

Boosted resonance searches at CMS and their long-term impact



Clemens Lange (CERN)

Fermilab Wine & Cheese Seminar 15th January 2021







Going to highest energies...

>The Large Hadron Collider (LHC) is the world's largest and highestenergy particle accelerator



Delivered \mathscr{L} = 163 fb⁻¹ of proton-proton collisions at \sqrt{s} = 13 TeV

Exceeded most of its design parameters, producing collisions 49% of the time in 2018

Image: Nathan Readioff

Clemens Lange - Boosted resonance searches at CMS and their long-term impact







... creating a big mess



CMS Experiment at the LHC, CERN Data recorded: 2016-Sep-08 08:30:28.497920 GMT Run / Event / LS: 280327 / 55711771 / 67

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

86 reconstructed vertices →"pileup"





My needle in the haystack

>Finding a very heavy and extremely rare resonance in events like that

Clemens Lange - Boosted resonance searches at CMS and their long-term impact





My needle in the haystack

>Finding a very heavy and extremely rare resonance in events like that





Clemens Lange - Boosted resonance searches at CMS and their long-term impact

Why?





Finding a very heavy and extremely rare resonance in events like that



force and gravity? $\mu^2 = \lambda v^2 = \frac{\lambda}{g^2} 4M_W^2 \sim 10^4$ Flect

- Referred to as hierarchy problem
- In other words why is the Higgs boson so light?
- the TeV scale: $\mu^2 \sim (\text{heavier scale})^2 \rightarrow \text{new particles}$

Why?

>Try to find a possible explanation to the big difference between weak

$$^{4} \, {
m GeV}^{2} \ll M_{Pl} \sim 10^{38} \, {
m GeV}^{2}$$

Electroweak scale

Planck scale

>,,Natural" explanation would be that SM is replaced by another theory at





15.01.2021







15.01.2021







15.01.2021





15.01.2021





More on the hierarchy problem



15.01.2021





More on the hierarchy problem

Need to test these experimentally!







Need to test these experimentally!





Examples of heavy resonances





h





Examples of heavy resonances



All the unknown particles X, W', b*, B etc. are very heavy: $m \ge 1$ TeV

15.01.2021

Clemens Lange - Boosted resonance searches at CMS and their long-term impact









- >In the examples shown, resonance decays to massive bosons (W, Z, H) or also top quarks
- >These decay further, since they are not stable
- >They predominantly decay to a pair of quarks (bosons) or three quarks (top)
- >This leaves a lot of energy to the quarks due to $E_{\rm decay} pprox m_X c^2$ and they will be close in space
- >The quarks hadronize and form socalled jets





- >In the examples shown, resonance decays to massive bosons (W, Z, H) or also top quarks
- >These decay further, since they are not stable
- >They predominantly decay to a pair of quarks (bosons) or three quarks (top)
- >This leaves a lot of energy to the quarks due to $E_{\rm decay} pprox m_X c^2$ and they will be close in space

>The quarks hadronize and form socalled jets

Note: leptonic decay channels bring along other challenges such as lepton isolation

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

 \boldsymbol{g} Jet 9 time W^{\star} Jet D







- >Object defined by jet clustering algorithm
- Provides link between theory predictions and experimental observations
- >Theory: quarks/partons \rightarrow hadronization \rightarrow particles (p, n, η, λ, π, ...)
- Experiment: sensor signals -> reconstruction -> tracker hits, calorimeter entries (\rightarrow particles (neutral/charged hadrons, χ , ...))







- >Jets are very messy objects
- >This is due to a phenomenon called confinement
- >A quark/gluon cannot exist on its own
- >It "pulls" in other quarks and gluons to form hadrons - this is due to the strong force
- >Several of the particles created this way decay further until stable particles have been created

$$(\tau_{\text{had}} = \Lambda_{\text{QCD}}^{-1} = 3 \cdot 10^{-24} \text{ s})$$



Example of a *single* proton-proton collision—now imagine more than 30 on average at the same time



[F. Krauss]





Clemens Lange - Boosted resonance searches at CMS and their long-term impact



- Standard" jets in CMS: use anti-k_T algorithm with R = 0.4 ("AK4 jets")
- >Approximation: all particles within
 - $\Delta R := \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = c$ will be
 - reconstructed as a single jet with cut-off parameter R = c



Clemens Lange - Boosted resonance searches at CMS and their long-term impact



- Standard jets in CMS: use anti-k_T algorithm with R = 0.4 ("AK4 jets")
- >Approximation: all particles within
 - $\Delta R := \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = c$ will be
 - reconstructed as a single jet with cut-off parameter R = c

For a W-boson decay: $R_{qq} \approx 2 \frac{m_W}{p_T^W}$ • Example: resonance of mass 1 TeV

bosons (m = 100 GeV) from the decay will on average have $p_T \sim 0.4 \text{ TeV} \rightarrow R_{qq} \approx 0.5$

An AK4 jet could miss parts of the decay





Clemens Lange - Boosted resonance searches at CMS and their long-term impact



- > "Standard" jets in CMS: use anti-k_T algorithm with R = 0.4 ("AK4 jets")
- >Approximation: all particles within
 - $\Delta R := \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = c$ will be
 - reconstructed as a single jet with cut-off parameter R = c
- For a W-boson decay: $R_{qq} \approx 2 \frac{m_W}{p_T^W}$
 - Example: resonance of mass 1 TeV
 P I bosons (m = 100 GeV) from the decay will on average have p_T~0.4 TeV → R_{qq} ≈ 0.5

An AK4 jet could miss parts of the decay

>Use larger jet "radius" to contain full decay (usually R = 0.8, → "AK8")









- >At first look, it's the same as any other jet
- However, if you look more closely, you can find differences





- >At first look, it's the same as any other jet
- However, if you look more **closely**, you can find differences
- >You will find that the energy **deposits** inside the jet are centred around two cores/axes instead of just one





- >At first look, it's the same as any other jet
- >However, if you look more closely, you can find differences
- >You will find that the energy **deposits** inside the jet are centred around two cores/axes instead of just one
- >The opening angle between two energy deposits relates to the mass of the original particle



Clemens Lange - Boosted resonance searches at CMS and their long-term impact



- >At first look, it's the same as any other jet
- >However, if you look more closely, you can find differences
- >You will find that the energy **deposits** inside the jet are centred around two cores/axes instead of just one
- >The opening angle between two energy deposits relates to the mass of the original particle
- >However, to be able to see that, we need to clean up the jet



Clemens Lange - Boosted resonance searches at CMS and their long-term impact



Cleaning up a jet

Phys. Rev. D 97 (2018) 072006



JINST 15 (2020) P09018



- >When looking at our "massive" jets, we are mostly interested in the initial quarks
- >Attempt to throw away low energy and large angle radiation → "undo" hadronization
- >This process is called grooming

There are a few different algorithms to do that

Phys. Rev. D 97 (2018) 072006



JINST 15 (2020) P09018



- >When looking at our "massive" jets, we are mostly interested in the initial quarks
- >Attempt to throw away low energy and large angle radiation → "undo" hadronization
- >This process is called grooming

There are a few different algorithms to do that

>In CMS, the modified mass-drop tagger "soft drop" is commonly used Phys. Rev. D 97 (2018) 072006



Clemens Lange - Boosted resonance searches at CMS and their long-term impact

JINST 15 (2020) P09018

- Remember that I told you something about an opening angle between the decay products?
- >We can also measure the energy flow inside a jet
- >Estimate to what extent it is aligned along 2 or 3 momentum axes (or just 1)
- >Again, there are a range of different algorithms
 - •N-subjettiness (T_N) , energy correlation functions, machine-learned variables, ...
- >For T_N (by now somewhat outdated):
 - ■ratio T₂/T₁ → W/Z boson tagging
 - •ratio $\tau_3/\tau_2 \rightarrow$ top quark tagging



Clemens Lange - Boosted resonance searches at CMS and their long-term impact



Confirming that this works

- >Rather complicated variables, difficult to model \rightarrow need to validate in data
- >Clean sample of W-boson jets: topantitop quark pairs used for calibration (W-boson jet $p_T \sim 200 \text{ GeV}$)







Clemens Lange - Boosted resonance searches at CMS and their long-term impact





Using jet substructure in resonance searches

Two hot-off-the-press examples





>Topology:

- ■1 high-p_T bottom-quark jet
- ■1 high-p_T top-quark jet
- both are identified using dedicated deep neural network algorithms (DeepJet and DeepAK8)

>Selection:

- ■One AK4 jet, p_T > 550 GeV, b tagged
- One AK8 jet, p_T > 550 GeV, top tagged, 105 < m_{SD}
 < 210 GeV</p>
- Both well-separated from each other (back-toback)

Define control and validation regions by inverting b/top tag requirements and using mass sidebands



- >Final discriminant: invariant tb mass (m_{tb})
- >Background estimation of dominant multijet production:
 - In mass sideband ($m_{SD} < 105 \text{ GeV}$), extract **b** tag pass/fail ratio as function of n of the AK4 jet
 - In signal mass window, invert b tag requirement (SR') and multiply m_{tb} spectrum by pass/fail ratio to predict multijet in signal region (SR)
 - Validate procedure in top tag fail region (VR)



Clemens Lange - Boosted resonance searches at CMS and their long-term impact

- >Final discriminant: invariant tb mass (m_{tb})
- >Background estimation of dominant multijet production:
 - In mass sideband ($m_{SD} < 105 \text{ GeV}$), extract **b** tag pass/fail ratio as function of n of the AK4 jet
 - In signal mass window, invert b tag requirement (SR') and multiply m_{tb} spectrum by pass/fail ratio to predict multijet in signal region (SR)
 - Validate procedure in top tag fail region (VR)





>W' can have different chiralities (left- and right-handed)

- tagger efficiency
- Also taking into account interference with standard model



new

>Set upper limits on the production cross section of a W' decaying to tb

• Different chiralities lead to different angular distributions of top quark decay products affecting







>Topology:

- ■1 high-p_T W-boson jet
- ■1 high-p_T top-quark jet
- both are identified using N-subjettiness ratios (T₂/T₁ and T₃/T₂), additional subjet b tagging for top quark

>Selection:

- ■Two AK8 jets, p_T > 400 GeV
- Select W-boson (and top-quark) mass windows 65 (105) < m_{SD} < 105 (210) GeV</p>
- Both well-separated from each other (back-toback)





Search for b* → tW









>Parameterize bump hunt in (m_t, m_{tW})

- Smoothly falling nonresonant background (multijets, W+jets) estimated from data
- Resonant backgrounds (tt and tW) estimated via template fit from simulation

>Perform a 2D likelihood fit in m_t = [65, 285] GeV and m_{tW} = [1200, 4000] GeV



>Parameterize bump hunt in (m_t, m_{tW})

- Smoothly falling nonresonant background (multijets, W+jets) estimated from data
- Resonant backgrounds (tt and tW) estimated via template fit from simulation
- >Perform a 2D likelihood fit in m_t = [65, 285] GeV and m_{tW} = [1200, 4000] GeV

>Nonresonant background: top tag pass/fail ratio

Smooth ratio, fit using 2D polynomials

Resonant tt and tW background:

- Define dedicated measurement region by requiring a second top tag
- Estimate multijet background as for signal

>Constrain top mass scale simultaneously

b* can have different chiralities (left- and right-handed, vector-like) >Largest excess for left-handed b* at 2.4 TeV (2.3 σ local) Interpret result in these models

new

>Largest excess for left-handed b* at 2.4 TeV (2.3 σ local) >Interpret result in these models

new

>b* can have different chiralities (left- and right-handed, vector-like)

Limits — and now what?

Ensuring analysis long-term impact and collaborating with theory

Digitised results?

Note: Can also make Rivet analysis available

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

>The figures that are part of a paper are published as images

Reading off values can be tedious

Note: There are tools such as <u>WebPlotDigitizer</u> that help with that

Note: Can also make Rivet analysis available

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

- >The figures that are part of a paper are published as images
- Reading off values can be tedious
 - Note: There are tools such as <u>WebPlotDigitizer</u> that help with that
- >Better: Provide digitised versions of plots and tables
 - Analyst most likely the only person with those at hand
- **For High Energy Physics: HEPData** portal
 - Use e.g. <u>hepdata_lib</u> Python library to create a submission

Show All 625 values

Visualize

Showing 50 of 625 values

SQRT(S)		13 TeV
LUMINOSITY		137 fb ⁻¹
<i>m_{VLQ}</i> [GeV]	χ^2 /ndf	Jet Tag Efficiency
50.0	1.0	0.0
50.0	3.0	0.0
50.0	5.0	0.0
50.0	7.0	0.0
50.0	9.0	0.0
50.0	11.0	0.0
50.0	13.0	0.0

Note: Can also make Rivet analysis available

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

>The figures that are part of a paper are published as images Reading off values can be tedious Note: There are tools such as <u>WebPlotDigitizer</u> that help with that >Better: Provide digitised versions of plots and tables Analyst mo those at ha **For High Energy Physics: HEPData** portal

Use e.g. <u>hepdata_lib</u> Python library to create a submission

Anyone can reproduce the plot!

	Λ	
50.0	1.0	0.0
50.0	3.0	0.0
50.0	5.0	0.0
50.0	7.0	0.0
50.0	9.0	0.0
50.0	11.0	0.0
50.0	13.0	0.0
50.0	13.0	0.0

Visualize

Note: Can also make Rivet analysis available

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

models and frameworks

>However:

- New models are continuously being developed and could be quite different
- There can be surprises in the data that "motivate" to take another look (remember the 2015) vv excess?)

>We try to interpret our experimental results in a breadth of theoretical

We try to interpret our experimen models and frameworks

>However:

- New models are continuously being developed and could be quite different
- There can be surprises in the data that "motivate" to take another look (remember the 2015 yy excess?)

>We try to interpret our experimental results in a breadth of theoretical

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

- >A new theory model only changes the posited signal, less often the background processes
 - Do not need to process the data and backgrounds again
 - Probably background model won't have to be changed either
- >Need to estimate new signal's selection efficiency × acceptance
 - Does the paper publication contain enough information to do that?
- >Optimal approach: preserve overall analysis setup

original analysis (w.r.t model A)

original analysis (recast to **model B**)

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

>Tools for preservation/reusability:

Version-controlled software (Git) including automated checks **of git**

Containerised software images (Docker) docker

Automated workflows

>Training available via <u>HEP</u> Software Foundation, LPC hands-on training sessions (<u>HATS</u>), experiment-specific trainings

>Tools being developed at CERN:

- CERN Analysis Preservation Portal
- REANA (Reusable analysis platform)

HSF Training

Performing a reinterpretation

15.01.2021

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

Performing a reinterpretation

15.01.2021

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

- >Theorists (and anyone else) can also use CMS data
- >More than 2 PB available (mostly Run-1) from the <u>CERN Open Data</u> Portal
- >Using the data can be a challenge
 - Training available through CMS Open Data Workshop for Theorists (Sep/Oct 2020)
 - Includes using public cloud offerings for largescale data analysis

All experiments will make data for scientific use available starting five years after data taking

Summary and conclusions

- >Jet substructure enables the search for extremely rare and heavy resonances
- Several new CMS analyses using 2016-18 data becoming public pushing the boundaries
 - Only presented the latest two results
- > Performing reusable analyses has a long-term impact on the field

Recluster jet constituents using Cambridge-Aachen jet algorithm (based on spatial separation only)

>Iteratively break into two subjets

- Remove softer contribution (and) continue with harder one) if: $\frac{\min(p_{T_1}, p_{T_2})}{< 0.1 \text{ (CMS choice)}}$ $p_{T_1} + p_{T_2}$
- >Stop otherwise

Example: "modified mass-drop tagger"

Undoing hadronisation

Reconstructed jet

>Topology:

- Each B decays to a b quark and a Z or Higgs boson
- Investigate $H \rightarrow bb$ and $Z \rightarrow qq$ (incl. bb)

>Selection:

- Select at least 4 AK4 jets, $p_T > 30$ GeV
- If AK8 jet (pT > 200 GeV) exists close to AK4 jet ($\Delta R < 0.3$) and sufficiently far from a second AK4 jet AK4 jet ($\Delta R > 0.6$), use AK8 jet instead
- Use DeepJet b tag algorithm for AK4 jets
- Double-b tagger (incl. substructure info) for AK8 jets bHbZ nultiplicity եղբո

et multiplicity	lag	DHDH	DHD
4 jets	Single b	2	2
	Double b	1	1
E inte	Single b	3	3
5 jets	Double b	0	0
6 jets	Single b	4	4

Phys. Rev. D 102 (2020) 112004

Search for bottom-type, vector-like quark pair production

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

Systematic source Luminosity $t\bar{t}$ cross section Single top cross section JES JER B tagging scale factor top quark tagging scale factor Pileup Prefire correction tt normalization and slope Scale (μ_R and μ_F) PDF Pass-to-fail ratio Alternative function of pass-to-fail ratio Non-closure uncertainty B candidate m_{SD} correction

Correlation between different years	Impact
No correlation	0.5%
Full correlation	0.1%
Full correlation	0.2%
No correlation	0.8%
No correlation	< 0.1%
No correlation	< 0.1%
No correlation	0.2%
Full correlation	< 0.1%
No correlation	< 0.1%
No correlation	0.2%
Full correlation	0.3%
Full correlation	< 0.1%
No correlation	0.2%
No correlation	0.1%
No correlation	0.3%
No correlation	0.9%

Search for b^{*} → tW: additional plots

15.01.2021

Clemens Lange - Boosted resonance searches at CMS and their long-term impact

Search for b* → tW: uncertainties

Source	Uncertainty
tt cross section	$\pm 20\%$
Single top cross section	$\pm 30\%$
Integrated Luminosity	$\pm 1.8\%$
Pileup	Shape (σ_{mb})
Prefire	Shape $(p_{\rm T}, \eta)$
Jet energy scale	Shape $(p_{\rm T})$
Jet energy resolution	Shape (p_T, η)
Jet mass scale	Shape (m_W)
Jet mass resolution	Shape (m_W)
W tagging	Shape $(p_{\rm T})$
W tagging: p_{T} extrapolation	Shape (p_T)
Top tagging, merged	Shape (p_T)
Top tagging, semimerged	Shape (p_T)
Top tagging, not merged	Shape $(p_{\rm T})$
Trigger	Shape (H_T)
Top quark $p_{\rm T}$ correction - α	Shape $(p_{\rm T})$
Top quark $p_{\rm T}$ correction - β	Shape $(p_{\rm T})$
PDF	Shape (m_t, m_{tW})
KDE bandwidth	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{SR}p_0$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{SR}p_1$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{SR}p_2$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{SR}p_3$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{ m tar t}p_0$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{ m tar t} p_1$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{ m t\bar t}p_2$	Shape (m_t, m_{tW})
$R_{\rm ratio}(m_{\rm t},m_{\rm tW})^{\rm t\bar{t}}p_3$	Shape (m_t, m_{tW})

	Samples	Impact (up/down)
	tī	-4.6/+4.4%
	single top	+1.2/-1.4%
	tt, single top, signal	+1.6/-1.1%
	tt, signal, single top	+0.3/-0.2%
	$t\bar{t}$, signal, single top	+0.0/+0.1%
	$t\bar{t}$, signal, single top	+0.3/-0.6%
	$t\bar{t}$, signal, single top	-0.4/-0.5%
	tt, signal, single top	-0.1/ - 0.0%
1	$t\bar{t}$, signal, single top	+0.07 + 0.9%
	signal, single top	+0.9/-0.9%
$\left \right $	signal, single top	+4.9/-4.9%
	tt, signal, single top	+0.2/-0.2%
	tt, signal, single top	+1.1/-0.9%
IF	$t\bar{t}$, signal, single top	-0.1/+0.1%
	$t\bar{t}$, signal, single top	+0.3/-0.4%
	tī	-0.3/+0.3%
$\backslash \lor$	tī	-3.9/+3.5%
	signal	+0.1/-0.1%
	multijet (from simulation)	-1.2/+0.2%
	multijet (from data)	-4.4/+0.0%
	multijet (from data)	-2.0/+2.2%
	multijet (from data)	+0.9/-0.8%
	multijet (from data)	+18.6/-18.8%
	multijet (from data)	-0.4/+0.6%
	multijet (from data)	-0.4/+0.6%
	multijet (from data)	+0.5/-0.6%
	multijet (from data)	-0.6/+0.6%

