Five-flavour scheme predictions for $t\bar{t}b\bar{b}$ at next-to-leading order accuracy

Tetiana Moskalets
Southern Methodist University

F9 HEP Seminar
30 August 2024, Jozef Stefan Institute, Ljubljana

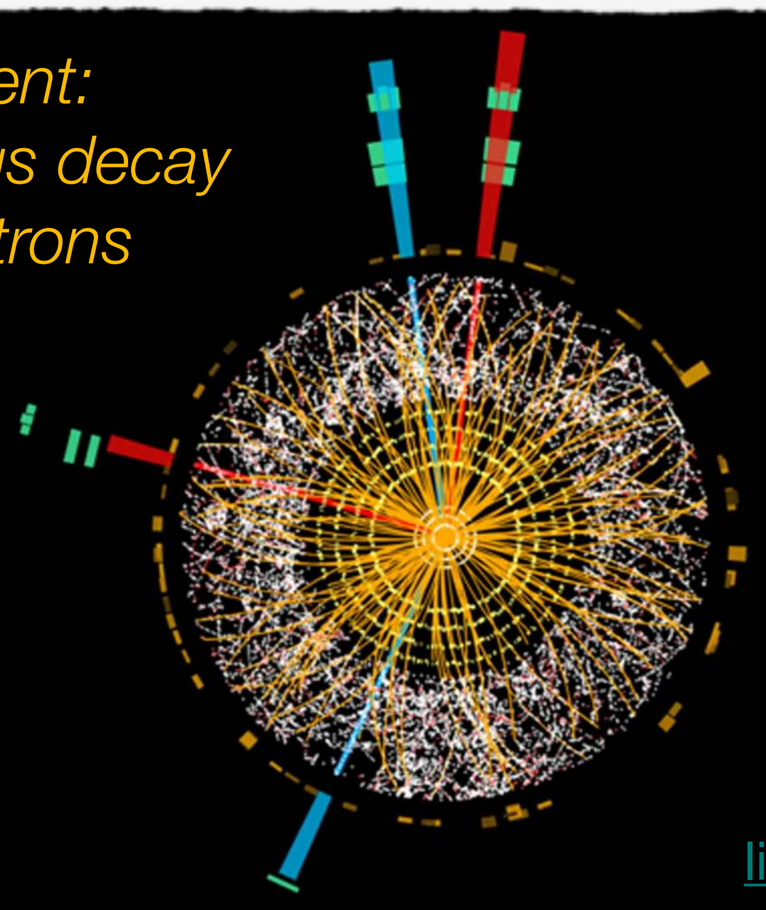
SMU®

Where to start from?

Higgs boson discovery

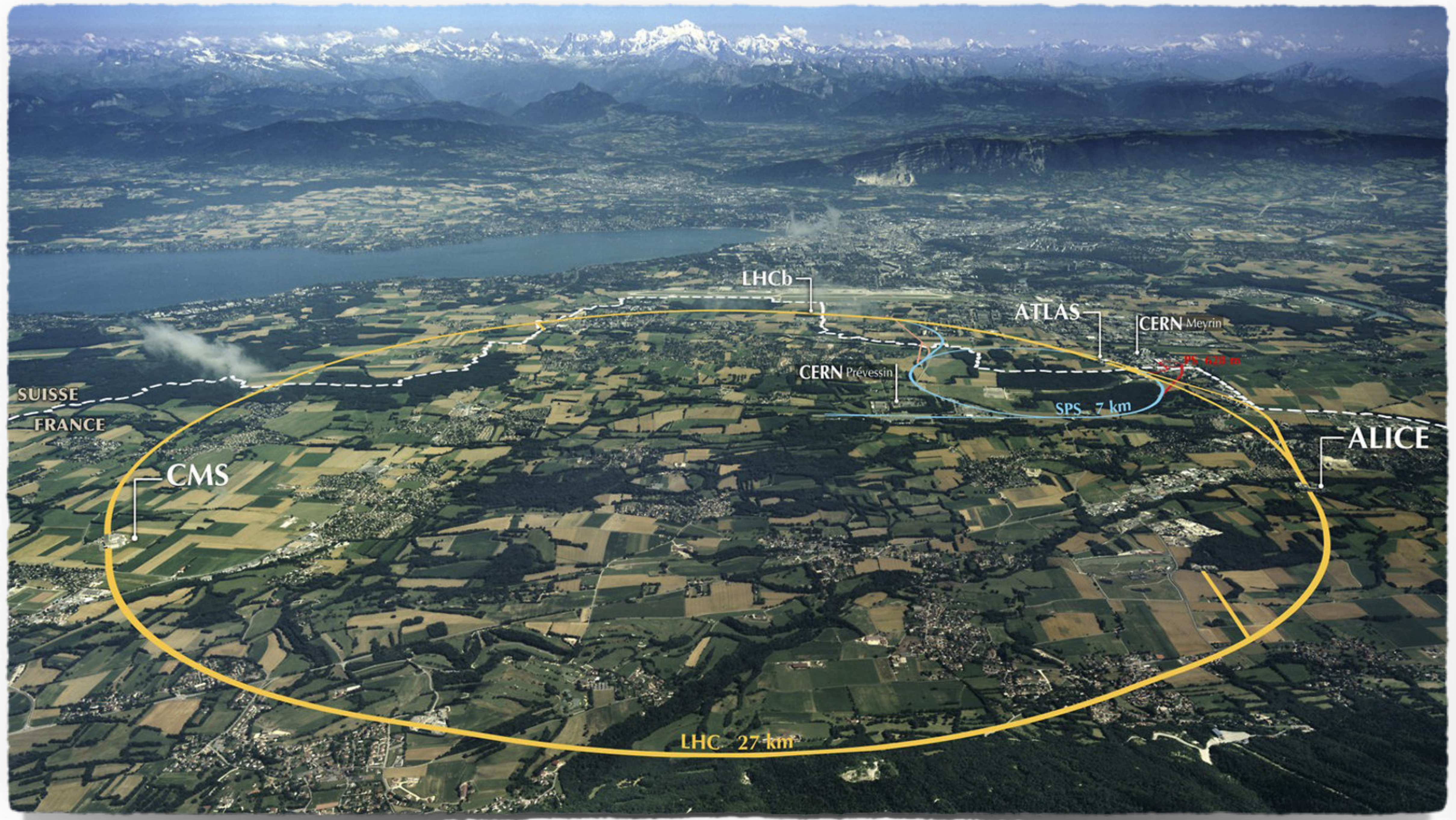
- It all started in 2012 with the Higgs boson discovery by ATLAS and CMS at the LHC
- Since then, a program of detailed measurements of Higgs boson properties and couplings has been launched

ATLAS event:
candidate Higgs decay
to four electrons



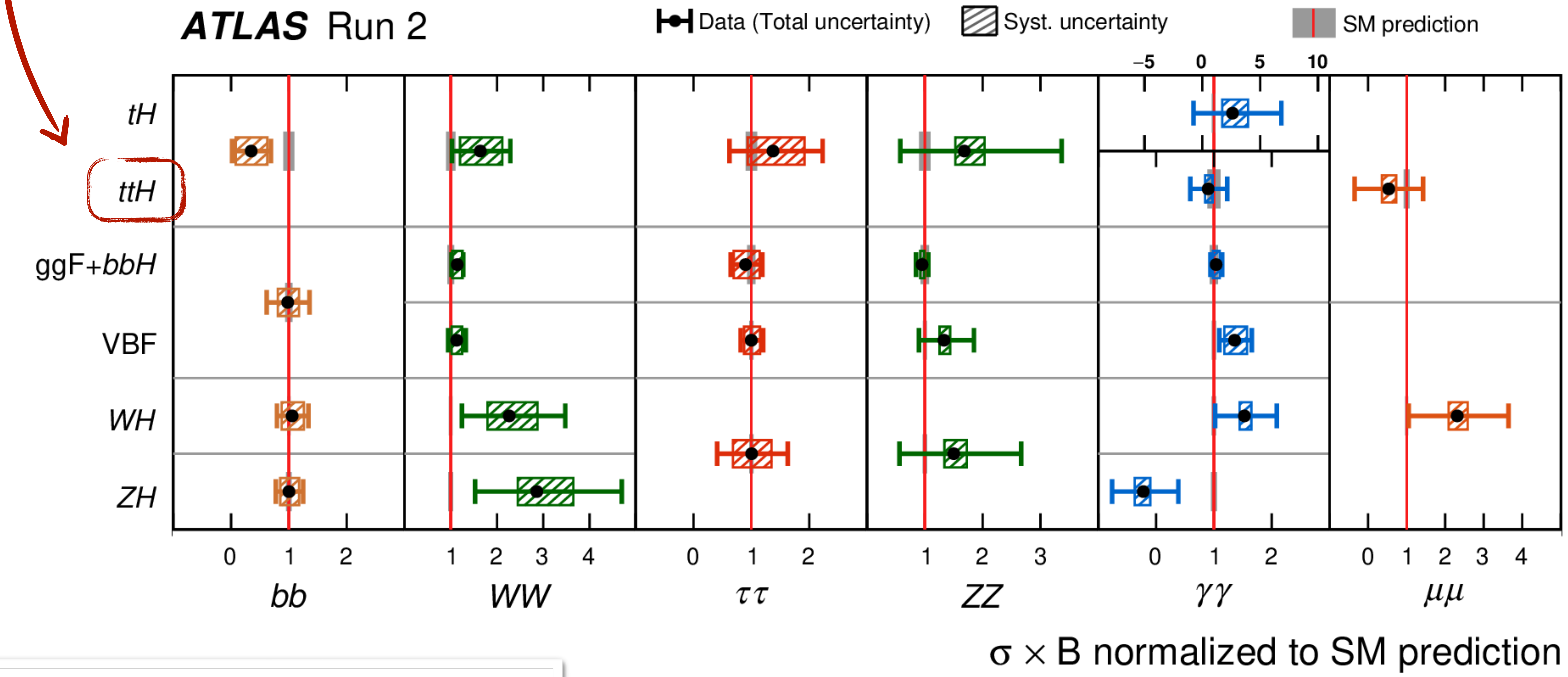
link

link

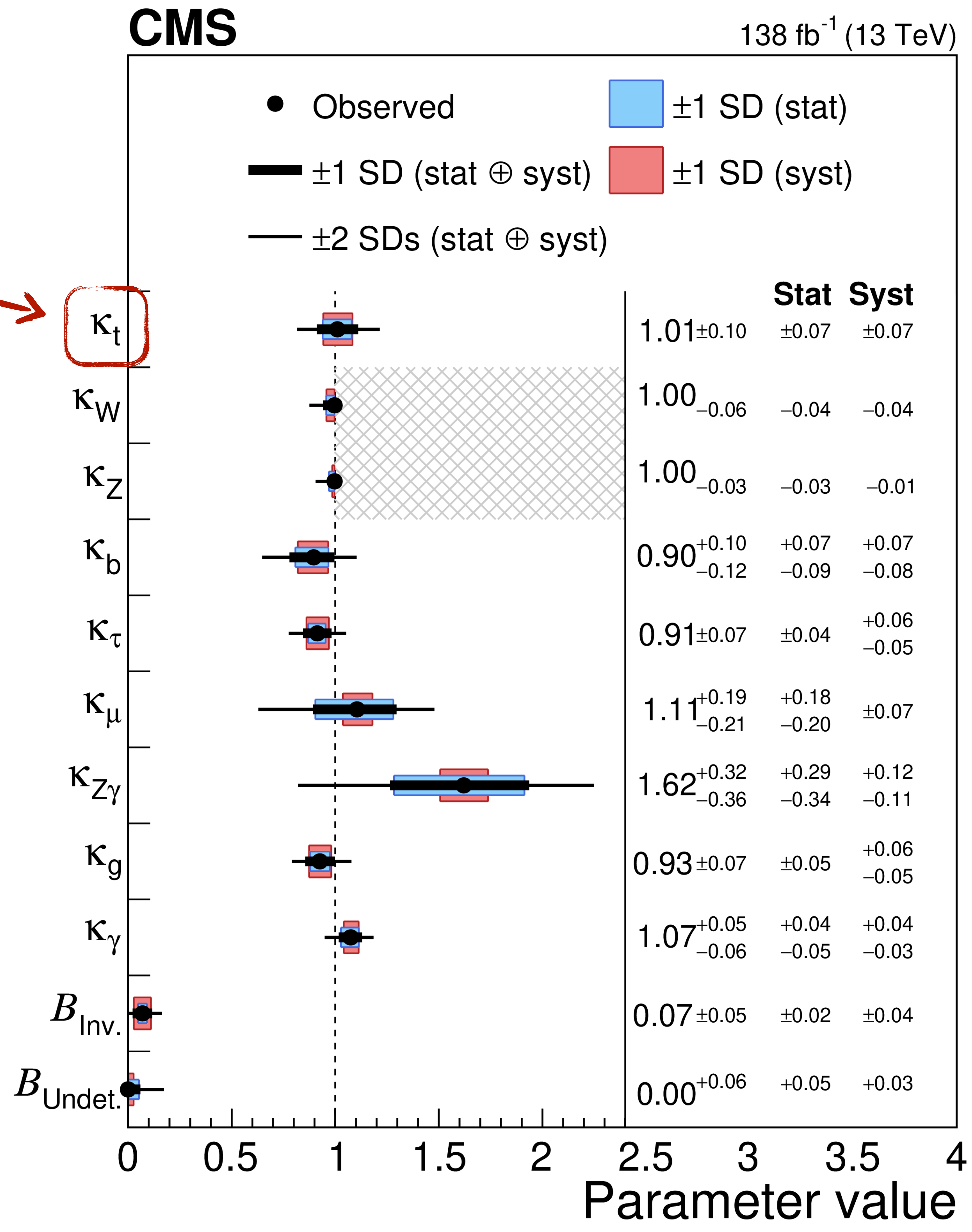


Scrutinising Higgs boson interactions

- Interactions of the Higgs boson with gauge bosons and third-generation matter particles are well measured
- In particular, **top-quark Yukawa coupling** can be extracted from the measurements of Higgs production in association with a $t\bar{t}$ pair
  - Can be also constrained from the $t\bar{t}t\bar{t}$ production



ATLAS Nature 607 (2022) 7917, 52-59

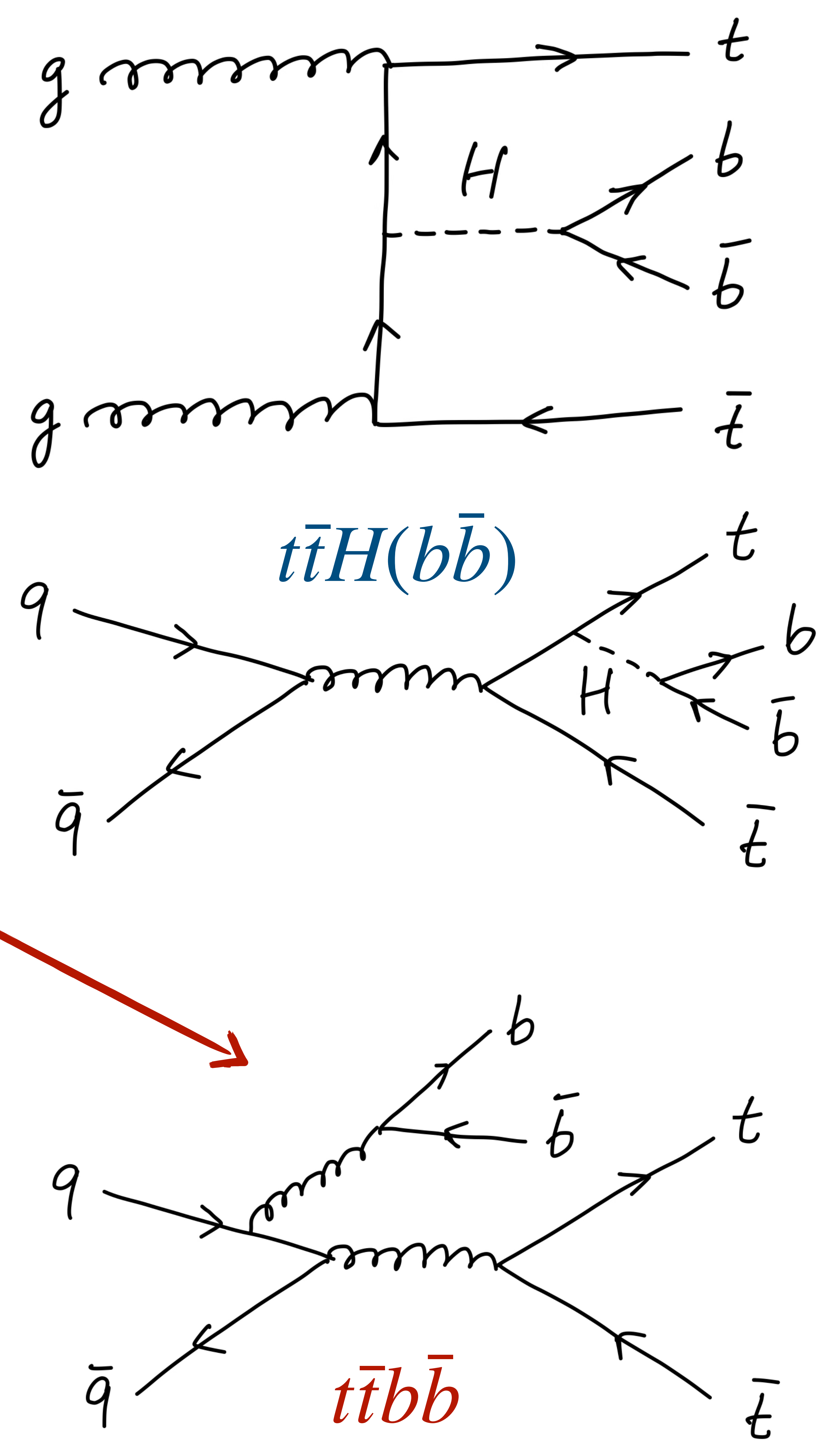


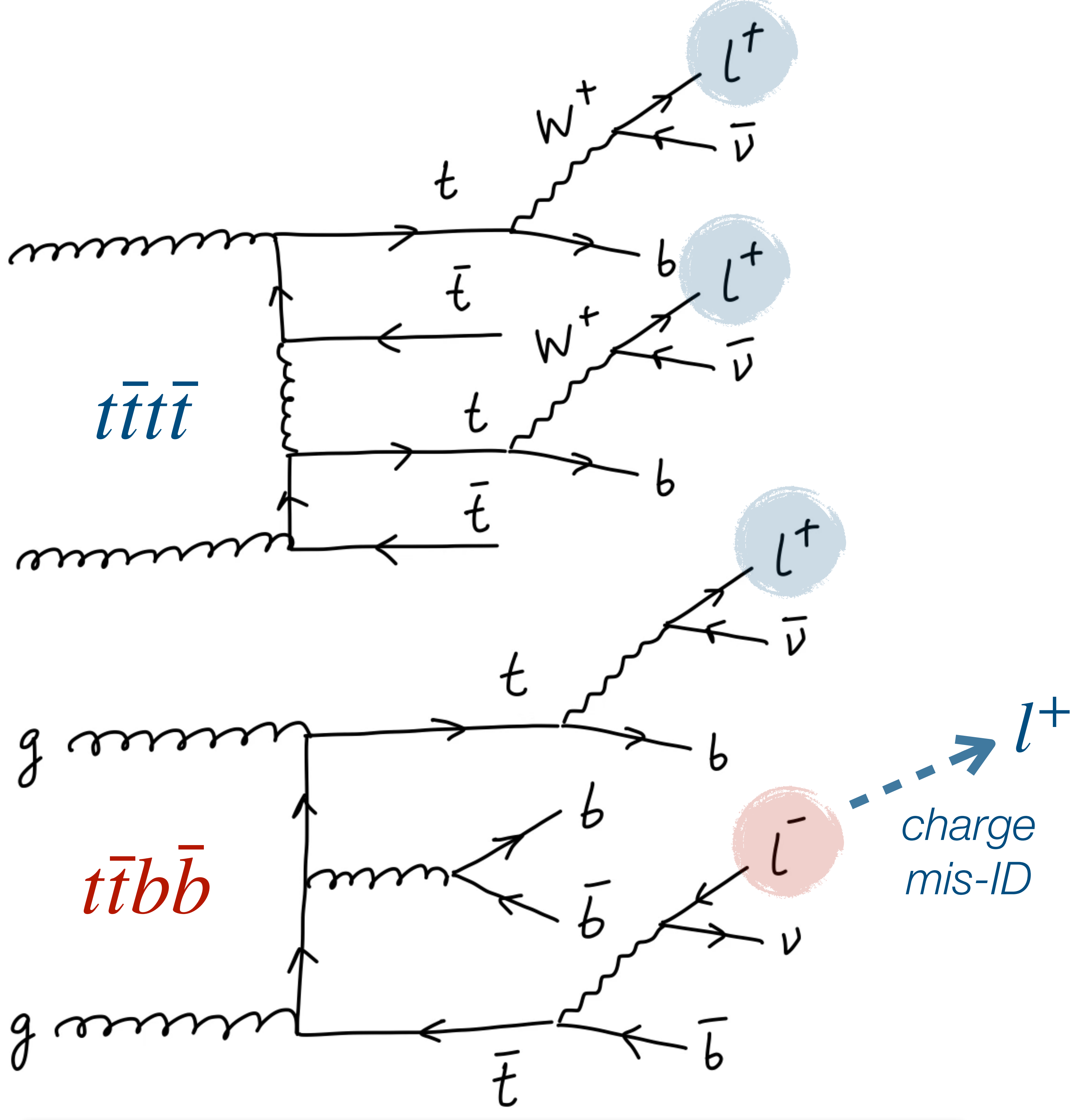
CMS Nature 607 (2022) 7917, 60-68

(One of the) experimental challenges of $t\bar{t}H(\rightarrow b\bar{b})$ and $t\bar{t}t\bar{t}$ analyses

latest $t\bar{t}H(\rightarrow b\bar{b})$ from ATLAS PLB 849 (2024)
and CMS JHEP 05 (2024) 042

- In $t\bar{t}H(\rightarrow b\bar{b})$ analysis one needs to discriminate between the signal process and the large background from $t\bar{t}$ + jets
  - In particular, with heavy-flavour jets
- QCD production of $t\bar{t}b\bar{b}$ is an **irreducible background**
  - ➡ Its precise Monte-Carlo (MC) simulation is of crucial importance!
- Uncertainty on the $t\bar{t}b\bar{b}$ modelling is currently a limitation for the $t\bar{t}H(\rightarrow b\bar{b})$ measurements





latest $t\bar{t}t\bar{t}$ from ATLAS EPJC 83 (2023) 6, 496
and CMS PLB 847 (2023) 138290

- In ATLAS $t\bar{t}t\bar{t}$ analysis, events are required to have one same-sign lepton pair
- $t\bar{t}b\bar{b}$ events can fake the $t\bar{t}t\bar{t}$ events if the charge of one lepton from the top decay is mis-identified in the detector

A little digression… Let’s introduce some MC-related terminology

An LHC collision as a MC event

Matrix element (ME):
MC integration
over phase-space
(MadGraph, Powheg, Sherpa, ...)

Parton shower (PS):
Markov chain evolution
(Pythia8, Herwig7, Sherpa, ...)

Matching:
combination of perturbative
QCD results with the PS
(MC@NLO, Powheg, ...)

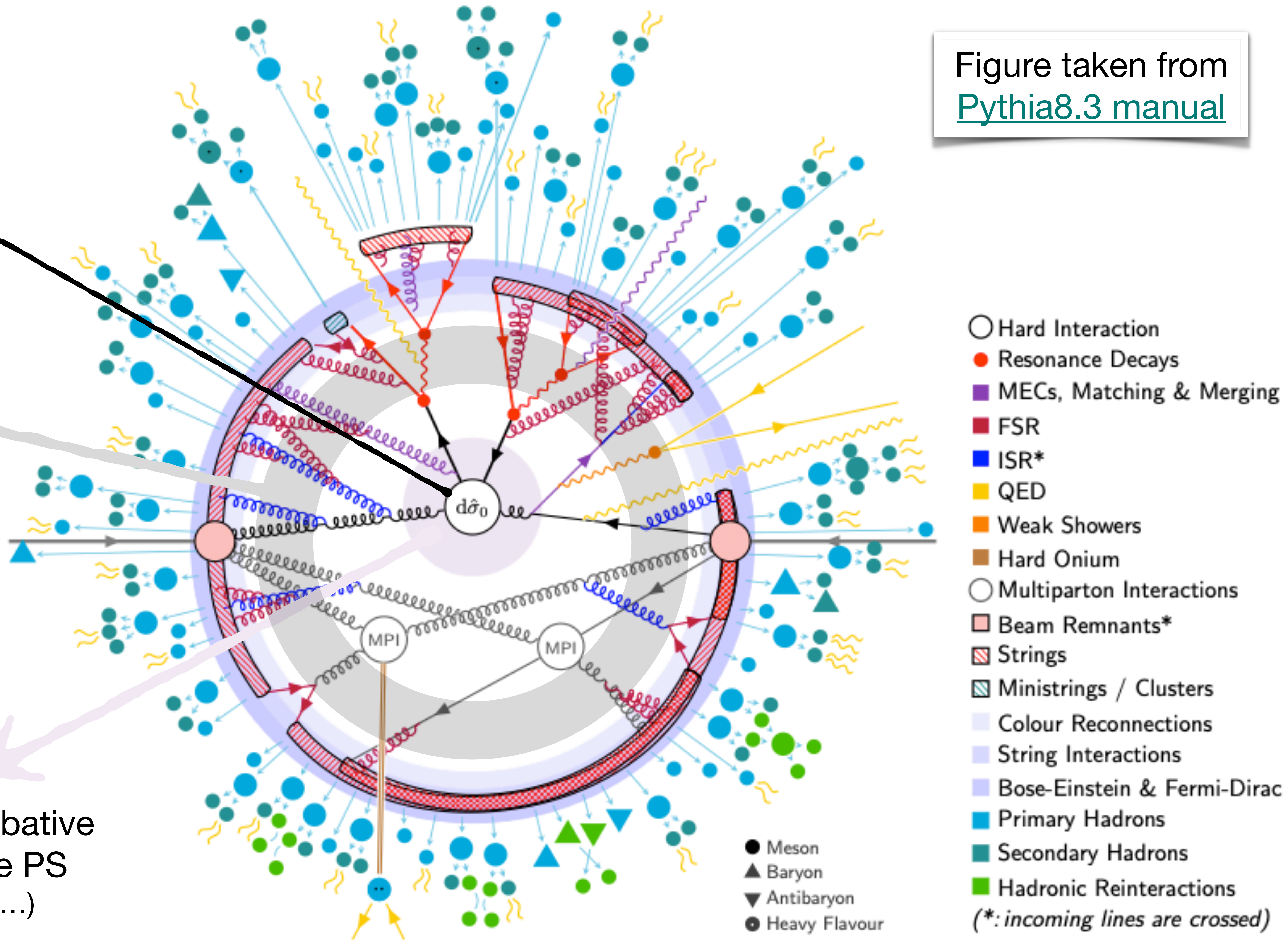


Figure taken from Pythia8.3 manual

schematic of the structure of a $pp \to t\bar{t}$ *event*

Fixed order calculations

Master formula for hadron collisions:

$$\sum_{a,b} \int dx_1 dx_2\, d\Phi_{\mathrm{FS}}\; f_a(x_1, \mu_\mathrm{F}) f_b(x_2, \mu_\mathrm{F})\; \sigma_{ab\to X}(s, \mu_\mathrm{F}, \mu_\mathrm{R})$$

phase-space integral

factorisation scale:
scale for absorbing IR divergent parton emissions into the PDF

parton density functions (PDFs)

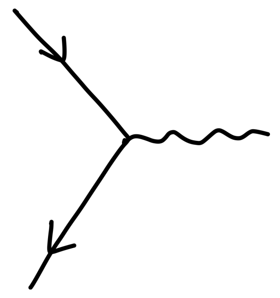
renormalisation scale:
scale for renormalisation of UV loop divergencies due to truncation of perturbartive series

perturbative expansion in α_S

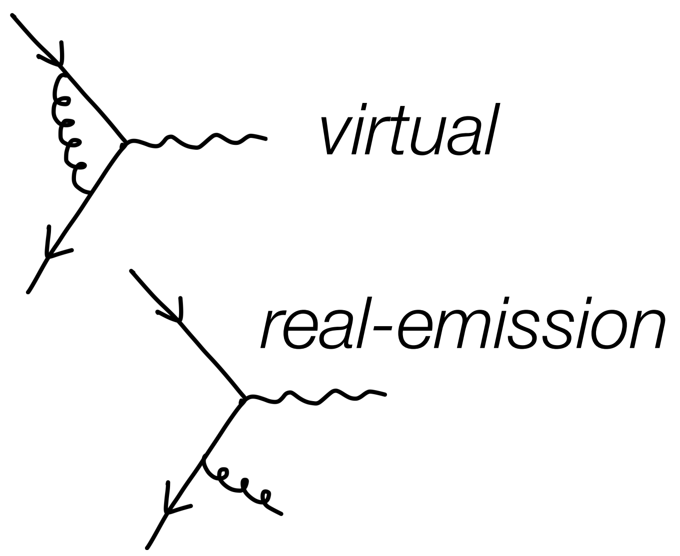
parton-level cross section

$$\sigma = \sigma^{\mathrm{Born}} \left(1 + \frac{\alpha_S}{2\pi}\sigma^{(1)} + \left(\frac{\alpha_S}{2\pi}\right)^2 \sigma^{(2)} + \dots \right)$$

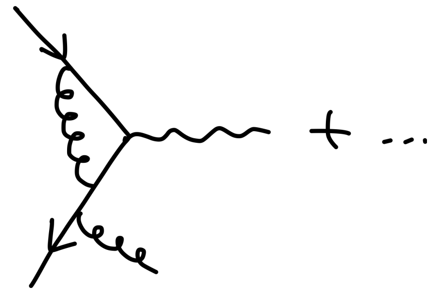
LO prediction

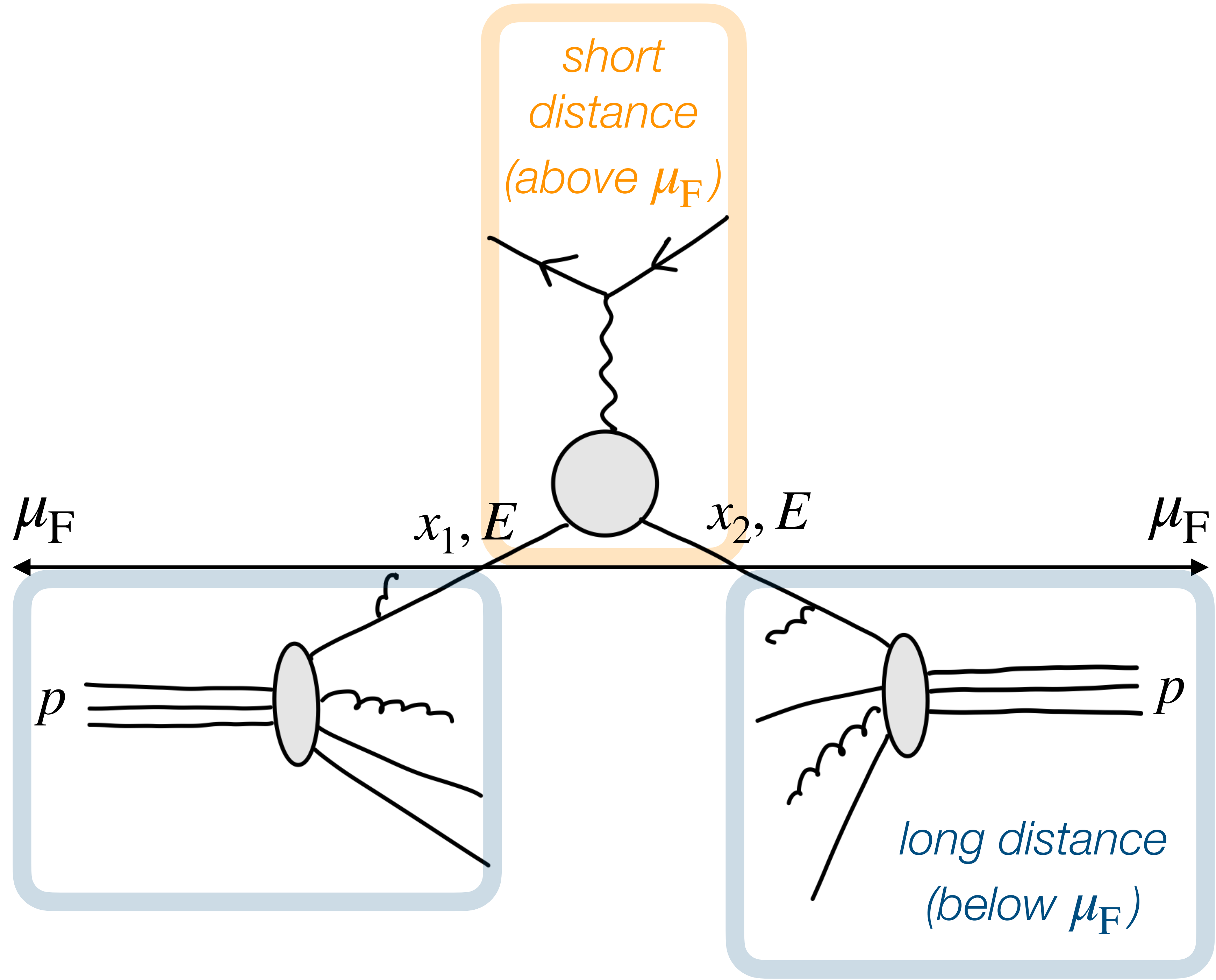


NLO corrections



NNLO corrections





- **Fixed order predictions** give an inclusive result for that process + anything else (any extra radiation) below μ_F
- **Parton shower** makes all the radiation exclusive
  - Also, resums soft/collinear radiation, adds hadronisation, etc, etc…
- We want to combine the two

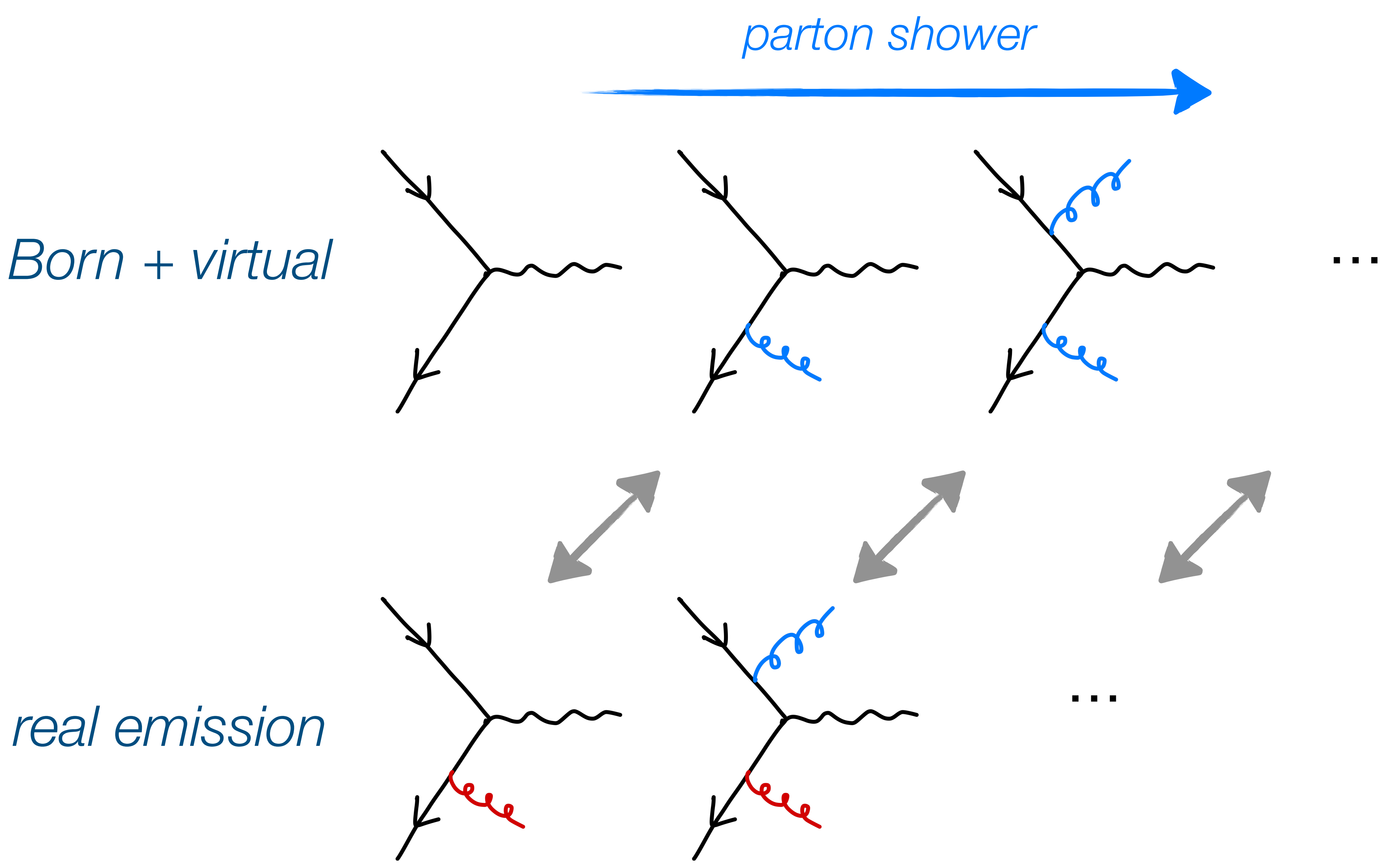
Matching NLO calculations to the parton shower

- **NLO calculation** includes real-emission and virtual corrections
  - Both corrections are separately divergent, but their sum is finite for the sufficiently inclusive *(IR-safe)* observables

- **Parton shower** also has both types of radiation
  - Soft and collinear radiation is resummed to all orders
    ➡ *Sudakov suppression*
  - Real and virtual/unresolved corrections are also assumed to cancel after integration over phase space
    - Like in the NLO matrix element

Sudakov form factor:
- Describes the no-emission probability
- Used by all PS generators (in analytic or numerical form)

- **Sources of double counting** between the ME and PS:
  - PS can produce the same extra radiation as the real emission ME
  - There is also an overlap between the virtual corrections in the ME and the Sudakov suppression in the PS



Matching NLO calculations to the parton shower

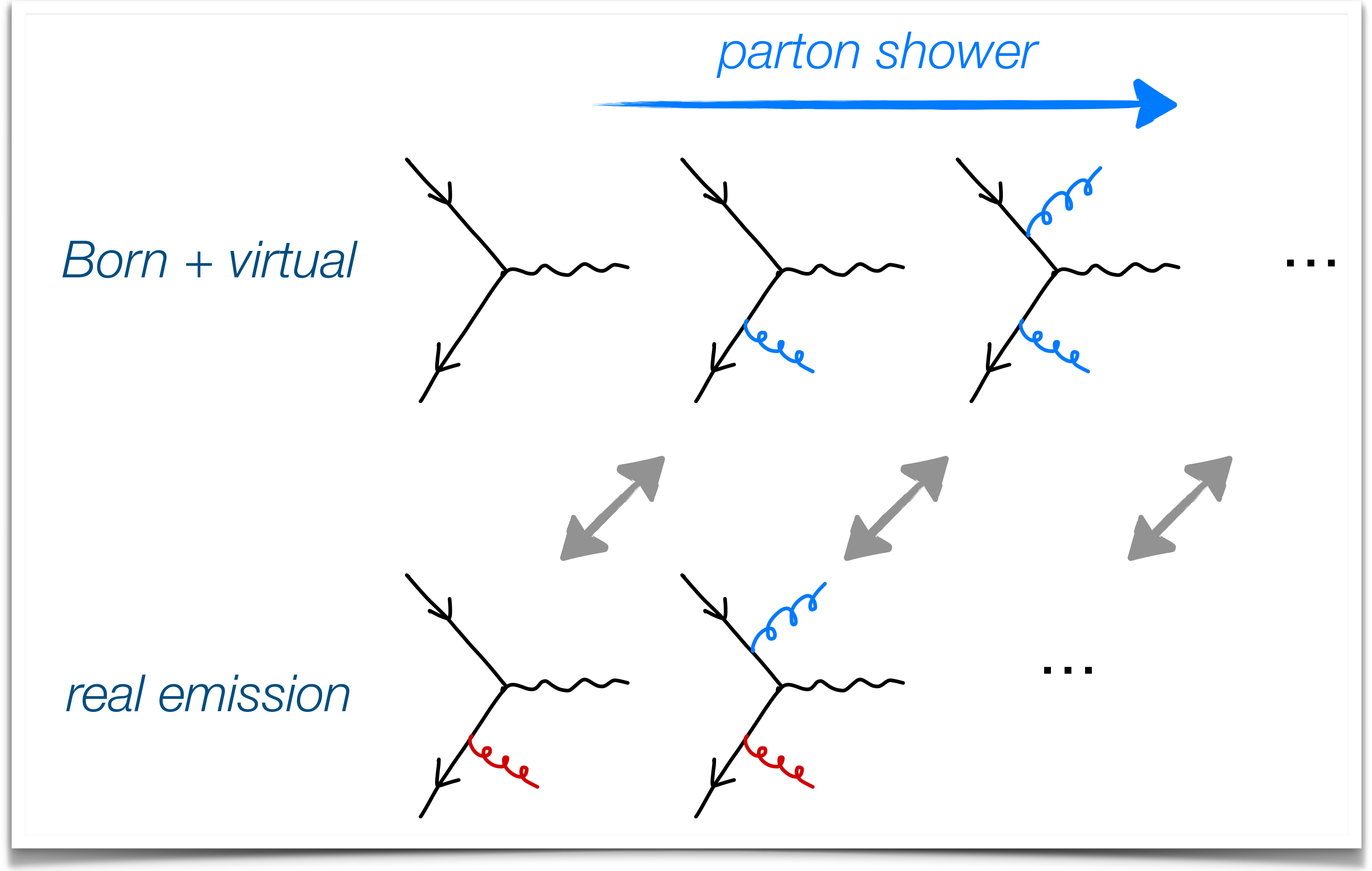
- **MC@NLO procedure** Frixione, Webber (2002)

$$\frac{d\sigma_{\mathrm{NLOwPS}}}{dO} = \left[d\Phi_m \left(B + \int_{\mathrm{loop}} V + \int \Phi_1 \, MC \right) \right] I_{\mathrm{MC}}^{(m)}(O)$$

$$+ \left[d\Phi_{m+1} \left(R - MC \right) \right] I_{\mathrm{MC}}^{(m+1)}(O)$$

"standard ($\mathbb{S}$) events"

"hard ($\mathbb{H}$) events"



  - Double counting is explicitely removed by including the shower subtraction "MC" terms

PS-specific

- For comparison, in **POWHEG matching** double counting is removed by modifying the first PS emission
  - The first emission is radiated according to the real-emission diagram
  - Inclusive NLO corrections are also added to each given event

Nason (2004)

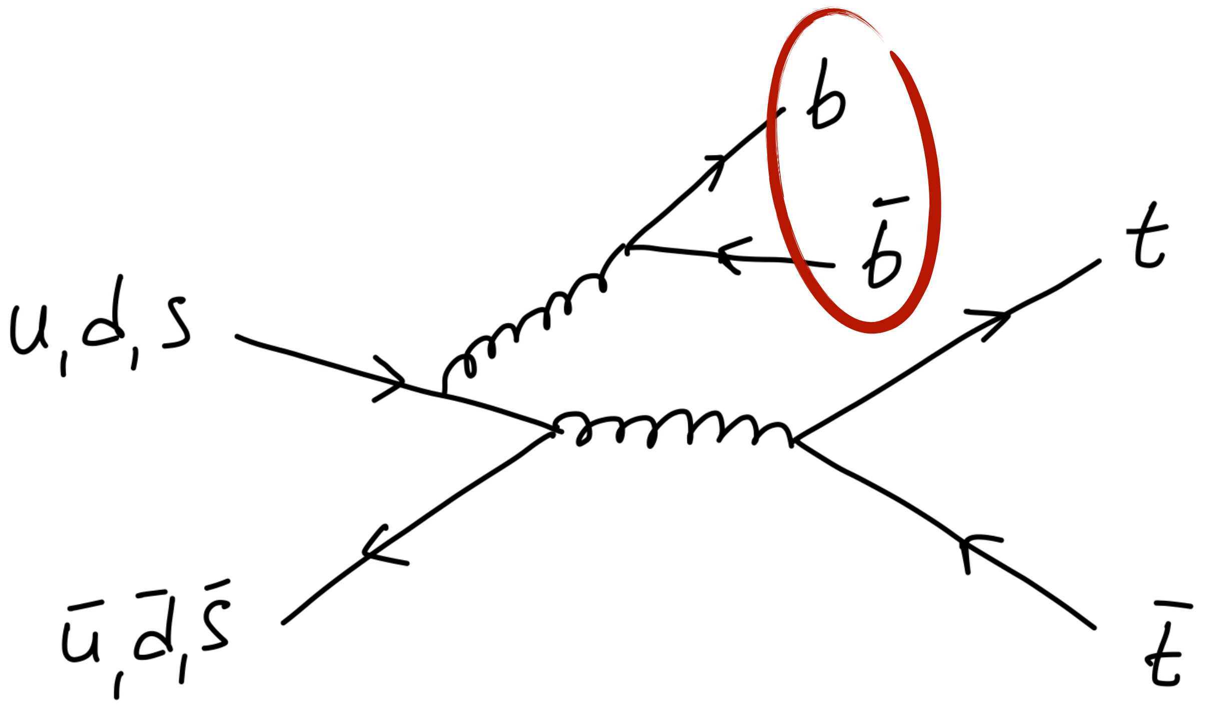
… now, back to $t\bar{t}b\bar{b}$

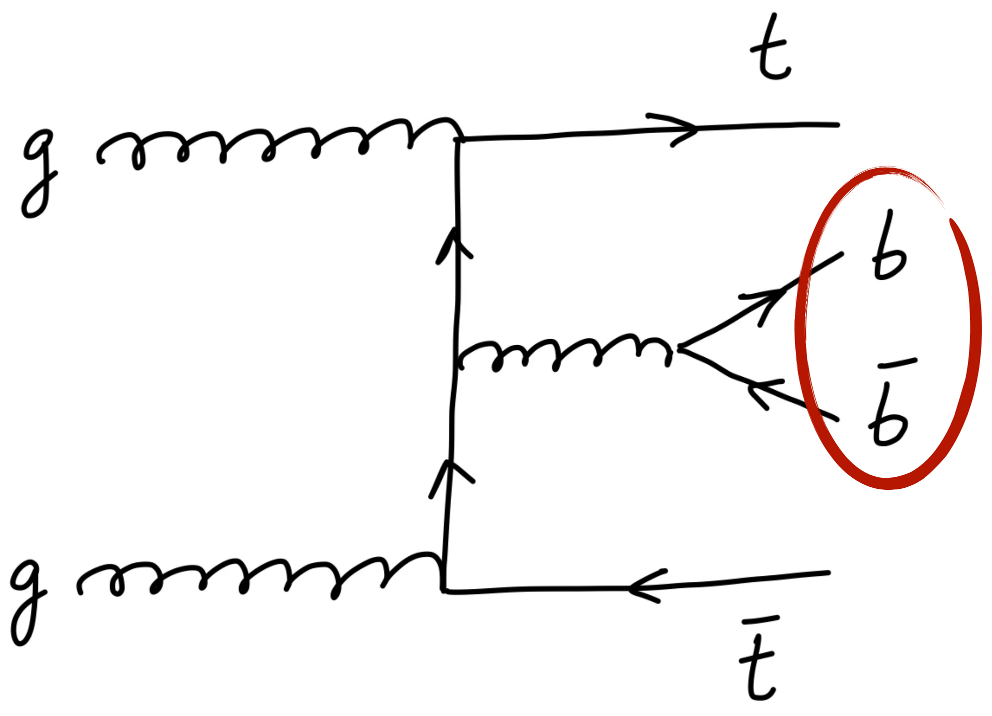
Simulation of the $t\bar{t}b\bar{b}$ process

- **Two primary theoretical frameworks:** four-flavour scheme (4FS) and five-flavour scheme (5FS)

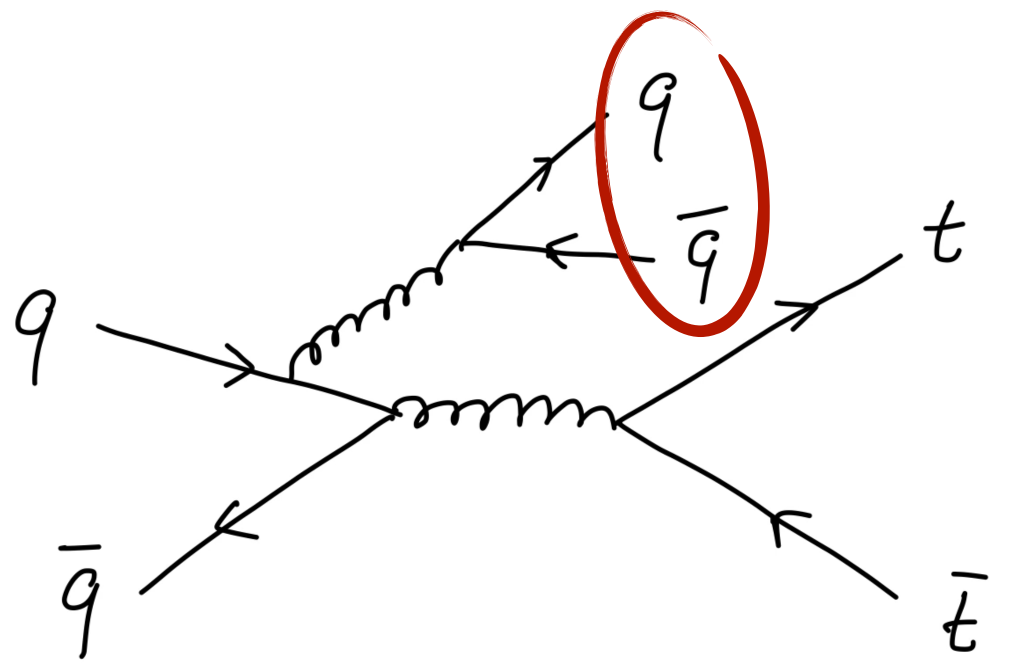
- **Alternative:** "fusion" method (or variable flavor number scheme)
  - Merges aspects of both the 4FS and 5FS calculations
  - Currently, the additional jets are only computed at LO

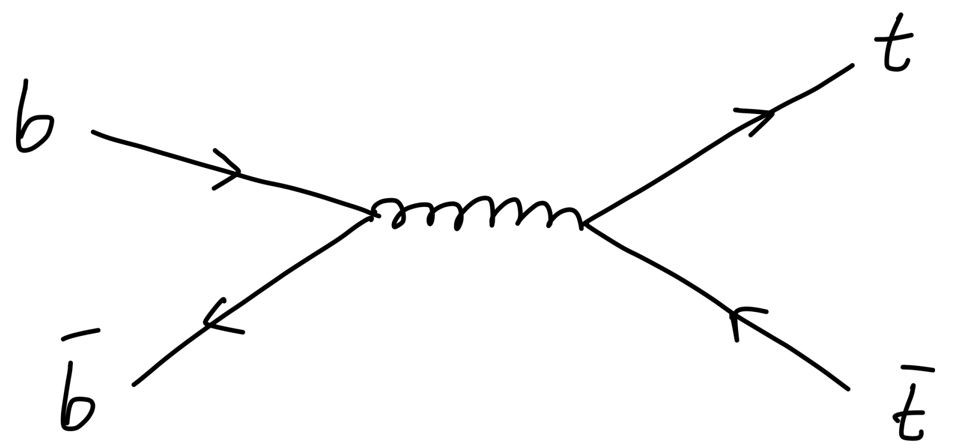
Höche, Krause, Siegert (2019)
Ferencz, Höche, Katzy, Siegert (2024)

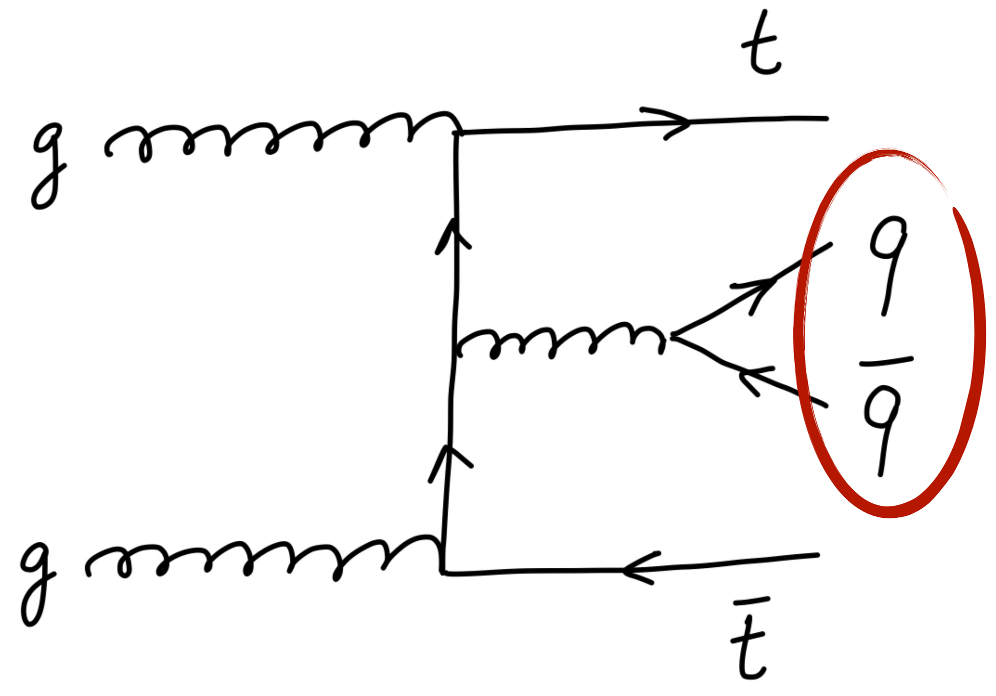




	4FS	5FS
***b*-quarks in the matrix element**	massive	massless
***b*-quarks included in the PDF?**	no	yes
renormalisation scheme	on-shell	$\overline{\mathrm{MS}}$
final state	exclusively $t\bar{t}b\bar{b}$	inclusive $t\bar{t}$ + jets







Simulation of the $t\bar{t}b\bar{b}$ process in 4FS

- 4FS calculations are the most precise at fixed order *(i.e. w/o a parton shower)*
  - b-quark mass effects taken into account
  - The processs can be generated down to any energies

Bredenstein, Denner, Dittmaier, Pozzorini (2008)
Bredenstein, Denner, Dittmaier, Pozzorini (2009)
Bevilacqua, Czakon, Papadopoulos, Pittau, Worek (2009)
Buccioni, Kallweit, Pozzorini, Zoller (2019)
Bredenstein, Denner, Dittmaier, Pozzorini (2010)
Denner, Lang, Pellen (2021)
Bevilacqua, Bi, Hartanto, Kraus, Lupattelli, Worek (2021)
Bevilacqua, Bi, Hartanto, Kraus, Lupattelli, Worek (2023)

- Calculation with a certain number of jets at fixed order is reliable only if there are no scale hierarchies
  - $t\bar{t}b\bar{b}$ production is a multi-scale process
  - Large mass difference between the t and b-quarks → **large logarithms** $\log^n(m_b/\sqrt{s})$
  - Difficult to choose optimal renormalisation and factorisation scales
    - Need a very small μ_{R} and a small $\mu_{\mathrm{F}} \neq \mu_{\mathrm{R}}$

see discussion in the LHC Higgs Xsec WG report arXiv:1610.07922

- Challenges arise when matching to a parton shower
  - PS radiation can produce additional b-quarks
  - Jets generated by the shower can be harder than the ME-level bottom quarks
  - Need only the subleading b-quarks to come from the PS but not the leading ones
    - Poorly understood how the PS radiation should be constraint

Cascioli, Maierhöfer, Moretti, Pozzorini, Siegert (2014)
Ježo, Lindert, Moretti, Pozzorini (2018)

Simulation of the $t\bar{t}b\bar{b}$ process in 5FS

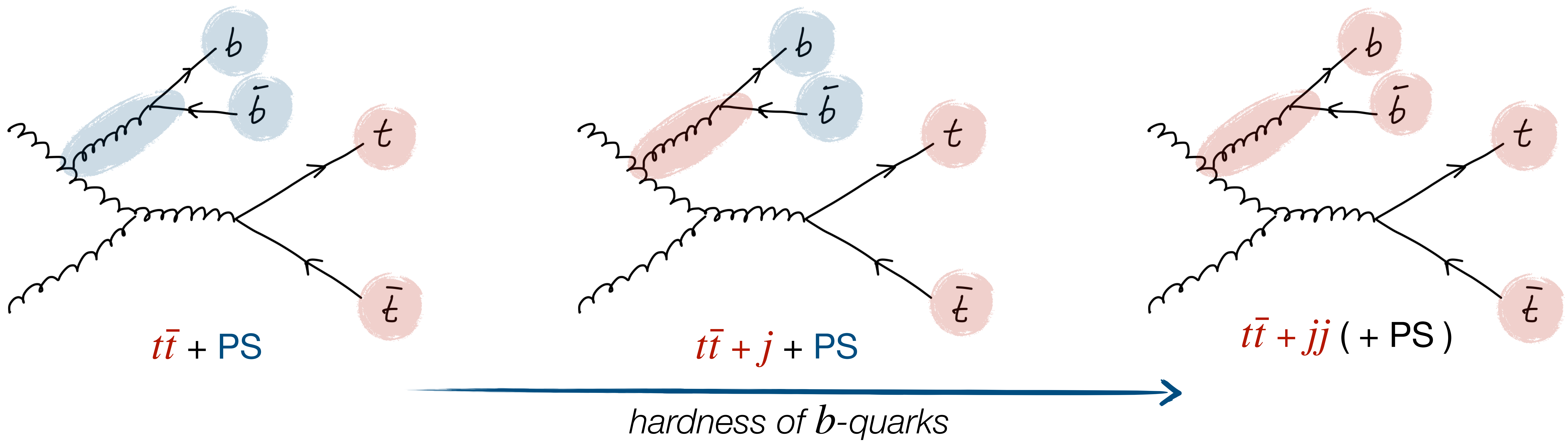
- In 5FS one has to generate an inclusive $t\bar{t}$ + jets sample
  - b-jets are selected only after parton showering

Frixione, Nason, Webber (2003)
Frixione, Nason, Ridolfi (2007)
Hoeche, Krauss, Maierhoefer, Pozzorini, Schonherr, Siegert (2015)
Mazzitelli, Monni, Nason, Re, Wiesemann, Zanderighi (2022)

- 5FS: massless b-quarks → large logarithms do not arise in the ME
- Large scale hierarchies between the top quarks and the jets can be resummed by a *multi-jet merging* procedure

- **Multi-jet merging**: combination of events with different jet multiplicities
  - For example, FxFx merging in MadGraph5_aMC@NLO Frederix, Frixione (2012)

the b-quarks will be produced either in the ME or by the PS, depending on their p_{T}



Why merging?

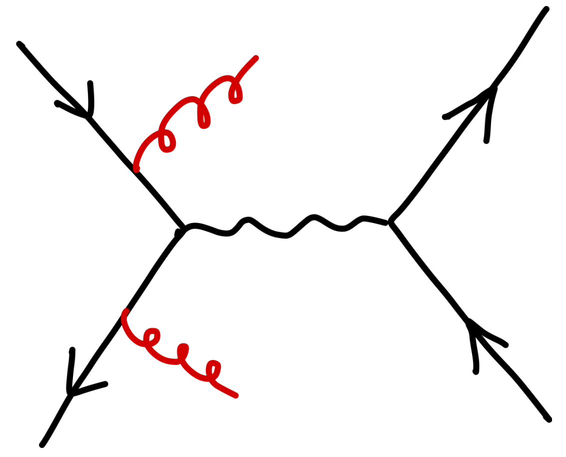
Why merging?

- PS is only correct in the collinear approximation
  - Cannot generate hard extra jets correctly (i.e. jets beyond the first)
  - Associated theory uncertainty is large

➡ **Need to add higher multiplicity fixed-order calculations → merging allows to combine them**

Merging schemes implemented in various MC generators differ in their use of PS, Sudakov factors (analytic/numerical) and in details concerning the jet vetoing in the PS

- **The task is the same as in matching, just with more real emissions**: is this diagram →
  - a Born contribution of $t\bar{t} + 2$ jets
  - or a real-emission correction to $t\bar{t} + 1$ jet?



- The easiest way to combine the two w/o double counting is to consider a jet cut at some *merging scale* μ_{Q}
  - If the 2nd emission is harder than μ_{Q}, use $t\bar{t} + 2$ jets, otherwise use $t\bar{t} + 1$ jet ← *shower starting scale(s) should reflect this: PS shouldn't have radiation harder than μ_{Q}*
  - NB: a too small μ_{Q} can lead to large logarithms → $\log(m_Q/\sqrt{s})$

- MC@NLO matching:
  - Recall, we have $\mathbb{S}$ and $\mathbb{H}$ events and MC counter terms which assure no double counting
  - After the μ_Q cut, there must still be a cancelation of MC terms within a given multiplicity

Merging

- Next step: resumming the higher-order corrections to maintain the overall logarithmic accuracy of the PS
  - Sudakov suppression ⟵ *resumming unresolved and virtual corrections*
    - multiply the matrix elements by the Sudakov factors (CKKW) Catani, Krauss, Kuhn, Webber (2001)
    - or reject events for which some PS jets do not match the ME partons (MLM) Mangano, Moretti, Pittau (2002)
  - α_S reweighting ⟵ *resumming higher-order corrections to soft-gluon radiation*

- **FxFx merging:** the procedures above are based on the "most-likely parton-shower history" Frederix, Frixione (2012)
  - Cluster partons into jets
  - α_S reweighting: set μ_R to the geometric mean of the cluster scales (w/o the first cluster for the $\mathbb{H}$ events)
  - Set μ_F to the first (second) cluster scale for the $\mathbb{S}$ ($\mathbb{H}$) events
  - Sudakov suppression: use a mixture of CKKW and MLM approaches
- This method cancels the leading- and next-to-leading-log dependence on μ_Q

- After merging and showering:
  - $t\bar{t}$ events → no jets harder than μ_Q
  - $t\bar{t} + 1$ jet events → exactly one jet harder than μ_Q (and "matched" to a parton)
  - …

… back to our 5FS simulation

Simulation of the $t\bar{t}b\bar{b}$ process in 5FS

- **In the 5FS all the logarithms are correctly resummed** in the
  - Parton shower
  - PDFs
  - Multi-jet merging procedure

- Accurate parton-shower approximation for all softer jets ← *not always the case in the 4FS*
  - Parton shower jets are always softer than the merging scale
    - *Except for jets coming from the highest-multiplicity sample*
  - Merging scale is smaller than the softest ME jets

- b-quark mass effects
  - Important in the collinear/IR region ← incorporated into parton shower splitting fuctions
  - Missing in the matrix element, but they are not so relevant for the hard b-quarks

Simulation of the $t\bar{t}b\bar{b}$ process in 5FS

- But one has to generate an inclusive $t\bar{t}$ + jets sample and select b-jets only after parton showering
- Generating $t\bar{t}$ + 0,1,2 jets @ NLO accuracy requires substantional computing resources

Hoeche, Krauss, Schonherr, Siegert (2013) | Frederix, Frixione (2012) | Plätzer (2013)

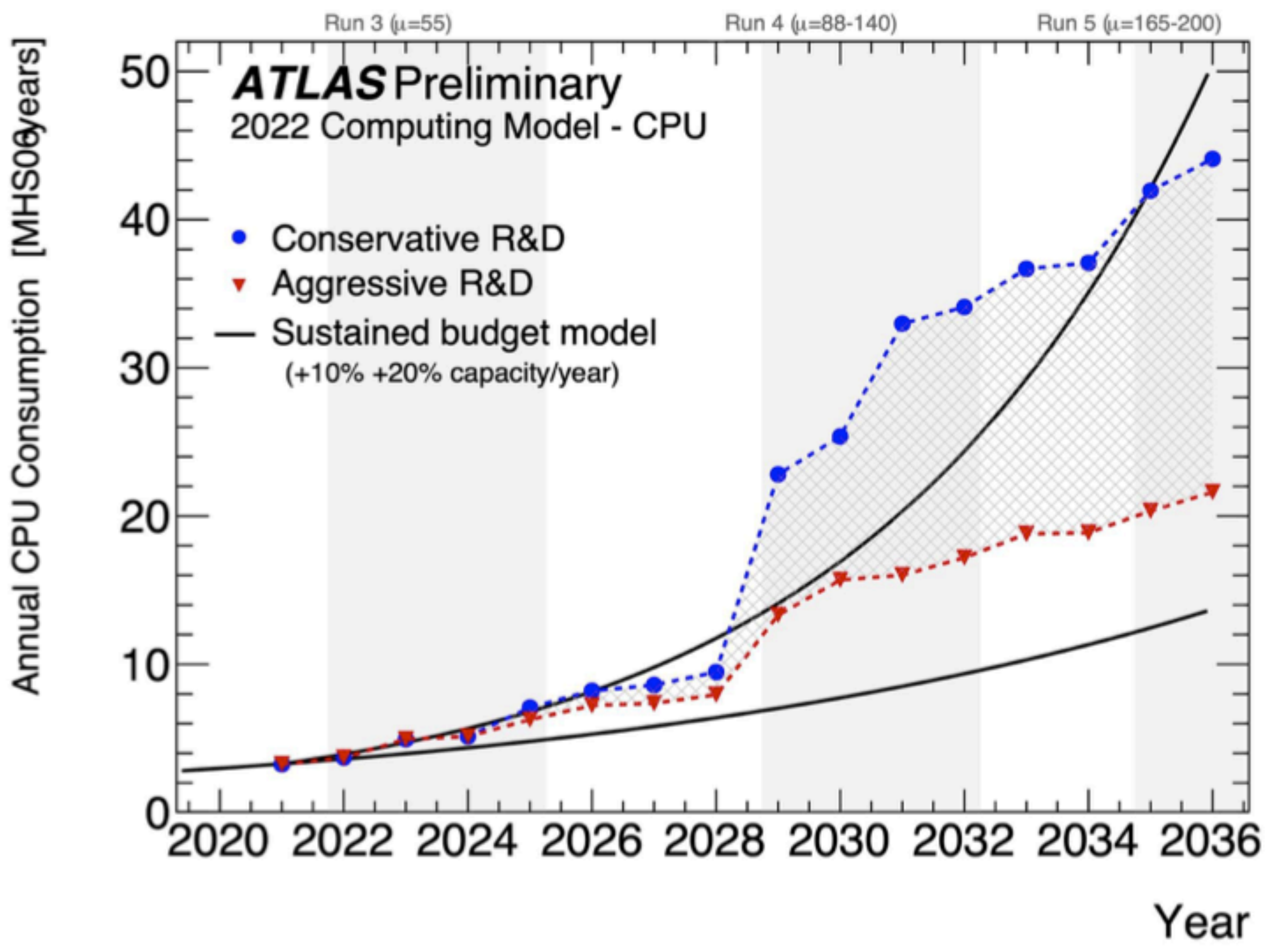
number of instructions to calculate a process in MadGraph5_aMC@NLO

	$gg \to t\bar{t}$	$gg \to t\bar{t}gg$	$gg \to t\bar{t}ggg$
madevent	13G	470G	11T
matrix1	3.1G (23%)	450G (96%)	11T (>99%)
└▸ ext	450M (3.4%)	3.3G (<1%)	7.3G (<1%)
└▸ int	1.9G (14%)	160G (35%)	2T (19%)
└▸ amp	530M (4.0%)	210G (44%)	5.5T (51%)

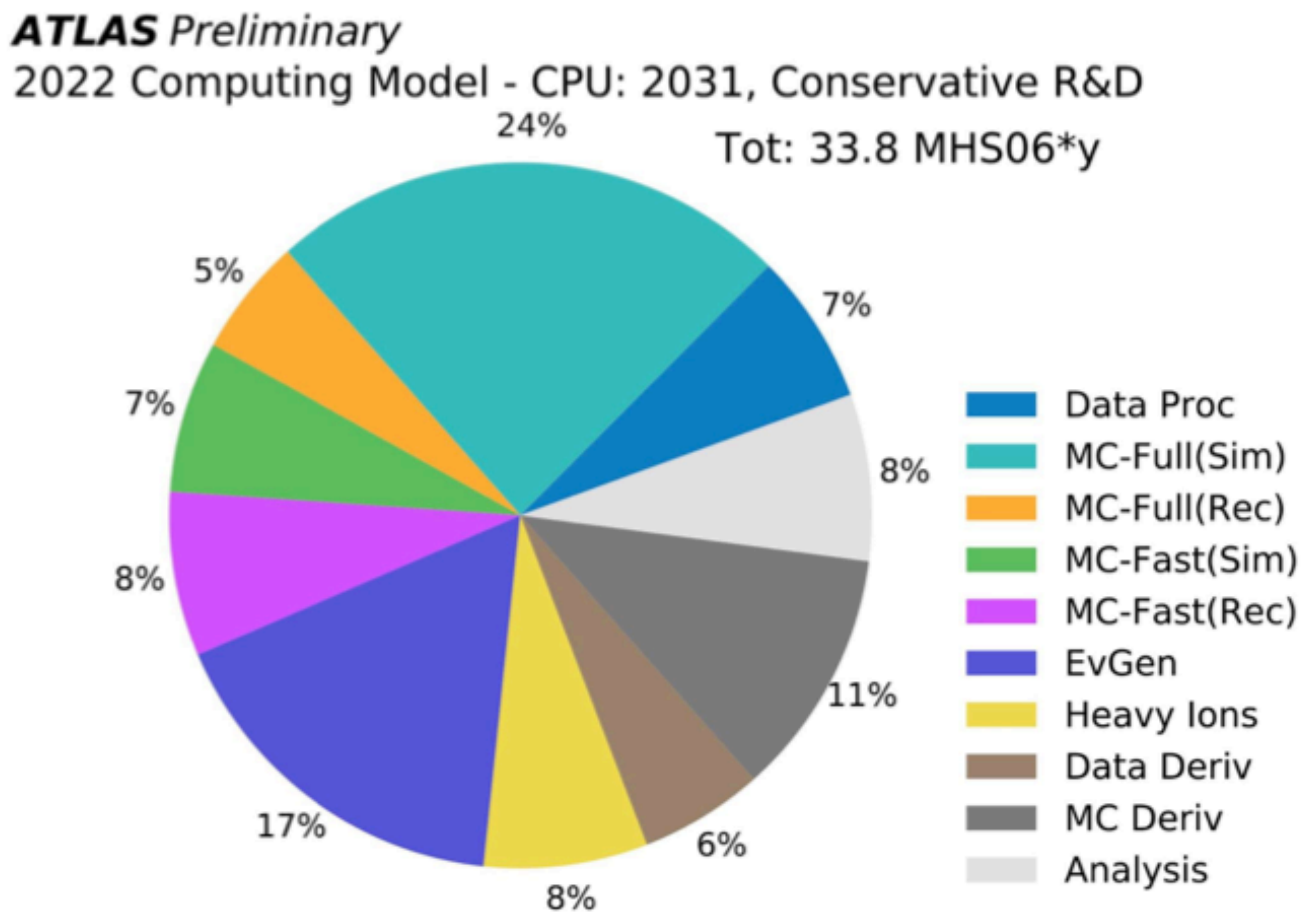
from O. Mattelaer's talk

- Selection efficiency of $t\bar{t}b\bar{b}$ is low (percent level)
  - $gg \to t\bar{t}gg$ dominates

➡ **5FS approach is computationally demanding!**
  - This will become even more relevant when producing MC for the HL-LHC era



CERN-LHCC-2022-005



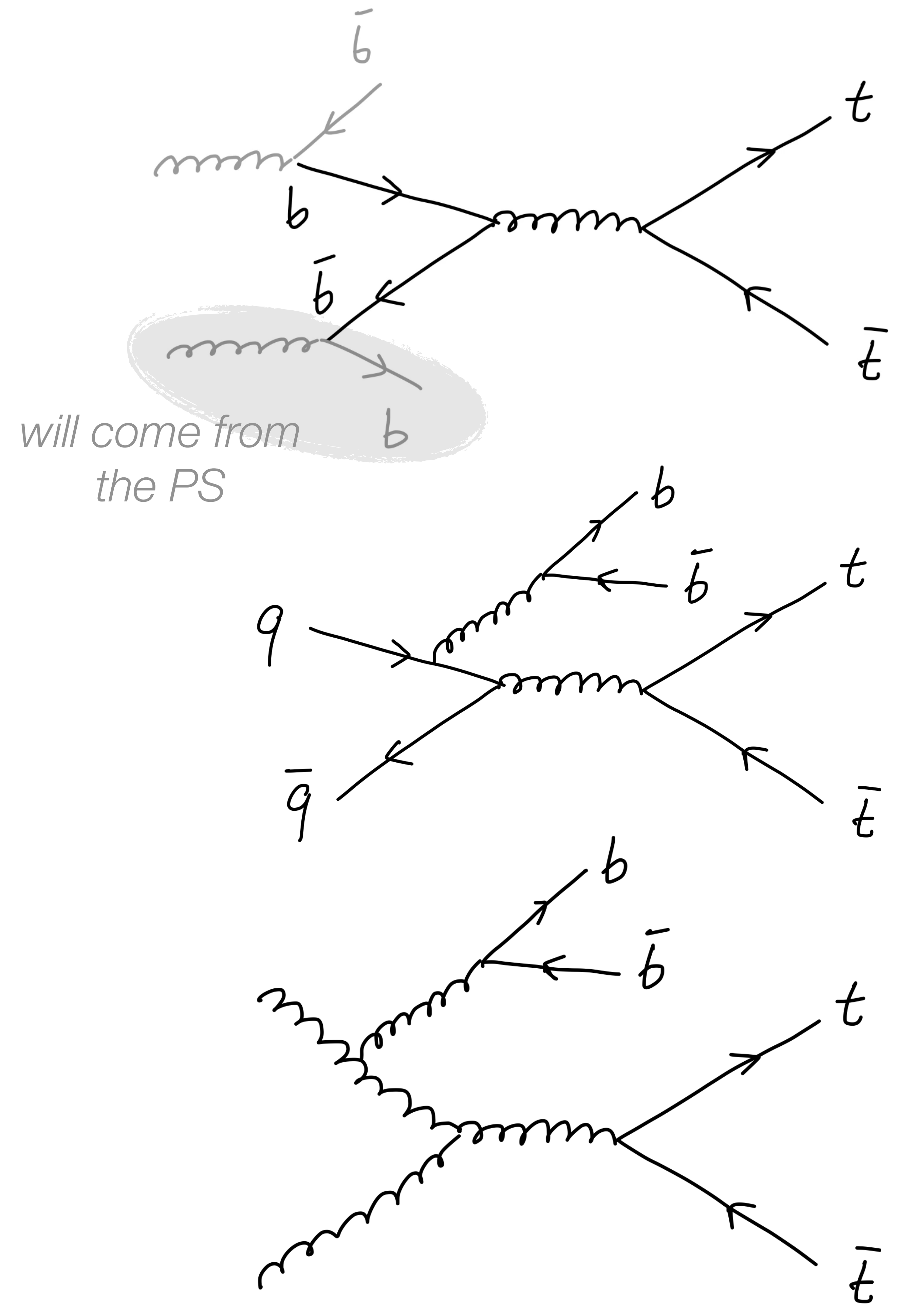
Can we improve the $t\bar{t}b\bar{b}$ statistics without changing the total number of events?

b-flavour enhancement in the matrix element

✓ We proposed a novel method to enhance the b-jet selection efficiency in the 5FS approach

R. Frederix, TM
EPJC 84, 763 (2024)

- Augment the generation probability of bottom quark flavour in the short-distance event generation
  - During phase-space integration and unweighting, multiply the weight of each contribution containing external b-quarks by w_{enh}
  - For bottom quarks can be in initial or final state → enhancing all
    - $gg \to t\bar{t}b\bar{b}(g)$ ← (g) *is an additional radiation from NLO*
    - $gb \to t\bar{t}bg(\to b\bar{b})$
    - $bb \to t\bar{t}q\bar{q}(g)$
    - …
- To compensate for this and to preserve the cross-section, multiply the weight of events with external b-quarks by $1/w_{\text{enh}}$



examples of the enhanced subprocesses

b-flavour enhancement in the matrix element

✓ We proposed a novel method to enhance the b-jet selection efficiency in the 5FS approach

R. Frederix, TM
EPJC 84, 763 (2024)

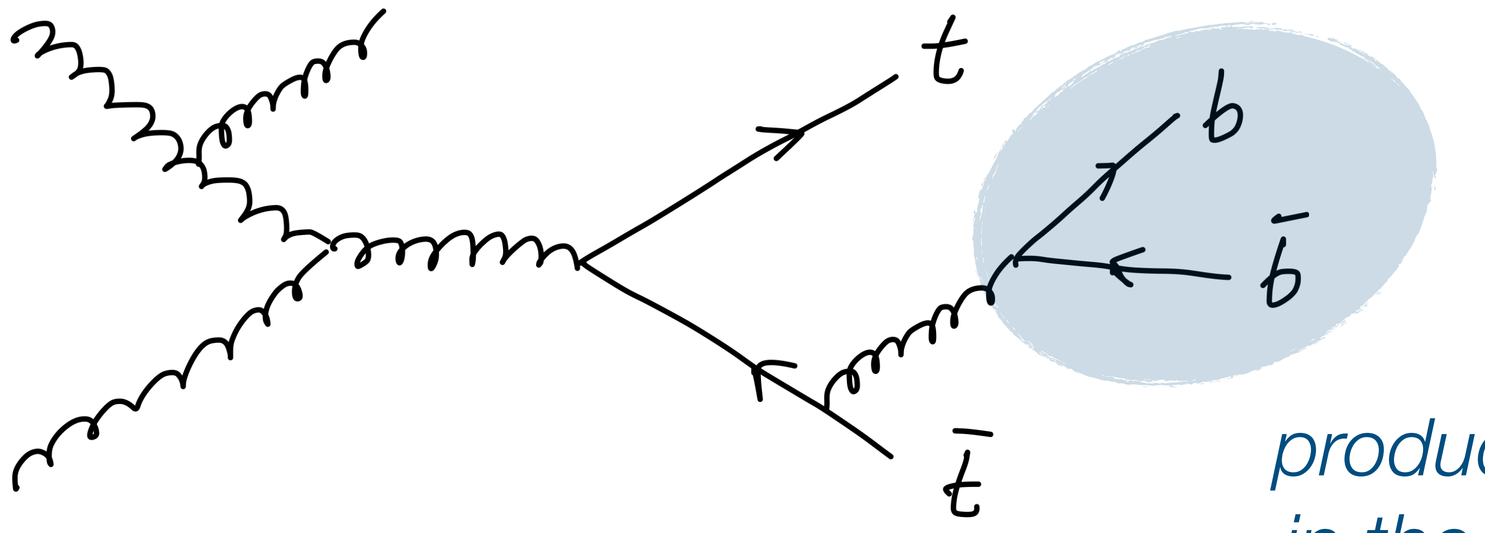
- This procedure is implemented in the MadGraph5_aMC@NLO
  - enhancement factor w_{enh} can be set by a new parameter, `bflav_enhancement`, in the `run_card.dat` file
  - **The new feature will become part of an upcoming release**

```
#*********************************************************************
#                       MadGraph5_aMC@NLO                            *
#                                                                    *
#                     run_card.dat MadEvent                          *
#                                                                    *
#  This file is used to set the parameters of the run.               *
#                                                                    *
#  Some notation/conventions:                                        *
#                                                                    *
#   Lines starting with a '# ' are info or comments                  *
#                                                                    *
```

```
#*********************************************************************
100 = bflav_enhancement
#*********************************************************************
```

✳ ***NB:*** *the hard processes like $gg \to t\bar{t}gg$ which can yield a $t\bar{t}b\bar{b}$ event after a $g \to b\bar{b}$ splitting in the parton shower will not get enhanced $\Rightarrow$ effectively, the fraction of $t\bar{t}b\bar{b}$ events is increased by a factor smaller than w_{enh}*

- *Also, too high enhancement factors (>100) cause instabilities which result in large statistical fluctuations*

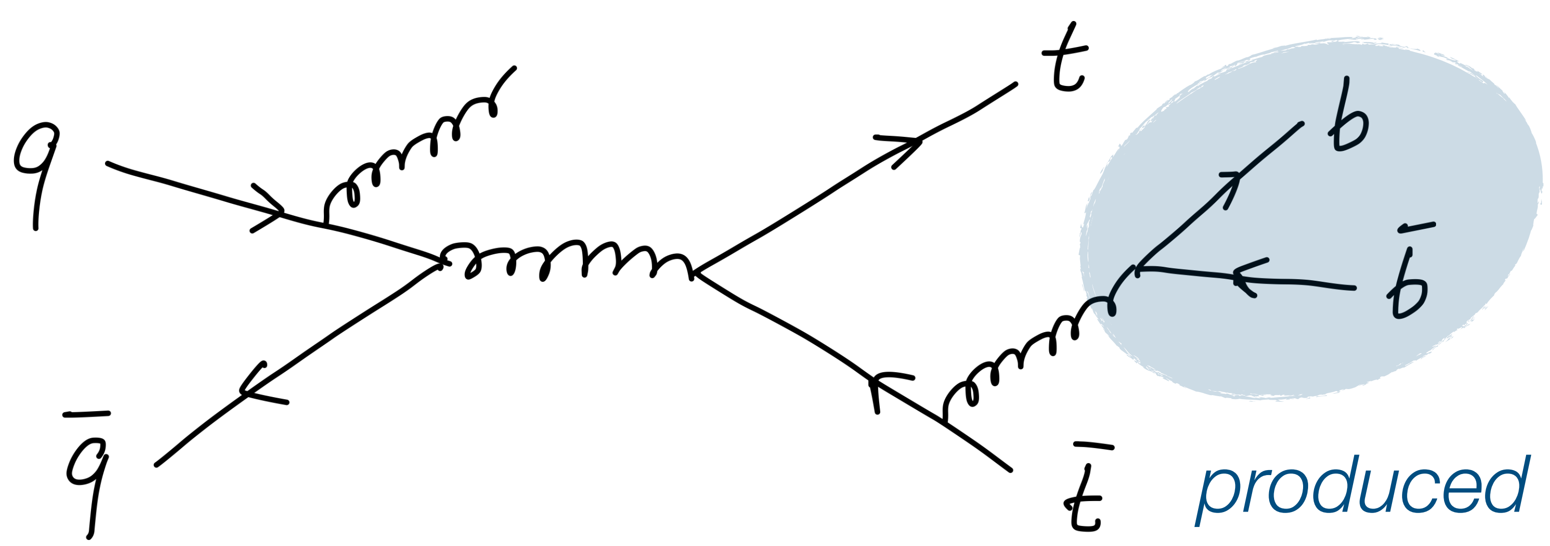


produced in the PS

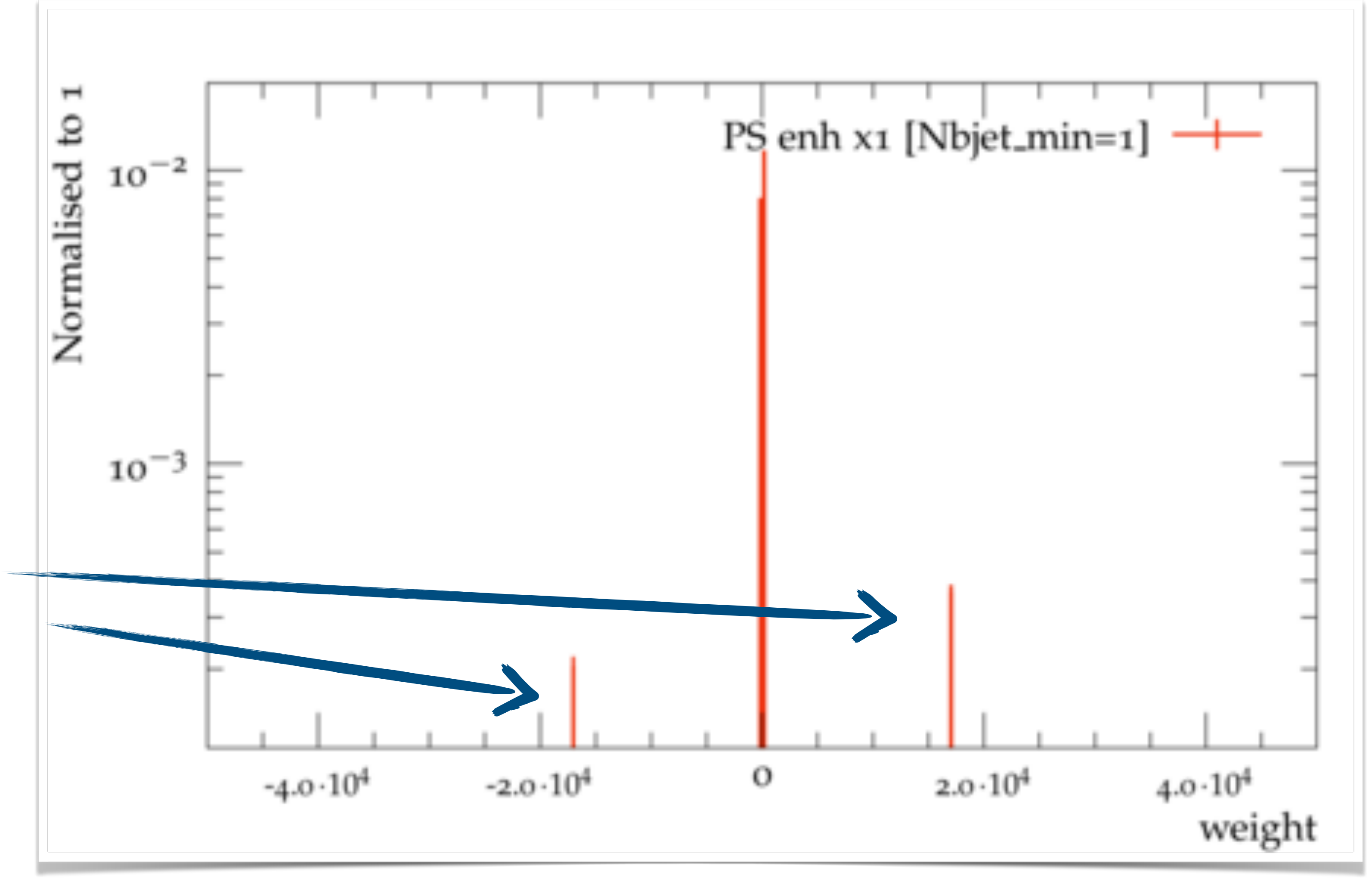
this diagram is not enhanced

b-flavour enhancement in the parton shower?

- A similar biasing strategy can be potentially applied in the parton shower
- Pythia8 has a built-in mechanism for enhancing splitting probabilities, in particular $g \to b\bar{b}$ ones
  - In versions ≤8.303 and ≥8.311
- In practice, we have found significant trade-offs :(
  - Even moderate enhancement in the PS causes significant widening of the event weight distribution
  - **Large weights deteriorate the statistics** → cancels the improvement from the b-enhancement completely



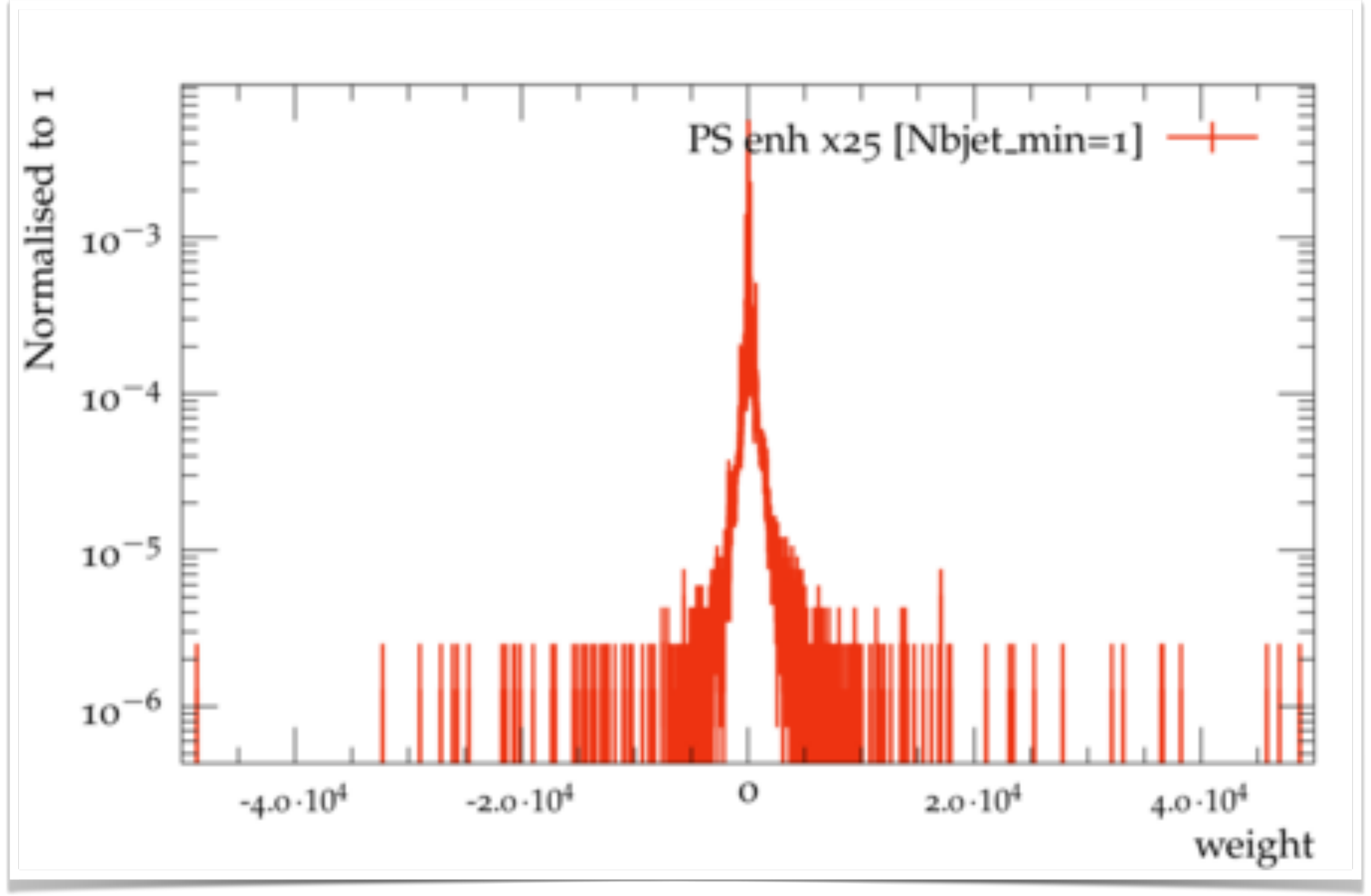
event weights w/o enhancement in the PS



those large weights are from the PS

$g \to b\bar{b}$ enhancement

event weights with enhancement in the PS



Setup for the 5FS sample and comparison to the 4FS

5FS sample with b-enhancement in the ME

R. Frederix, TM
EPJC 84, 763 (2024)

- **MadGraph5_aMC@NLO $t\bar{t} + 0{,}1{,}2$ jets @NLO sample, FxFx merged**
- **Enhancement factor $w_{\mathrm{enh}} = 100$**
- **Renormalisation/factorisation scales:** central values for are taken from the FxFx merging
  - 7-point variations
- **Merging scale:** 40 GeV
  - variations: 70 and 100 GeV
- **Shower starting scale: $H_T/2$**
  - variation: $H_T/4$

taking the envelope of those as a total uncertainty

- Generation-level cut of 20 GeV on jet p_T

softer jets would be thrown away during the matching/merging anyways

- **Matched to the Pythia8 parton shower**
- **Not including**:
  - hadronisation
  - underlying events
  - top quark decay

to reduce the generation time and to simplify the analysis, and because we focus on the differences in the ME

Truth-level analysis:

- anti-k_T jets ($R > 0.4$)
  - $p_T > 25$ GeV
  - $|\eta| < 2.5$
- jets containing at least one bottom quark are identified as b-jets
- consider two scenarios:
  - at least 1 b-jet
  - at least 2 b-jets

4FS sample

R. Frederix, TM
EPJC 84, 763 (2024)

- **MadGraph5_aMC@NLO+Pythia8** NLO+PS $t\bar{t}b\bar{b}$ sample
- **Renormalisation/factorisation scales**:
  - central values:

$$\mu_R = (E_{T,t}E_{T,\bar{t}}E_{T,b}E_{T,\bar{b}})^{1/4}$$
$$\mu_F = \frac{1}{2}(E_{T,t} + E_{T,\bar{t}} + E_{T,b} + E_{T,\bar{b}})$$

  - 7-point variations

following the recommendations in the LHC Higgs Xsec WG report arXiv:1610.07922

- **Shower starting scale:** $H_T/2$
- Generation-level cut of 20 GeV on jet p_T
- **Matched to the Pythia8 parton shower**
- **Not including**:
  - shower starting scale uncertainty *(small)*
  - matching scheme uncertainty
  - hadronisation
  - underlying events
  - top quark decay

expected to be sizeable, (see the LHCHXSWG report) but is non-trivial to assess exactly

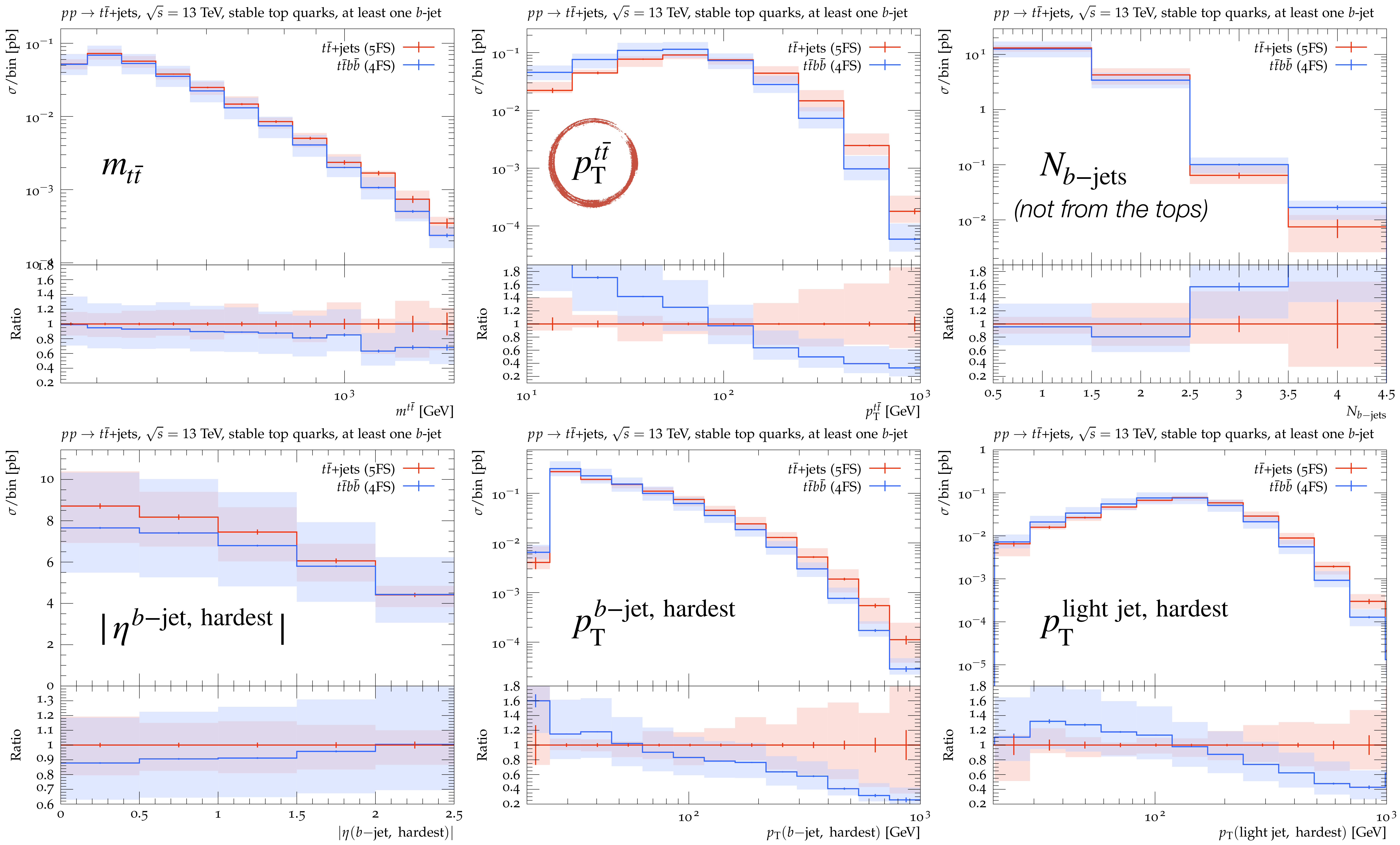
Truth-level analysis:

- anti-k_T jets ($R > 0.4$)
  - $p_T > 25$ GeV
  - $|\eta| < 2.5$
- jets containing at least one bottom quark are identified as b-jets
- consider two scenarios:
  - at least 1 b-jet
  - at least 2 b-jets

5FS vs 4FS: at least 1 b-jet scenario

R. Frederix, TM
EPJC 84, 763 (2024)

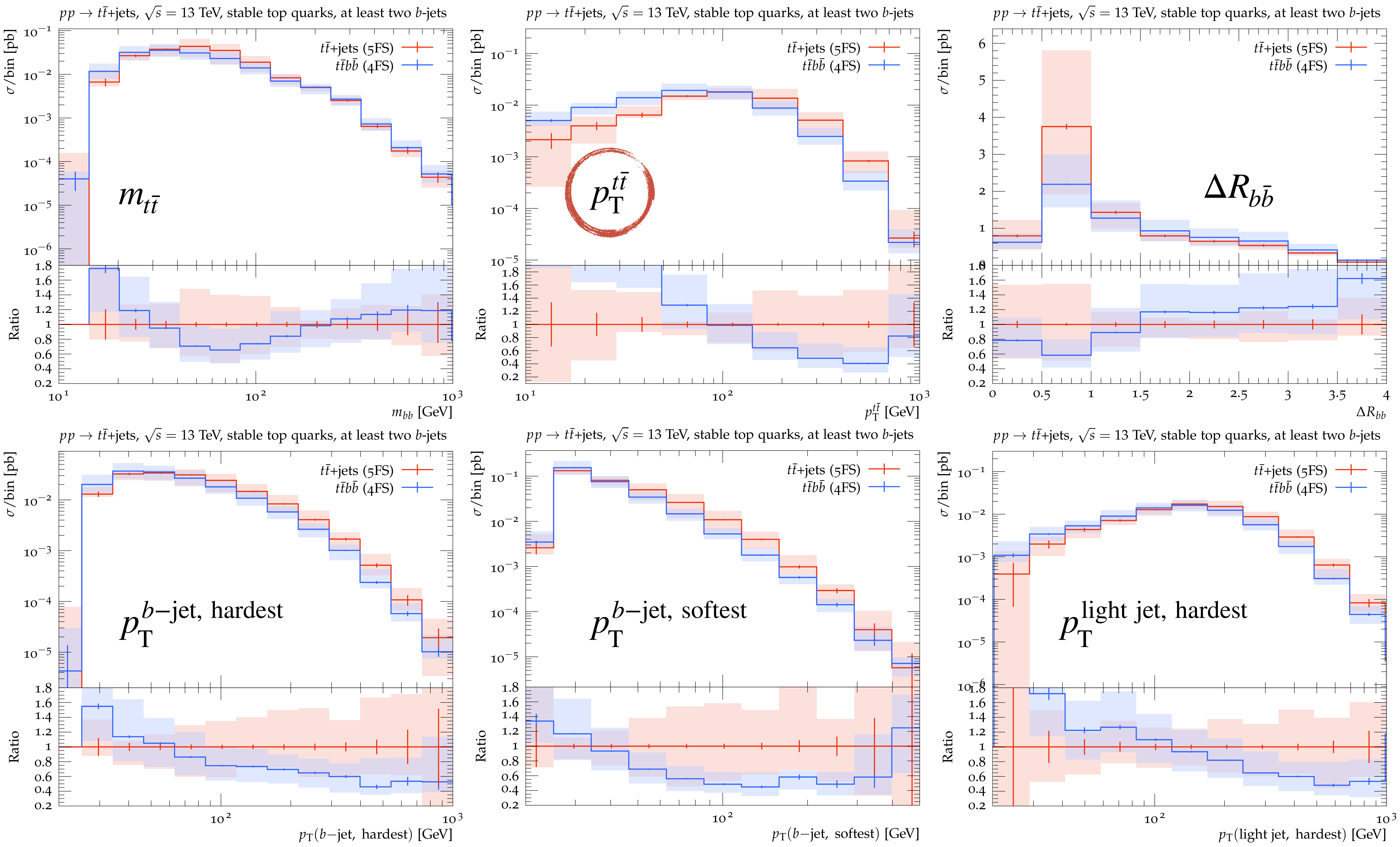
- For most of the variables, 4FS and 5FS predictions are compatible within the uncertainty bands
- 5FS uncertainty is more reliable than the 4FS one, since the 4FS matching uncertainty is expected to be significant but is not included
- The $p_T^{t\bar{t}}$ differs quite a lot, **the 5FS predicts a much harder spectrum than 4FS**
  - ➡ We investigated it further, see next slides



5FS vs 4FS: at least 2 b-jets scenario

R. Frederix, TM
EPJC 84, 763 (2024)

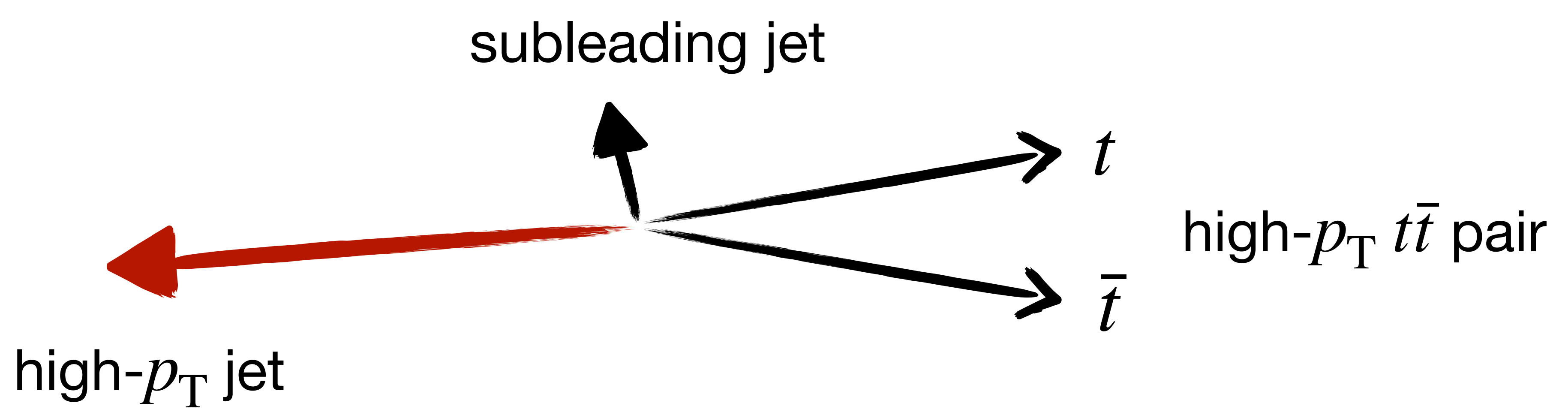
- Similar picture as for the ≥ 1 b-jet selection
- $p_T^{t\bar{t}}$ spectrum differs again
- The rest of the variables are in agreement



5FS vs 4FS: differences in the $p_T^{t\bar{t}}$ distribution

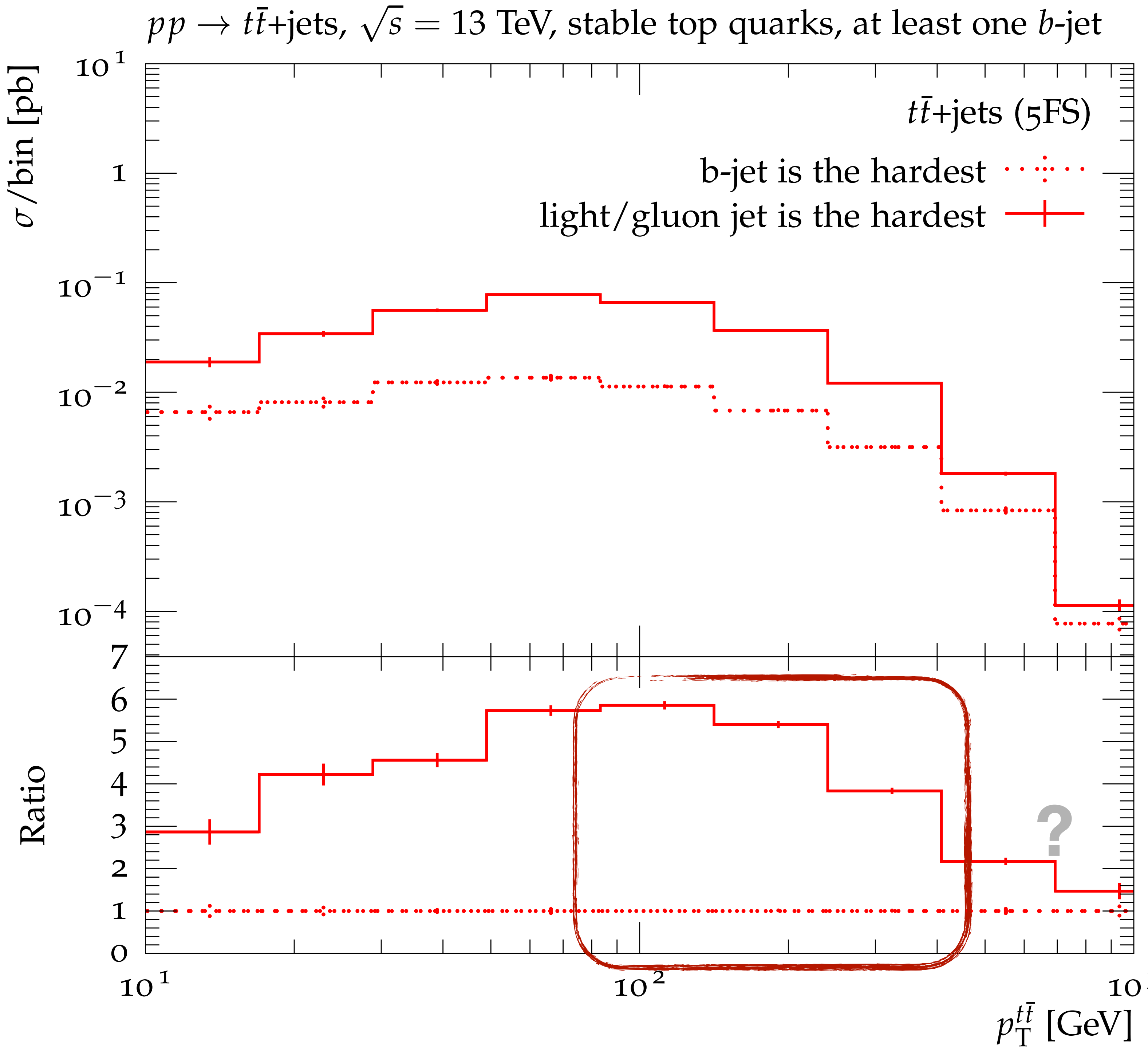
R. Frederix, TM
EPJC 84, 763 (2024)

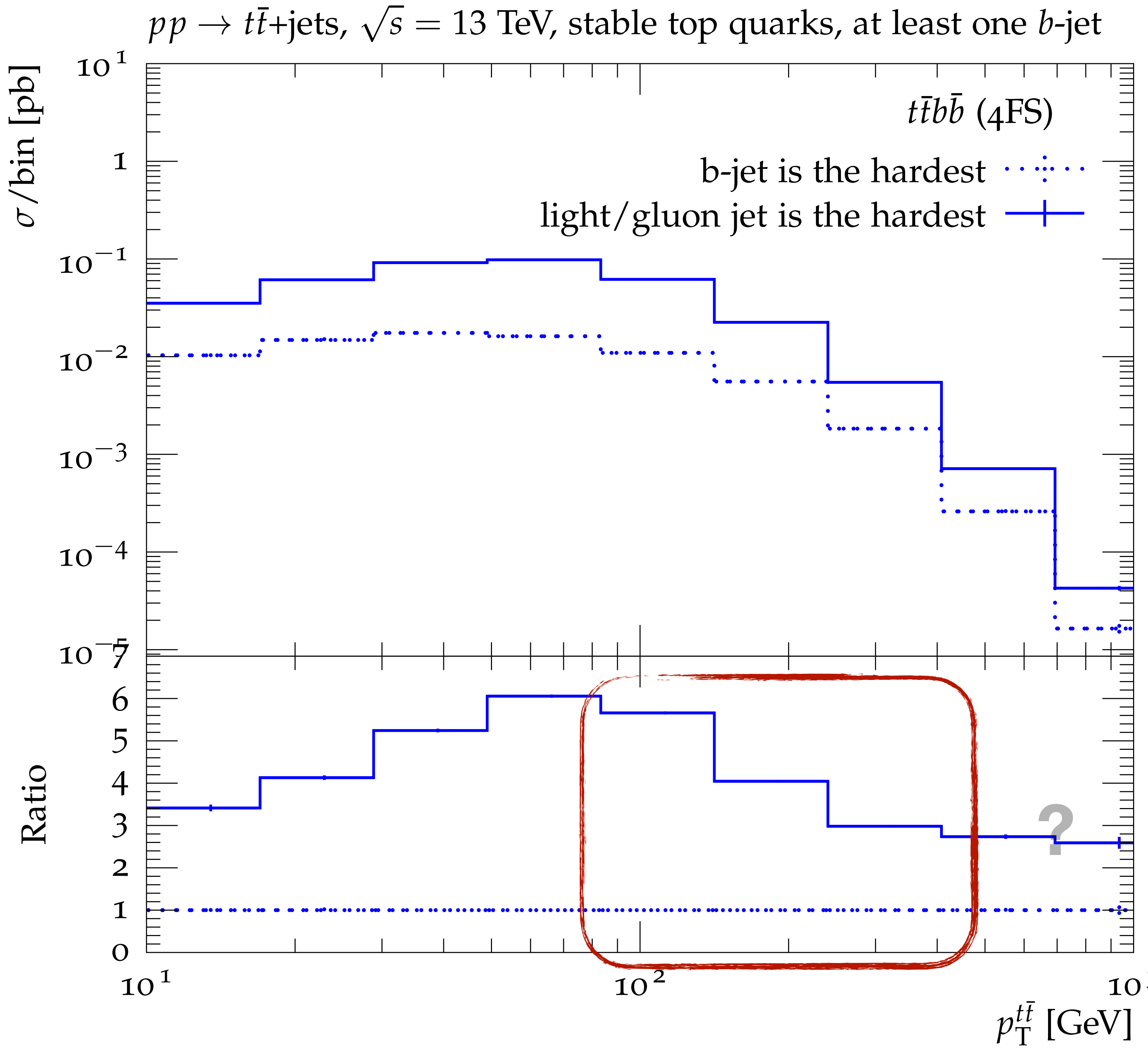
- At large $p_T^{t\bar{t}}$, it is kinematically most-likely that the $t\bar{t}$ pair recoils agains a single hard jet
- If the hardest jet is a light jet:
  - 5FS: described at NLO (most likely it is a gluon jet)
  - 4FS: described at LO or by the PS
    - No $t\bar{t}gg$ events from the ME
    - There is no hard gluon to recoil from



- For high $p_T^{t\bar{t}}$, the fraction of events with hardest jet being light-flavoured is indeed larger in the 5FS
- But after $p_T^{t\bar{t}} \sim 500$ GeV the situation is opposite — why?
  - Let's look again at the jet p_T distributions…

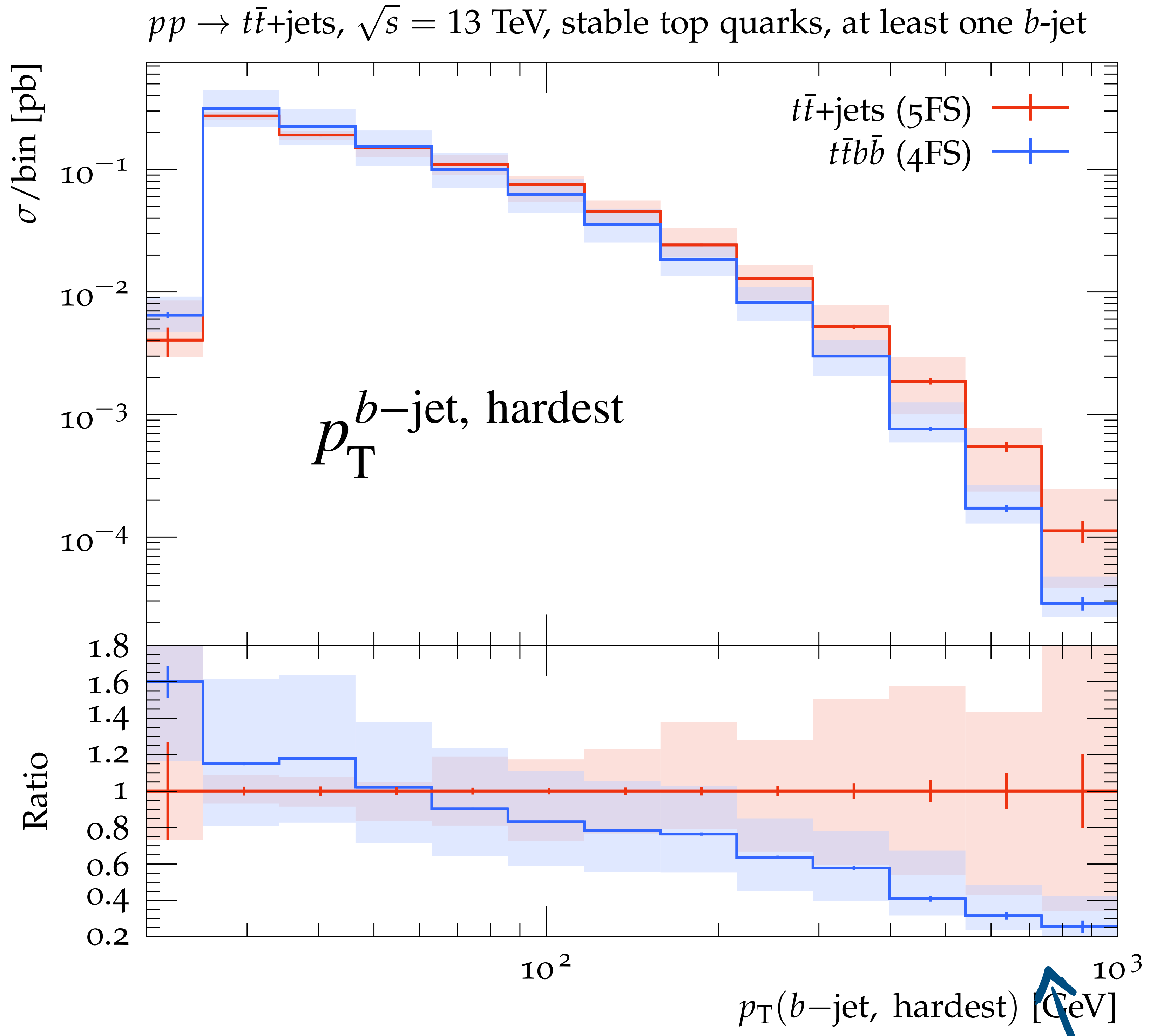
at least 1 b-jet selection

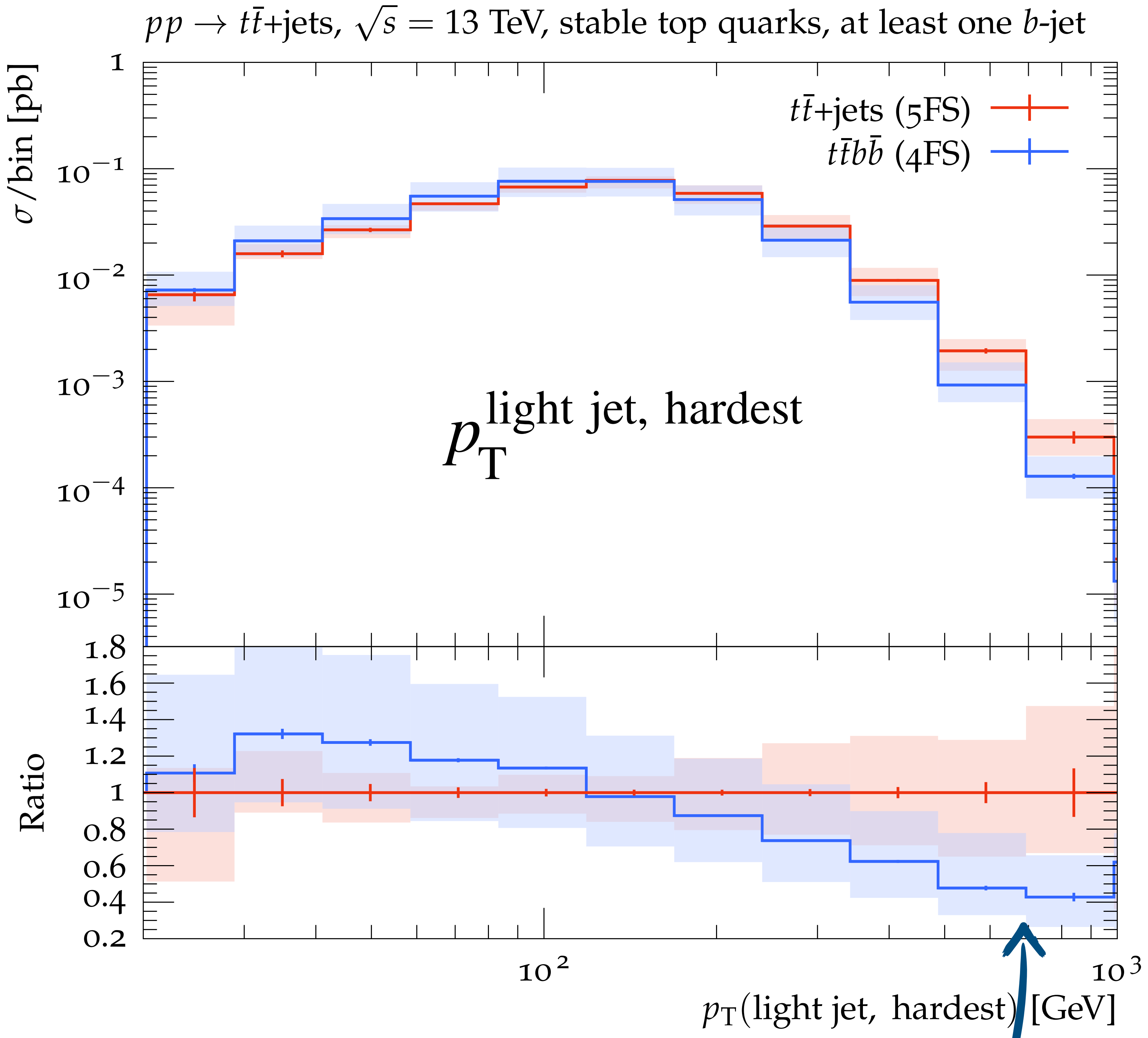




5FS vs 4FS: differences in the $p_T^{t\bar{t}}$ distribution

R. Frederix, TM
EPJC 84, 763 (2024)





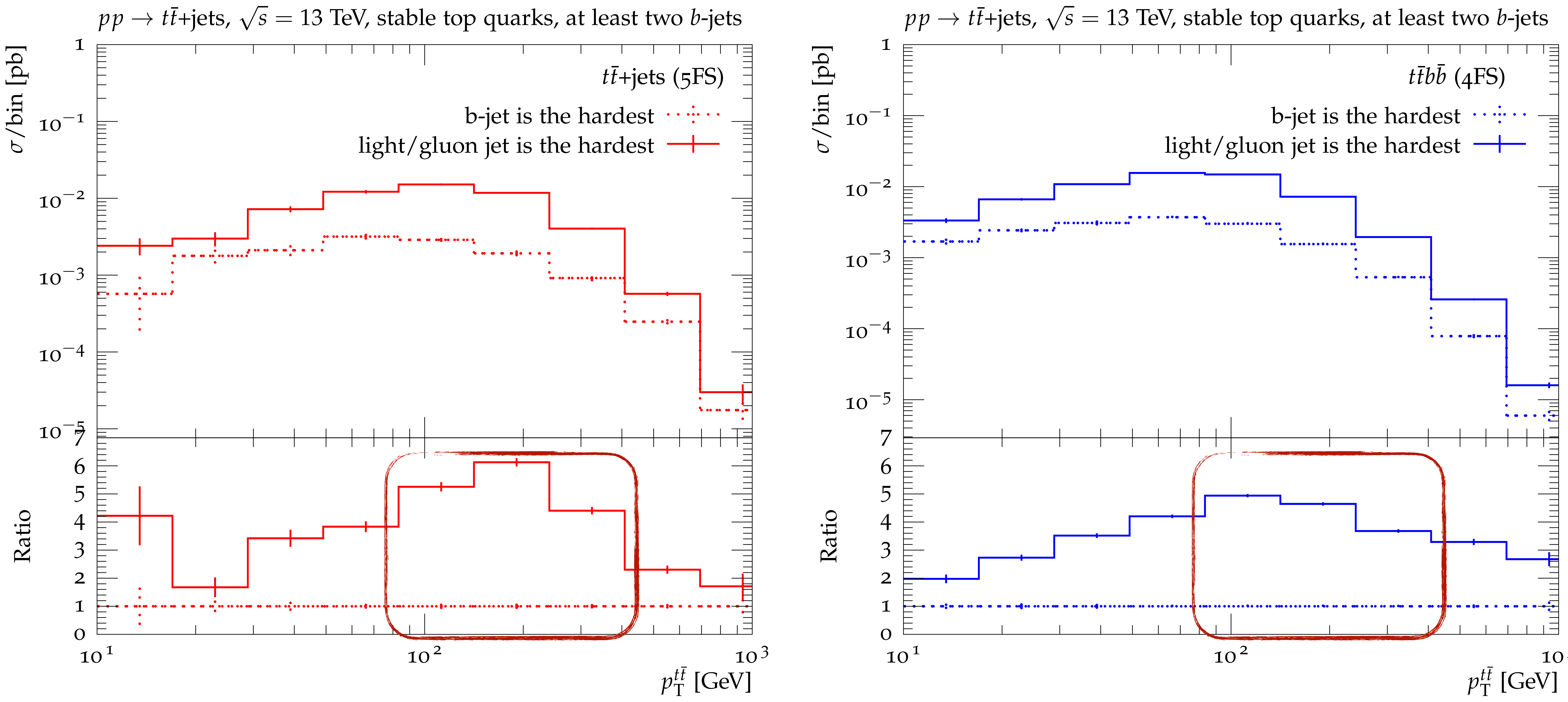
at very high p_T, the ratio of the 5FS over the 4FS predictions is larger for b-jets than for light jets

5FS vs 4FS: differences in the $p_T^{t\bar{t}}$ distribution

R. Frederix, TM
EPJC 84, 763 (2024)

- The difference in the fraction of the hardest light jets in even more pronouced in the ≥ 2 b-jet selection

at least 2 b-jets selection



➡ The reason for the large 5FS–4FS difference in the $p_T^{t\bar{t}}$ spectrum at large momenta is

- The correlation between $p_T^{t\bar{t}}$ and $p_T^{\text{light jet, hardest}}$
- Expected 5FS–4FS difference between the fraction of events with the hardest jet being light-flavoured

To summarise…

To summarise:

- $t\bar{t}b\bar{b}$ production serves as a significant (often irreducible) background process across various high-energy physics phenomena
- 5FS calculation of $t\bar{t}b\bar{b}$ at NLO yields the most accurate prediction for this process to date
  - no large logarithms appearing in the matrix element calculation
  - no complications when matching to a parton shower
- We compute the $t\bar{t}$ + jets process with up to 2 jets at NLO using the FxFx merging prescription and match it to the Pythia8 shower
- To improve the efficiency of selecting events with additional b-jets we enhance the probability of producing short-distance events with additional b-quarks using a newly implemented feature in the MadGraph5_aMC@NLO generator
  - This makes producing the $t\bar{t}b\bar{b}$ in the 5FS at NLO more viable, given the computational demands of the 5FS approach
- Similar heavy-flavour enhancement could also be applied to the “fusion” method in Sherpa
  - Which might help in increasing the accuracy of the computation for the additional jets

BACK-UP

Alternative approach: Sherpa fusion

[Höche, Krause, Siegert (2019)](#)
[Ferencz, Höche, Katzy, Siegert (2024)](#)

- $t\bar{t}b\bar{b}$ matched to $t\bar{t}$+jets in a variable flavour number scheme
- Should have at least the same precision as 5FS, if computed at the same order
- But up to now the additional jets in the 5FS component are only computed at LO