Towards an improved MadGraph for multi-jet final states

Timea Vitos
based on arXiv:2409.12128 (with Rikkert Frederix)

Uppsala University,
ELTE Eötvös Loránd Budapest

MadGraph meeting,
CERN, 3 February 2025

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Today's talk

1. Introducing the problem: multi-jet processes

2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
3.1 Generating events at LC accuracy
3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Today's talk

1. Introducing the problem: multi-jet processes

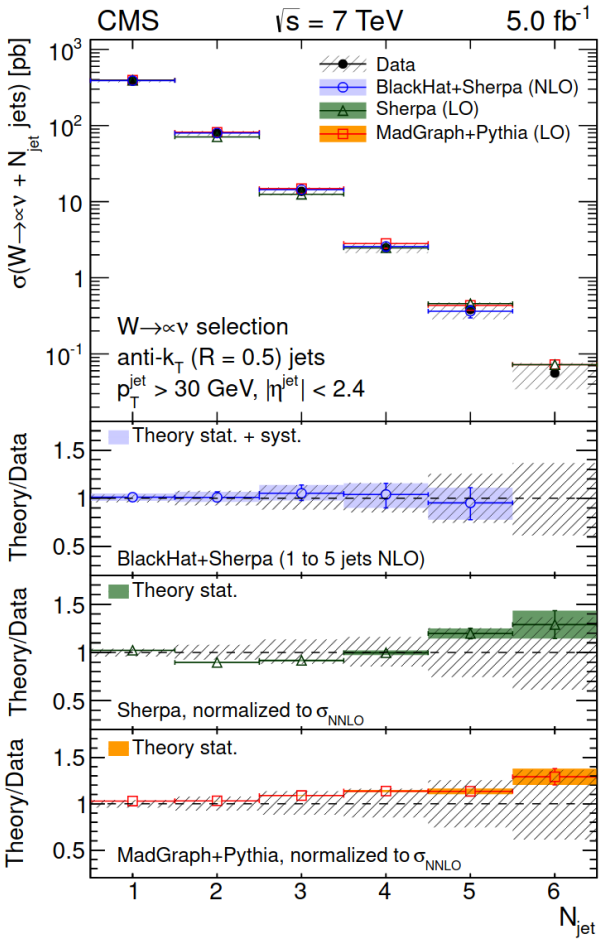
2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
3.1 Generating events at LC accuracy
3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

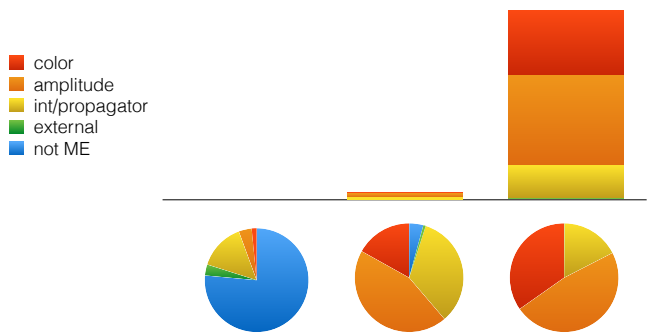
High-multiplicity processes: why bother?

- Hadronic decays of Standard Model background: leads to **multi-jet signatures**
  $\rightarrow \geq$ 5 well-separated final-state jets in the detector
- (Almost) any new physics search includes decays leading to such signatures



Code Complexity

	$gg \to t\bar{t}$	$gg \to t\bar{t}gg$	$gg \to t\bar{t}ggg$
`madevent`	13G	470G	11T
`matrix1`	3.1G (23%)	450G (96%)	11T (>99%)



Complexity raise factorially square

Mattelaer Olivier MadGraph 10

High-multiplicity processes in matrix element generators

- Current colour treatment: **colour decomposition** of amplitudes:

$$\langle|\mathcal{M}|^2\rangle \propto \sum_{\sigma_k,\sigma_l} \underbrace{C(\sigma_k,\sigma_l)}_{\text{colour matrix}} \underbrace{\mathcal{A}(\sigma_k)(\mathcal{A}(\sigma_l))^*}_{\text{dual amplitudes}} \tag{1}$$

  where $\sigma_{k,l}$ are some permutations of the external particles (each dual amplitude $\mathcal{A}(\sigma_k)$ corresponds to a **colour order**)

- Colour matrix size for n final state particles: $\sim n! \times n!$
- This factorial bottleneck is present already at **tree-level**: place for a first-step improvement

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Today's talk

1. Introducing the problem: multi-jet processes

2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
3.1 Generating events at LC accuracy
3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Colour decomposition of amplitudes

$$\langle|\mathcal{M}|^2\rangle \propto \sum_{\sigma_k,\sigma_l} C(\sigma_k,\sigma_l)\mathcal{A}(\sigma_k)(\mathcal{A}(\sigma_l))^*$$

- Expand in N_C^{-2} (large-N_C limit)

$$C(\sigma_k,\sigma_l) = \underbrace{a_0 N_C^x}_{\text{Leading colour (LC)}} + \underbrace{a_1 N_C^{x-2}}_{\text{Next-to-leading colour (NLC)}} + \mathcal{O}(N_C^{x-4}) \quad \forall k,l \quad (2)$$

  keeping all terms yields full colour (FC)

- Schematically, we obtain a sparse colour matrix in this NLC approximation:

$$C(\sigma_k,\sigma_l) \to \begin{pmatrix} \text{LC} & 0 & 0 & 0 & 0 & \text{NLC} \\ 0 & \text{LC} & 0 & \text{NLC} & 0 & 0 \\ 0 & 0 & \text{LC} & 0 & 0 & 0 \\ 0 & \text{NLC} & 0 & \text{LC} & 0 & 0 \\ 0 & 0 & 0 & 0 & \text{LC} & 0 \\ \text{NLC} & 0 & 0 & 0 & 0 & \text{LC} \end{pmatrix}$$

- Truncation at NLC greatly improves the number of terms in the colour sum[1]

[1] R. Frederix, T. Vitos, arXiv:2109.10377

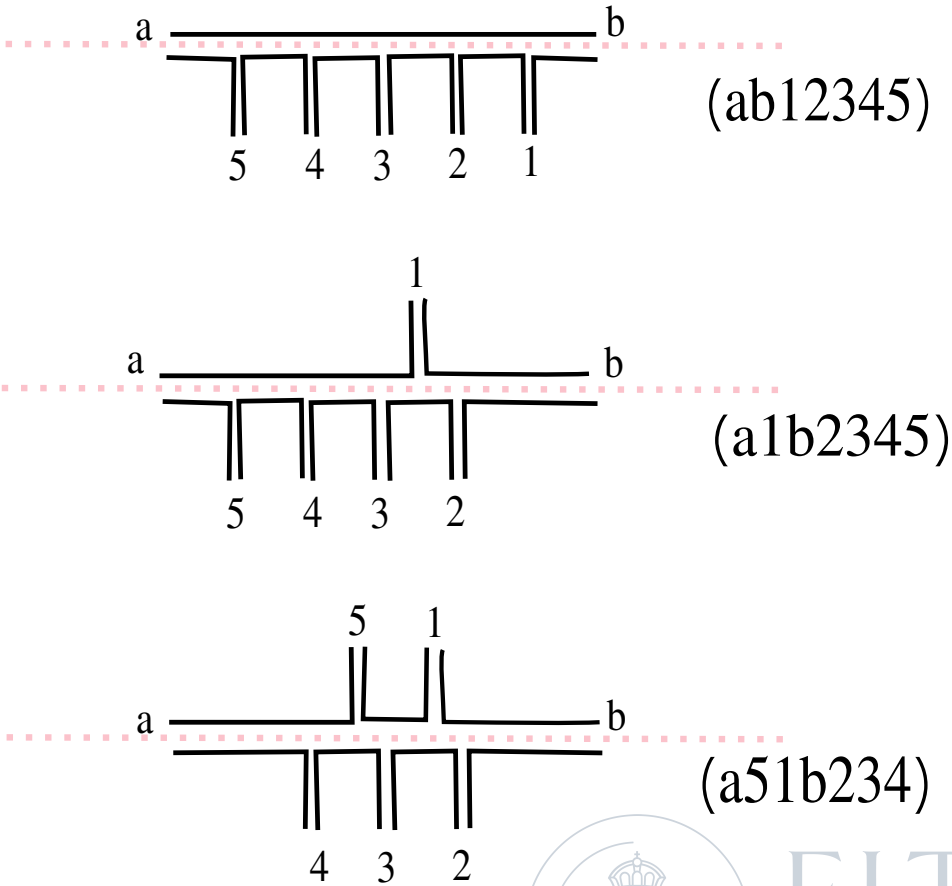
Some conventions for colour ordering

- Colour order classified by the string of particles in the process

$$ab \to 1 \ldots n$$

- With quarks lines, same convention:
  each quark line being independent with respect to the overall order

$$(\underbrace{q_1 g_1 g_2 g_3 \bar{q}_1}_{\text{quark line 1}} \underbrace{q_2 g_4 g_5 \bar{q}_2}_{\text{quark line 2}})$$



Today's talk

1. Introducing the problem: multi-jet processes

2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
   - 3.1 Generating events at LC accuracy
   - 3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Proposal for efficient event generation

- The colour-summed and averaged matrix element using the colour decomposition

$$\langle|\mathcal{M}|^2\rangle \propto \sum_{\{c\},\{h\}} \sum_{\sigma_k,\sigma_l} C(\sigma_k,\sigma_l)\mathcal{A}(\sigma_k,\{h\})(\mathcal{A}(\sigma_l,\{h\}))^* \tag{3}$$

- Expansion in N_C^{-2} (for all-gluon process):

$$\langle|\mathcal{M}|^2\rangle = \frac{N_C^n(N_C^2-1)}{2^{(n-2)}} \sum_{\sigma_k} \left(\underbrace{|\mathcal{A}(\sigma_k,\{h\})|^2}_{\text{LC}} + \mathcal{O}(N_C^{-2}) \right) \tag{4}$$

- Dominant peak structure from squared MHV amplitudes:

$$\mathcal{A}(\sigma_k,h)(\mathcal{A}(\sigma_k,h))^* \sim \frac{1}{s_{\sigma_k(1)\sigma_k(2)} \cdot s_{\sigma_k(2)\sigma_k(3)} \cdots s_{\sigma_k(n)\sigma_k(1)}} \tag{5}$$

with $s_{\sigma(i)\sigma(j)} = (p_{\sigma(i)} + p_{\sigma(i+1)})^2$

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Proposal for efficient event generation

- Re-write the integrated cross section as:

$$\int \sum_{\{c\}} |\mathcal{M}|^2 \mathrm{d}\Phi_n = \sum_{\sigma_k} \int |\mathcal{A}_{\sigma_k}|^2 \underbrace{\frac{\sum_{\{c\}} \sum_{\sigma_k,\sigma_l} \mathcal{A}_{\sigma_k} C_{\sigma_k\sigma_l} \mathcal{A}^*_{\sigma_l}}{\sum_{\sigma_k} |\mathcal{A}_{\sigma_k}|^2}}_{\text{reweight factor}} \mathrm{d}\Phi_n \tag{6}$$

- **Idea**: generate events with approximated integrand
  $\rightarrow$ LC approximation of integrand

- Generate FC-accurate events in two steps:

  1. **Integration**: approximate peak structure of integrand with LC accuracy
  2. **Reweighting**: reweight LC events to FC with reweight factor

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Today's talk

1. Introducing the problem: multi-jet processes

2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
   3.1 Generating events at LC accuracy
   3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Phase-space integration

- Integration over the Lorentz-invariant phase space

$$\mathrm{d}\Phi_n \sim \prod_{i=1}^{n} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i}$$

- For hadron collisions, we have an additional integral

$$\int \mathrm{d}x_1 \mathrm{d}x_2 \tag{7}$$

- We investigate two setups:

LHC setup

- Collision with protons at $\sqrt{s} = 14$ TeV
- Cuts: $p_T > 30$ GeV, $|\eta| < 6.0$, $\Delta R > 0.4$

Fixed-energy setup

- Partonic collision at $\sqrt{\hat{s}} = 1$ TeV
- Cuts: $\sqrt{|2p_i \cdot p_j|} \geq 30$ GeV

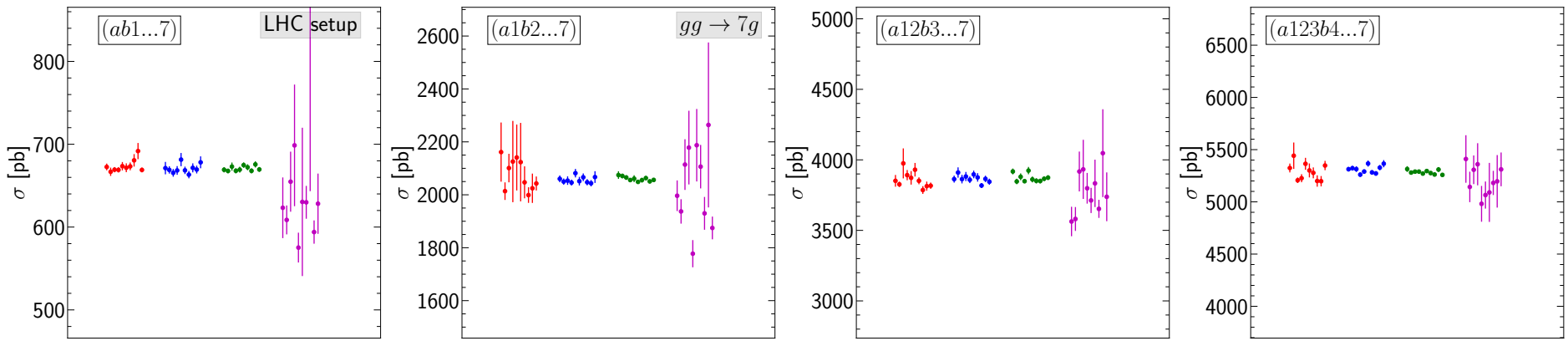
ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Phase-space integration

- Any variable change from d^3p works: but some work better than others!
- We make a comparison of four choices of integration variables:

HAAG[2] t-channel[3] $2 \to 3$[4] p_T-based[5]

- Compare the Monte Carlo error for the cross section and spread of result for 10 random seeds, with fixed number of grid points

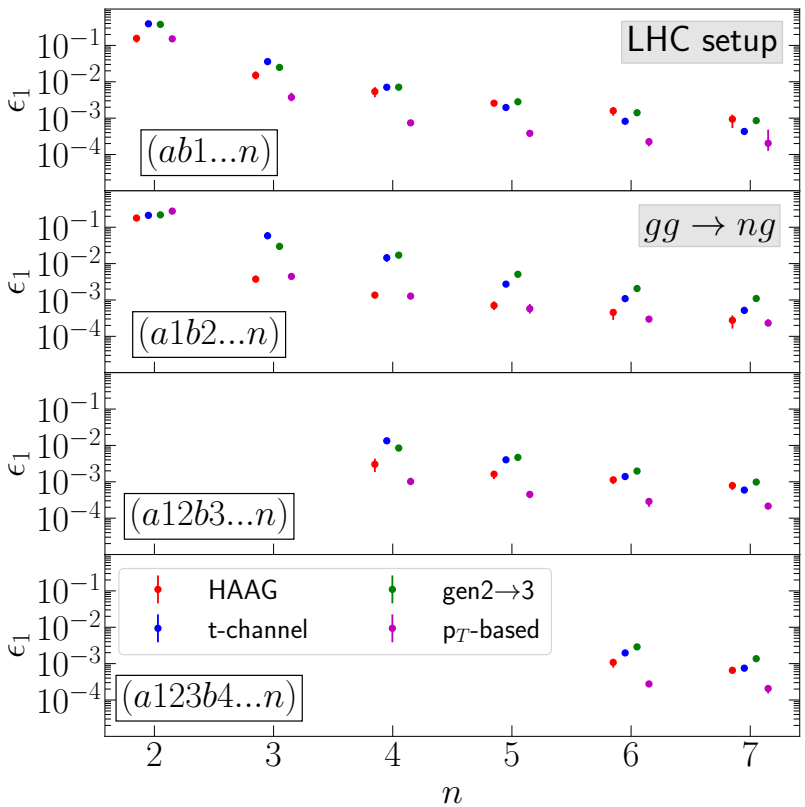


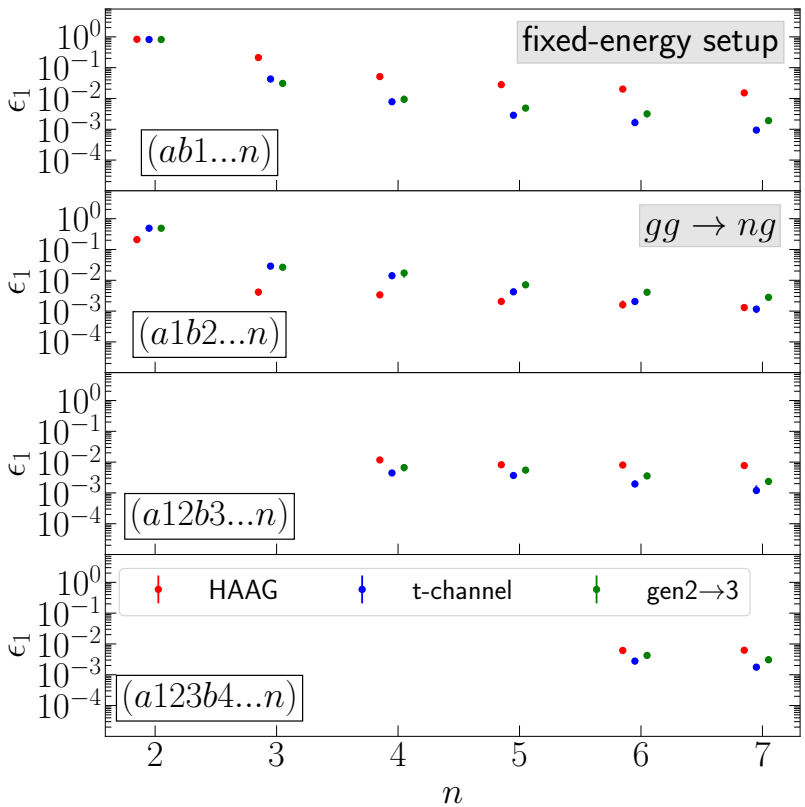
[2] A. van Hameren, C.G. Papadopoulos, arXiv:hep-ph/0204055
[3] E. Byckling, K. Kajantie, 10.1016/0550-3213(69)90271-5
[4] E. Byckling, K. Kajantie, 10.1103/PhysRev.187.2008
[5] Bothmann et al., arXiv:2302.10449

Phase-space integration

- Efficiency measure of LC events:

$$\epsilon_1 = \frac{N_{\text{events}}}{N_{\text{total}}} \tag{8}$$





Today's talk

1. Introducing the problem: multi-jet processes

2. Colour decomposition of amplitudes

3. Proposing a new colour-aware generator
   - 3.1 Generating events at LC accuracy
   - 3.2 Reweighting events to FC and NLC accuracy

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Reweighting the LC events to FC and NLC

- From the set of LC-accurate events, generating FC accuracy, one reweights each event with $r_{\mathrm{LC}\to\mathrm{FC}}$ according to

$$\sum_{\sigma_k}\int |\mathcal{A}_{\sigma_k}|^2 \underbrace{\frac{\sum_{\{c\}}\sum_{\sigma_k,\sigma_l}\mathcal{A}_{\sigma_k}C_{\sigma_k\sigma_l}\mathcal{A}^*_{\sigma_l}}{\sum_{\sigma_k}|\mathcal{A}_{\sigma_k}|^2}}_{\text{reweight factor } r_{\mathrm{LC}\to\mathrm{FC}}} \mathrm{d}\Phi_n$$

- An alternative is to reweight to NLC accuracy, reducing significantly the amount of terms needed in the colour sum

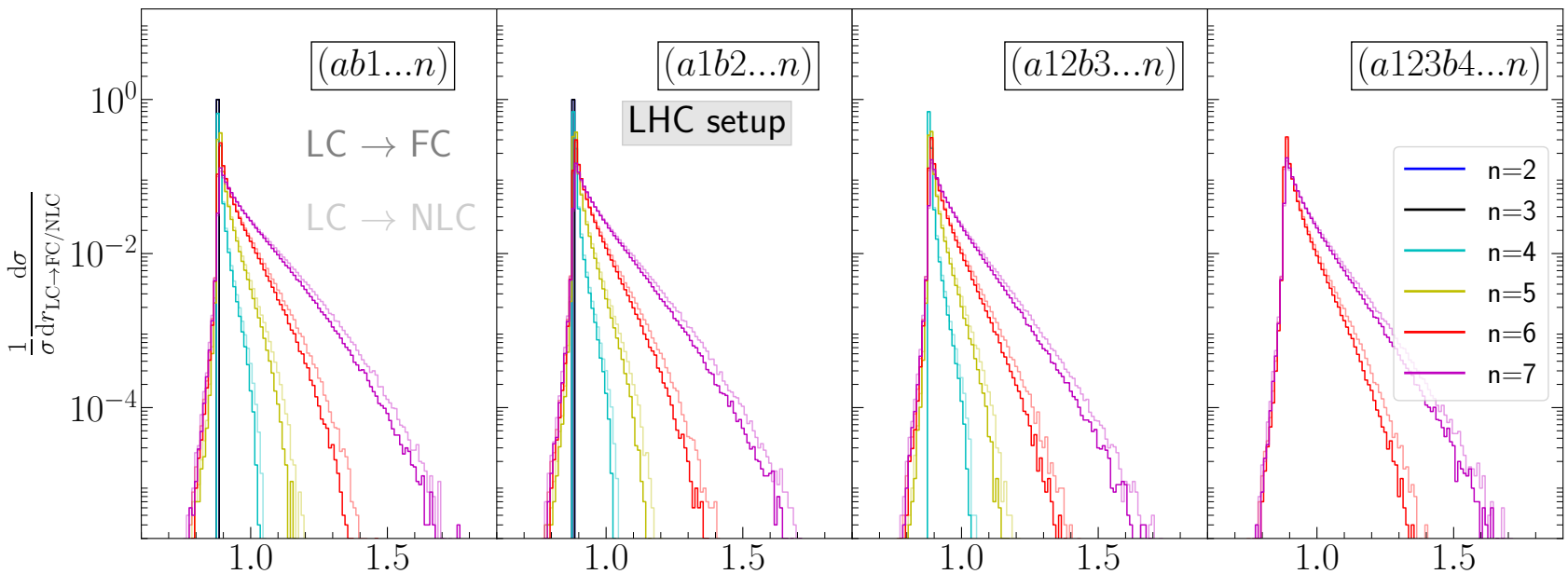
$$\frac{\sum_{\{c\}}\sum_{\sigma_k,\sigma_l}\mathcal{A}_{\sigma_k}C_{\sigma_k\sigma_l}\mathcal{A}^*_{\sigma_l}}{\sum_{\sigma_k}|\mathcal{A}_{\sigma_k}|^2} \to \underbrace{\frac{1}{\sum_{\sigma_k}|\mathcal{A}_{\sigma_k}|^2}\left(\sum_{\{c\}}\sum_{\sigma_k}\mathcal{A}_{\sigma_k}\times\left[\sum_w C_w\mathcal{A}^*_w\right]\right)}_{r_{\mathrm{LC}\to\mathrm{NLC}}} \quad (9)$$

where all $C_w \sim$ NLC

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

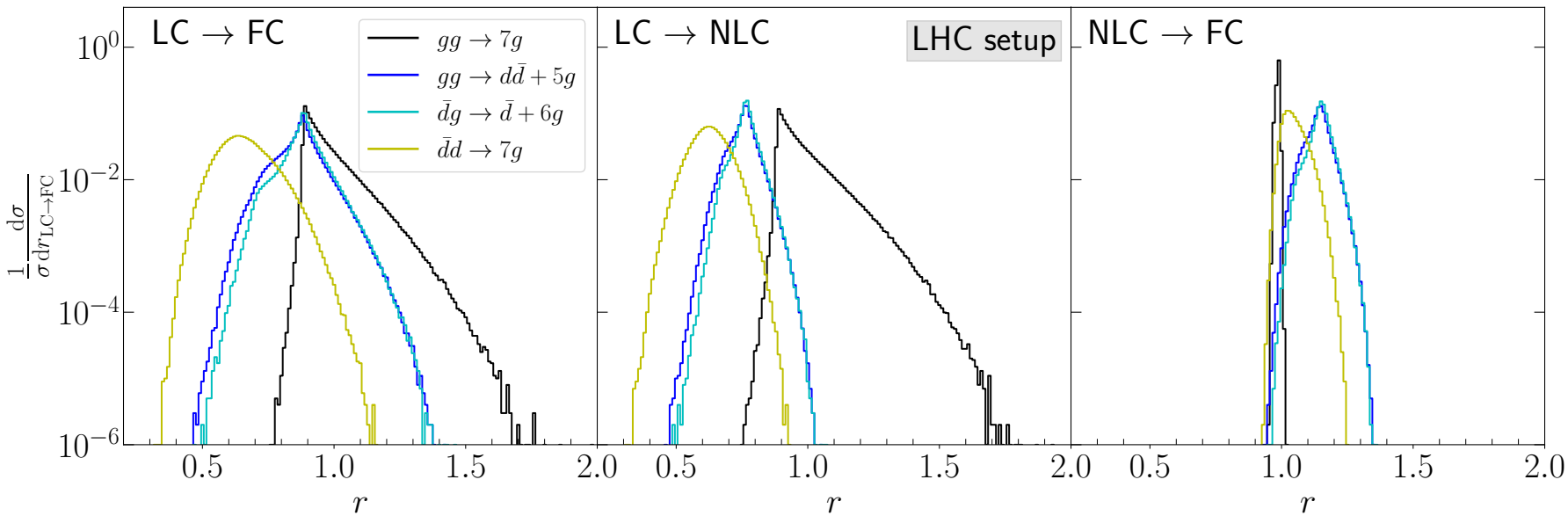
Weight spread of reweight procedure

- For a **secondary unweighting**, we need events which are kept with high efficiency and no very high weights → narrow weight spread
- Multiplicity dependence of this weight spread for all-gluon process shows peaked distribution for the reweight factors



Weight spread of reweight procedure

- For processes with one quark line the reweighting efficiency is similar:

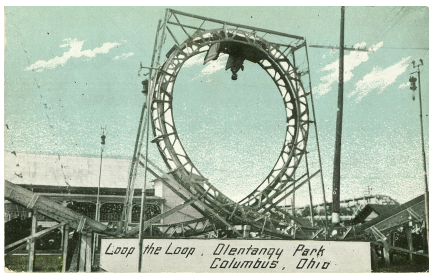


Summary and outlook

- Comparison of integration parametrisations:
  $\rightarrow$ overall, the $2\rightarrow3$ method is the **best-performing**
  $\rightarrow$ for higher multiplicities, for partonic processes and for simple colour orders, HAAG performs better
  $\rightarrow$ a p_T-based integration is **less efficient**, even with physical cuts
- Reweigh technique is promising:
  $\rightarrow$ relatively small weight spread for all process types

Future prospects

- Extend to physical processes and LHC physics: add singlets, massive bosons, massive quarks
- Extend to NLO computations (colour treatment in loops)



Thank you for listening!

Appendix

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Colour matrix: fundamental decomposition

For n-gluon amplitudes

- Matrix-element

$$\mathcal{M} = g^{n-2} \sum_{\sigma \in S_{n-1}} \mathrm{Tr}[T^{a_1} T^{a_{\sigma(1)}} \ldots T^{a_{\sigma(n-1)}}] \mathcal{A}(1, \sigma(1), \ldots, \sigma(n-1)) \tag{10}$$

Not a minimal set

Dual amplitudes are related by the Kleiss-Kuijf relation (dual Ward identity)

$$\mathcal{A}(1,2,3,4\ldots,n) + \mathcal{A}(2,1,3,4,\ldots,n) + \ldots + \mathcal{A}(2,3,4,\ldots,1,n) = 0 \tag{11}$$

- Squared matrix-element

$$|\mathcal{M}|^2 = (g^2)^{n-2} \sum_{k,l=1}^{(n-1)!} C_{kl} \mathcal{A}(1, \sigma_k(1), \ldots, \sigma_k(n-1)) \left(\mathcal{A}(1, \sigma_l(1), \ldots, \sigma_l(n-1))\right)^*$$

- Colour matrix (size $(n-1)! \times (n-1)!$):

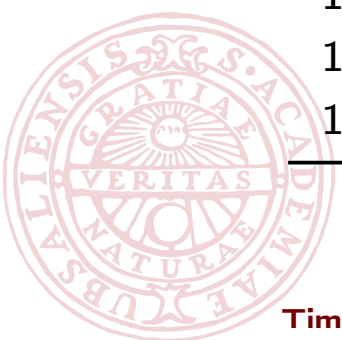
$$C_{kl} = \sum_{\mathrm{col.}} \mathrm{Tr}[T^{a_1} T^{a_{\sigma_k(1)}} \ldots T^{a_{\sigma_k(n-1)}}] \left(\mathrm{Tr}[T^{a_1} T^{a_{\sigma_l(1)}} \ldots T^{a_{\sigma_l(n-1)}}]\right)^* \tag{12}$$

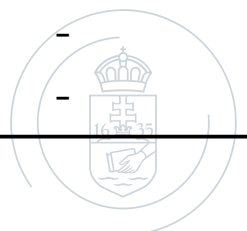
Colour matrix: results

For n-gluon amplitudes

Including the adjoint decomposition: matrix size $(n-2)! \times (n-2)!$

all-gluon			
n	Fundamental	Colour-flow	Adjoint
4	6 (6)	6 (6)	2 (2)
5	11 (24)	16 (24)	5 (6)
6	24 (120)	36 (120)	18 (24)
7	50 (720)	71 (720)	93 (120)
8	95 (5040)	127 (5040)	583 (720)
9	166 (40320)	211 (40320)	4162 (5040)
10	271 (362880)	331 (362880)	31649 (40320)
11	419 (3628800)	496 (3628800)	-
12	620 (39916800)	716 (39916800)	-
13	885 (479001600)	1002 (479001600)	-
14	1226 (6227020800)	1366 (6227020800)	-





ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Adjoint decomposition

For n-gluon amplitudes

- The amplitude is now

$$\mathcal{M} = \sum_{\sigma \in S_{n-\mathbf{2}}} \left(F^{a_{\sigma(\mathbf{2})}} \ldots F^{a_{\sigma(n-\mathbf{1})}}\right)_{a_{\mathbf{1}} a_n} \mathcal{A}(1, \sigma(1), \ldots, \sigma(n), n), \tag{13}$$

  with $(F^a)_{bc} = i f^{abc}$
- Minimal basis: $(n-2)!$ independent dual amplitudes
- Smaller colour matrix: but **LC not only on diagonal!**
- No found algorithm (yet) to get NLC elements

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Colour-flow decompositions

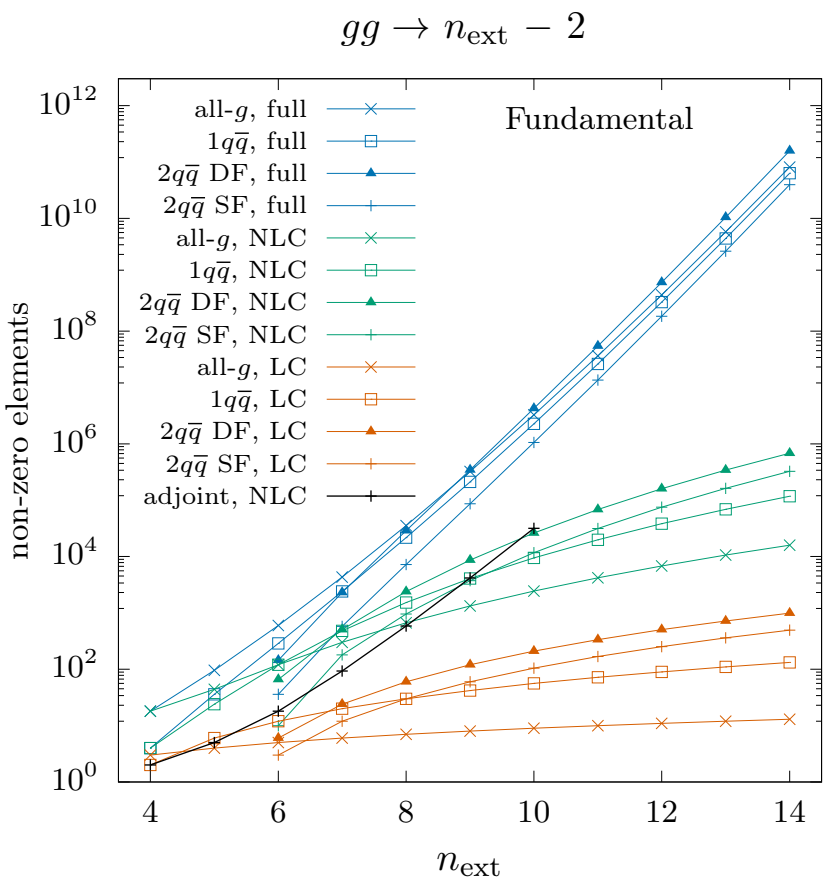
For one quark line plus n-gluon amplitudes: the full projection of U(1) gluons

$$
\begin{aligned}
\mathcal{M}_{1qq} = & \sum_{\sigma\in S_n} \delta^{i_q}_{j_{\sigma(1)}} \delta^{i_{\sigma(1)}}_{j_{\sigma(2)}} \dots \delta^{i_{\sigma(n-1)}}_{j_{\sigma(n)}} \delta^{i_{\sigma(n)}}_{j_q} \mathcal{A}_{1qq}(q,\sigma(1),\dots,\sigma(n),\overline{q}) \\
& + \left(\frac{-1}{N}\right) \sum_{\sigma\in S_n} \delta^{i_q}_{j_{\sigma(1)}} \delta^{i_{\sigma(1)}}_{j_{\sigma(2)}} \dots \delta^{i_{\sigma(n-1)}}_{j_q} \delta^{i_{\sigma(n)}}_{j_{\sigma(n)}} \mathcal{A}_{1qq}(q,\sigma(1),\dots,\sigma(n-1),\overline{q},\sigma(n)) \\
& + \left(\frac{-1}{N}\right)^2 \frac{1}{2!} \sum_{\sigma\in S_n} \delta^{i_q}_{j_{\sigma(1)}} \delta^{i_{\sigma(1)}}_{j_{\sigma(2)}} \dots \delta^{i_{\sigma(n-2)}}_{j_q} \delta^{i_{\sigma(n-1)}}_{j_{\sigma(n-1)}} \delta^{i_{\sigma(n)}}_{j_{\sigma(n)}} \\
& \qquad \mathcal{A}_{1qq}(q,\sigma(1),\dots,\sigma(n-2),\overline{q},\sigma(n-1),\sigma(n)) \\
& + \dots \\
& + \left(\frac{-1}{N}\right)^n \frac{1}{n!} \sum_{\sigma\in S_n} \delta^{i_q}_{j_q} \delta^{i_{\sigma(1)}}_{j_{\sigma(1)}} \dots \delta^{i_{\sigma(n-1)}}_{j_{\sigma(n-1)}} \delta^{i_{\sigma(n)}}_{j_{\sigma(n)}} \mathcal{A}_{1qq}(q,\overline{q},\sigma(1),\dots,\sigma(n)).
\end{aligned}
$$

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Results for gg initiated processes[6]

- For external particles $n_{\text{ext}} \in [4, 14]$
- Blue: full colour Green: NLC Red: LC



[6] R. Frederix, T. Vitos arXiv:2109.10377

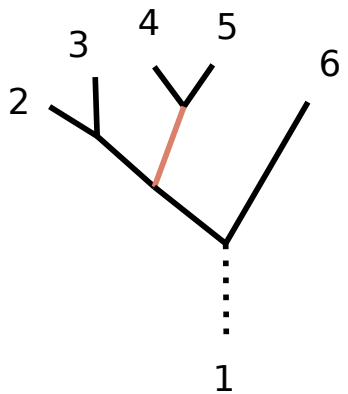
Integration

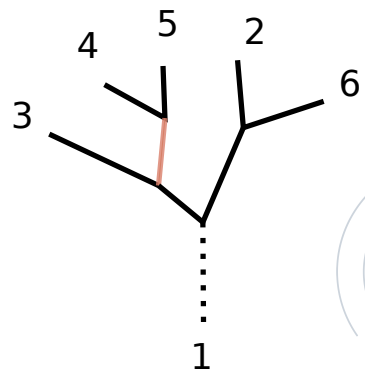
Amplitude evaluation

- Each of the $\mathcal{A}_{\sigma_k}$ computed in a recursive fashion
  $\rightarrow$ recycle parts of the amplitudes
- For a fixed colour-ordering (example for $n = 6$):

$$\begin{gathered}
\{1,2\} \quad \{2,3\} \quad \{3,4\} \quad \{4,5\} \\
\{(1,2),(2,3)\} \quad \{(2,3),(3,4)\} \quad \{(3,4),(4,5)\} \\
\{((1,2),(2,3)),((2,3),(3,4))\} \quad \{((2,3),(3,4)),((3,4),(4,5))\} \\
\{(((1,2),(2,3)),((2,3),(3,4))),(((2,3),(3,4)),((3,4),(4,5)))\}
\end{gathered} \tag{14}$$

- Example: same current (4,5) appears in the two colour-orderings
  $\sigma = (1,2,3,4,5,6)$ and $\sigma = (1,3,4,5,2,6)$

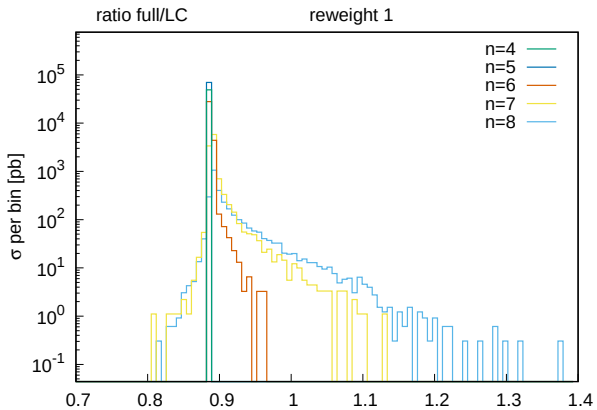


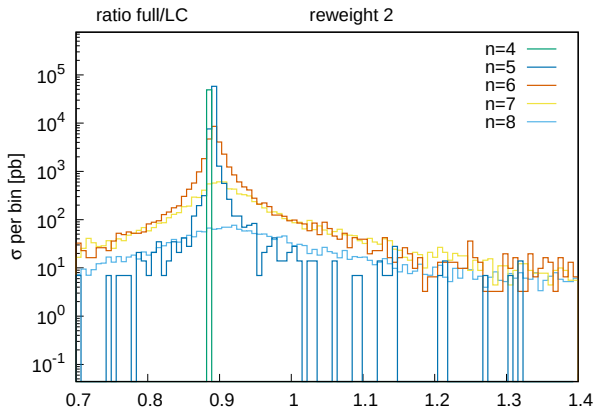


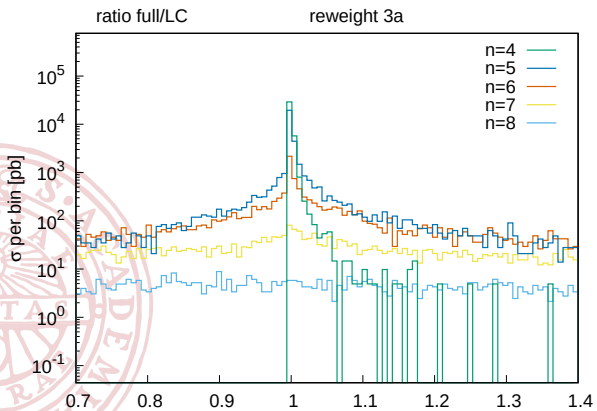
Reweighting the LC events to FC

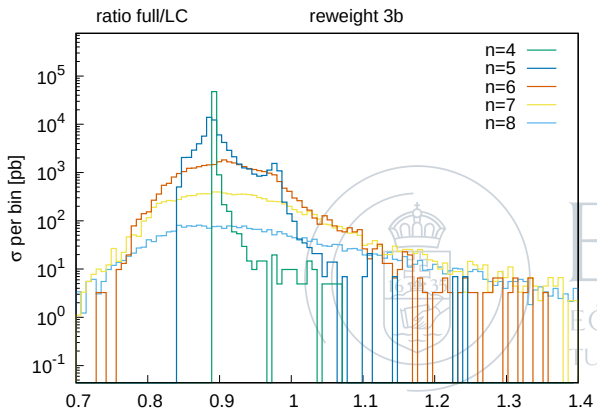
Obtaining NLC and FC accurate predictions from the LC events can be done in several ways:

1. Perform FC sum (sum over colour configurations)
2. Pick subset of permutations (using phase space symmetry)
3. a) Pick specific colour configuration randomly (consistent with colour ordering) or b) use an average of all colour configurations





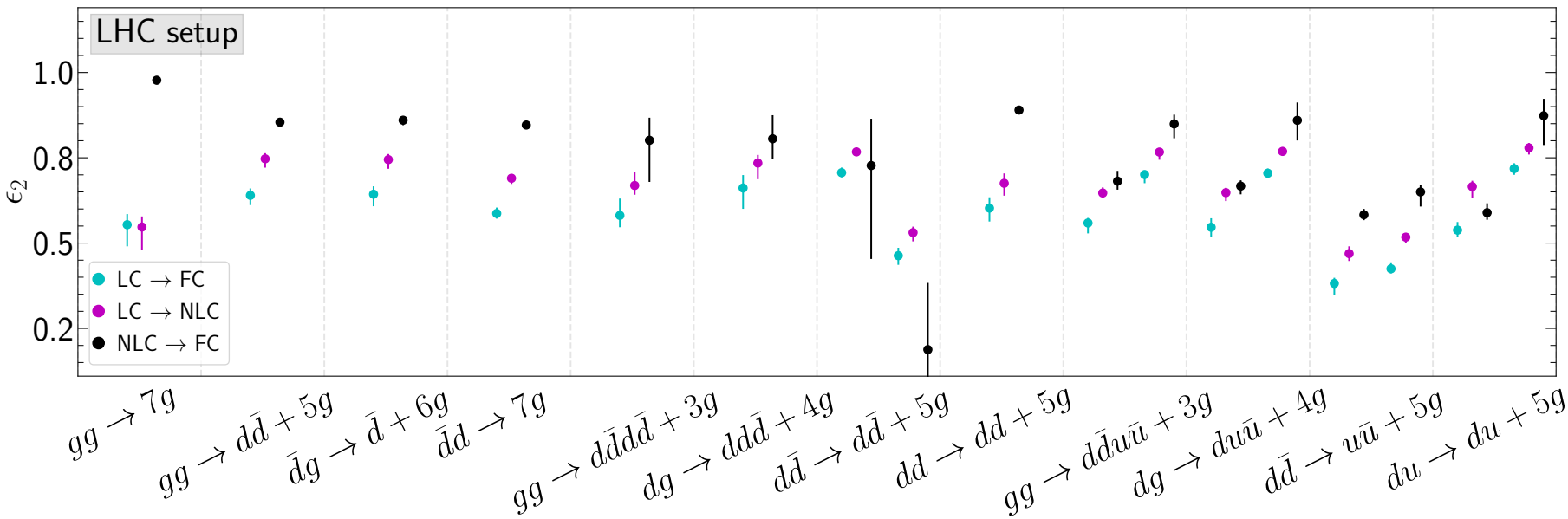




Weight spread of reweighting procedure

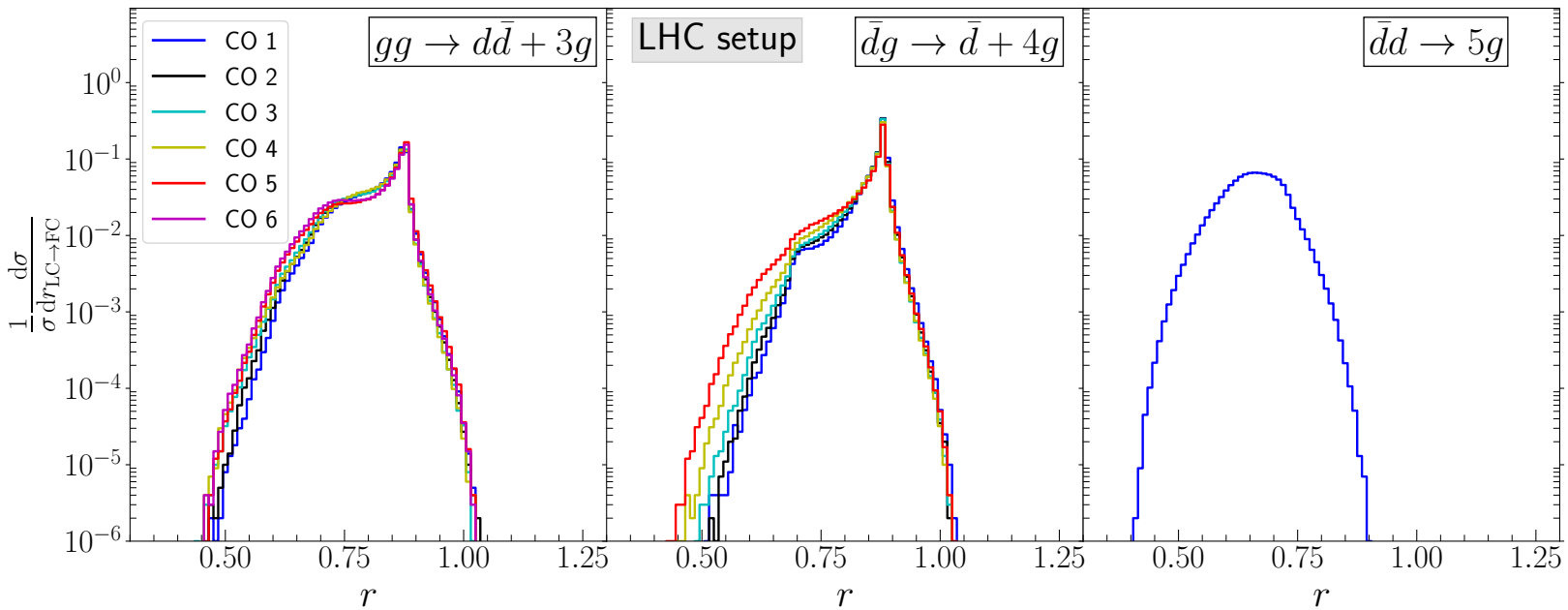
- Use a secondary unweighting efficiency to measure the spread of weights:

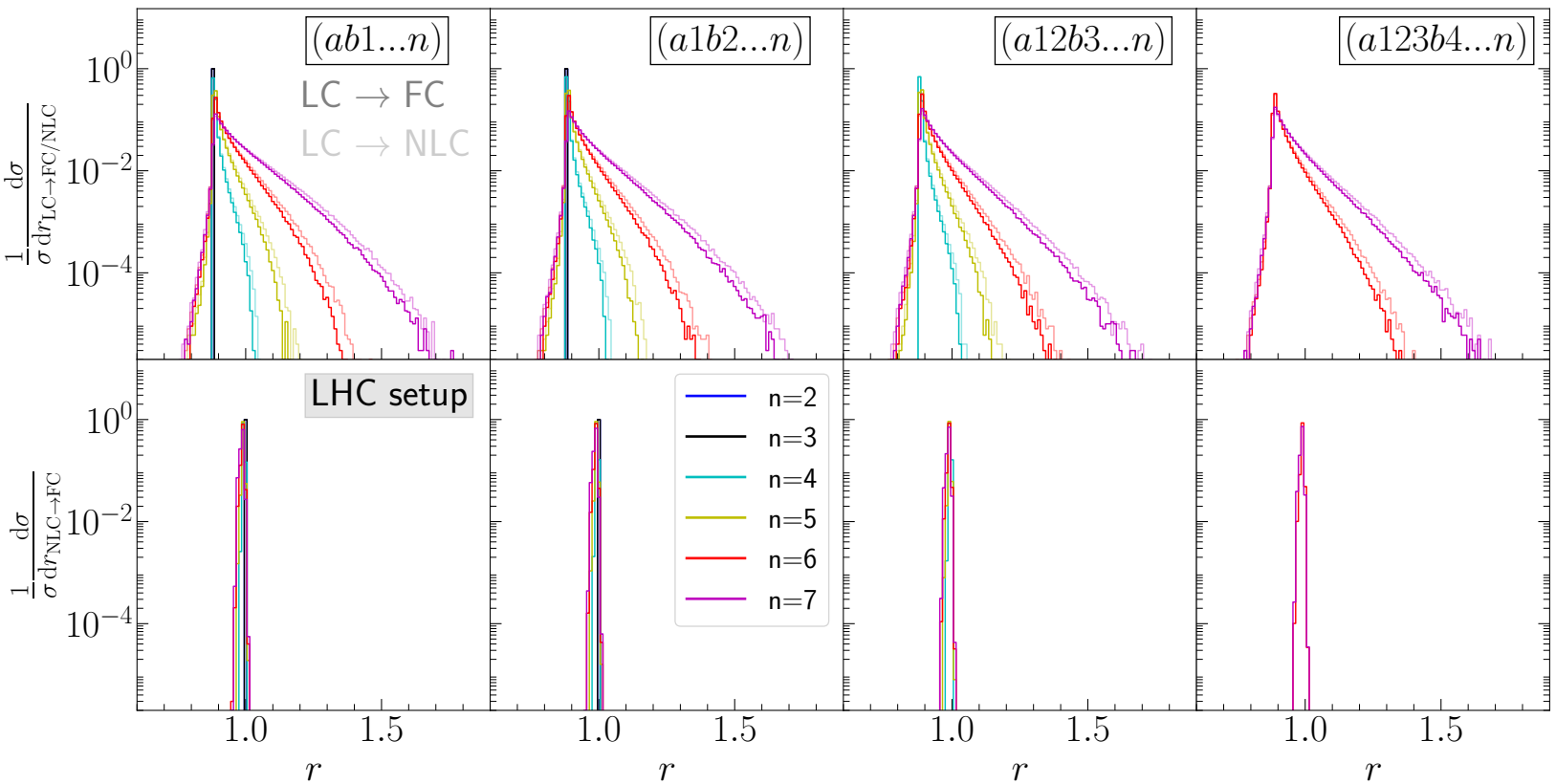
$$\epsilon_2 = \frac{1}{N} \sum_i \frac{r_i}{r_{\max}} \tag{15}$$



Weight spread of reweighting procedure

- Reweight factor for varying colour ordering for various one quark line processes:



Weight spread of reweighting procedure

- Reweight factor for varying colour ordering for all-gluons, including the NLC→FC reweight factors