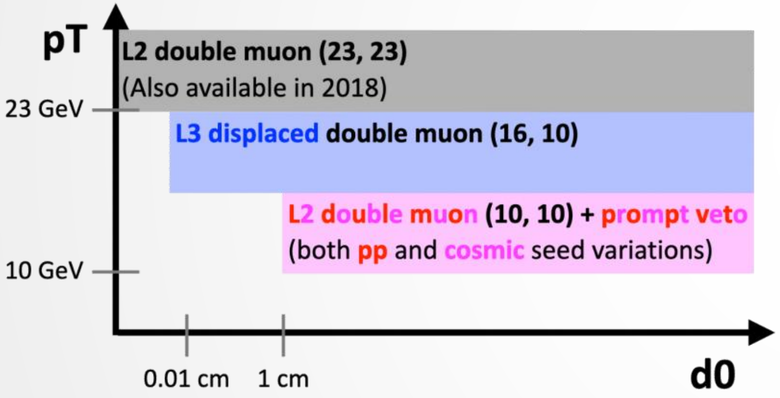




  - Displaced dimuons in inner tracker (do > 0.01 cm)
  - Displaced dimuons in muon system without prompt muons in tracker (do > 1 cm)
- factor 2-4 gain in signal efficiency (depending on $c\tau$)
  - Potential to improve the sensitivity already with 2022 data (38 fb^{-1} recorded)

DISPLACED DIMUONS: RUN 3 IMPROVEMENTS

- Used Run 2 L1-experience for new HLT paths for Run 3
  1. Remove beamspot constraint in p_T measurement at L1
  2. Lower p_T thresholds at HLT (p_T> 23 GeV in Run 2)
  3. Prompt veto

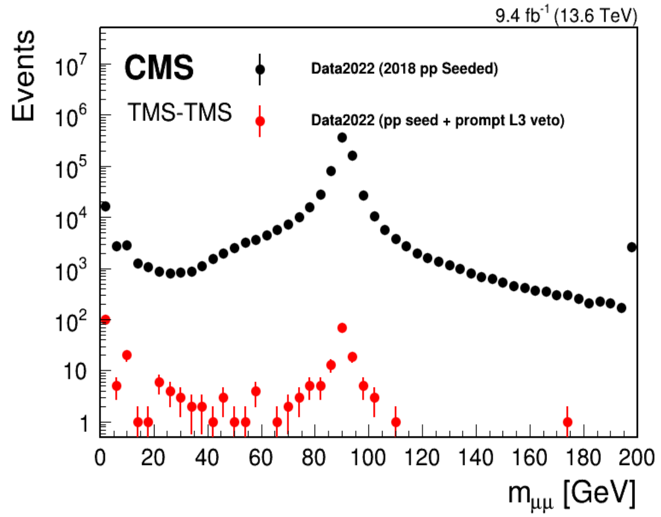


  - Displaced dimuons in inner tracker (do > 0.01 cm)
  - Displaced dimuons in muon system without prompt muons in tracker (do > 1 cm)

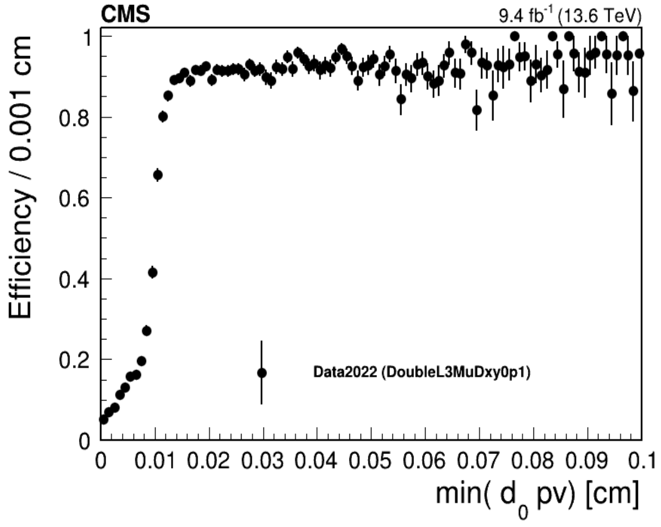
- factor 2-4 gain in signal efficiency (depending on cτ)
  - Potential to improve the already with 2022 data (38 fb^{-1} recorded)

Run III, 13.6 TeV

10^{-4} prompt dimuon suppression with d_0 < 1 cm



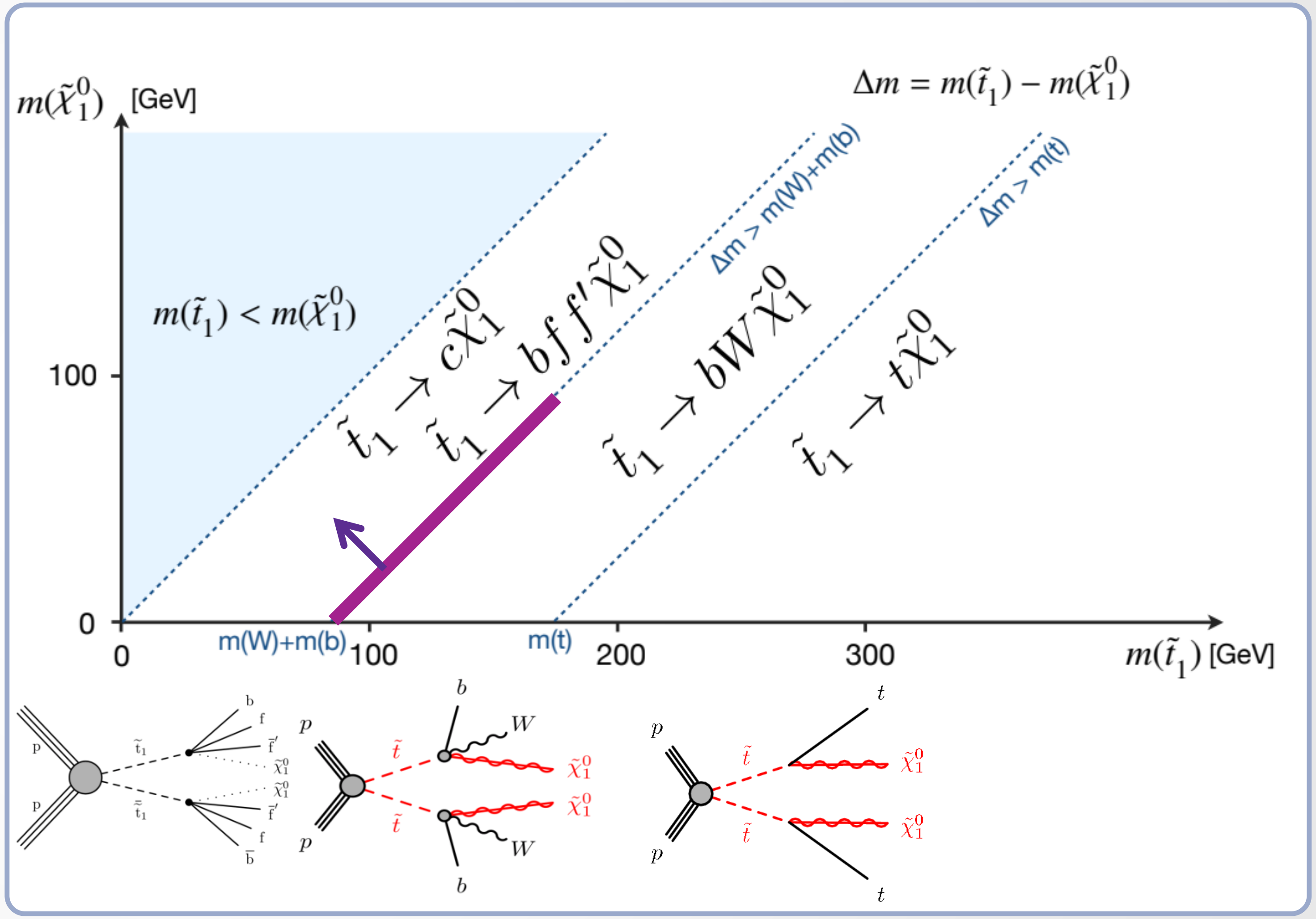
Efficient pair selection with d_0 > 0.01 cm



SUPERSYMMETRY: COMPRESSED TOP SQUARKS

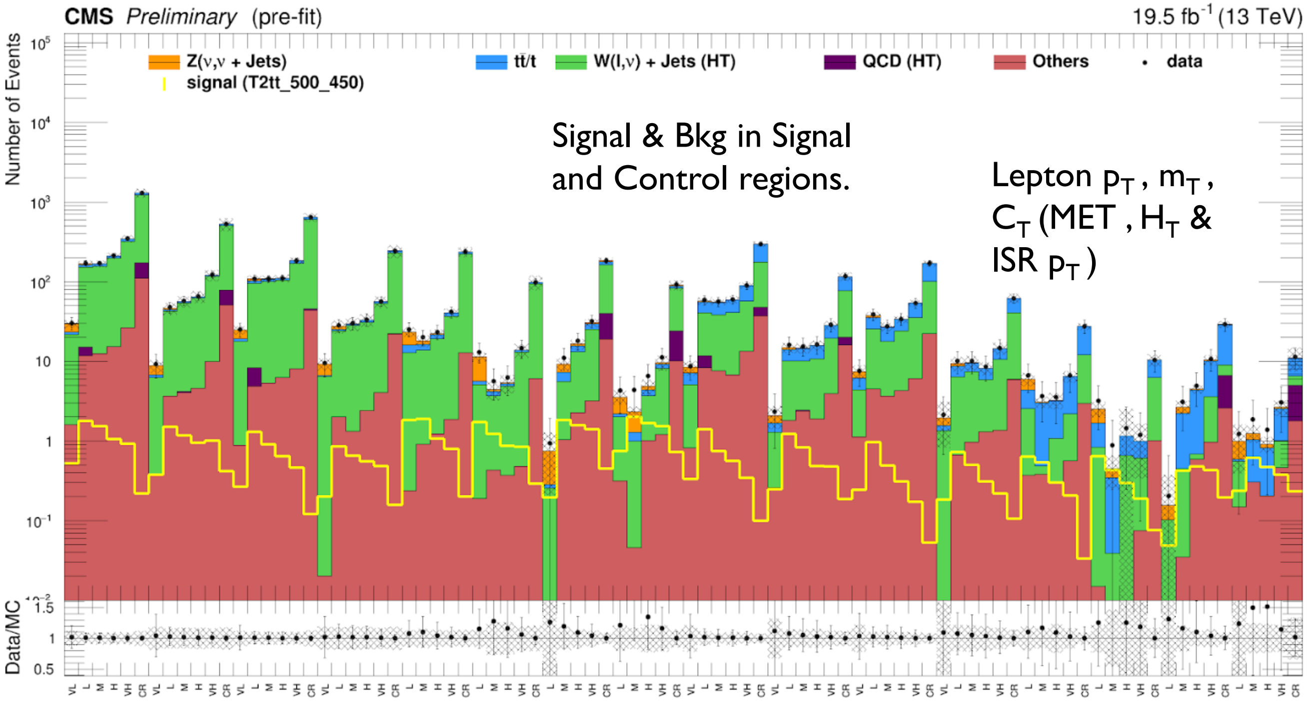
P. Hussein, I. Mikulec, D. Liko, W. Adam

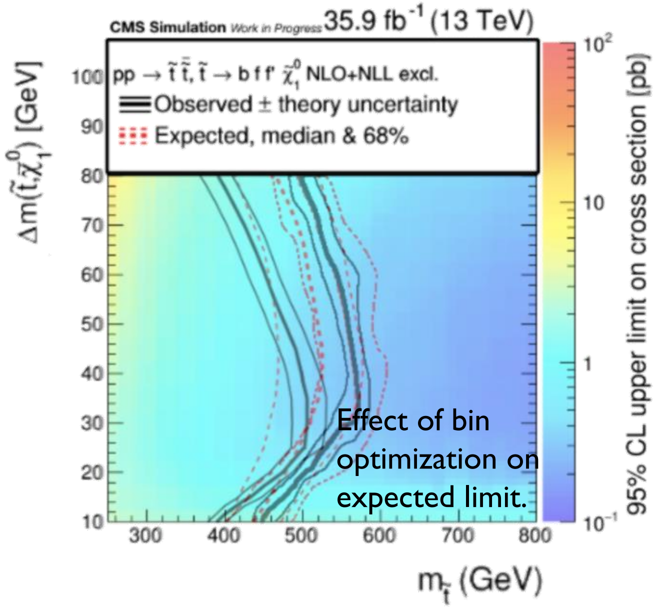
- top squark decay-phenomenology strongly dependent on mass hierarchies
- Focus on 4-body decay of top squark, 1ℓ
  - Target very compressed scenarios $\Delta m = m_{stop} - m_{LSP} < m_W$
  - Boost sensitivity with high-p_T ISR jet
  - Including LL scenario (100% BR)
- Collaboration with ELTE
  - Based on 2016 cut-and-count approach AN-17-165)
- Search regions, generically defined
  - lepton p_T, m_T, E_T^{miss}, H_T and ISR jet p_T
  - Background processes
    - Prompt lepton (WJets, tt/t)
    - Fake lepton (QCD, Z-Inv)

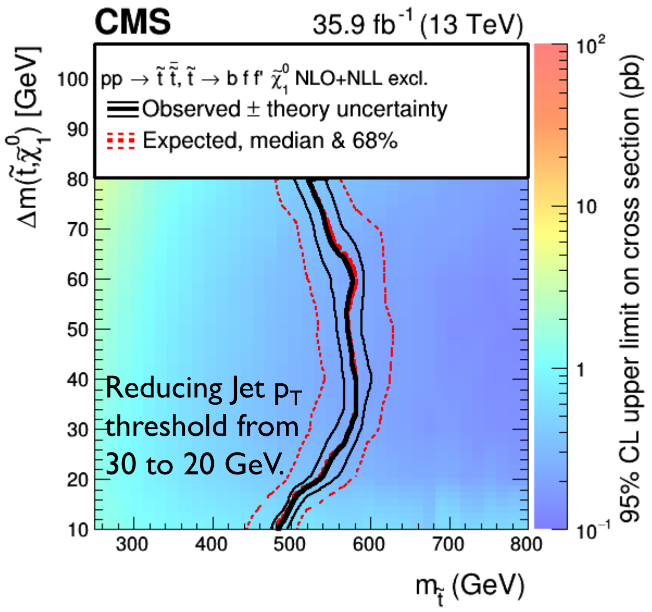


COMPRESSED TOP SQUARKS

- Extend to LL scenario for Δm(stop, neutralino) ≤ 30 GeV
  - Higher lepton impact parameter ~10cm
  - Common strategy with prompt search
- Pushing the limit on various fronts
  1. More bins to the search region with higher lepton p_T and m_T to thresholds
  2. Reducing jet p_T threshold
  3. Changing anti-QCD cut from $\Delta\Phi(j1, j2)$ to $\min\{\Delta\Phi(j_1, E_T^{miss}), \Delta\Phi(j_2, E_T^{miss})\}$
  4. Using secondary vertex to capture the soft b (< 20 GeV) coming from the signal
  5. Utilizing new low p_T electron reconstruction to gain the efficiency at higher impact parameter
     - First SUSY analysis with ≥ 3 GeV e/μ
- Should conclude this year







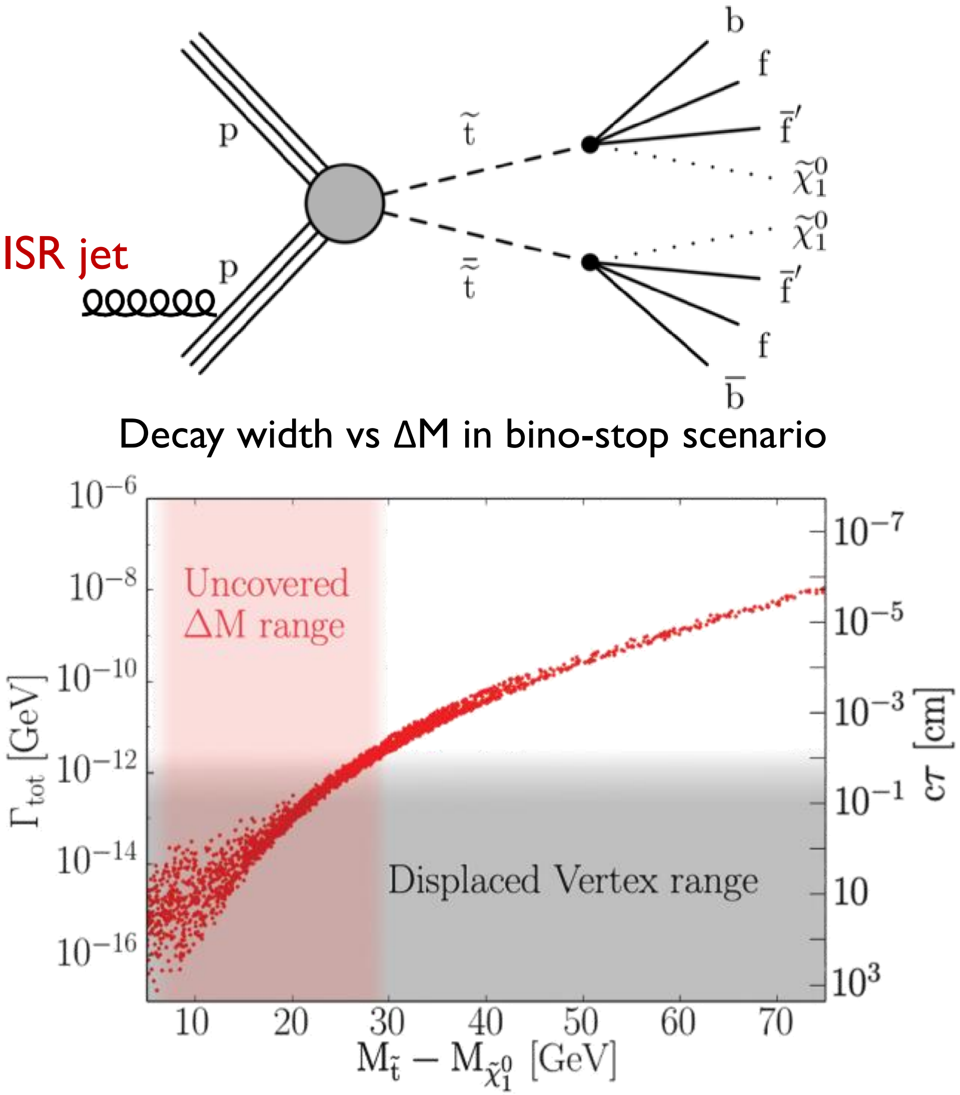
SOFT VERTICES: PHYSICS GOAL

I. Mikulec + 2x N.N.

- Explore DM co-annihilation scenarios for mass gaps from from few GeV to few tens of GeV
  - Bino-stop [1408.4662], Bino-wino [1506.08206], Singlet-triplet Higgs portal [1812.04628], Extra-dimensions and composite models [1702.00750]
- Experimental signature
  - Similar to classical collider DM searches: E_T^{miss} +ISR jet, but with addition of soft displaced vertices (DV)
  - Stops Compressed can "only" go to 3 GeV
  - Exploring DVs up to a few cm's displacement
- Make use of objects to their limit of detectability:
  - Tracks with p_T > ~0.5 GeV
  - DV formed by at least two selected tracks
  - Unprecedented at LHC

Stop decay in bino-stop scenario

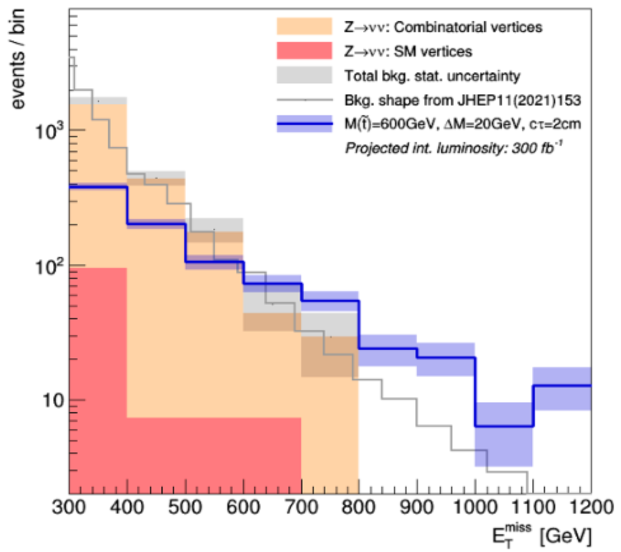
Decay width vs ΔM in bino-stop scenario

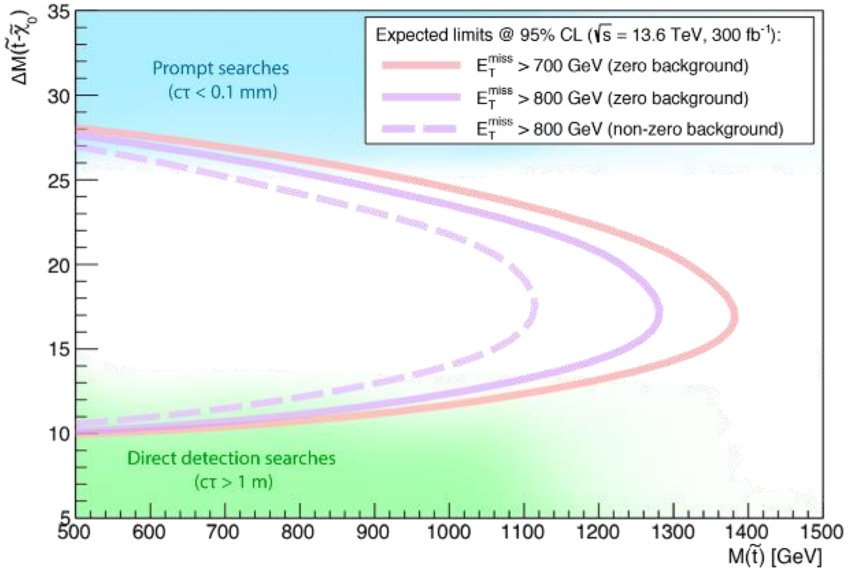


SOFT VERTICES: STATUS

I. Mikulec + 2x N.N.

- MC-based feasibility study promising:
  - Reconstructed E_T^{miss} distribution and projected exclusion ranges
  - Closes, e.g.,the gap between “mono-jet” and MET+ISR+soft high-level object signatures
- Plan for the start of project
  - Apply ML techniques for both track and DV selection
  - Use both Run 2 and Run 3 data
    - Use existing MET triggers
    - Make use of pixelless track reconstruction improvements in Run 3
- Hiring Postdoc + PhD (FWF)





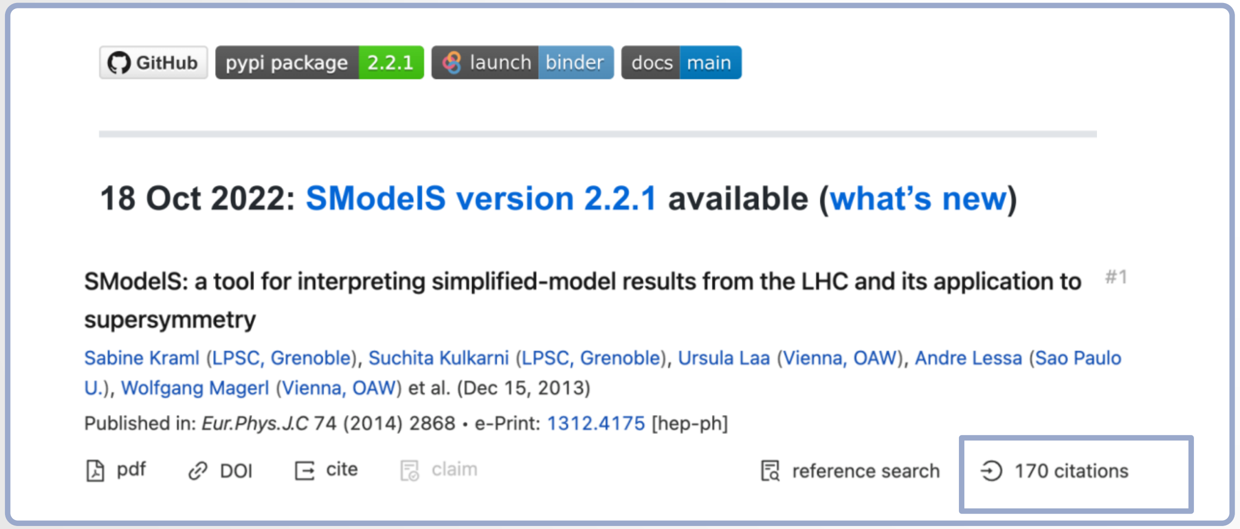
SMODELS INTERPRETATION

SMODELS

W. Waltenberger,
S. Narasimha

https://smodels.github.io/

- A tool for quickly comparing a theory with a database of experimental results
- Decomposes theory automatically into its simplified model spectrum
  - Matches against results
  - Obtains new limits



GitHub | pypi package 2.2.1 | launch binder | docs main

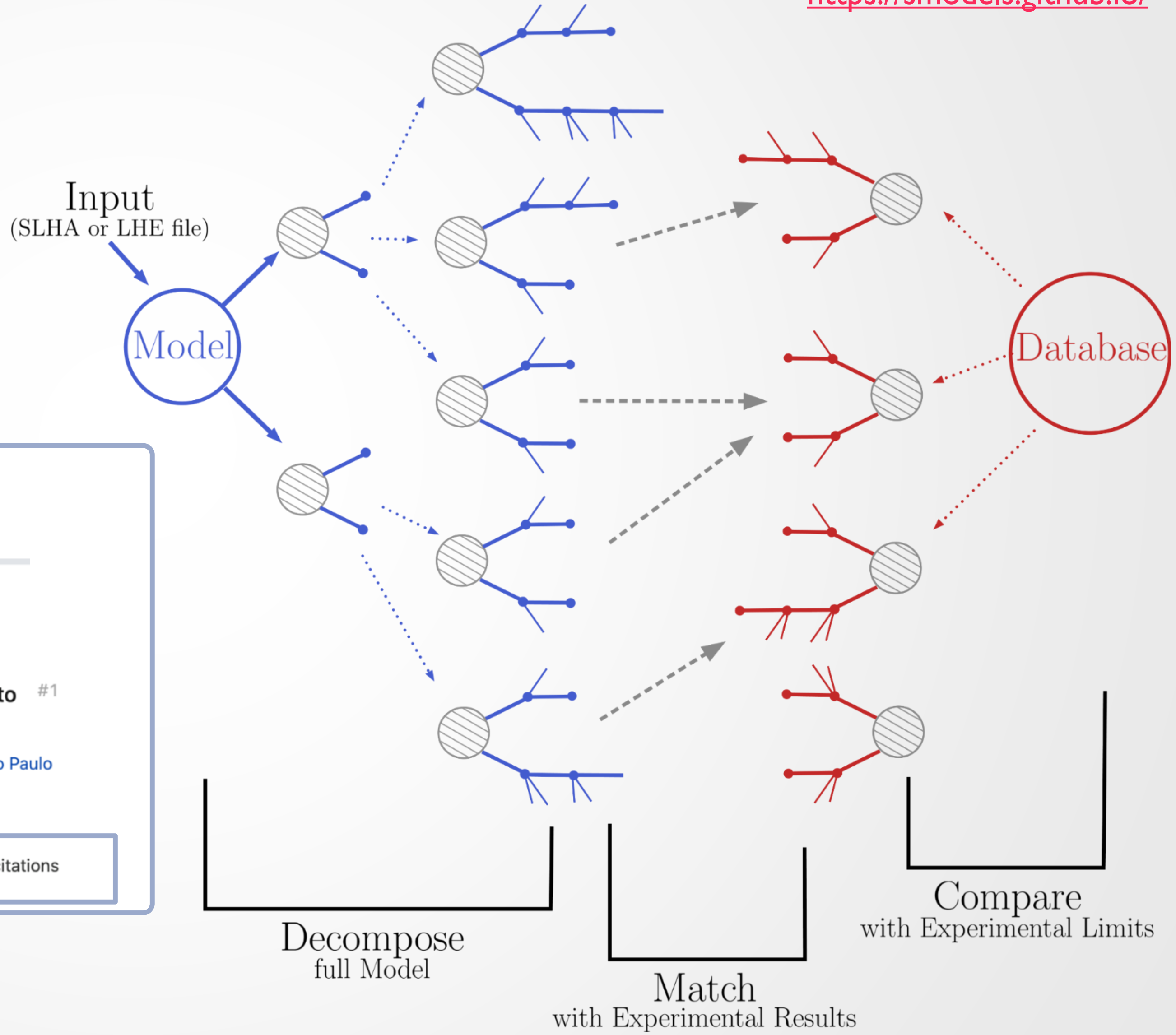
18 Oct 2022: SModelS version 2.2.1 available (what's new)

SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry #1

Sabine Kraml (LPSC, Grenoble), Suchita Kulkarni (LPSC, Grenoble), Ursula Laa (Vienna, OAW), Andre Lessa (Sao Paulo U.), Wolfgang Magerl (Vienna, OAW) et al. (Dec 15, 2013)

Published in: *Eur.Phys.J.C* 74 (2014) 2868 • e-Print: 1312.4175 [hep-ph]

pdf | DOI | cite | claim | reference search | 170 citations

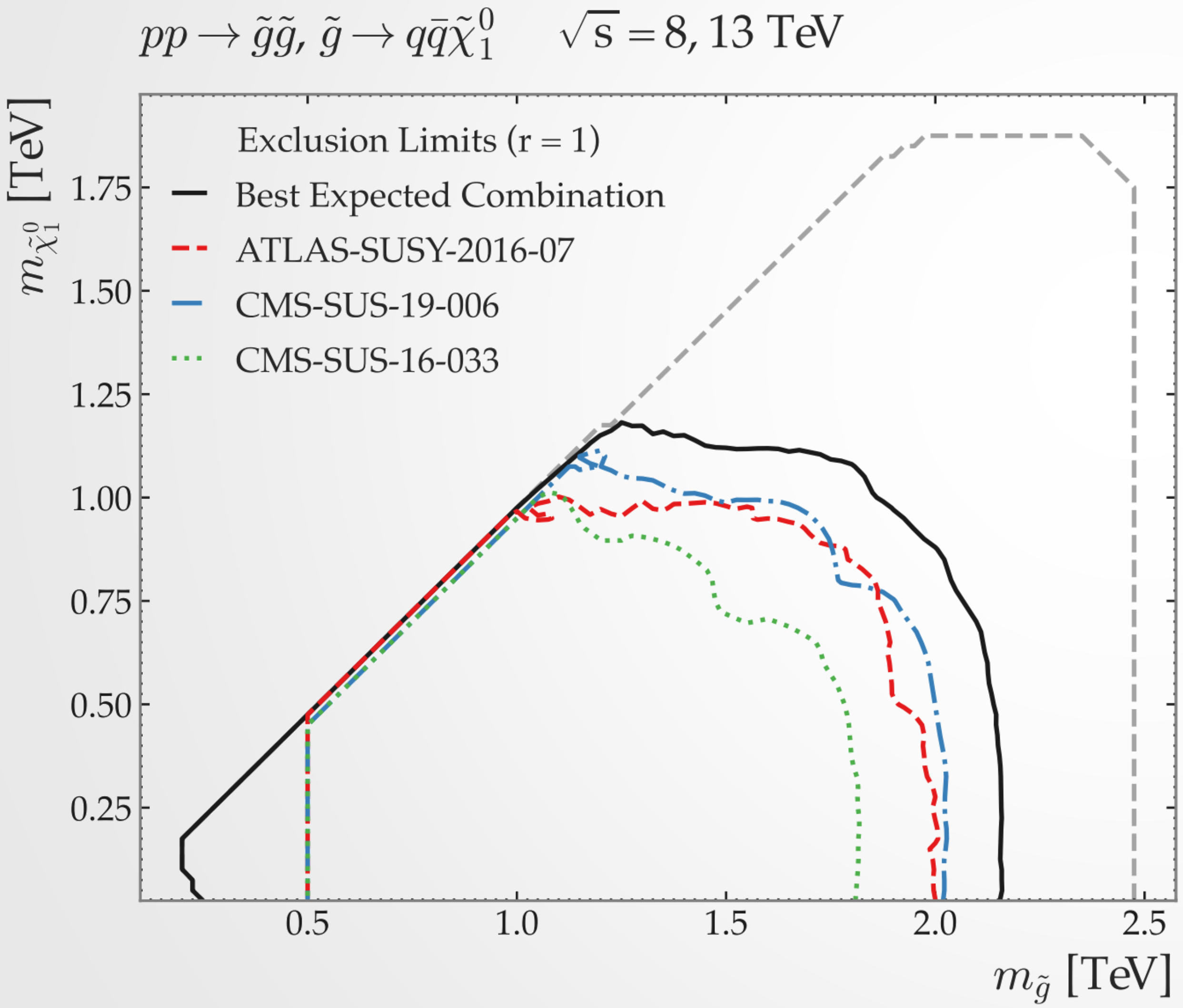


- 10yrs onwards, database contains results from > 100 CMS and ATLAS publications

NEW FEATURE: COMBINATIONS

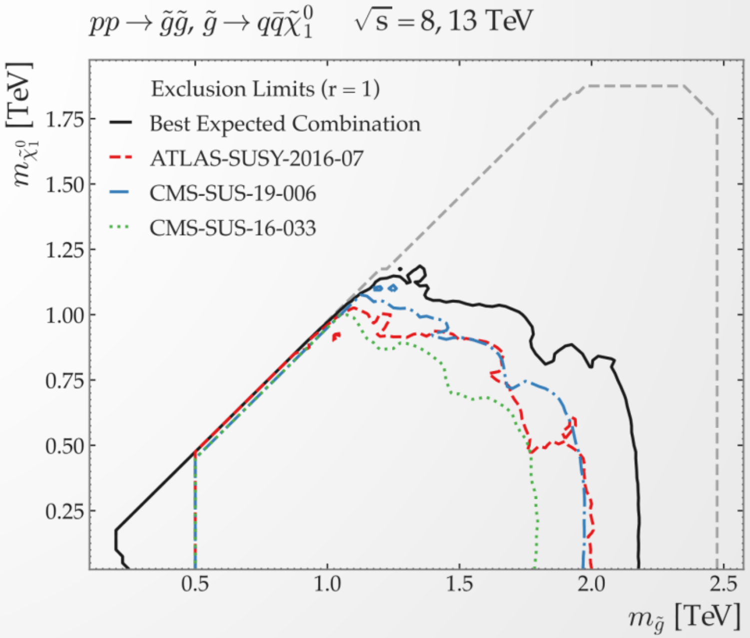
W. Waltenberger,
S. Narasimha

Best combination based on expected exclusion reach



- can now **combine** several **results** to a single, more constraining result
- often **results in ~ 200 GeV gain** in terms of exclusion
  - some **approximations** necessary

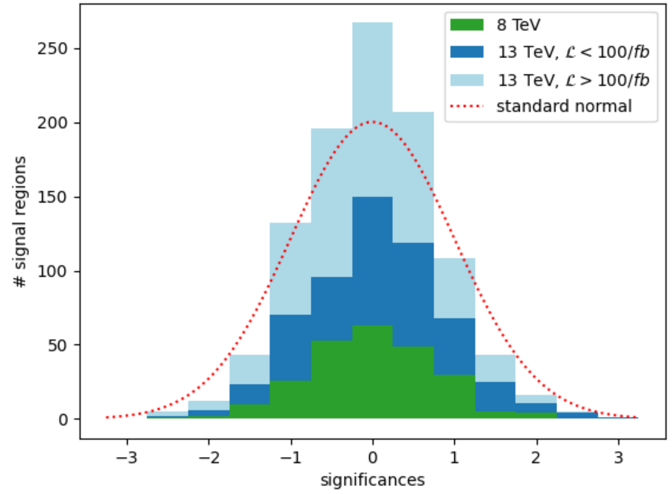
Observed exclusion (new result)



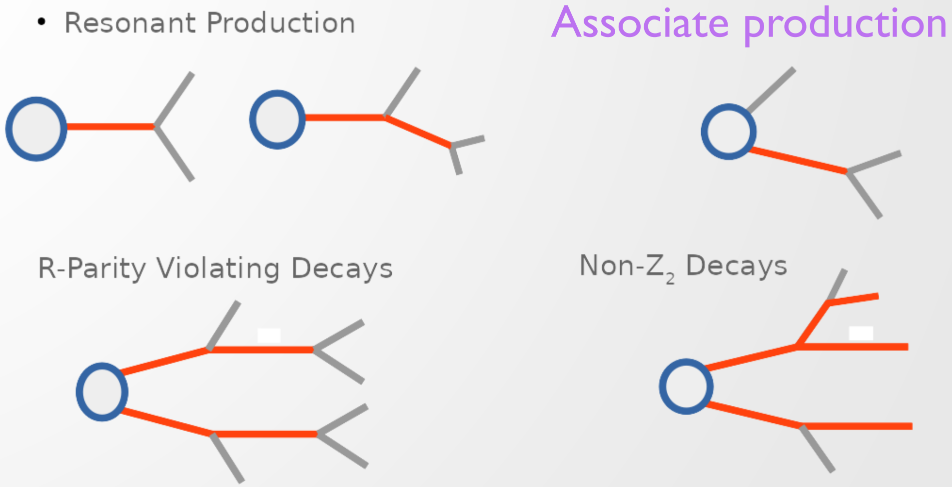
META-ANALYSIS & PLANS

- Meta-Analysis of all SRs across all analyses
  - Standard Model hypothesis: significances ~ N(0,1)
  - holds true to a remarkable degree
    - no reproducibility crisis in LHC physics!

LHC posterior (meta analyses)



- next update: SModelS v3
  - > 1000 signal regions from > 100 publications
  - going beyond SUSY-like topologies
  - Cover new topologies with more general, graph-like topologies
  - SModelS v3 will cover most of the amenable theory landscape

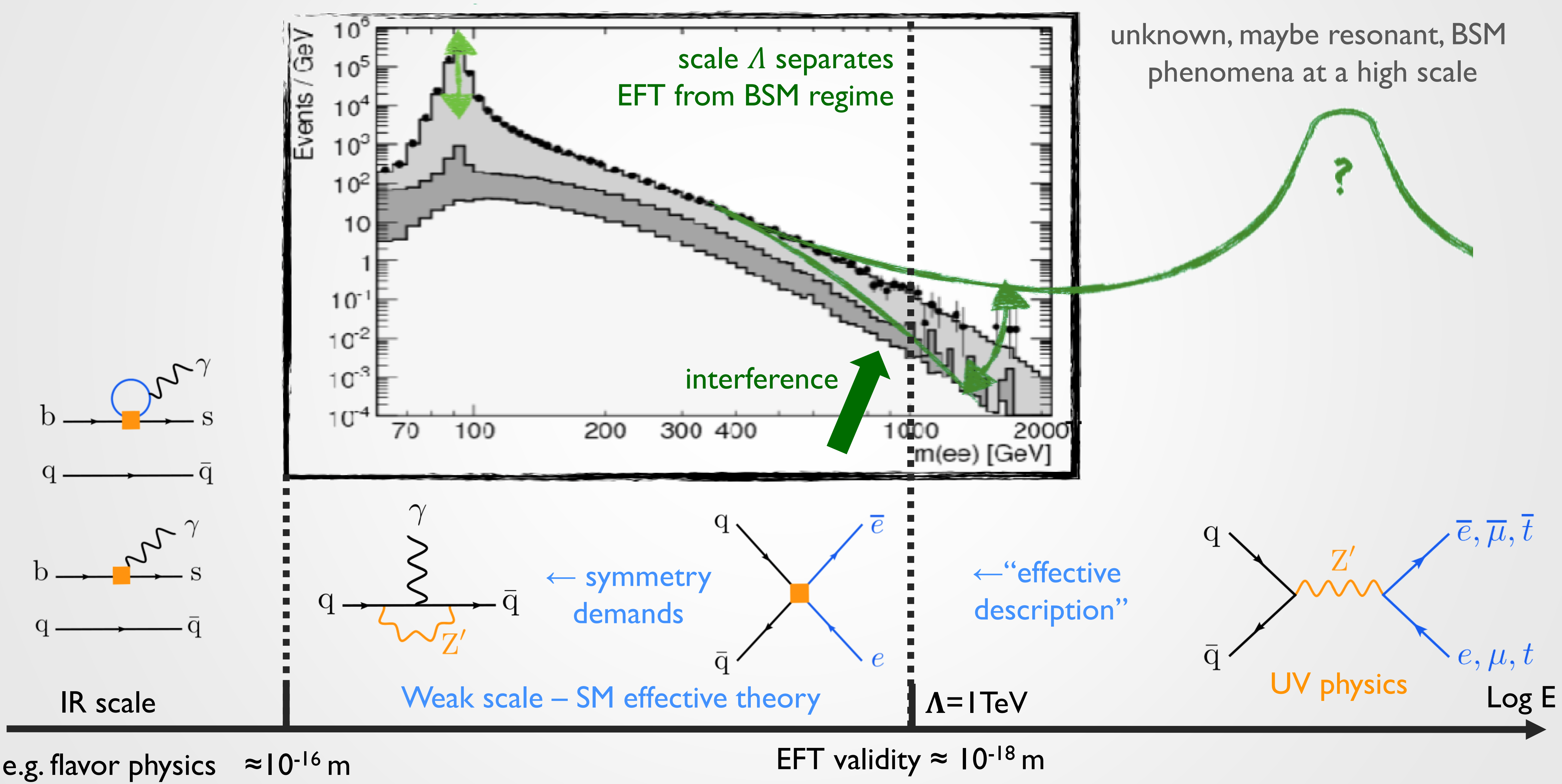


MEASUREMENTS,
EFFECTIVE INTERACTIONS & ML

CATCHING NEW PHYSICS BY THE TAIL

Sketch from F. Riva

Catchphrase from [PRL 120, 101801(2018)]



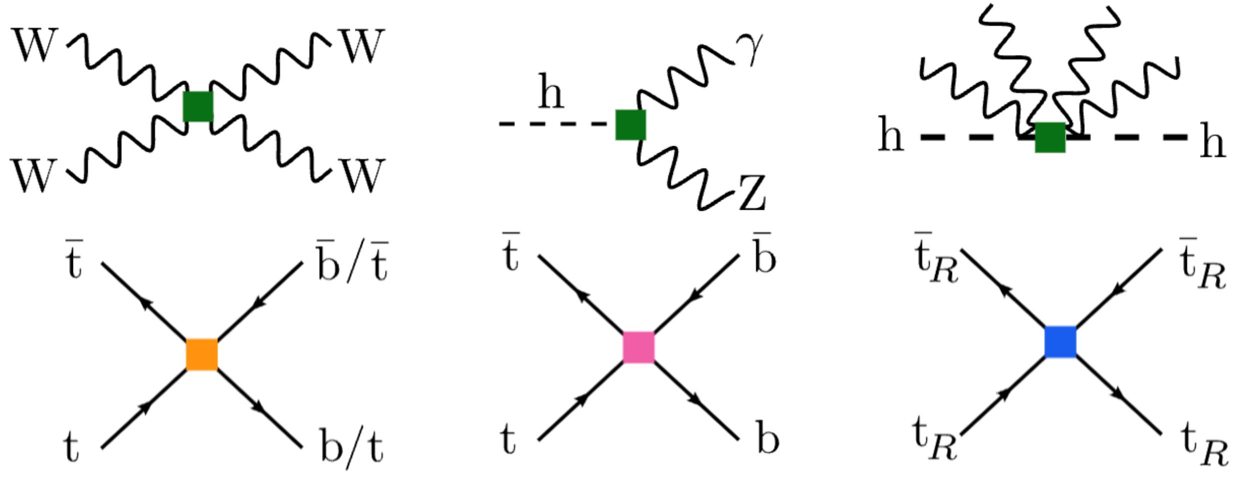
THE STANDARD MODEL EFFECTIVE FIELD THEORY

- organizing principle: mass dimension

$$\mathcal{L}_{eff} = \mathcal{L}_{SM}^{(4)} + \sum \frac{C_x}{\Lambda^2} O_{6,x} + h.c.$$

- Keep SM symmetries
  - $SU(3)_c \otimes SU(2)_L \otimes U(1)$
  - 59 operators at d=6 [JHEP10(2010)085]
- operators affect all SM predictions

Anomalous couplings & new interactions (tiny selection!)



$$\frac{C_{\phi W}}{\Lambda^2} \left(\phi^\dagger \phi\right) W_I^{\mu\nu} W_{\mu\nu}^I$$

$$\frac{C_{qq}^{(8)}}{\Lambda^2} (\bar{q}\, \gamma^\mu T^A\, q)\, (\bar{q}\, \gamma_\mu T^A\, q)$$

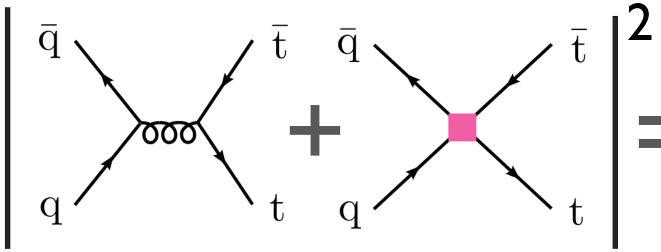
$$\frac{C_{qq}^{(3)}}{\Lambda^2} (\bar{q}\, \gamma^\mu \tau^I\, q)\, (\bar{q}\, \gamma_\mu \tau^I\, q)$$

unknown coefficients

known SM particles

known SM symmetries

- We predict rates from "squared" diagrams:

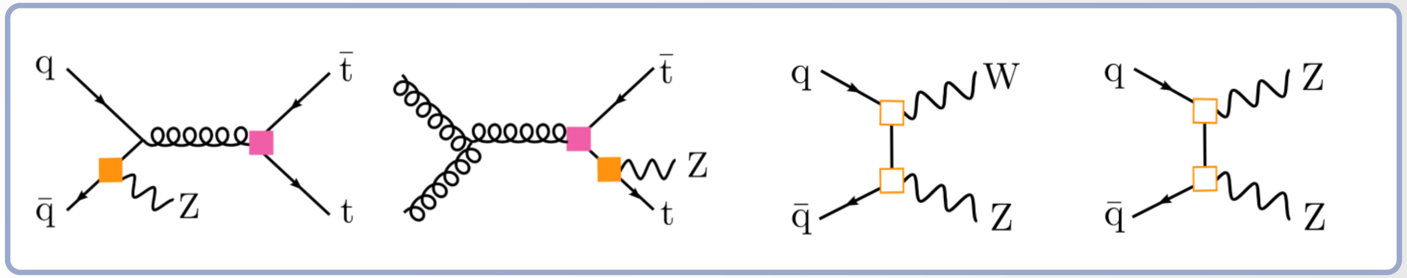


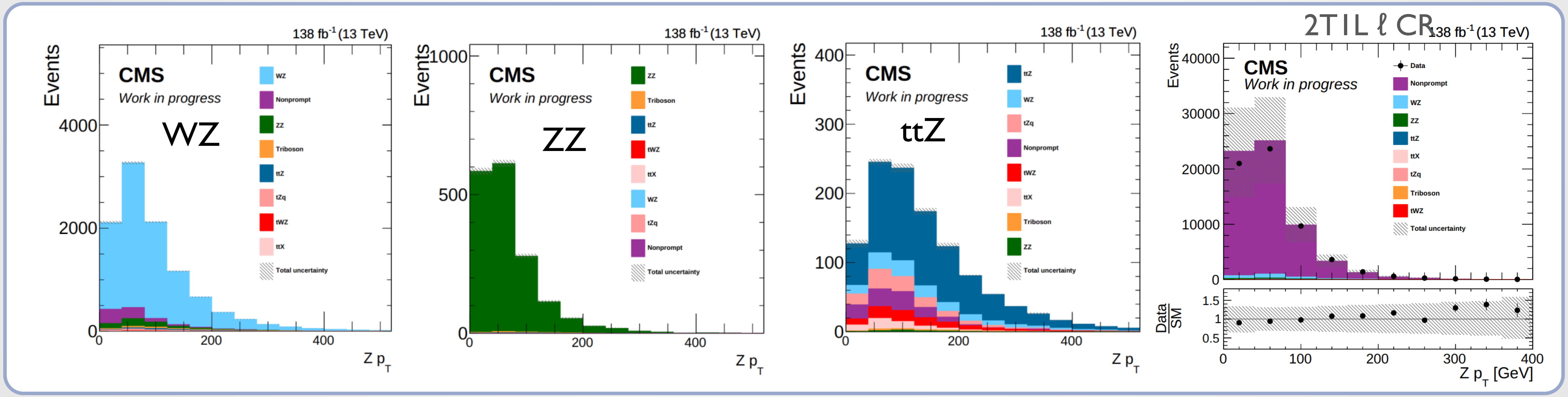
$$= \sigma^{SM} + \frac{C}{\Lambda^2}\sigma^{int} + \frac{C^2}{\Lambda^4}\sigma^{quad}$$

- Quite exceptional simplification!

EFT FLAVOR STRUCTURE OF THE Z COUPLING

- Measure Z-quark coupling in multilepton (3 or 4) final states
- Resolve BSM flavor structure
  - 3rd generation in ttZ
  - 1st+2nd generation in WZ/ZZ/ttZ
- WZ and ttZ are mutual backgrounds
- flavor dependent EFT effects simultaneously affect all processes: extract simultaneously

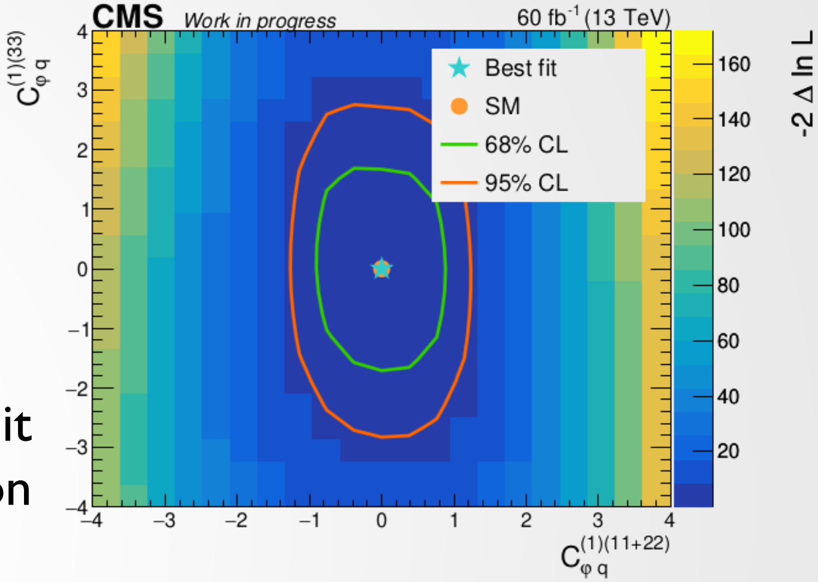




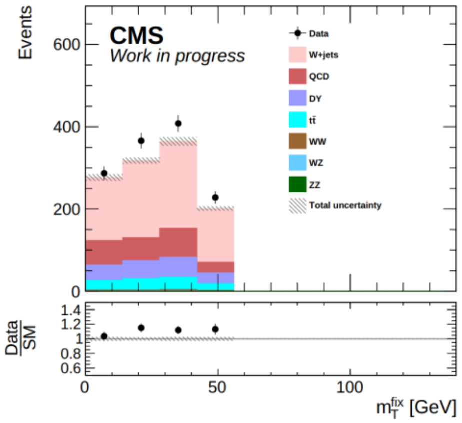
EFT FLAVOR STRUCTURE OF THE Z COUPLING: EXP. LIMITS

- Preliminary expected results with partial systematic uncertainties
  - Focus on $O^{(1)}_{\phi q}$ and $O^{(3)}_{\phi q}$ operators (vector coupling)
  - $1^{st}+2^{nd}$ generation constrained by ZZ, WZ and ttZ (ISR) ⟶ x-axis
  - 3^{rd} generation constrained by ttZ ⟶ y-axis
- Lepton fake rate measured in dedicated single lepton region
  - Differential in lepton p_T, η, flavour
- Target pre-approval this year
- General issue in SM-EFT searches: Optimal test statistic exploiting the simple analytic structure?

Limit combination

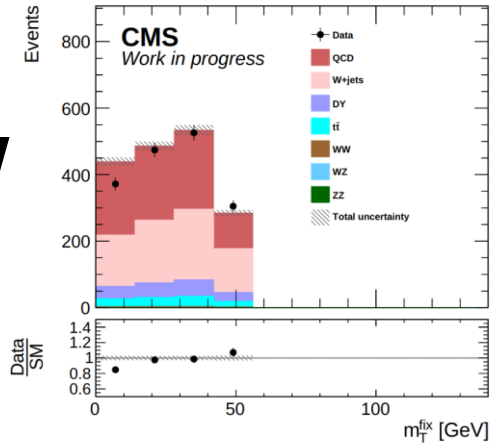


Tight



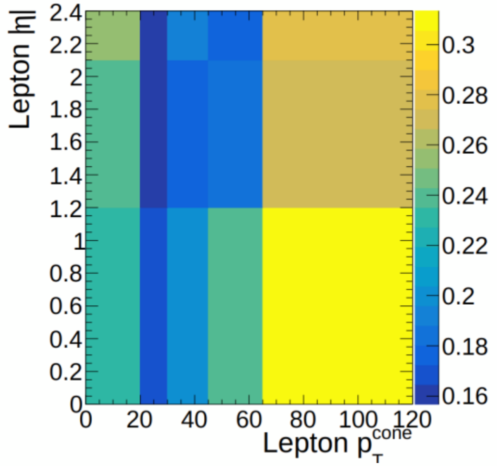
/

Loose

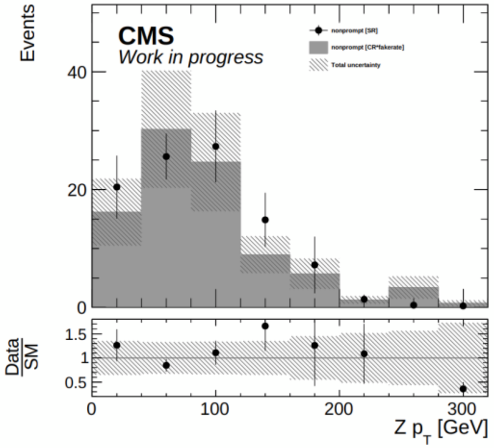


=

Fake rate map

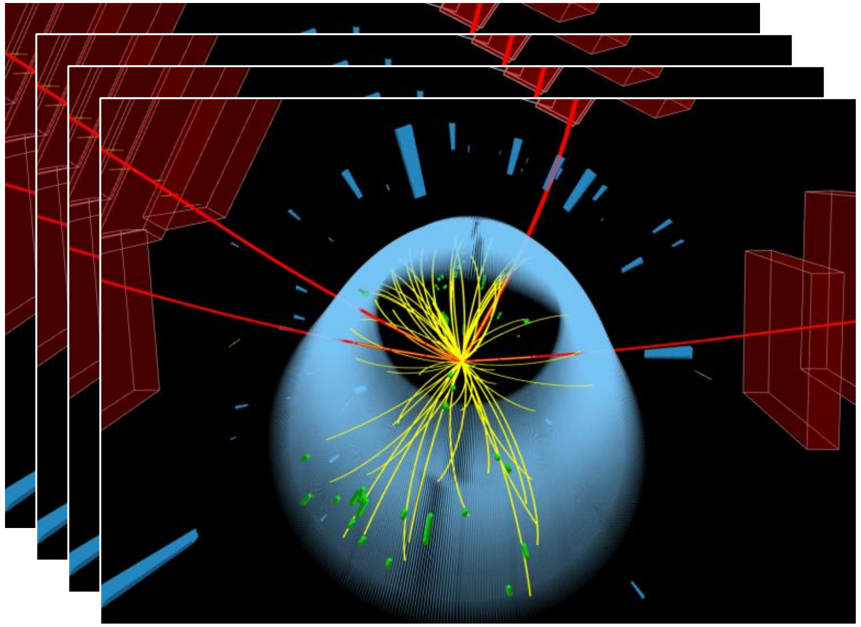


Closure



NEYMAN-PEARSON & LIKELIHOOD RATIO "TRICK"

arxiv:1503.0x7622



Neyman-Pearson Lemma: The *likelihood* ***ratio*** *test statistic is optimal*

data-set with feature vectors **x**

diff xsec ratio

$$q_{\boldsymbol{\theta}}(\mathcal{D}) \sim -\sum_{\boldsymbol{x}_i \in \mathcal{D}} \log \frac{\sigma(\boldsymbol{\theta})}{\sigma(\mathrm{SM})} \frac{p(\boldsymbol{x}_i|\boldsymbol{\theta})}{p(\boldsymbol{x}_i|\mathrm{SM})}$$

theory parameters

Likelihood ratio "trick": label two values: $\boldsymbol{\theta}$, SM

$$L = \int \mathrm{d}\boldsymbol{x} \sum_{z\in\{0,1\}} p(\boldsymbol{x}, z) \left(z - \hat{f}(\boldsymbol{x})\right)^2$$

training samples — truth (supervised) — classifier

$$f^*(\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \mathrm{SM})}{p(\boldsymbol{x}, \mathrm{SM}) + p(\boldsymbol{x}, \boldsymbol{\theta})} = \frac{1}{1 + \frac{\sigma(\boldsymbol{\theta})}{\sigma(\mathrm{SM})} r(\boldsymbol{x})}$$

supervised learning provides (close-to) optimal test statisticsc

What to do with the parameter dependence?

PARAMETRIZED CLASSIFIERS: NETS & TREES

Wulzer et.al. [JHEP 05 (2021) 247]

RS et. al., [2107.10859], [2205.12976]

$$L = \sum_{\boldsymbol{\theta}\in\mathcal{B}} \int \mathrm{d}\boldsymbol{x} \left(p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\theta})\hat{f}(\boldsymbol{x};\boldsymbol{\theta})^2 + p(\boldsymbol{x},\boldsymbol{z}|\mathrm{SM})(1-\hat{f}(\boldsymbol{x};\boldsymbol{\theta}))^2 \right)$$

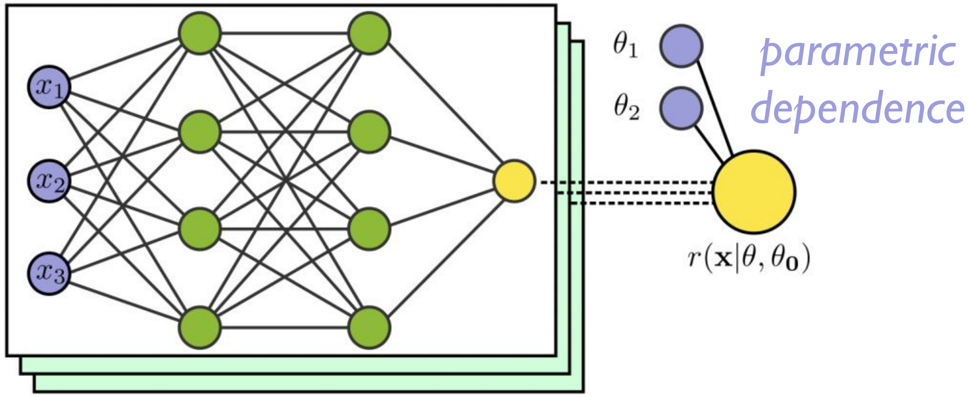
Make predictor aware of analytic SMEFT structure

Invert likelihood trick
with positive polynomial of NN -outputs

$$\hat{f}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{1+\hat{r}(\boldsymbol{x};\boldsymbol{\theta})}$$

$$\hat{r}(\boldsymbol{x};\boldsymbol{\theta}) = \left(1+\sum_a \theta_a \hat{n}_a(\boldsymbol{x})\right)^2 + \sum_a \left(\sum_{b\geq a} \theta_b \hat{n}_{ab}(\boldsymbol{x})\right)^2$$

Fit NNs simultaneously



$$L = \sum_{\boldsymbol{\theta}\in\mathcal{B}} \int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x},\boldsymbol{z}|\mathrm{SM}) \left(r(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\theta},\mathrm{SM}) - \hat{F}(\boldsymbol{x},\boldsymbol{\theta}) \right)^2$$

Tree ansatz with polynomial SMEFT dependence

$$\hat{F}(\boldsymbol{x},\boldsymbol{\theta}) = \sum_{j\in\mathcal{J}} \mathbb{1}_j(\boldsymbol{x}) F_j(\boldsymbol{\theta})$$

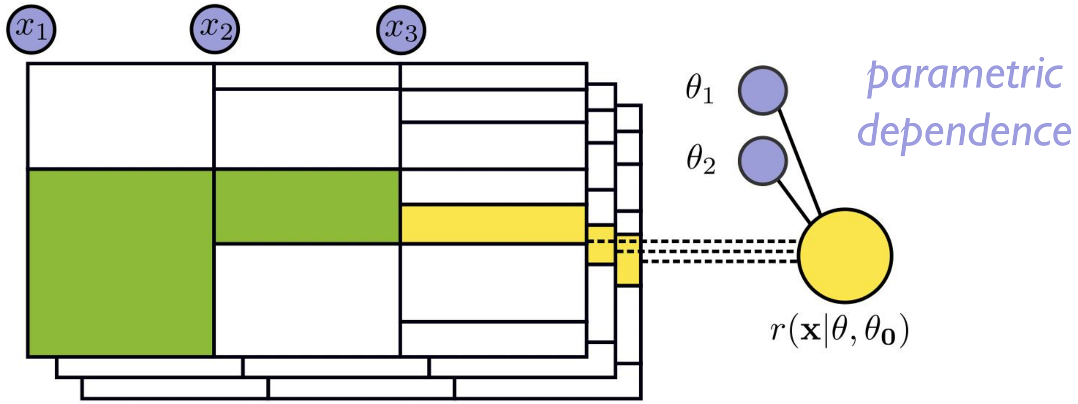
Can solve for trainable parameters of the predictor
→ *Large training speedup*

$$F_j(\boldsymbol{\theta}) = \frac{\sum_{i\in j} w_i(\boldsymbol{\theta})}{\sum_{i\in j} w_i(\boldsymbol{\theta}_0)} \equiv \frac{w_j(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

Obtain loss function for optimal partitioning, solved by e.g. CART algorithm → Boost

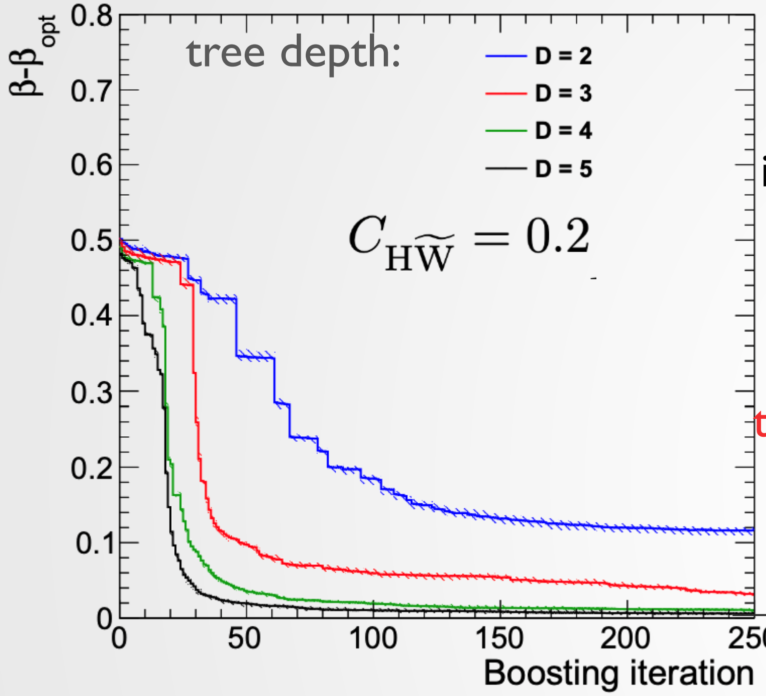
$$L = -\sum_{\boldsymbol{\theta}\in\mathcal{B}} \sum_{j\in\mathcal{J}} \frac{w_j^2(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

linear truncation: optimize Fisher information



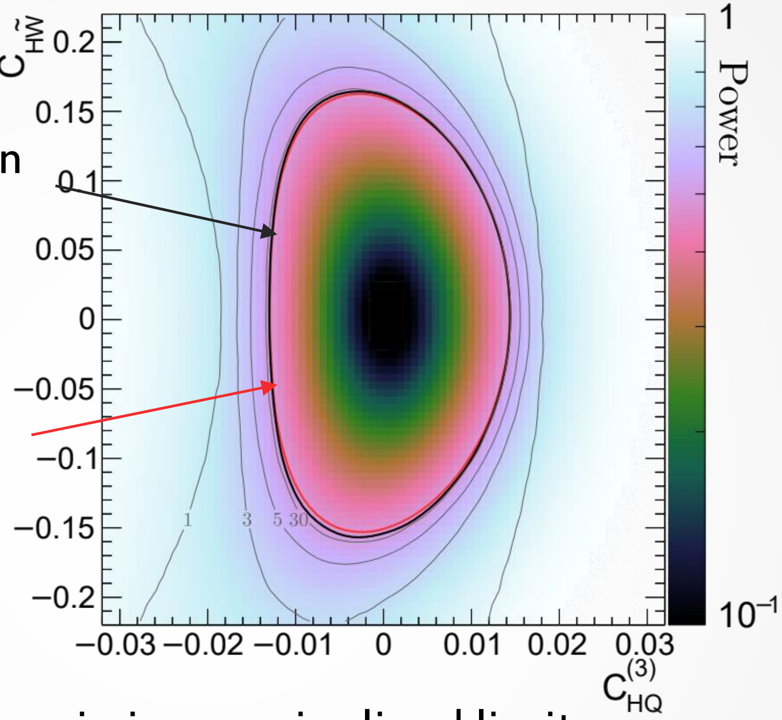
OPTIMALITY IN TEST CASES

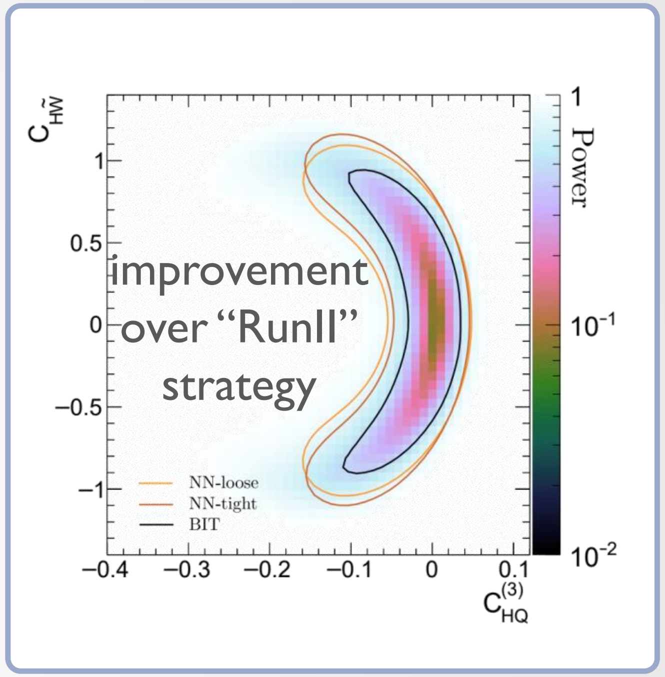
[arXiv:2107.10859, arXiv:2205:12976]



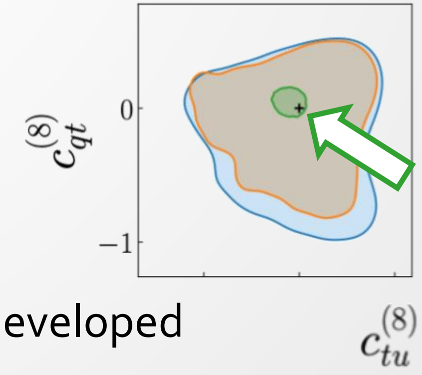
Boosted information tree (BIT)

theoretical optimum





- 20-40% improvements in 2D toy cases; more gain in marginalized limits
- Explored by: VH analysis with ETH, tW scattering (Boston)
- No free lunch – Analysis dependent choices are needed
  - A case-by-case compromise if background estimation is CPU intensive
  - Systematics treatment for unbinned analyses (beyond Higgs $M_{4\ell}$) less far developed
  - Simulation: PhD student with CERN-IT & Olivier Mattelaer (Louvain) on GPU for MG



[ML4EFT, Rojo et.al]

Binned $(p_T^{\ell\bar{\ell}}, \eta_\ell)$

Unbinned ML $(p_T^{\ell\bar{\ell}}, \eta_\ell)$

Unbinned ML (18 features)

SM

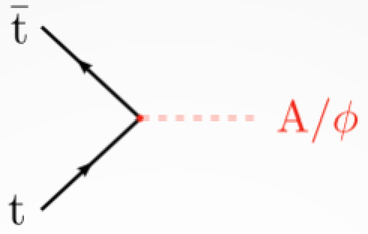
Large gain for constraining forces among top quarks

TTTT: NEW FORCES BETWEEN HEAVY QUARKS?

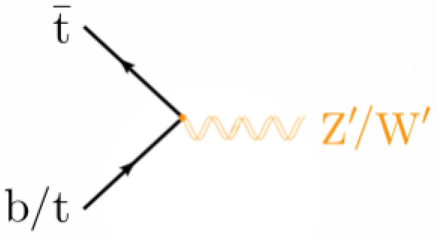
C. Giordano, M. Shooshtari
L.Wild, RS

- Hypothetical new models

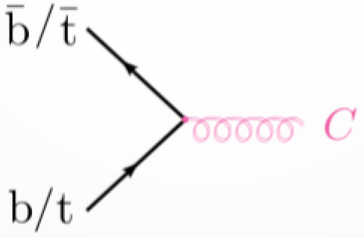
- Extended Higgs sectors “two Higgs doublet models” from SUSY or other BSM physics [review]



- High-mass force carriers similar to the W and Z bosons : Z’ and W’ bosons [review]



- Massive ”chiral” colored force carriers, otherwise similar to the gluon: axigluons [Mimasu et.al.]

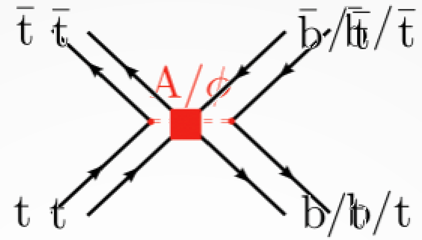


- Composite sector whose bound states mix with the SM particles: (right-handed) top-quark and/or Higgs compositness [review]

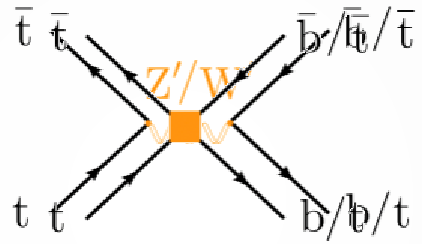
C. Giordano, M. Shooshtari
L.Wild, RS

TTTT: NEW FORCES BETWEEN HEAVY QUARKS?

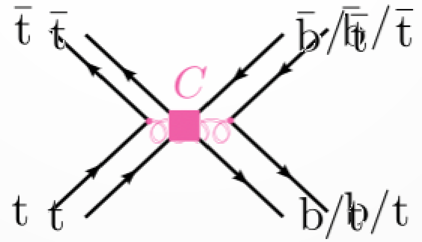
- Extended Higgs sectors "two Higgs doublet models" from SUSY or other BSM physics [review]



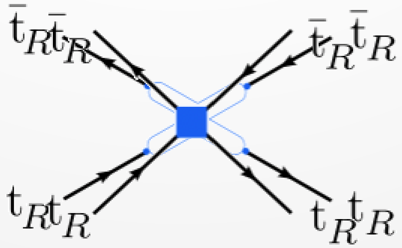
- High-mass force carriers similar to the W and Z bosons : Z' and W' bosons [review]



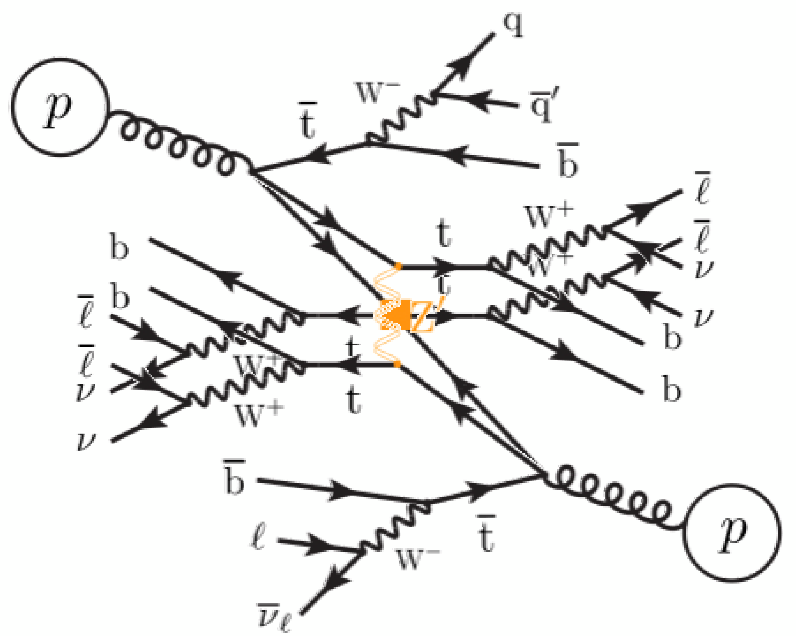
- Massive "chiral" colored force carriers, otherwise similar to the gluon: axigluons [Mimasu et.al.]



- Composite sector whose bound states mix with the SM particles: (right-handed) top-quark and/or Higgs compositness [review]



- Hypothetical new models
  - predict force-carrier exchange
  - modify predictions for LHC processes
  - described by "effective theory"



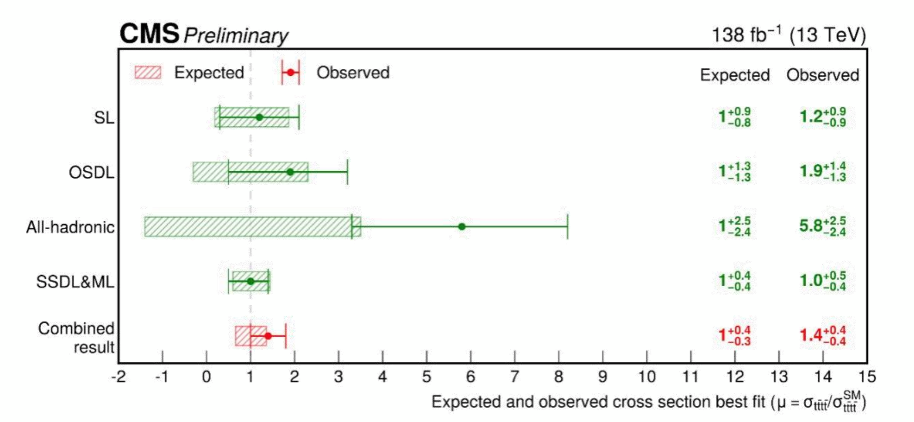
- Combine t vs. t & t vs. b & t vs. light quarks

[EPJC 80 (2020) 1085, arXiv:2211.01136]

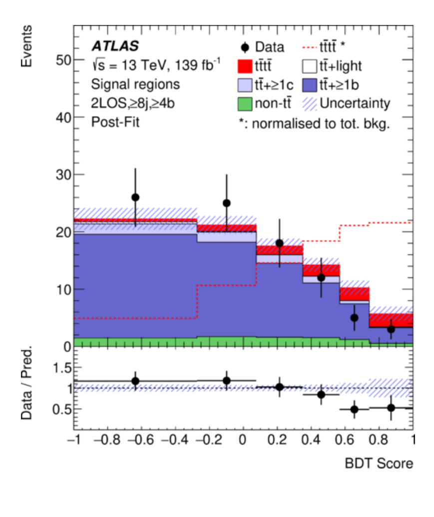
FOUR-TOP QUARK PRODUCTION: STATUS

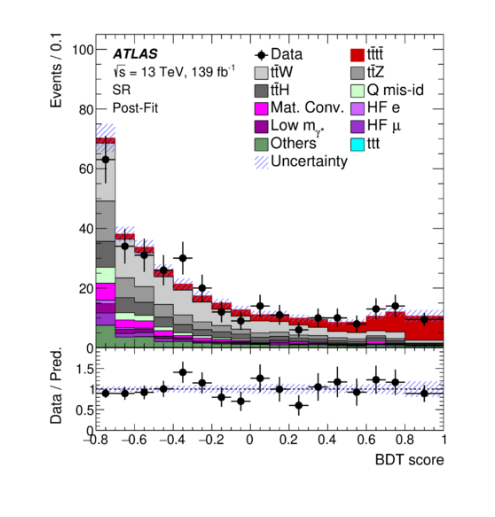
[TOP-21-005, EPJC 80 (2020) 75]

- Extracting inclusive cross section (SM: 12.2 ± 2.2 (scale) fb)
  - CMS: **μ(tttt) = 1.4 ± 0.4**, **4σ** above zero, limited by stat
  - ATLAS: 24^{+7}_{-6} fb: 4.7**σ** close to "discovery"
  - factor **μ(tttt) = 2.0 (+0.8-0.6)** high, but within 2**σ** of the SM prediction

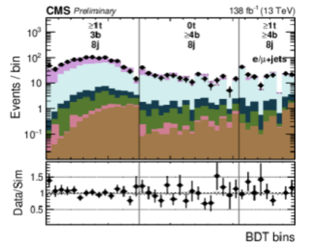
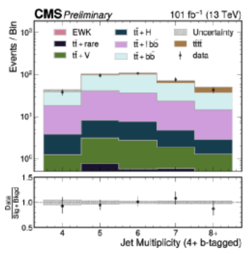
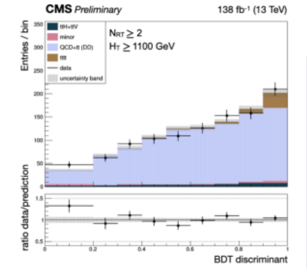


ATLAS	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+
		JHEP 11 (2021) 118		EPJC 80 (2020) 1085	
2018					
2017					
2016					

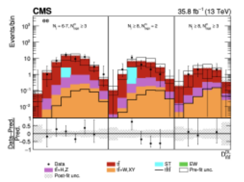
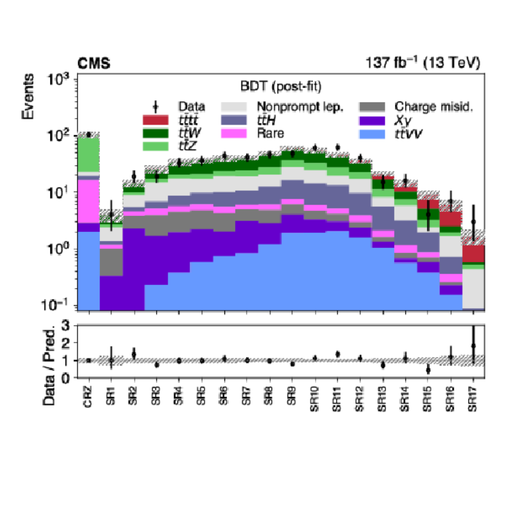




CMS	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+
	TOP-21-005			EPJC 80 (2020) 75	
2018					
2017					
2016					

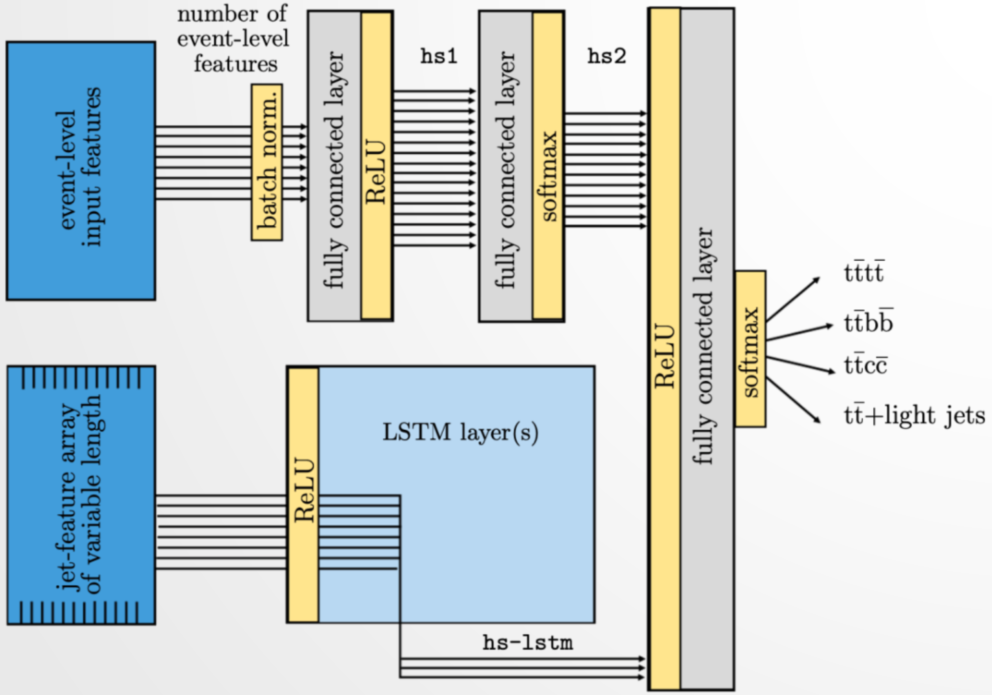
JHEP 11 (2019) 082

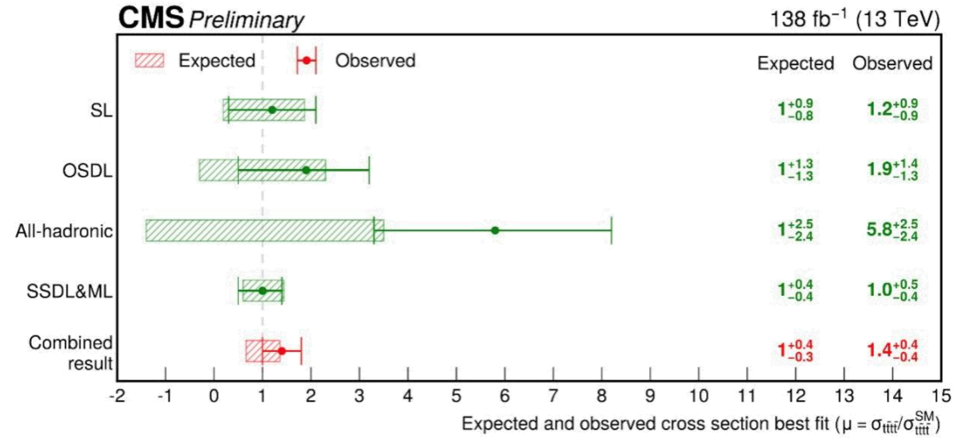



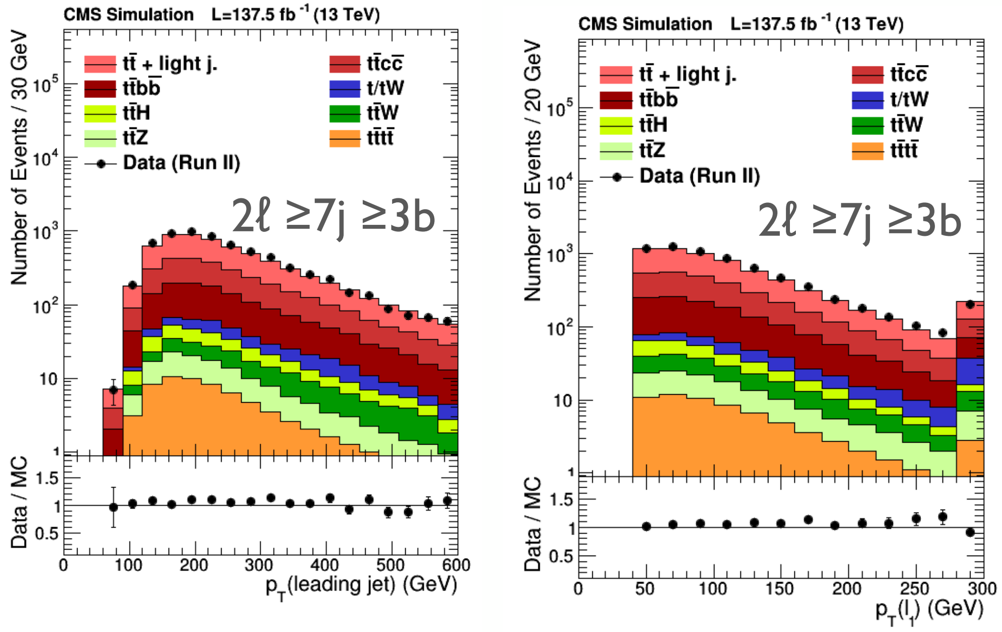
FOUR-TOP QUARK PRODUCTION: PROJECT STATUS

C. Giordano, M. Shooshtari
L.Wild, RS

- New project started with UGhent and UNIGE (Theory)
  - ML-assisted analysis strategy, combining
    - 1ℓ, 2ℓOS, 2ℓSS, 3ℓ, 4ℓ
  - status: systematics, SR definition, EFT sensitivity
  - Developed multi-classification with RNNs on jet system: +20% efficiency







TOP QUARK PAIR + HEAVY FLAVOR

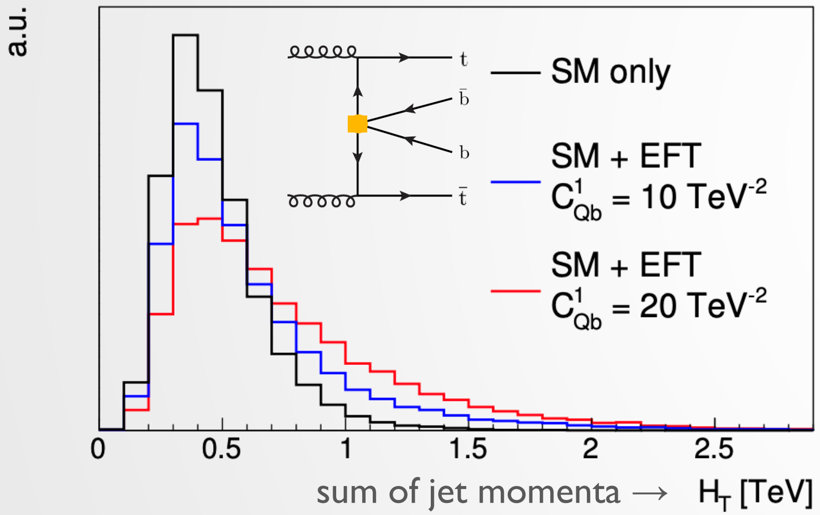
[JHEP 07 (2020) 125]
[PLB (2020) 135285]

C. Giordano, M. Shooshtari
L.Wild, RS

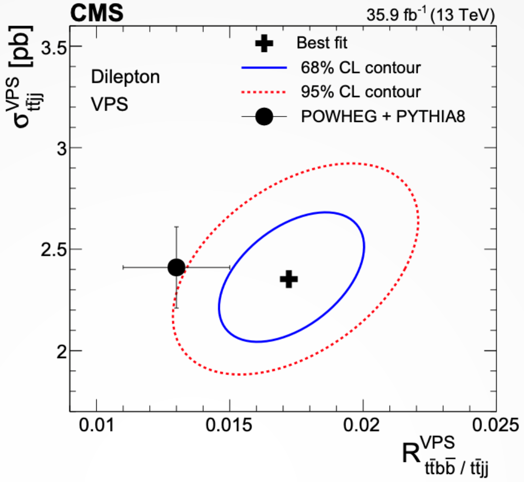
[JHEP 04 (2019) 046]

- An example of how EFT shapes interest: So far, tt+bb studied mostly in "generator-tuning" context and as bkg

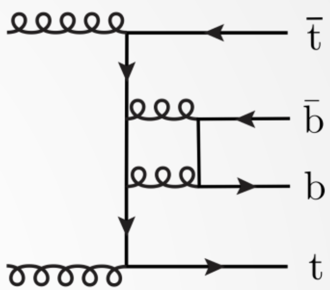
[Mimasu et.al. JHEP 11 (2018) 131]



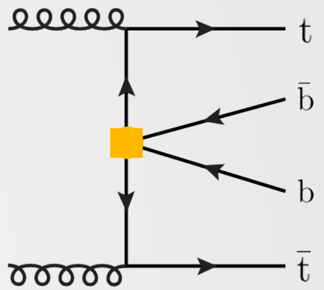
- complementary EFT information to 4-top
  - parametrized EFT classifiers with (never tried)



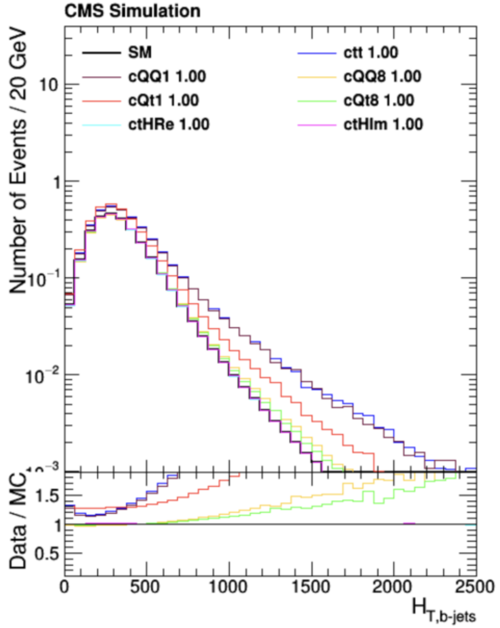
difficult modelling of SM bb production



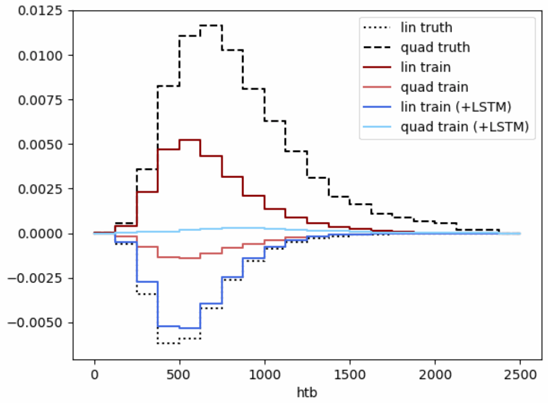
hypothetical EFT ttbb interactions



EFT effects in TTTT



Closure test of likelihood-free inference of parametrized classifiers

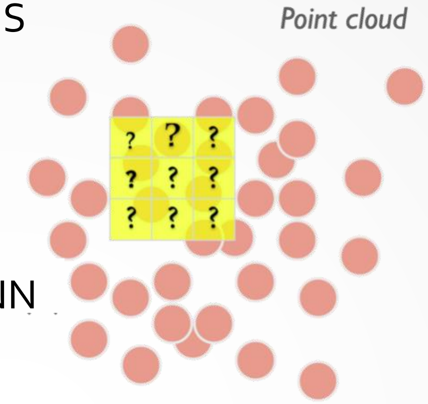


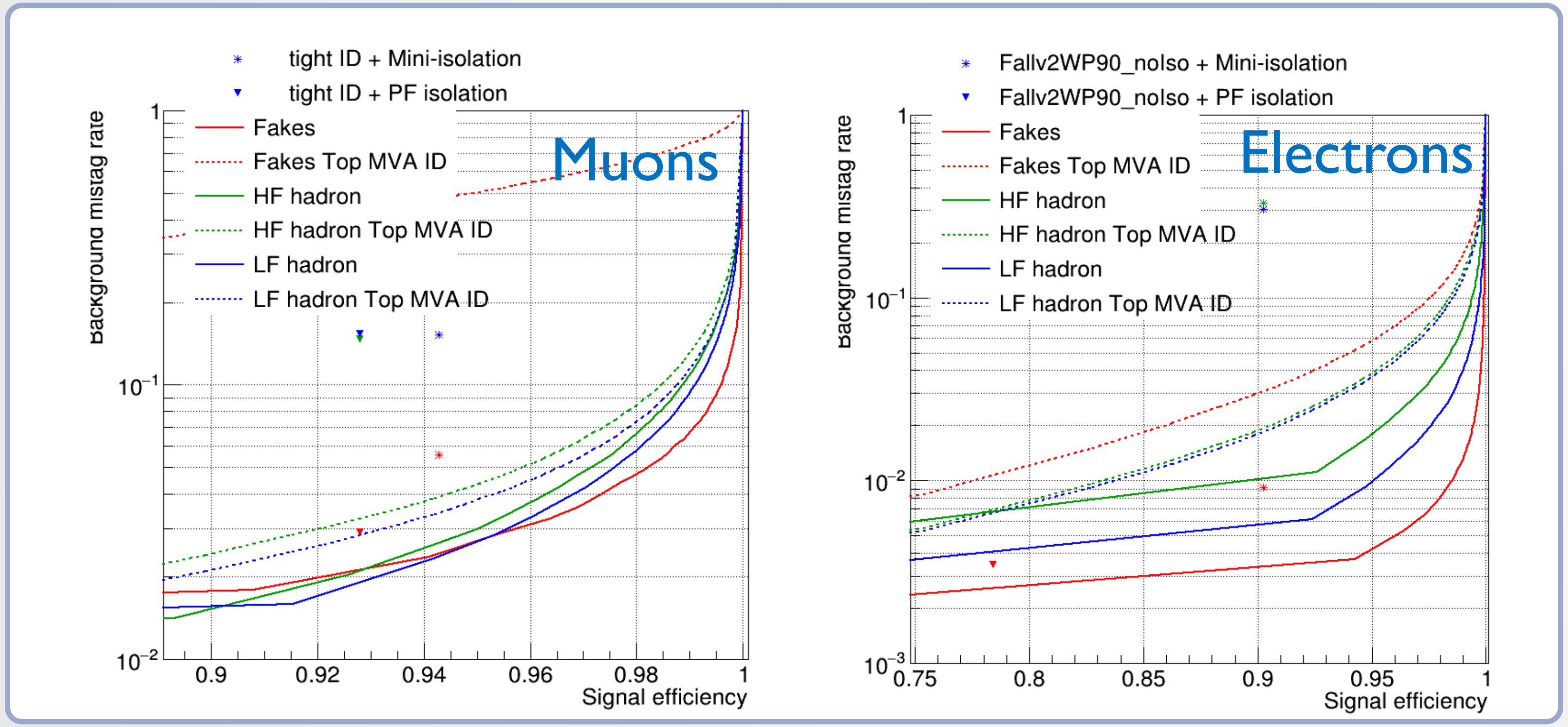
Target BSM characterization of TTbb
Started collaboration with KIT

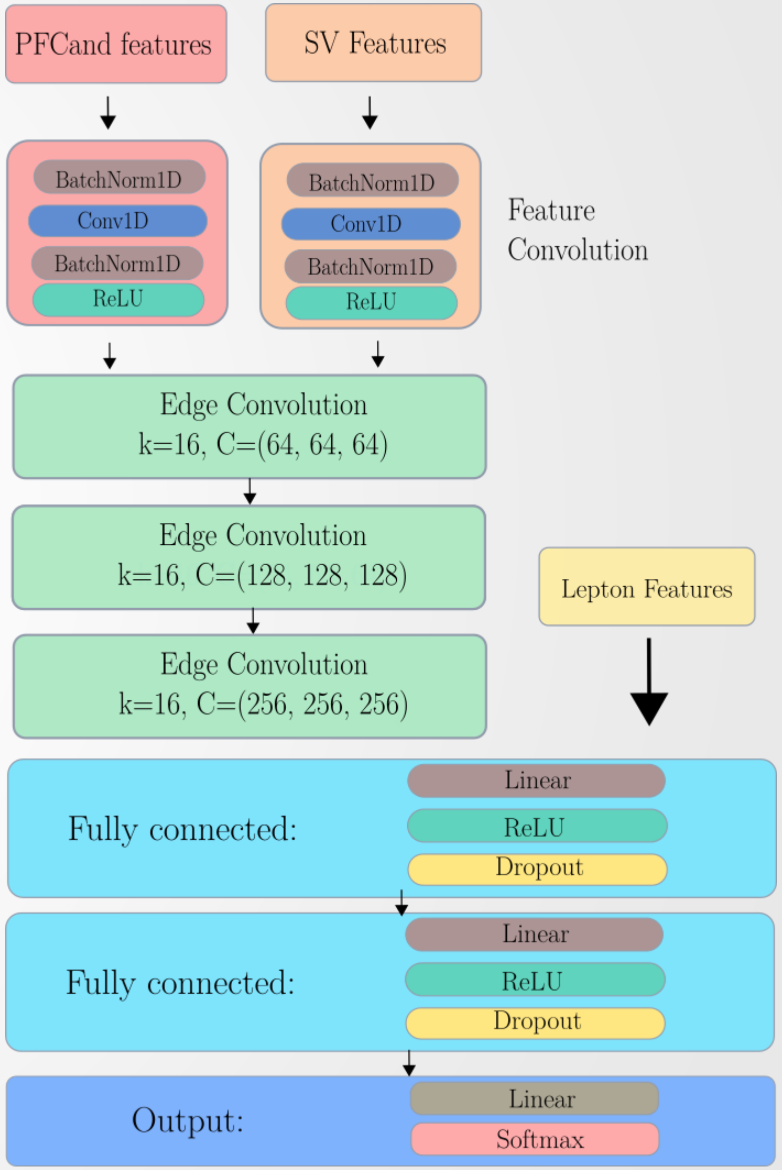
GRAPH-NN FOR LEPTON ID: PARTICLENET

S. Chatterjee, A. Gruber

- Adaptation of ParticleNet algorithm used for jet tagging in CMS
  [Y. Wang et al. (2018) L. Gouskos, H. Qu (2019)]
  - Fully exploit data on activity around the lepton candidate
  - Treat particles around lepton as point cloud: unordered, permutation-invariant set of particles
  - Learn particle correlation with EdgeConv operation & DNN
- Prompt, heavy-flavor, light-flavor, fake categories





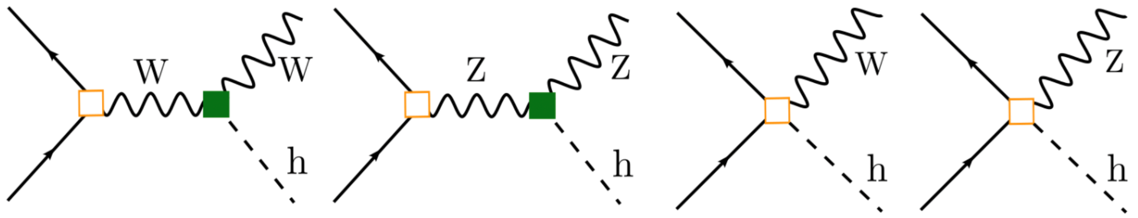


SMEFT ANALYSIS OF HIGGS-STRAHLUNG

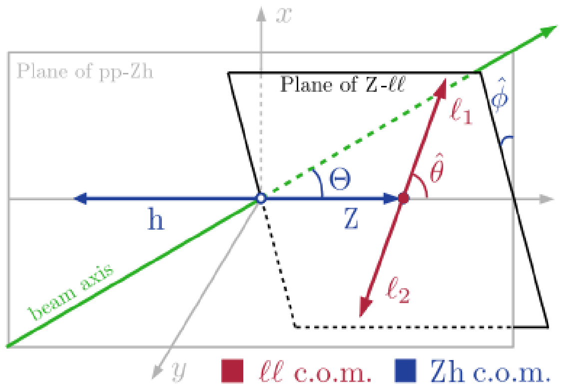
S. Chatterjee
Collaboration with ETH Zurich

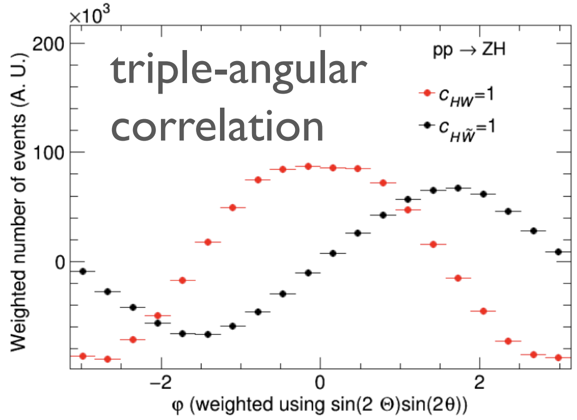
- Probing eight SMEFT operators in ZH and WH production
  - Final state: 1ℓ/2ℓ + H→bb
  - boosted & resolved
  - H→bb: Mass-decorr. ParticleNet tagger + M_{SD}
- In collaboration with ETH team

1. Exploit energy growth of 4-point functions
   - Unique sensitivity to vector current couplings
2. Interference resurrection
   [Spannowsky, JHEP 09 (2020)170]
   - Recover sensitivity from full angular analysis
     - triple-variable correlations boost sensitivity
     - recover CP structure of BSM couplings



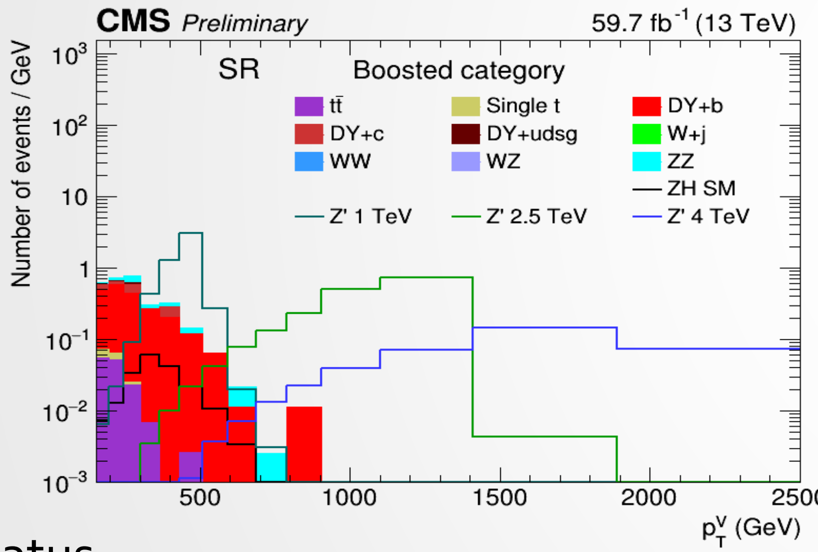
vector coupling modifications		aTGC/aQGC	
$\mathcal{O}_{Hq}^{(1)}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{q}\gamma^\mu q$	$\mathcal{O}_{HWB}$	$H^\dagger \sigma^a H W^a_{\mu\nu} B^{\mu\nu}$
$\mathcal{O}_{Hq}^{(3)}$	$iH^\dagger \sigma^a \overleftrightarrow{D}_\mu H \bar{q}\sigma^a\gamma^\mu q$	$\mathcal{O}_{H\widetilde{W}B}$	$H^\dagger \sigma^a H W^a_{\mu\nu} \widetilde{B}^{\mu\nu}$
$\mathcal{O}_{Hu}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{u}_R\gamma^\mu u_R$	$\mathcal{O}_{HW}$	$(H^\dagger H)\, W_{\mu\nu} W^{\mu\nu}$
$\mathcal{O}_{Hd}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{d}_R\gamma^\mu d_R$	$\mathcal{O}_{H\widetilde{W}}$	$(H^\dagger H)\, W^a_{\mu\nu} \widetilde{W}^{a\mu\nu}$

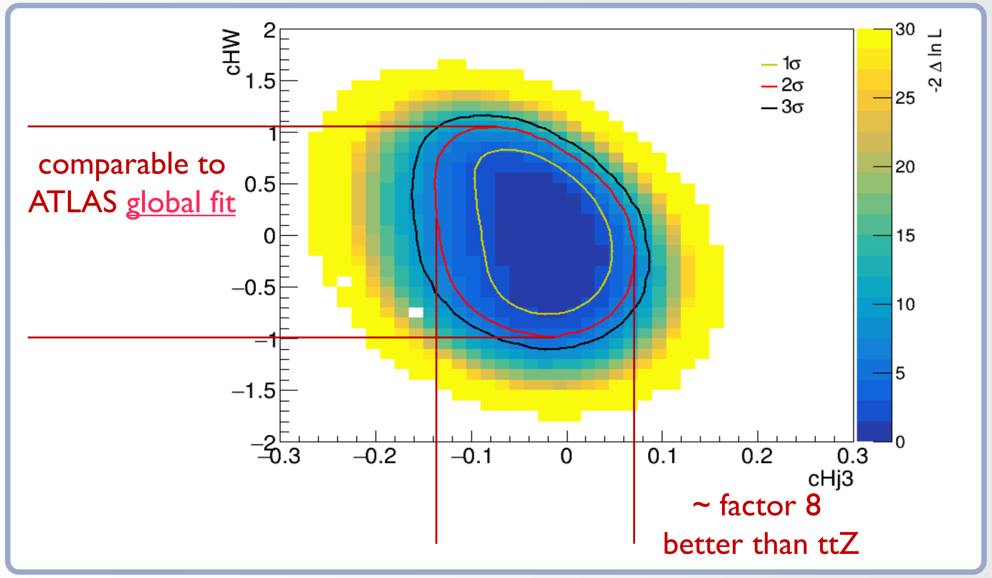


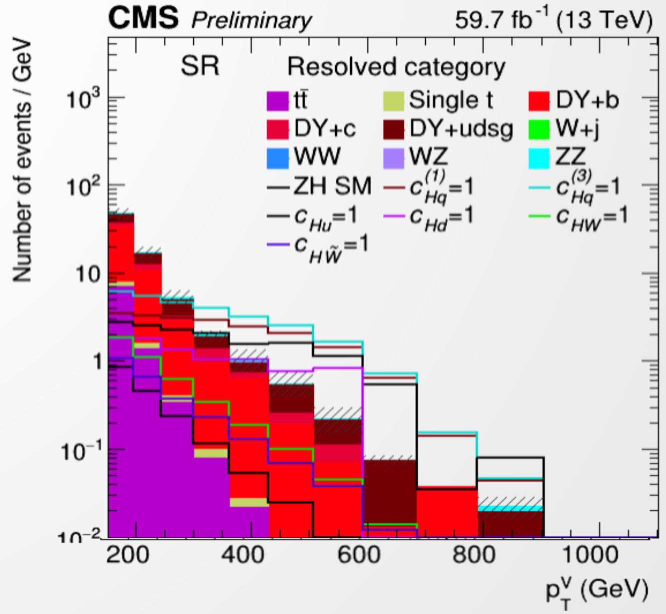


SMEFT ANALYSIS OF HIGGS-STRAHLUNG

- Extract BSM sensitivity with Boosted Information Tree
  [CPC 277(2022)10859, arxiv:2205:12976]
  - 1-D & 2-D constraints on ~8 SMEFT coefficients
  - Also sensitive to UV model (Z' production





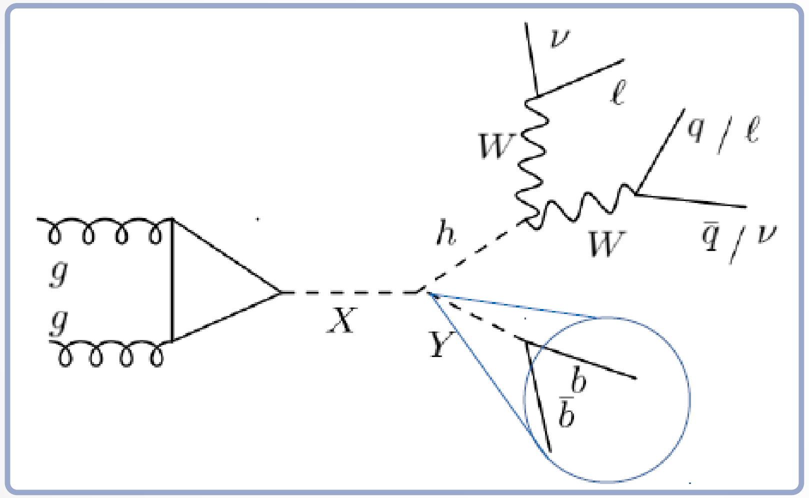


- Status
  - Finalize systematic uncertainty
  - Finalizing extracting of SMEFT constraints
  - FWF proposal: Two rounds with reviewers

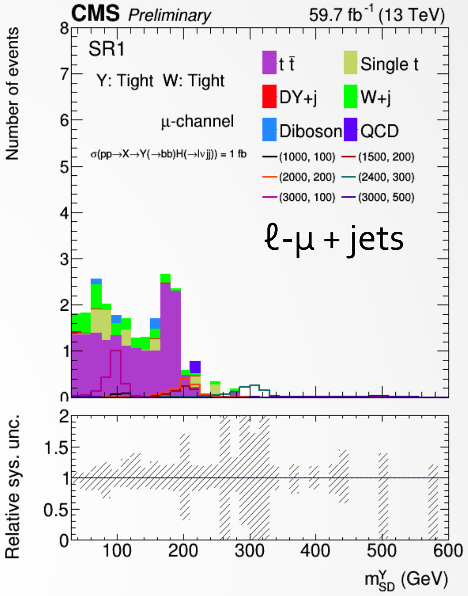
SEARCH FOR X→ Y (→ BB) H (→ WW*)

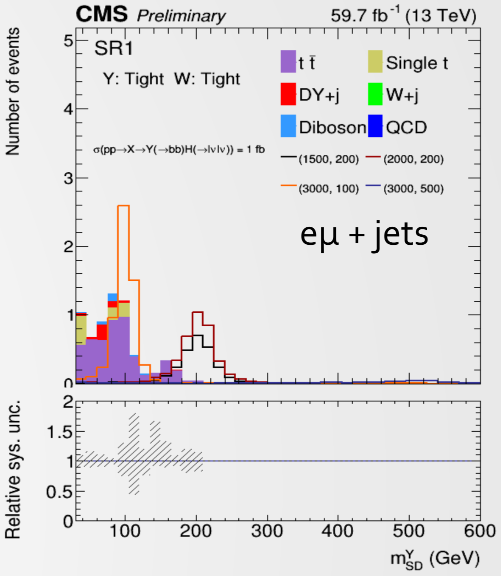
S. Chatterjee, collaboration
with TIFR Mumbai

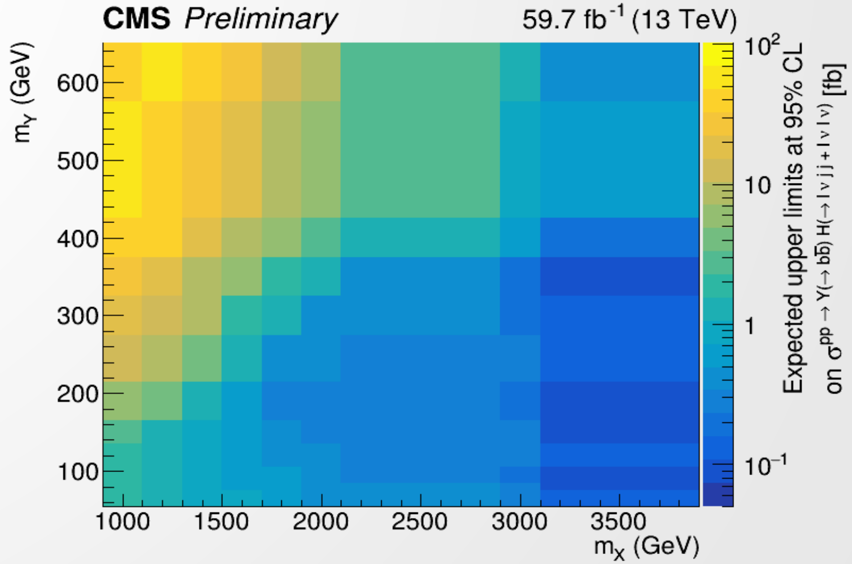
- Resonant search for extended Higgs sectors
  - NMSSM or Two-real-singlet-scalar extension of SM (TRSM)
  - 3 CP-even Higgs bosons - X, Y & SM h
- Final states:
  1ℓ+jets, 2ℓ+jets
- Technology:
  mass-decorrelated
  ParticleNET tagger



- Status:
  - Sensitivity estimation with simulation & relevant systematics
- Work in progress: Background estimation using data



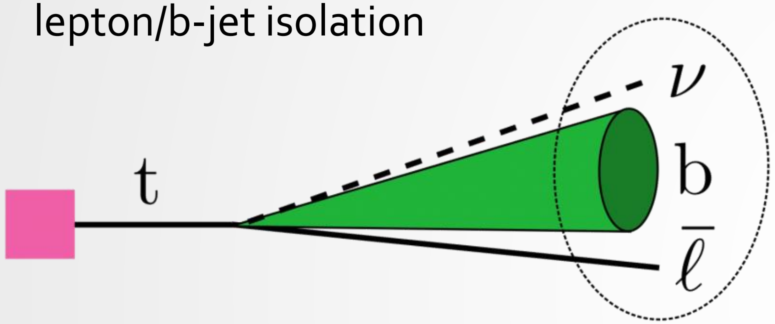




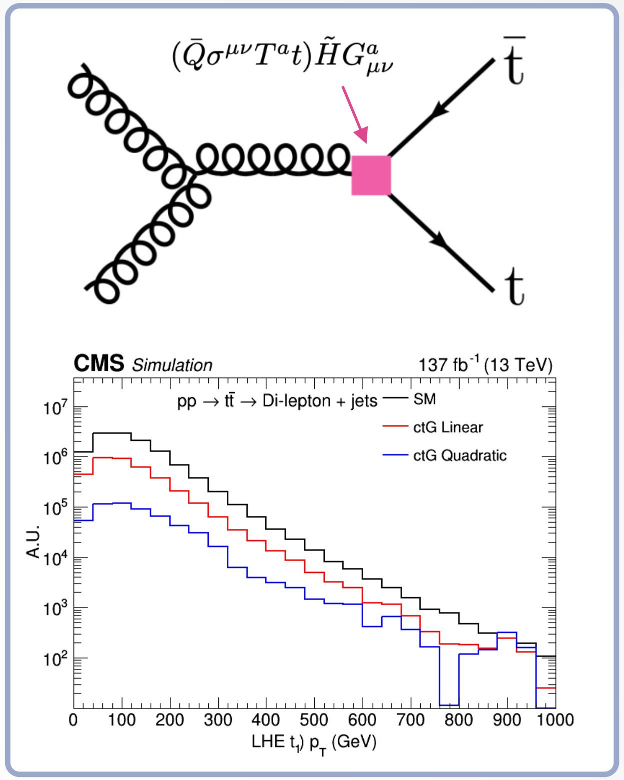
SMEFT ANALYSIS IN DILEPTONIC T-TBAR

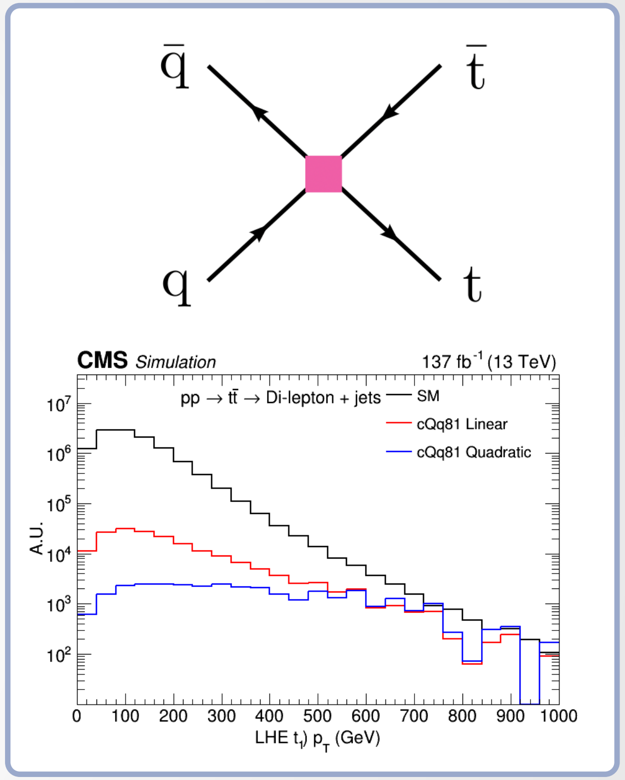
S. Chatterjee, collaboration
with TIFR Mumbai

- Top quark chromomagnetic EFT coupling constrained in spin-correlation analysis
  - Effects grow with energy → sensitivity loss from lepton/b-jet isolation



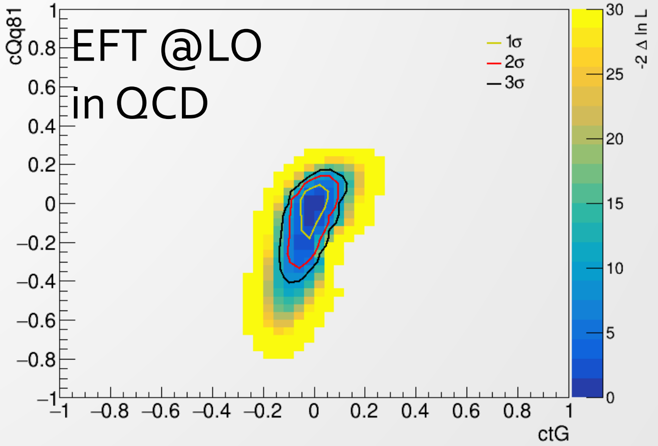
- SMEFT analysis of boosted 2ℓ tt production
  - Extend interpretation: 2-heavy-2-light "forces"
- Dedicated 'lepton in jet' fat-jet tagger
  - S. Chatterjee [JHEP01(2020)170]
- In tandem with SM unfolded x-sec measurement
- Status: Early stage
  - EFT sensitivity at LO and NLO
  - Selection & observable- optimization





recover unique high-p_T sensitivity

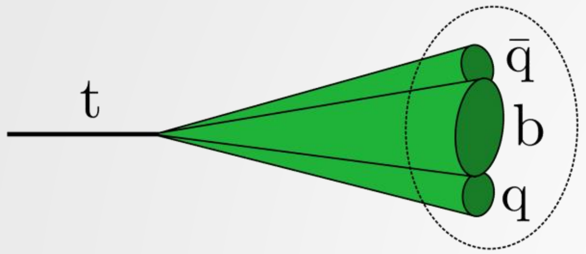
independent from top-A_C measurements



D. Schwarz
with UHH & DESY

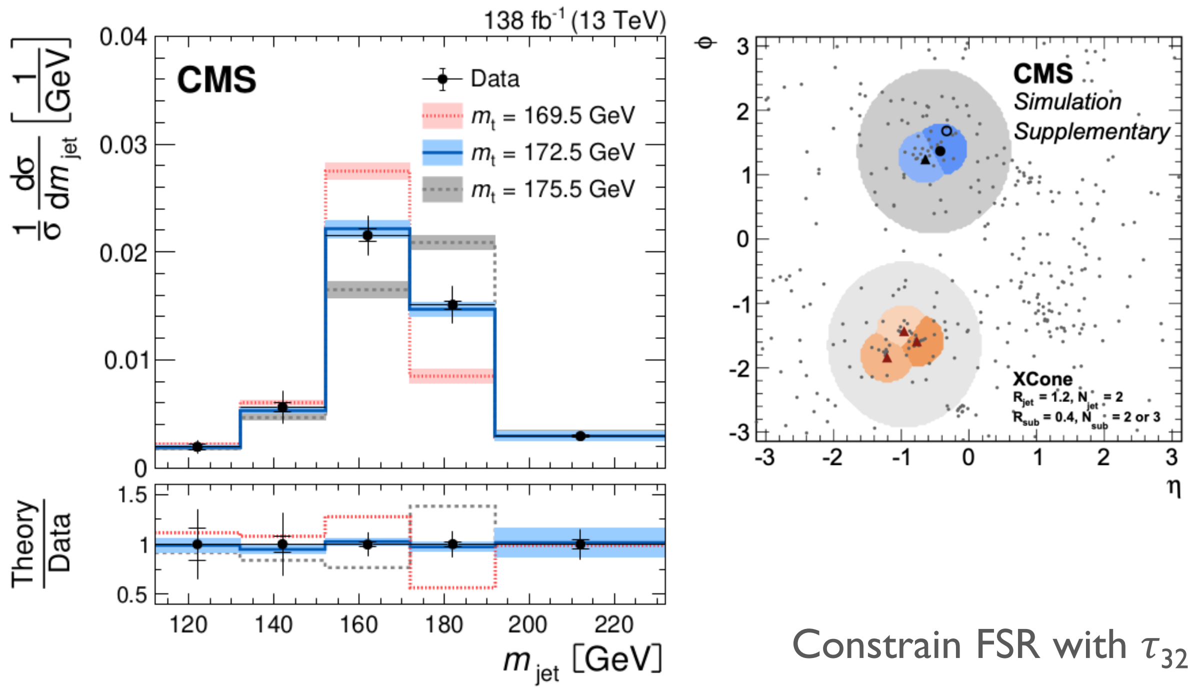
JET MASS IN BOOSTED TOP QUARK DECAYS

- Highly boosted top quarks ⟶ decay products merge

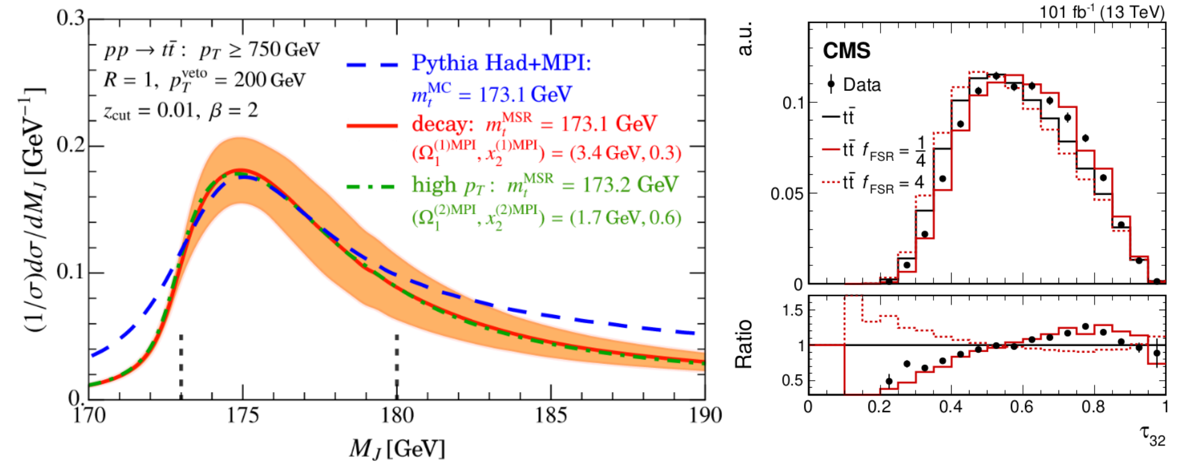


  - Jet mass sensitive to top quark mass M_t
  - Compute M_{jet} from XCone subjets
- Jet mass can be calculated *analytically* and allows an extraction of pole mass
  - Theory phase space (p_T>750) not yet accessible
  - For now: direct measurement (p_T>400)
- Calibration of jet mass scale and FSR modelling improve sensitivity to 800 MeV [TOP-21-012, EPJC sub.]

m_t = 172.76 ± 0.81 GeV

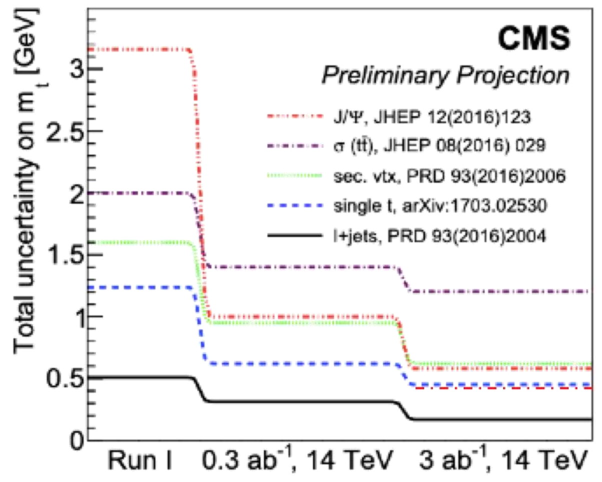


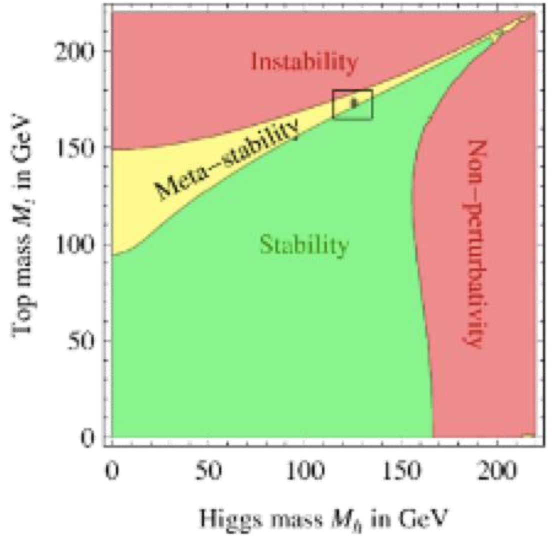
Constrain FSR with τ_{32}

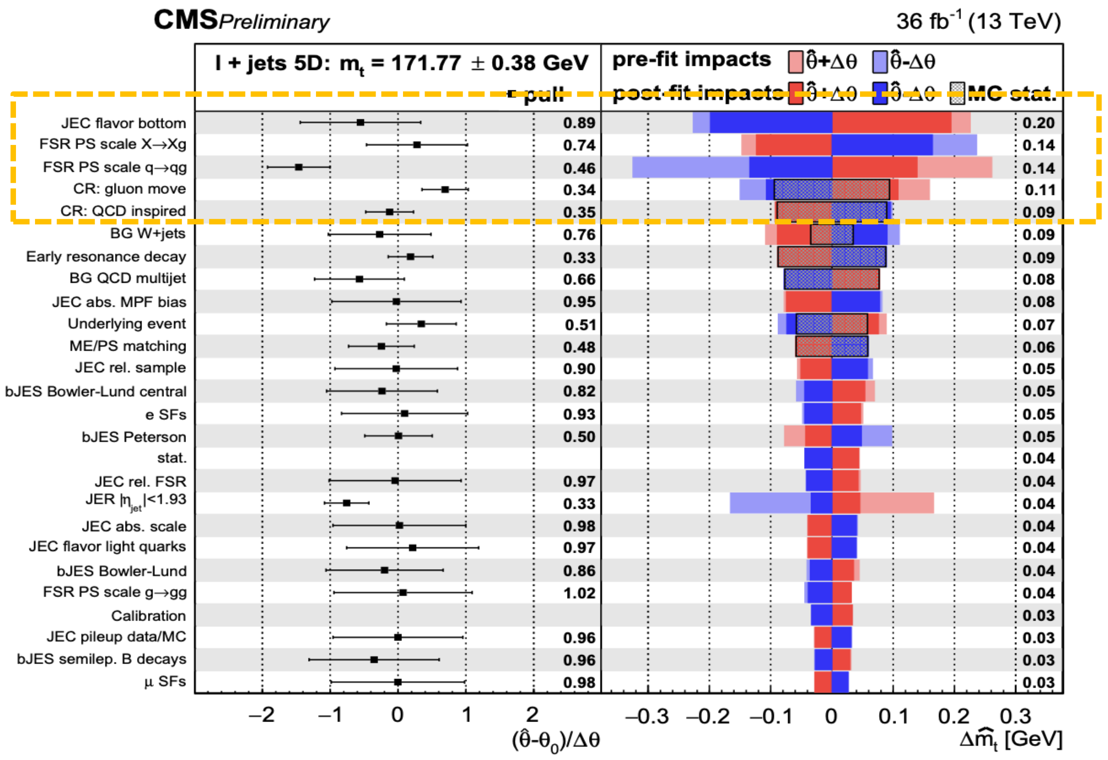


THE MASS OF THE TOP QUARK

- Precision M_t measurement is an *important HL-LHC target*
- HL-LHC projected 0.1% within factor of 2 with 36/fb (!!!)
  - 380 MeV in TOP-20-008 with 5D LL method
  - Top mass is a proxy to the state of pp physics at large
  - Theory & exp. developments towards common goal
- Winning experimental strategy so far:
  resolved jets & in-situ JEC calibration on m_W
  - Exp: uncertainties: bottom vs. light JES
    - plateau for any m_W calibration strategy
    - "tracking vs. calorimetry" response ratio differs for light jets and b jets
  - If M_t(MC) is measured and M_t(MS-bar) is desired:
    - O(1GeV) non-perturbative uncertainties Review by A. Hoang
- Further improvements require *strategic change*, while building on what is known
  - CMS Phase II tracking plays an important role



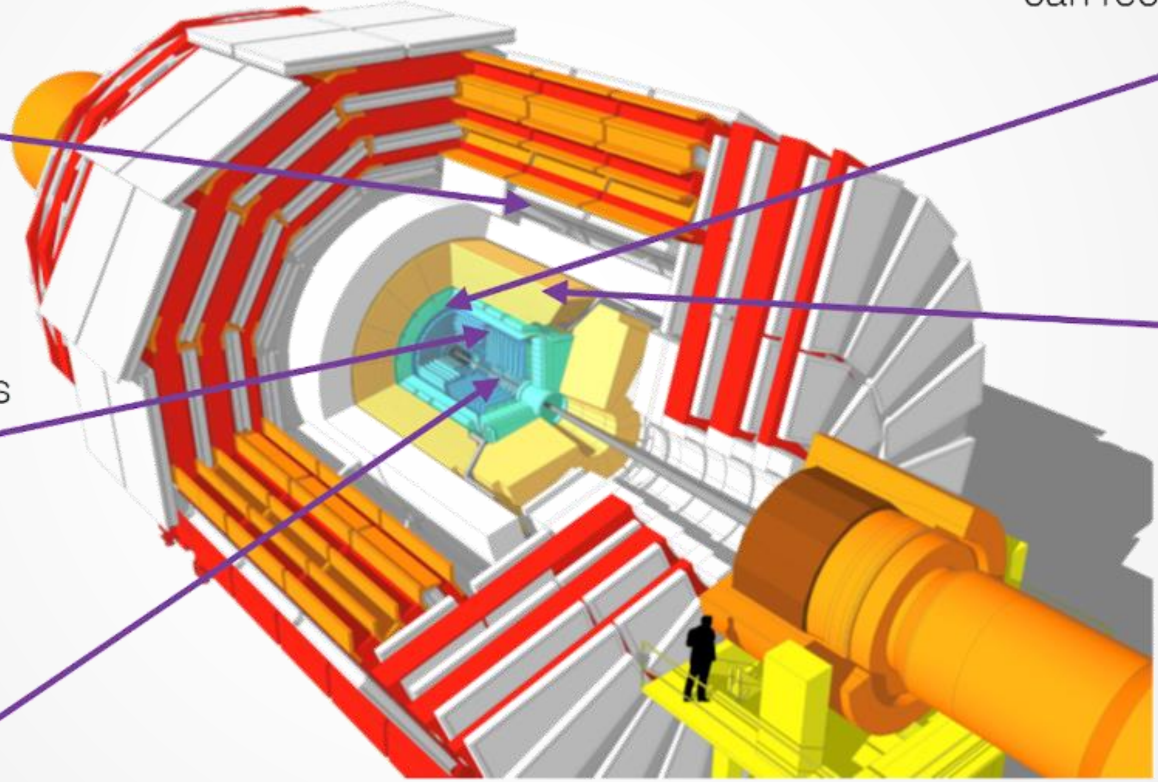




CMS UPGRADES FOR HL-LHC

Improved muon coverage and trigger
increased RPC coverage ($1.5 < |\eta| < 2.4$)
new electronics
[CMS-TDR-016]

New endcap calorimeters
high granularity
can reconstruct showers in 3D
CMS-TDR-019



New precision timing detector
Timing resolution of 30-40 ps for MIPs
full coverage of $|\eta| < 3.0$
[CMS-TDR-020]

Updates to calorimeter and trigger
higher granularity
electronics for trigger

New inner tracker
all silicon tracker
4 layers of pixels
5 layers of strips
coverage to $|\eta| < 4$ [CMS-TDR-014]

Upgrade to trigger and DAQ
L1 rate increased to 750 kHz
High Level trigger rate to 7.5 kHz
Track information at L1
[L1: CMS-TDR-021]
[DAQ/HLT: CMS-TDR-022]

TOP MASS FROM ENERGY CORRELATORS

D. Schwarz, M. Kettner, RS

- M. Procura (UNIVIE), J. Holguin (École Poly.) et. al [2201.08393]
  - 3-point energy correlators (EEEC) computed with *tracks* (!) in boosted hadronic top jets - same [TOP-21-012]
  - Ensembles of triples ↔ many per event, each is weighted
- Track-based m_t measurement with complete theoretical control

- HEPHY will keep playing an essential role in the CMS experiment and will contribute, besides the analysis efforts, to hardware responsibilities within the experiment. The collaboration between the CMS Tracker and the CMS analysis group should be further intensified to open new possibilities for young researchers to gain visibility within CMS and the community. In particular, participation in the commissioning and operation of the Tracker will be pursued as well as the contribution to possible upgrades after phase 2. Further details have to be discussed among the CMS members of HEPHY and the directorate.

CMS

CERN-PH-EP/2013-037
~~2014/10/29~~
2028/03/01

~~CMS-TRK-11-001~~
CMS-TRK-28-001

Description and performance of track and primary-vertex reconstruction with the CMS tracker
during high-luminosity LHC operations

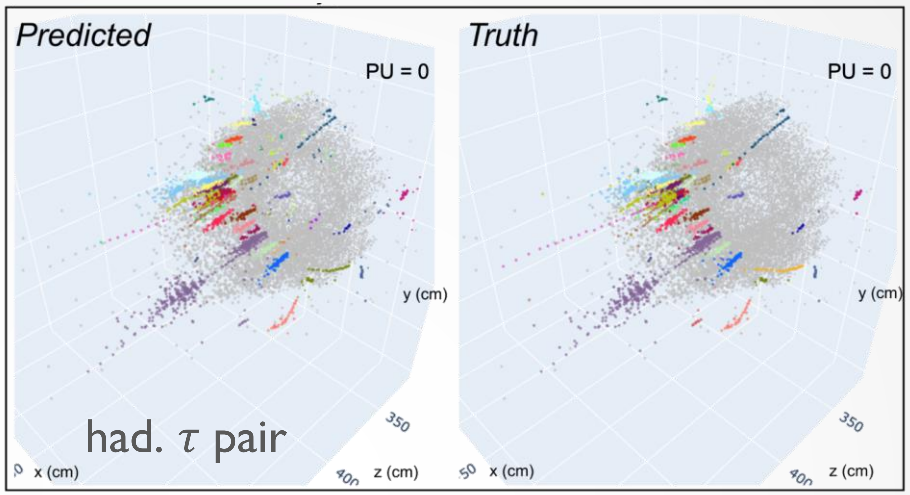
The CMS Collaboration*

- Experimental: tracking in dense jets
  - Example: Sensor-level dead channels modelling→ non-linear high p_T effects
  - beyond the PF requirements
  - EPR task started
- Status: gen-/reco level studies
  - Rather intense exchange with theorists!
  - Master thesis M. Kettner
  - Simultaneously tackle leading 5 (!!) uncertainties of best current meas.
- Entirely new M_t handle for HL-LHC!
- SM-EFT extensions possible [2207.03511]

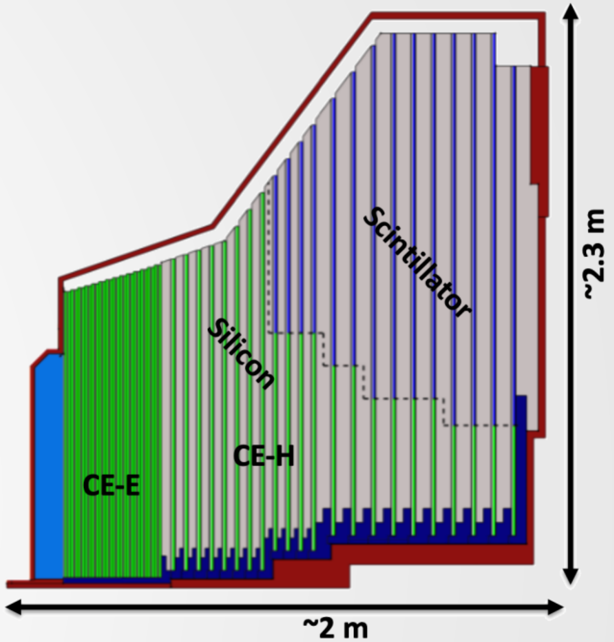
NEW OPPORTUNITIES AT HL-LHC

[DP-2022-004]

- HEP progress often tools-driven
  - High-granularity calorimetry a major opportunity on the +10 years timescale
- Resolve shower particles 200 PU
  - Status: ML based local reco., boosted object IDs, etc.
  - Example fig: gNN based reconstruction of hadronic τ lepton pair



- Run-4 opportunities building on all the HEPHY strengths
  - Resolve (e.g.) hadronic BSM effects using state-of-the-art ML
    - Exploring VBS in 1ℓ with spatially resolved orientation of substructure
    - Exploring semileptonic spin-correlation with top quark pairs
  - Start early with building expertise on reconstruction/experimental systematics
    - new PhD with E. Brondolin (CERN Austrian doctoral program) on HL-LHC reco
  - Emphasize Run II efforts with HL-LHC prospects & relevance for theory



Key Parameters:

- HGCAL covers $1.5 < \eta < 3.0$
- Full system maintained at -30°C
- **~600m²** of silicon sensors
- **~500m²** of scintillators
- **6M Si channels**, 0.5 or 1.1 cm² cell size
  - Data readout from all layers
  - Trigger readout from alternate layers in CE-E and all layers in CE-H
- ~27000 Si modules
- ~140 kW per endcap

WRAP UP

- Core projects
  - displaced 2μ is out, capitalize on expertise in Run 3
  - Flavor/EFT in 3ℓ in full swing
  - VH angular analysis in full swing
  - Stops compressed in full swing / going out
  - Soft vertices ramping up
  - 4t+ttbb ramping up
  - SModelS is maturing
- Rich involvements based on core projects
  Building blocks for the future
  - Boosted hadronic top mass (with DESY)
  - Boosted Information Tree
  - Energy correlators for the top mass (UNIVIE) ↔ Phase 2 tracking
  - SM-EFT in boosted 2ℓ ttbar (TIFR)
  - X→Yh resonance search
  - Madgraph@GPU (with CERN IT & Louvain)
  - Tau reco at HGCal (with CERN/E. Brondolin) ↔ HGCal reco
  - Lepton ID with gNN

Convenerships

Alberto: LHC LLP WG CMS convener (2022-24)

Dennis: CMS TOP ttX convener (2022-24)

Robert: LHC EFT WG convener (2022-24)
LPC distinguished researcher (2022)

Wolfgang: CMS Deputy Spokesperson

Claudia: CMS Collaboration Board Chair

List of talks

Next slide

THE HEPHY CMS DATA ANALYSIS GROUP

[data analysis group]



















TALKS SINCE 2022

Conference	Speaker	Type	Title
DAE-BRNS 2022	S. Chatterjee	Parallel	Effective field theory results in Higgs and top sector from the CMS experiment
DAE-BRNS 2022	S. Chatterjee	Plenary	Summary of Higgs working group activities
LHC EFT WG	R. Schöfbeck	Plenary	Exploring EFT with ML at the LHC
TOP2022	D. Schwarz	Plenary	Studies of top quark properties in CMS
ICHEP2022	S. Chatterjee	Parallel	Studies of anomalous couplings of the Higgs boson and its CP structure at CMS
ICHEP2022	D. Schwarz	Parallel	Recent studies on top quark properties and mass in CMS
Humbolt Conf.	M. Sanowane	Poster	Search for Long Lived Particles in Higgs Decays with the CMS experiment
ML at GGI	R. Schöfbeck	Sem.+Plen.	Exploring EFT with ML at the LHC
LHCP2022	A. Escalante	Plenary	Searches with displaced particles (covering CMS, ATLAS, LHCb, NA62)
LHCP2022	S. Chatterjee	Parallel	Top: BSM searches ATLAS+CMS
DIS2022	R. Schöfbeck	Plenary	Prospects for QCD, EW and Top Physics at the (HL-)LHC
Moriond EW 22	M. Sanowane	Parallel	Search for long-lived particles decaying to a pair of muons at 13 TeV
LP2021	A. Escalante	Parallel	Search for Dark Matter and new physics with LL and unconventional sign. in CMS
LLP workshop	A. Escalante	Plenary	Gaps, overlaps, and complementarity in recent ATLAS, CMS, and LHCb LLP searches
TIFR Mumbai	S. Chatterjee	Seminar	Learning likelihood with tree boosting for extracting EFT parameters
IIT Mumbai	S. Chatterjee	Seminar	Machine learning based methods for extracting EFT parameters
DESY	R. Schöfbeck	Colloquium	Top quarks, effective interactions, and future LHC measurements

[Madminer 1805.00020]

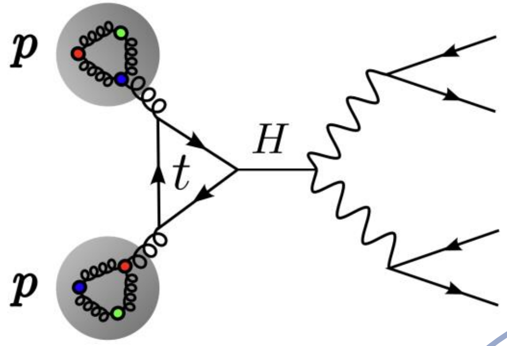
full list of references: backup

HEP MODELLING FROM *ML* POINT OF VIEW

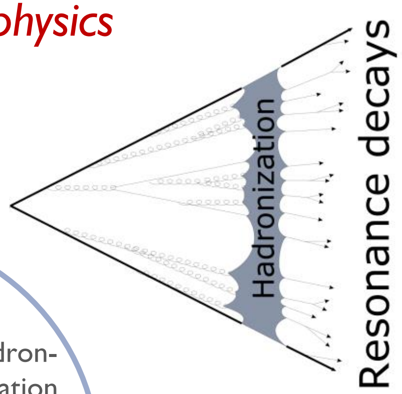
Short-distance physics

- *generative models* for MEs from perturbative QFT, sample parton level, and provide likelihood ratios via

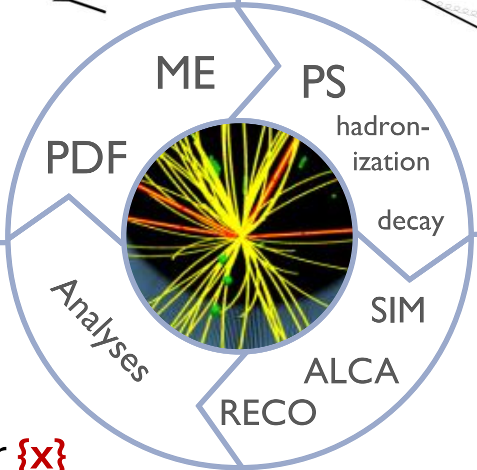
$$\frac{|\mathcal{M}(z_p|\theta\ \ )|^2}{|\mathcal{M}(z_p|\mathrm{SM})|^2}$$



Long-distance physics



- parton shower
- factorization / hadronization models
- decay branchings, calculated & measured
- large *latent space*

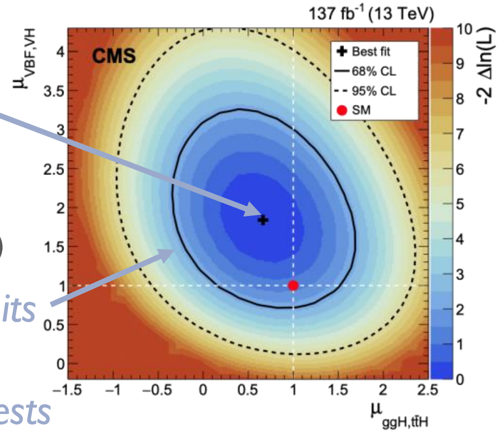


Event reconstruction & data analysis

Maximum-likelihood estimate

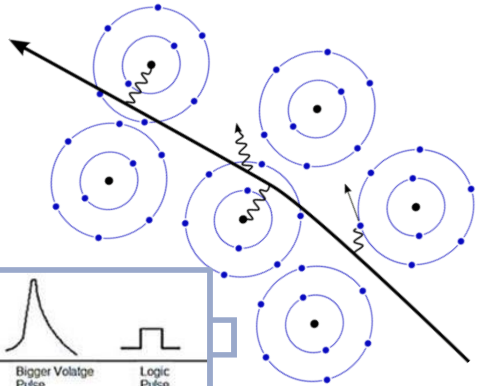
$\theta_{MLE} = \mathrm{argmax}_\theta\ L(\mathcal{D},\theta)$

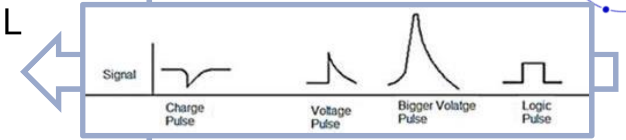
Confidence limits based on likelihood-ratio tests

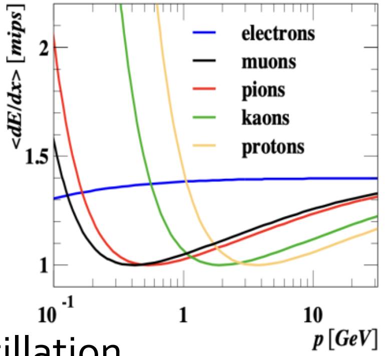


- observed feature vector **{x}**
  - used in MLE
  - hypo.-tests & CL
- Simulation: sampling of **{x,z}** for ML training & test design

Detector interactions







- ionization (Bethe-Bloch), szintillation, brem., transition-radiation … more latent variables

LIKELIHOOD-FREE INFERENCE FOR SM-EFT

[Madminer 1805.00020]

full list of references: backup

Simulation: Sampling of $p(\boldsymbol{x}, \boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{s}, \boldsymbol{z}_\mathrm{p} | \boldsymbol{\theta})$

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \underbrace{\int \mathrm{d}\boldsymbol{z}_\mathrm{d}\, \mathrm{d}\boldsymbol{z}_\mathrm{s}\, \mathrm{d}\boldsymbol{z}_\mathrm{p}\; p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d})\, p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s})\, p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p})}_{\text{intractable}}\, p(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta})$$

The joint space is simpler:

$$r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \mathrm{SM}) \equiv \frac{p(\boldsymbol{x}, \boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{s}, \boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}\quad)}{p(\boldsymbol{x}, \boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{s}, \boldsymbol{z}_\mathrm{p}|\mathrm{SM})} = \frac{p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d})}{p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d})} \frac{p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s})}{p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s})} \frac{p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p})}{p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p})} \frac{p(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}\quad)}{p(\boldsymbol{z}_\mathrm{p}|\mathrm{SM})} \propto \frac{|\mathcal{M}(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}\quad)|^2}{|\mathcal{M}(\boldsymbol{z}_\mathrm{p}|\mathrm{SM})|^2}$$

Change in likelihood of simulated observation x with latent "history" z going from "SM" to θ

staged simulation in forward mode: Intractable factors cancel

re-calcuable theory prediction

Minimize loss on (simulated) joint distribution: $L = \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\mathrm{SM}) \left(r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \mathrm{SM}) - \hat{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) \right)^2 \longrightarrow \min$

$$f^*_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{\sigma(\theta)}{\sigma(\theta_0)} \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}) r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\theta}_0)}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z})} = \frac{\sigma(\theta)}{\sigma(\theta_0)} \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{\sigma(\theta)}{\sigma(\theta_0)} \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta}_0)} = r(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

Available from simulation

Latent space is integrated

what we actually want: change in likelihood of a specific observation

LIKELIHOOD-FREE INFERENCE WITHB TREES

[arXiv:2107.10859, arXiv:2205:12976]

Regress in R, including its the polynomial θ dependence

$$R(\boldsymbol{x}|\boldsymbol{\theta}, \mathrm{SM}) = \frac{\mathrm{d}\sigma(\boldsymbol{x}, \boldsymbol{\theta})/\mathrm{d}\boldsymbol{x}}{\mathrm{d}\sigma(\boldsymbol{x}, \mathrm{SM})/\mathrm{d}\boldsymbol{x}}$$

→ will allow to compute the optimal LLR test statistic q($\mathcal{D}$)

$$L = \sum_{\boldsymbol{\theta}\in\mathcal{B}} \int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\mathrm{SM}) \left(R(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \mathrm{SM}) - \hat{F}(\boldsymbol{x}, \boldsymbol{\theta}) \right)^2$$

$$F^*(\boldsymbol{x}, \boldsymbol{\theta}) = R(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

Tree ansatz for each a, ab:
$F_j(\theta)$ polynomial with const. coeff. (per node)

$$\hat{F}(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{j\in\mathcal{J}} \mathbb{1}_j(\boldsymbol{x}) F_j(\boldsymbol{\theta})$$

find optimal partitioning — find optimal predictor

Solve for the predictor on the empirical distribution (simulated sample)

$$F_j(\boldsymbol{\theta}) = \frac{\sum_{i\in j} w_i(\boldsymbol{\theta})}{\sum_{i\in j} w_i(\boldsymbol{\theta}_0)} \equiv \frac{w_j(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

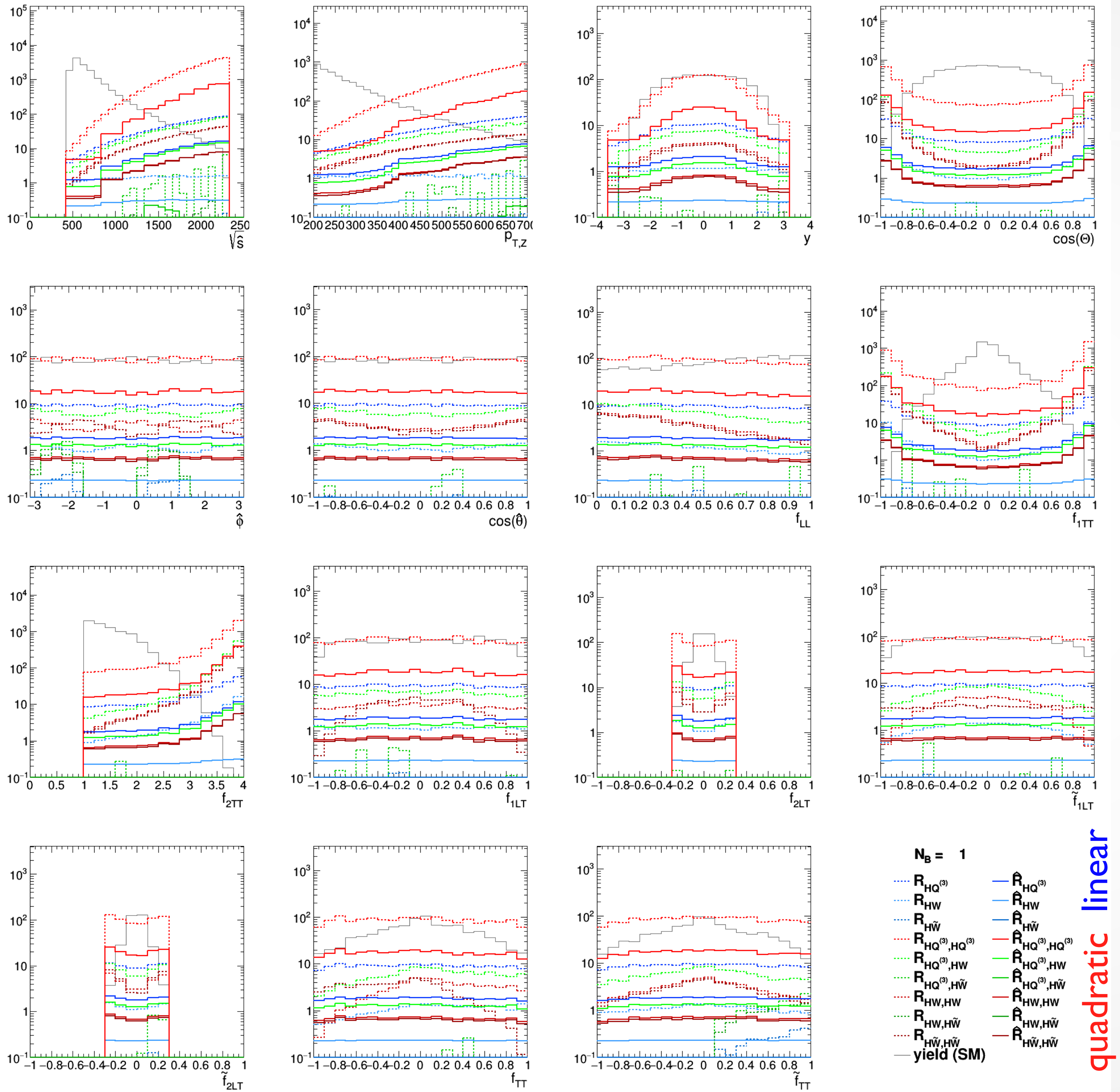
No trainable parameters in the predictor

Solve for optimal partitioning with greedy CART algorithm

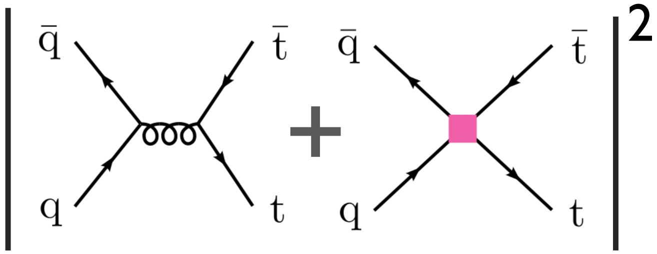
$$L = -\sum_{\boldsymbol{\theta}\in\mathcal{B}} \sum_{j\in\mathcal{J}} \frac{w_j^2(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

split only if $w_j(\theta)$ is *positive* ∀θ

"Boosted Information Tree"



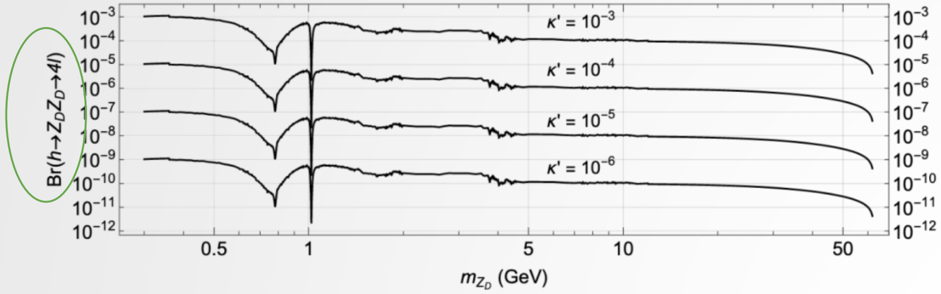
- Test-case: models of ZH and WZ
- Left: "Boosted Information Tree (BIT)"
  - NN are equivalent
  - 3 WC, 9 DOF, 500k events, ZH
  - 200 trees, D=5, 9 minutes of training
  - also more realistic study, including backgrounds [2107.10859], [2205.12976]
- Learning coefficient functions to compute parametrized optimal oberables

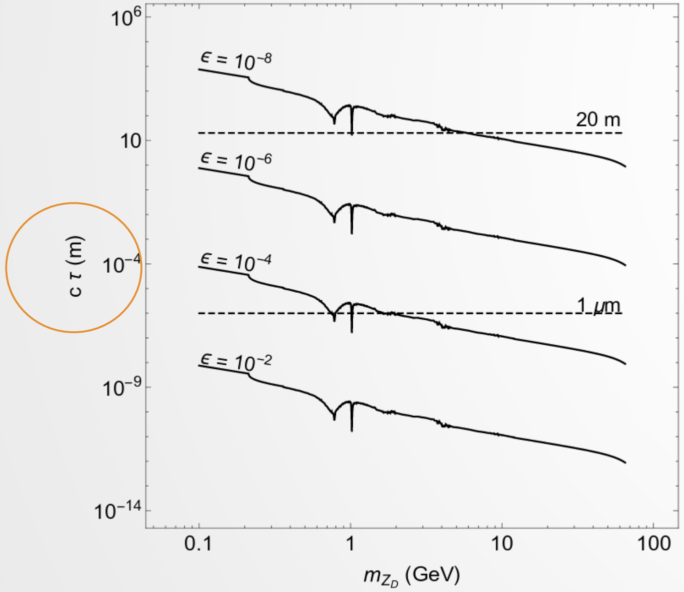


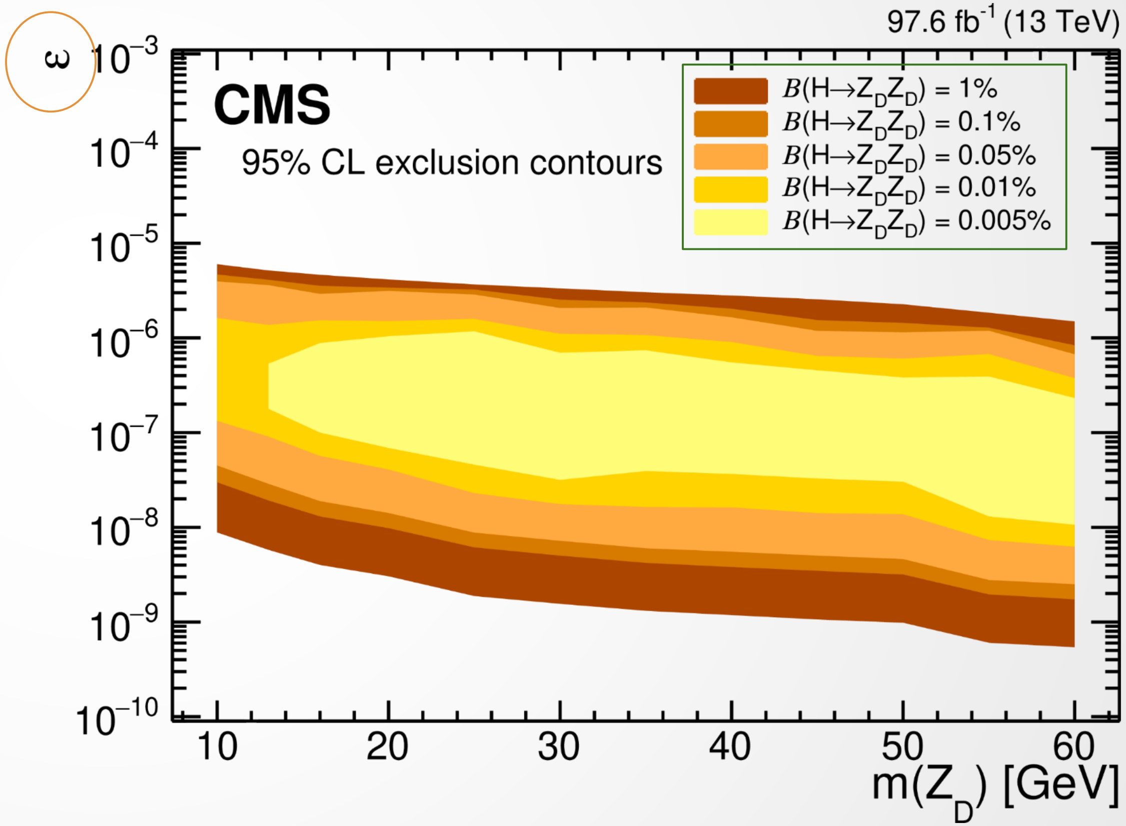
$$= \sigma^{\mathrm{SM}} + \frac{\theta}{\Lambda^2}\sigma^{\mathrm{int}} + \frac{\theta^2}{\Lambda^4}\sigma^{\mathrm{quad}}$$

$\rightarrow$ parametrized $q_{\boldsymbol{\theta}}(\mathcal{D})$

EXCLUSION IN DARK SECTOR MODEL







LOW P_T ELECTRON RECONSTRUCTION

- Standard electron: low efficiency at high IP (~zero at d_{xy}= 5 cm)
  - Decrease in displaced signal acceptance
- Employ special low p_T electro-reco developed with B-parking data
- Tangible reco efficiency in higher IP
  - Beneficial also for prompt
  - Study presented in EGM POG