Recasting LHC analyses with MADANALYSIS 5

Problems & solutions

Sam Bein, Monika Blanke, Guillaume Chalons, Eric Conte,
Karen de Causmaecker, Beranger Dumont, Benjamin Fuks,
Iftah Galon, Sabine Kraml, Suchita Kulkarni, Kentarou Mawatari
& Dipan Sengupta (*et al.*)

CERN - IPHC - LPSC - University of Strasbourg - Upper Alsace University - VUB

Mini-workshop on recasting ATLAS and CMS new physics searches

LPSC Grenoble - September 8-12, 2014

Outline

1. The MA5-tune of DELPHES 3

2. Selected items on the implementation and the validation of LHC analyses

3. Summary

The MA5-tune of DELPHES 3: technicalities

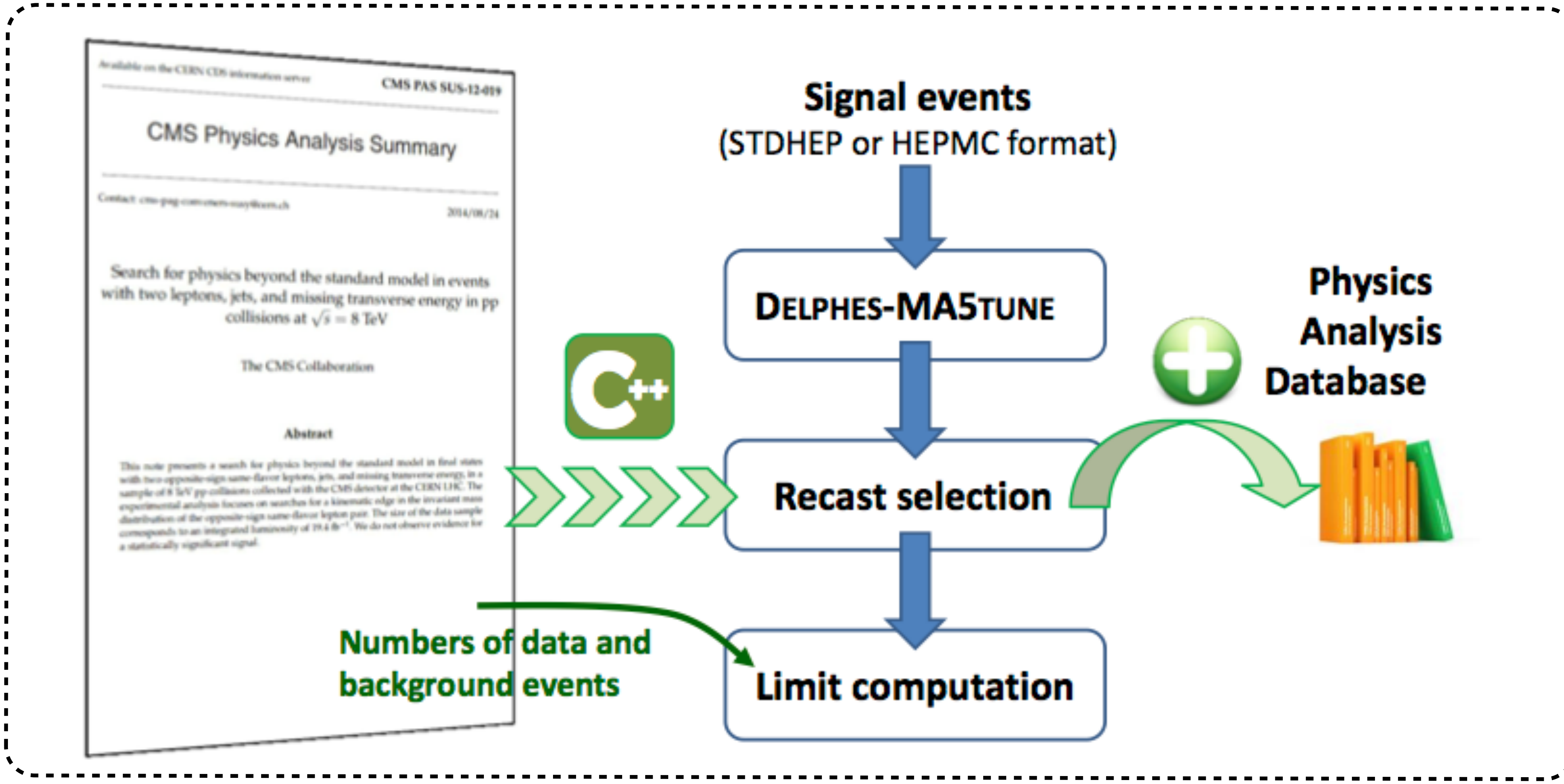
Recasting and designing LHC physics analyses

- Determining the sensitivity of the LHC to a new physics model by analyzing specific signatures
  - Recasting an existing LHC analysis
    - Generation of signal Monte Carlo event samples
    - Event generation: LHE (parton level) - STDHEP/HEPMC (hadron level) - ROOT (reconstructed level)
    - Analysis by mimicking as much as possible the experimental cuts
  - Designing a novel LHC analysis
    - Generation of background and signal Monte Carlo event samples
    - Event generation: LHE (parton level) - STDHEP/HEPMC (hadron level) - ROOT (reconstructed level)
    - Analysis by tuning the thresholds to increase a factor of merit of choice

- Monte Carlo production: time and disk space consuming
  - In order to only do it once:
    - Inclusion in the output ROOT file all information needed for all foreseen works
    - Analysis-dependent issues (like lepton isolation) to be achieved at the analysis level
  - The size of the ROOT file must be kept reasonable

A compromise must be found

The MADANALYSIS 5 strategy for recasting an LHC analysis



✦ Why a tune of DELPHES?

- ✤ The output ROOT file must be flexible and general enough to address any LHC analysis
  - ★ Optimal choice: one ROOT file for any (ATLAS or CMS) analysis
  - ★ Likely option: one CMS and one ATLAS ROOT file
  - ★ Current option: one CMS and many ATLAS ROOT files
- ✤ The standard DELPHES output cannot be used for that purpose

One of the goals of this workshop

Cleaning the ROOT output file

- Generated particles at all levels (parton, hadron, reconstructed)
  - Useful
  - Heavy (disk space consuming)

- The MA5-tune solution
  - All reconstructed objects are stored, including track information
  - Keeping all initial-state and hard-process particles
  - The particle history for beauty and charm hadrons (starting from the hard process)
  - Final-state electrons and muons are linked to their mother, …, grand-grand-mother
    - ★ Useful to determine whether the lepton is prompt, issued from a tau, *etc*.
  - All other particles are discarded

- Linking reconstructed and generated objects
  - Leptons: reconstructed leptons are smeared generated leptons
    - ★ Easy to link without any mismatch (included in the output file)
  - Jets: reconstructed jets can be matched to generated partons (via a standard ΔR method)
    - ★ To be implemented at the analysis level by the user
    - ★ Private methods exist for the moment (to be public one day?)
    - ★ Limitations: light partons arising from the showering are not stored in the output file
    - ★ Could be included at the DELPHES level

Lepton isolation

- **Lepton isolation is analysis-dependent**
  - The output ROOT file must contain the necessary information
  - Storing track, calorimeter, *etc.*, information is too heavy

- **The MA5-tune solution**
  - Storing isolation variables that are computed at the DELPHES level
  - Sum of the p_T of all tracks and of all calorimetric deposits in a given cone around the lepton
  - Number of tracks in this cone
  - Particle flow isolation information is also retained (currently under study)

	DR=0.2	DR=0.3	DR=0.4	DR=0.5
Tracker isolation				
Calorimeter isolation				
ParticleFlow isolation				

  - The p_T threshold is fixed at the level of DELPHES

- **Cleaning of the jet collection**
  - In standard DELPHES, isolated leptons are removed from the jet collection (the UNIQUEOBJECTFINDER method)
  - By-passed in the MA5-tune
  - Jet cleaning must be done at the analysis level (e.g., a ΔR method)
  - The H_T variable must be recalculated

B-tagging

✦ Two strategies are implemented in DELPHES

- ✤ Parameterization (in the detector card): several benchmark (or cards) are thus required
- ✤ Track-counting algorithm

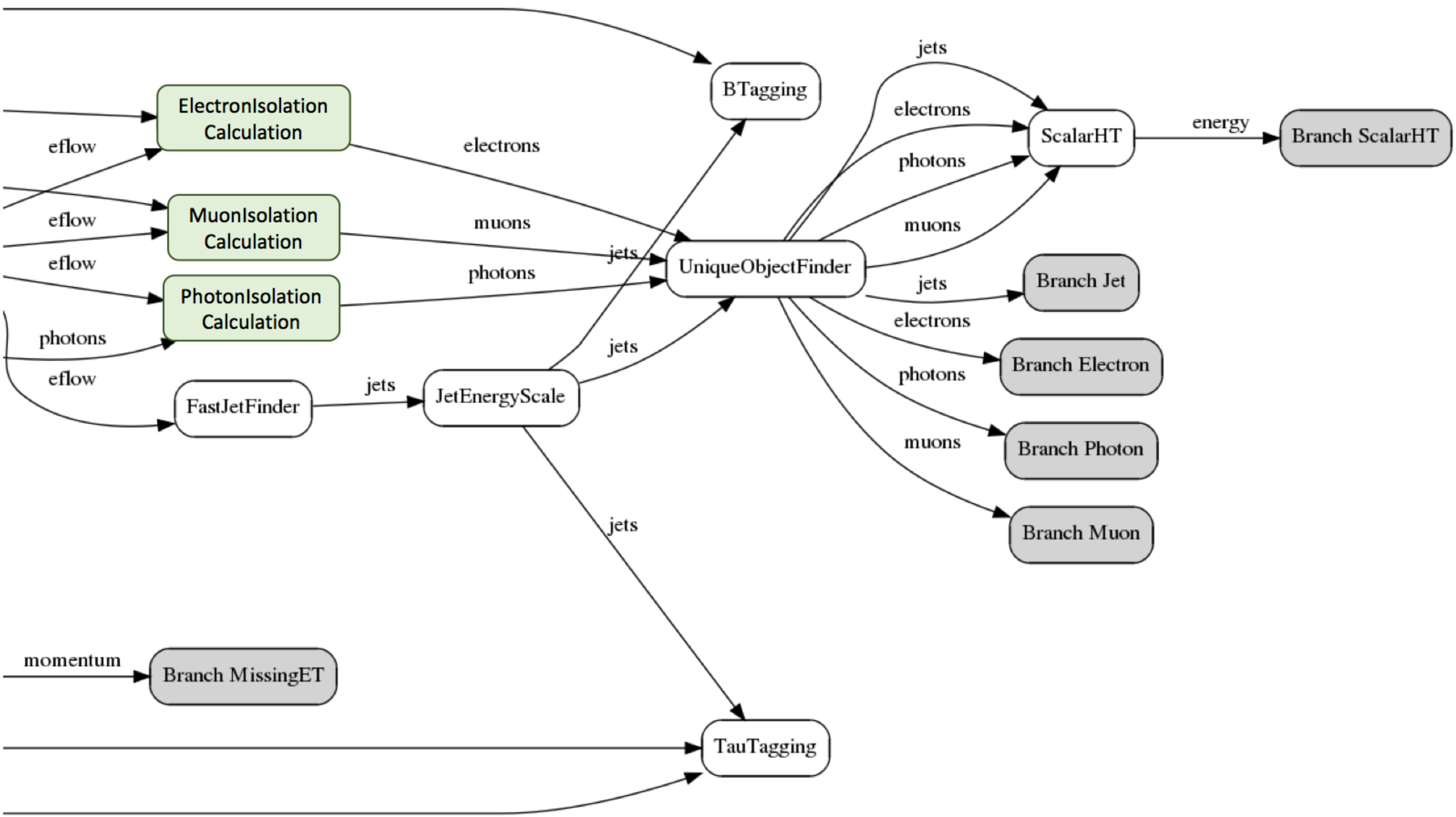
✦ The (old) MA5-tune solution

- ✤ A parametric approach included in the detector card

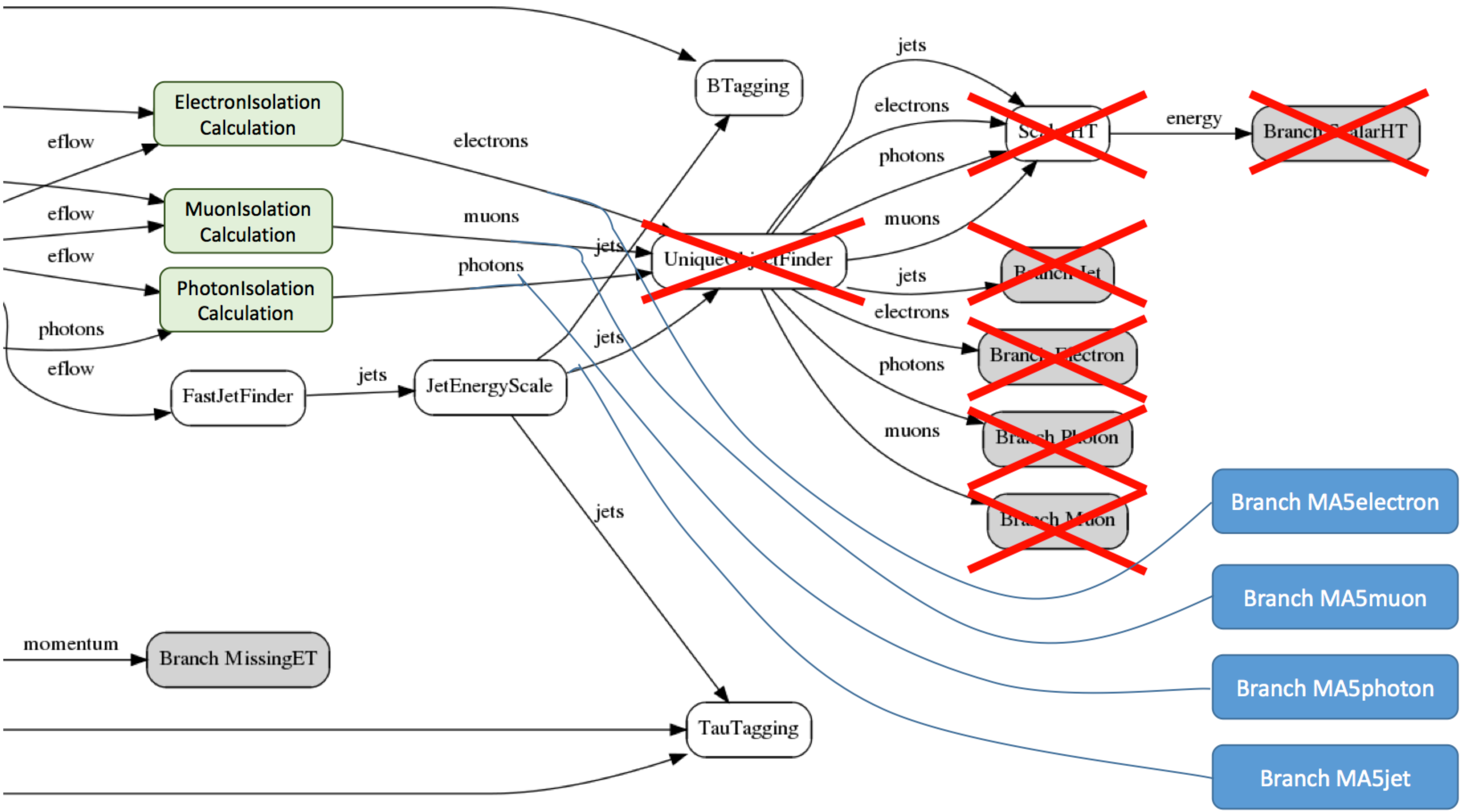
✦ The (new) MA5-tune solution

- ✤ A parametric approach to be performed at the analysis level
- ✤ A matching between jets and partons is required
  - ★ CMS: currently developed by CMS b-taggging experts themselves in Strasbourg
  - ★ ATLAS: nothing is done at the moment

Inside the code: isolation



Inside the code: output



Future developments

✦ c-tagging
- ✤ On-going studies
- ✤ Status: no implementation so far

✦ Displaced vertices
- ✤ Account for particle lifetimes
- ✤ Efficiency for track reconstruction depending on its impact parameters and pseudorapidity
- ✤ Status: existing old private DELPHES 2 code

✦ Muon electric charge mis-identification
- ✤ Important for analysis based on same-sign dileptons
- ✤ Status: existing old private DELPHES 2 code

✦ Identification of hadronically decaying tau
- ✤ Current strategy: a tau is a jet identified as a tau
- ✤ Improvement strategy 1: dedicated algorithm like in DELPHES 2
- ✤ Improvement strategy 2: efficiency and resolution effects applied on the generated taus
- ✤ Status: really needed soon

Outline

1. The MA5-tune of DELPHES 3

2. Selected items on the implementation and the validation of LHC analyses

3. Summary

Implementation and validation of ATLAS analyses

Published or to-be published	Problems with the validation	Starting validation

1. ATLAS-SUS-2013-04 (1308.1841): multijet + missing energy
   [Blanke, Fuks, Galon]
2. ATLAS-SUS-2013-05 (1308.2631): two b-jets + missing energy
   [Chalons]
3. ATLAS-SUS-2013-11 (1403.5294): two leptons + missing energy
   [Dumont]
4. ATLAS-SUS-2013-12 (1402.7029): three leptons + missing energy
   [de Causmaecker, Fuks, Mawatari]
5. ATLAS-SUS-2013-13 (1405.5086): at least four leptons + missing energy
   [Mawatari]
6. ATLAS-SUS-2013-14 (1407.0350): two hadronic taus + missing energy
   [de Causmaecker, Fuks]
7. ATLAS-SUS-2013-18 (1403.4853): at least three b-jets + missing energy
   [Mitzka, Fuks]
8. ATLAS-SUS-2013-19 (1407.0600): two leptons + b-jet(s) + missing energy
   [Chalons]

Implementation and validation of CMS analyses

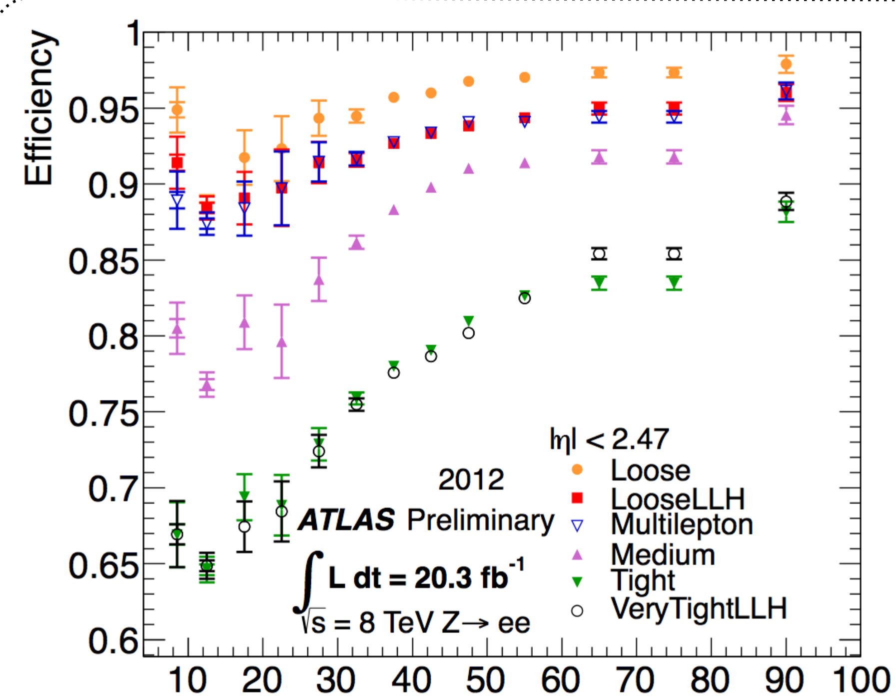
Published or to-be published | Problems with the validation | Starting validation

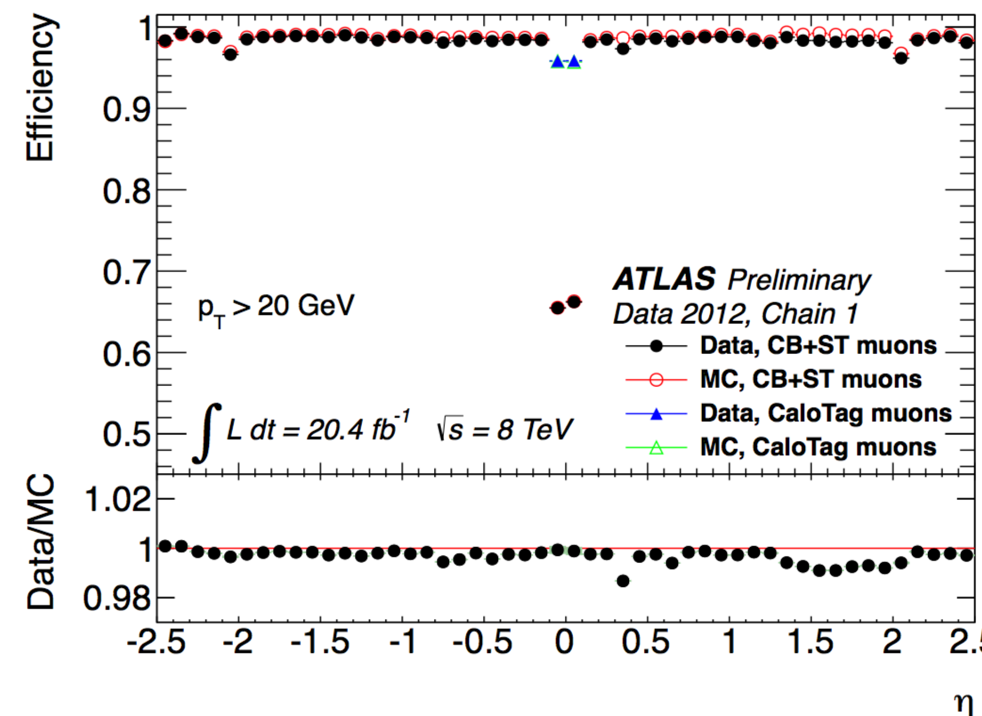
1. CMS-B2G-13-003: vector-like quarks in multileptonic events
   [Alloul, Basso, Fuks]
2. CMS-SUS-12-028 (1303.2985): α_T
   [Blanke, Fuks, Galon]
3. CMS-SUS-13-002 (1404.5801): three or more lepton + missing energy
   [Alloul, Basso, Fuks]
4. CMS-SUS-13-007 (1311.4937): single lepton + b-jets + missing energy
   [Kraml, Laa, Sengupta]
5. CMS-SUS-13-008: three leptons + 1 b-jet + missing energy
   [Alloul, Basso, Fuks]
6. CMS-SUS-13-011 (1308.1586): stops in the single lepton mode
   [Dumont, Fuks, Wymant]
7. CMS-SUS-13-012 (1402.4770): multijets and missing energy
   [Bein, Sengupta]
8. CMS-SUS-13-013 (1311.6736): same-sign dilepton + jets
   [Alloul, Basso, Fuks]
9. CMS-SUS-13-015: stops in multijet + missing energy
   [Bein, Kraml]
10. CMS-SUS-13-016: opposite sign dilepton, b-jets and missing energy
    [Sengupta, Kulkarni]

Lepton identification in ATLAS analyses

[Chalons]

- Original implementation of ATLAS-SUS-2013-05: large differences with ATLAS results
  - Discussions with ATLAS people
  - Update of the lepton efficiencies (from ATLAS-CONF-2014-032)
  - Update of the b-tagging identification and misidentification rates (see next slides too)





$$\varepsilon_e^{\rm med.} = \begin{cases} p_T < 90 : 1.66.10^{-11} p_T^6 - 5.71.10^{-9} p_T^5 + 8.09.10^{-7} p_T^4 - 5.88.10^{-5} p_T^3 \\ \qquad + 2.2.10^{-3} p_T^2 - 3.46.10^{-2} p_T + 0.96 \\ p_T > 90 : 0.95 \end{cases}$$

$$\varepsilon_e^{\rm med.} / \varepsilon_e^{\rm tight} = \begin{cases} p_T < 90 : 2.34.10^{-7} p_T^3 - 4.09.10^{-5} p_T^2 + 3.15.10^{-3} p_T + 0.81 \\ p_T > 90 : 0.93 \end{cases}$$

$$\varepsilon_\mu = \begin{cases} |\eta| \leq 0.1 \ \& \ p_T > 6 : 0.65 \\ 0.1 < |\eta| \leq 2.5 \ \& \ p_T > 6 : 0.99 \\ |\eta| > 2.5 : 0 \end{cases}$$

Description of ATLAS-SUS-2013-05 (I)

[Chalons]

Description	Signal Regions	
	SRA	SRB
Event cleaning	Common to all SR	
Lepton veto	No e/μ after overlap removal with $p_{\mathrm{T}} > 7(6)$ GeV for $e(\mu)$	
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 150 GeV	> 250 GeV
Leading jet $p_{\mathrm{T}}(j_1)$	> 130 GeV	> 150 GeV
Second jet $p_{\mathrm{T}}(j_2)$	> 50 GeV,	> 30 GeV
Third jet $p_{\mathrm{T}}(j_3)$	veto if > 50 GeV	> 30 GeV
$\Delta\phi(\boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}}, j_1)$	-	> 2.5
b-tagging	leading 2 jets ($p_{\mathrm{T}} > 50$ GeV, $\|\eta\| < 2.5$)	2nd- and 3rd-leading jets ($p_{\mathrm{T}} > 30$ GeV, $\|\eta\| < 2.5$)
	$n_{b\text{-jets}} = 2$	
$\Delta\phi_{\mathrm{min}}$	> 0.4	> 0.4
$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(k)$	$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(2) > 0.25$	$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(3) > 0.25$
m_{CT}	$> 150, 200, 250, 300, 350$ GeV	-
$H_{\mathrm{T},3}$	-	< 50 GeV
m_{bb}	> 200 GeV	-

Table 1. Summary of the event selection in each signal region.

$$\Delta\phi_{min} = \min(|\phi_1 - \phi_{\not{p}_T}|, |\phi_2 - \phi_{\not{p}_T}|, |\phi_3 - \phi_{\not{p}_T}|)$$

$$m_{eff}(k) = \sum_{i=1}^{k} (p_T^{jets})_i + \not{E}_T$$

$$H_{T,3} = \sum_{i=4}^{n} (p_T^{jets})_i$$

Description of ATLAS-SUS-2013-05 (2)

- No LHE input files were provided by the ATLAS collab.
- Simulate the signal sample through MadGraph5_v1.4.8+PYTHIA6 then passed to DELPHESMA5TUNE using generic official SLHA files provided by the ATLAS SUSY conveners

- For SRA four benchmark point are given for validation
  - Two for the cutflows
    - $(m_{\tilde{b}_1}, m_{\tilde{\chi}_1^0}) = (500, 1)$ GeV
    - $(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^\pm}, m_{\tilde{\chi}_1^0}) = (500, 120, 100)$ GeV
  - Two for the distributions
    - $(m_{\tilde{b}_1}, m_{\tilde{\chi}_1^0}) = (500, 1)$ GeV
    - $(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^\pm}, m_{\tilde{\chi}_1^0}) = (500, 105, 100)$ GeV

- For SRB four benchmark point are given for validation
  - Two for the cutflows
    - $(m_{\tilde{b}_1}, m_{\tilde{\chi}_1^0}) = (350, 320)$ GeV
    - $(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^\pm}, m_{\tilde{\chi}_1^0}) = (500, 420, 400)$ GeV
  - Two for the distributions
    - $(m_{\tilde{b}_1}, m_{\tilde{\chi}_1^0}) = (300, 200)$ GeV
    - $(m_{\tilde{t}_1}, m_{\tilde{\chi}_1^\pm}, m_{\tilde{\chi}_1^0}) = (250, 155, 150)$ GeV

Some results for ATLAS-SUS-2013-05

[Chalons]

✦Exemplary cut-flow tables

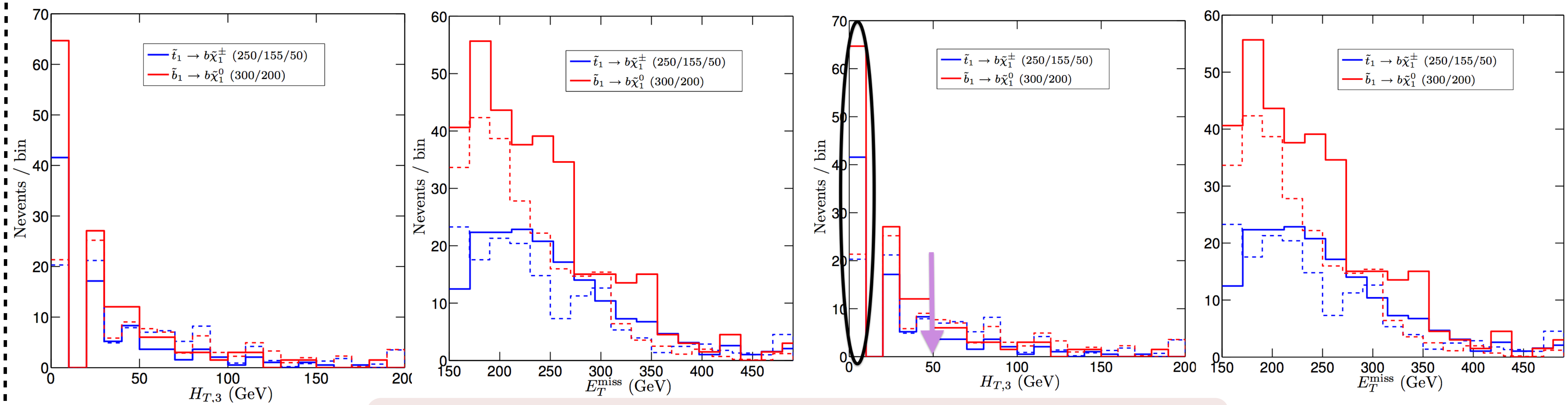
cut	$m_{\tilde{b}_1} = 500$ GeV ATLAS	MA 5	rel.	$m_{\tilde{t}_1} = 500$ GeV ATLAS	MA5	rel.
$E_T^{\rm miss} > 80$ GeV filter	1606.0	1628.2	+1.38%	1632.0	1585.2	−2.87%
+ Lepton veto	1505.0	1223.5	−18.7%	1061.0	863.2	−18.6%
+ $E_T^{\rm miss} > 150$ GeV	1323.0	1052.2	−20.5%	859.0	696.3	−18.9%
+ Jet Selection	119.0	142.3	+19.6%	39.0	47.6	+22.0%
+ $M_{bb} > 200$ GeV	96.0	116.5	+21.4%	32.0	38.8	+21.5%
+ $M_{CT} > 150$ GeV	82.0	97.5	+18.9%	26.8	31.7	+18.3%
+ $M_{CT} > 200$ GeV	67.0	80.7	+20.4%	20.2	24.5	+21.3%
+ $M_{CT} > 250$ GeV	51.0	60.8	+19.0%	13.2	16.6	+25.8%
+ $M_{CT} > 300$ GeV	35.0	42.3	+20.9%	7.7	9.2	+19.5%

cut	$m_{\tilde{b}_1} = 350$ GeV ATLAS	MA 5	rel.	$m_{\tilde{t}_1} = 500$ GeV ATLAS	MA5	rel.
$E_T^{\rm miss} > 80$ GeV filter	6221.0	5963.7	−4.13%	1329.0	1117.9	−15.9%
+ Lepton veto	4069.0	4450.4	+9.37%	669.0	702.9	+5.07%
+ $E_T^{\rm miss} > 250$ GeV	757.0	724.5	−4.29%	93.0	86.8	−6.67%
+ Jet Selection	7.9	7.5	−5.06%	6.2	5.7	−8.06%
+ $H_{T,3} < 50$ GeV	5.2	6.6	+26.9%	3.0	4.6	+53.3%

Lepton veto issues:

★ Might be related to lepton isolation (see next slides too)

✦Exemplary figures



Not enough jet activity for SRB

★ Might be related to JES (see next slides)

Limit settings for ATLAS-SUS-2013-05

[Chalons]

benchmark	σ_{95} MA5	σ_{95} ATLAS	bestSR MA5	bestSR ATLAS
b300_n200	0.75 pb	1.17 pb	SRB	SRB
b350_n320	3.31 pb	4.24 pb	SRB	SRB
b500_n1	0.0166 pb	0.0298 pb	SRA MCT_{300}	SRA MCT_{300}
t250_c155_n150	1.30 pb	3.65 pb	SRB	SRB
t500_c105_n100	0.0253 pb	0.0284 pb	SRA MCT_{300}	SRA MCT_{200}
t500_c120_n100	0.0503 pb	0.117 pb	SRA MCT_{250}	SRA MCT_{300}
t500_c420_n400	0.51 pb	2.56 pb	SRB	SRB

benchmark	(1-CLs)% MA5	(1-CLs)% ATLAS
b300_n200	100.0%	99.8%
b350_n320	35.9%	32.5%
b500_n1	100.0%	99.7%
t250_c155_n150	100.0%	99.1%
t500_c105_n100	100.0%	98.8%
t500_c120_n100	99.8%	90.2%
t500_c420_n400	24.8%	14.9%

Overall our recasted analysis is more constraining than the ATLAS one.

B-tagging in ATLAS analyses: ATLAS-SUS-2103-19

[Chalons]

✦ Improved description of the b-taggging with the implementation of ATLAS-SUS-2013-19

- Update of the b-tagging efficiency from ATLAS-CONF-2014-004
- Foreseen update with the b-jet misidentification rates from ATLAS-CONF-2014-046

✦ Description of ATLAS-SUS-2013-19

Two OS 10 GeV preselected leptons				
lepton isolation				
$m_{\ell\ell} > 20$ GeV				
p_T leading lepton > 25 GeV				
Leptonic M_{T2}				Hadronic M_{T2}
Z veto $\Delta\phi > 1$ $\Delta\phi_b < 1.5$				# b-jets $= 2$ $M_{T2}^{\mathrm{b-jet}} > 160$ $M_{T2} < 90$ $1^{st}\ell$ $p_T < 60$
$M_{T2} > 90$	$M_{T2} > 100$	$M_{T2} > 110$	$M_{T2} > 120$	
-	1^{st} jet $p_T > 100$	1^{st} jet $p_T > 20$	-	-
-	2^{nd} jet $p_T > 50$	2^{nd} jet $p_T > 20$	-	-

✦ Exemplary cut-flow table

cut	# events (scaled to 100000.0)	rel.	# events (official)
Initial number of events	100000.0	100000.0	
Two 10 GeV SF preselected ℓ	2893.0	-97.1%	2212.9
lepton isolation	657.3	-77.3%	1646.1
opposite sign leptons	657.3	-0.0%	1594.0
$m_{ll} > 20$ GeV	607.5	-7.6%	1506.0
Trigger lepton p_T requirement	507.5	-16.5%	1319.0
2 b-jets	209.3	-58.8%	529.9
$m_{T2}^{\mathrm{b-jet}} > 160$ GeV	28.1	-86.6%	42.3
$m_{T2} < 90$ GeV	28.1	-0.0%	42.3
leading lepton $p_T < 60$ GeV	22.3	-20.6%	29.9

Lepton isolation issues

Jet energy scale in the ATLAS card

- Original implementation of ATLAS-SUS-2013-04: large differences with ATLAS
  - Discussions with ATLAS people
  - Design of JES correction functions based on the Monte Carlo truth and DELPHES-MA5TUNE

$$\begin{cases} p_T > 100.0 : 1.00/(1.00 - 0.015874205774624516 - 1.5596526607501018 * \ln(p_T)/p_T) \\ p_T \leq 100.0 : 1.00/(1.00 - 0.28148029547368\ \ 0.019155389997112204 * \ln(95.6961580995732 * p_T)) \end{cases}$$

- Exemplary cut-flow table [ATLAS-SUS-2013-14: Blanke, Fuks, Galon]

Cut	# events MADANALYSIS 5	# events (official)
Initial number of events	206.3	206.3
6 jets with $E_T > 45$ GeV	150.3	168
lepton veto	89.4	78
8 jets ($p_T > 50$ GeV)	15.0	16.3
$E_T/\sqrt{H_T} > 4$ GeV$^{1/2}$	12.4	14.1
→ without b-tags	0.51	0.85
→ with 1 b-tag	2.1	3
→ with 2 b-tags	9.7	11
7 jets ($p_T > 80$ GeV)	7.44	7.53
$E_T/\sqrt{H_T} > 4$ GeV$^{1/2}$	6.08	6.25
→ without b-tags	0.14	0.31
→ with 1 b-tag	1.0	1.3
→ with 2 b-tags	4.9	5.1
≥ 8 jets ($p_T > 80$ GeV)	2.6	3.2
$E_T/\sqrt{H_T} > 4$ GeV$^{1/2}$	2.0	2.6
→ without b-tags	0.02	0.13
→ with 1 b-tag	0.23	0.55
→ with 2 b-tags	1.7	2.1

Good agreement for many signal regions

- Not working for some ATLAS analyses; improvements for some other
  - Further studies are on-going, based on well-controlled Monte Carlo samples

Lepton isolation and overlap removal in ATLAS analyses

✦ Most of the analyses dealing with leptons have problems

cut	our nev	exp events
Initial number of events	55.0	55.0
4 leptons	28.1	36.7
0 tau	25.6	21.4
trigger	25.6	21.3
Z-veto SFOS	15.9	11.6
etmiss > 50	13.4	10.0

[ATLAS-SUS-2013-13: Mawatari]

cut	# events (scaled to 80000.0)	# events (official)
Initial number of events	80000.0	
Two 10 GeV SF preselected ℓ	1881.2	3811.0
lepton isolation	1031.8	3197.0
opposite sign leptons	1031.8	3167.0
$m_{ll} > 20$ GeV	1013.9	3144.0
Trigger lepton p_T requirement	1005.0	3253.5
Z veto	777.3	2463.6
$\Delta\phi > 1$	654.5	1834.9
$\Delta\phi_b < 1.5$	479.8	1402.8
$m_{T2} > 90$ GeV	155.1	396.5

[ATLAS-SUS-2013-19: Chalons]

cut	our nev	exp events
Initial number of events	28033.7	28033.7
triggers	15759.1	-
dR(leptons) > 0.3	15742.3	-
3 signal leptons	1823.6	-
1 electron or muon	1823.6	-
no taus	583.1	-
0 sfos	176.6	19.0
b jet veto	175.2	18.0
etmiss > 50	91.1	12.0
3rd lepton pt > 20	32.2	7.0
min dphi(llp) $<= 1.0$	15.4	5.0

[ATLAS-SUS-2013-12: de Causmaecker, Fuks, Mawatari]

✦ Lepton identification is a problem

- ✤ On-going discussions with ATLAS people
- ✤ Status: object overlap removal might be the issue

Hadronic taus

✦ Some of our recasted analyses employ hadronically-decaying tau

cut	our nev	exp events
Initial number of events	4384.8	4384.8
triggers	1187.8	-
dR(leptons) > 0.3	1186.7	-
3 signal leptons	114.4	-
1 electron or muon	114.4	-
two taus	32.2	48.0
b jet veto	31.8	46.0
etmiss > 50	23.7	35.0
mt2 max > 100	7.7	14.0

[ATLAS-SUS-2013-12: de Causmaecker, Fuks, Mawatari]

cut	sim events	exp events
Initial number of events	30000.0	29500.0
at least two leptons	4630.0	1499.1
at least two taus and e/mu veto	806.0	352.5
ditau trigger	576.0	175.6
exactly 2 os taus, jet/z-veto	34.0	22.4
mt2 > 30 gev	14.0	16.1
mttau1 + mttau2 > 250 gev	8.0	8.5

[ATLAS-SUS-2013-14: de Causmaecker, Fuks]

✦ Electron and muon isolation might be the problem

- ✤ Further investigations are necessary
- ✤ Status: object overlap removal might be the issue

Outline

1. The MA5-tune of DELPHES 3

2. Selected items on the implementation and the validation of LHC analyses

3. Summary

Summary

✦ The MA5-tune of MadAnalysis 5

- ✤ New lepton and photon isolation module
- ✤ Cleaning of the output file
- ✤ Future novel b-tagging startegy
- ✤ Developments in the pipeline: lepton charge misidentification, c and tau tagging, displaced vertices

✦ Applications to several ATLAS and CMS analyses

- ✤ Good agreement for CMS analyses
  - ★ Current problems with SUS-13-013 under investigation
- ✤ Lots of troubles with ATLAS analyses
  - ★ JES issues under investigation
  - ★ Lepton identification and object overlap removal issues under investigation
  - ★ Maybe tau issues under investigation