SCET 2022, Bern

April 19th 2022

Power Counting Energy Flow Polynomials

Pedro Cal

Based on: 2204.xxxxx

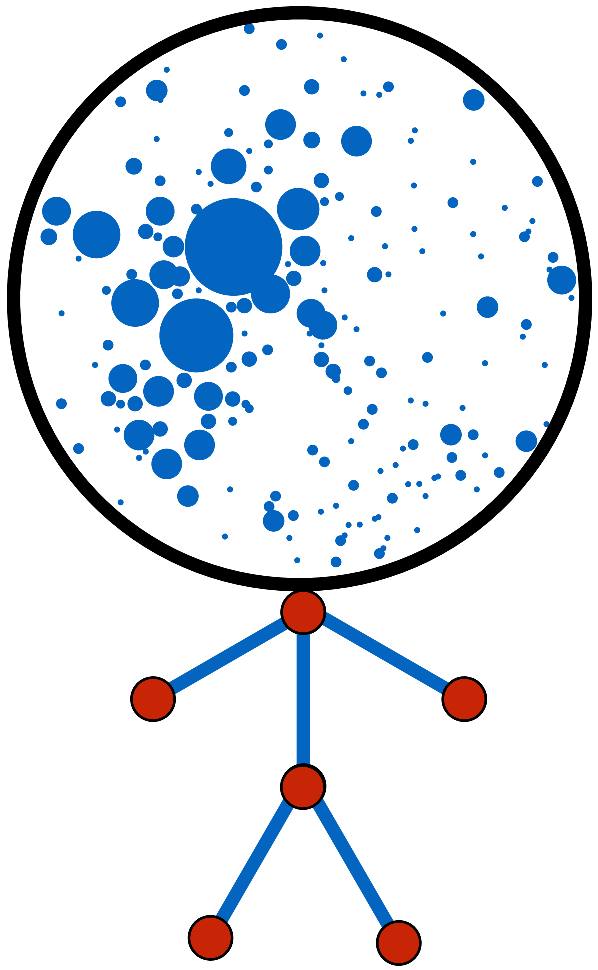
In collaboration with:

Jesse Thaler (MIT)

Wouter Waalewijn (Amsterdam)

Outline

- **Intro**
  - Energy Flow Polynomials (EFPs)
  - The power of power counting
- **Power counting Energy Flow Polynomials**
  - Strongly ordered approximation
  - 1-collinear approximation
  - 2-collinear approximation
  - The bases
  - Logistic regression for quark/gluon classification
- **Conclusions**



Intro

Energy Flow Polynomials

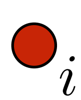
- Energy Flow Polynomials (EFPs): Discrete linear basis for all infrared- and collinear-safe observables

Komiske, Metodiev, Thaler ’17, ’19

- Can be represented in graph form:

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3}^2 \theta_{i_3 i_4}$$

- How this works:



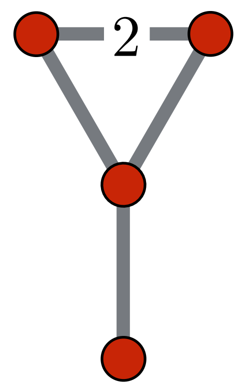
$$_i = \sum_{i=1}^{M} z_i$$

Sum over particles in the jet

$$= \theta_{jk}$$

Angular distance between particles j and k

Energy Flow Polynomials

- One more: 

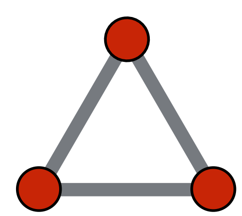
$$= \sum_{i_1,i_2,i_3,i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_3 i_4}^2 \theta_{i_4 i_2}$$

- **Degree**: how many lines in the EFP

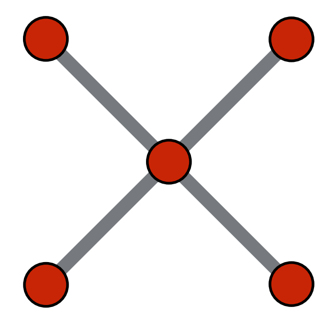
1

2

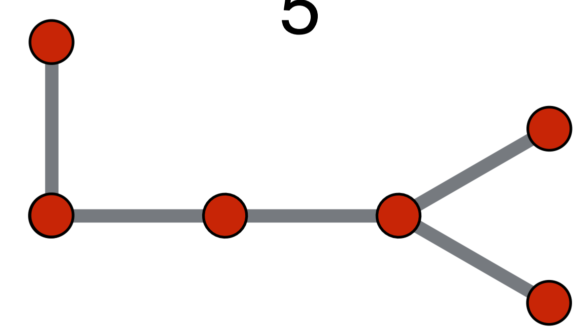
3



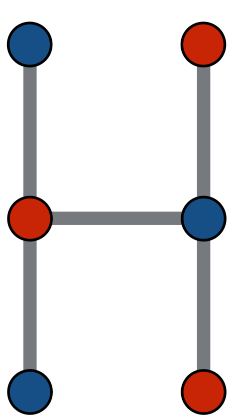
4



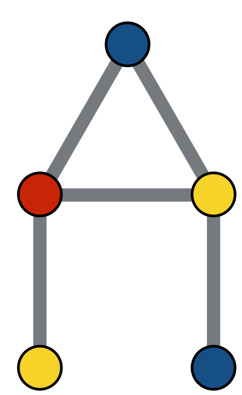
5



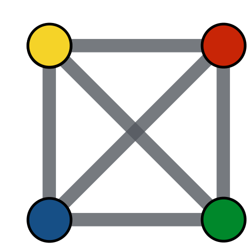
- **Chromatic number**: how many particles needed to have non-vanishing EFP



2



3

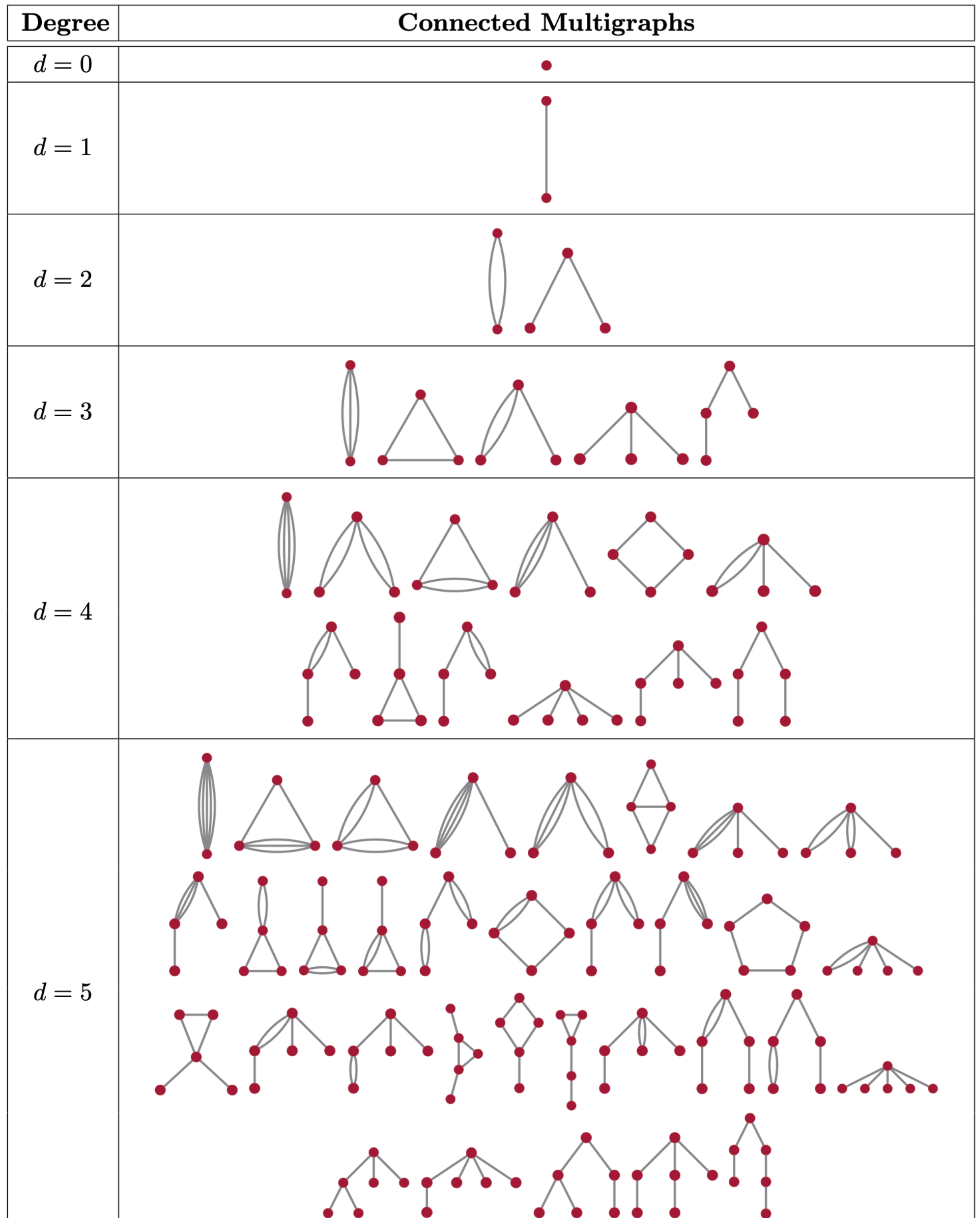


4

Energy Flow Polynomials

All prime (connected) EFPs up to degree 5

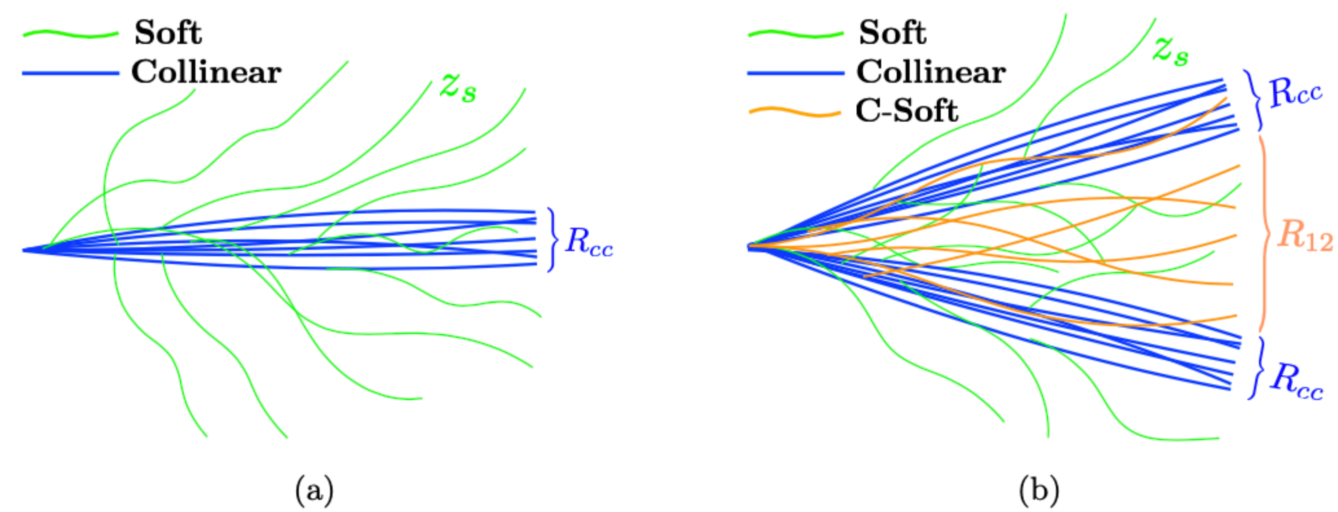
Figure from:
Komiske, Metodiev, Thaler ’17



The power of Power Counting

- Parametric power counting of the dynamics of QCD can be used to design substructure observables

- E.g. : QCD/Boosted-object jet discrimination Larkoski, Moult, Neill ‘14



$$R_{cc} \ll 1\,, \qquad R_{cs} \sim 1$$

$$\Longrightarrow \qquad D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}\,.$$

- We will use one-prong power counting - not to design observables - but to better understand relations between observables

This talk

- Power counting provides relations that greatly reduce the number of independent EFPs at a given level of accuracy

- Using the smaller EFP basis does not affect tagging performance

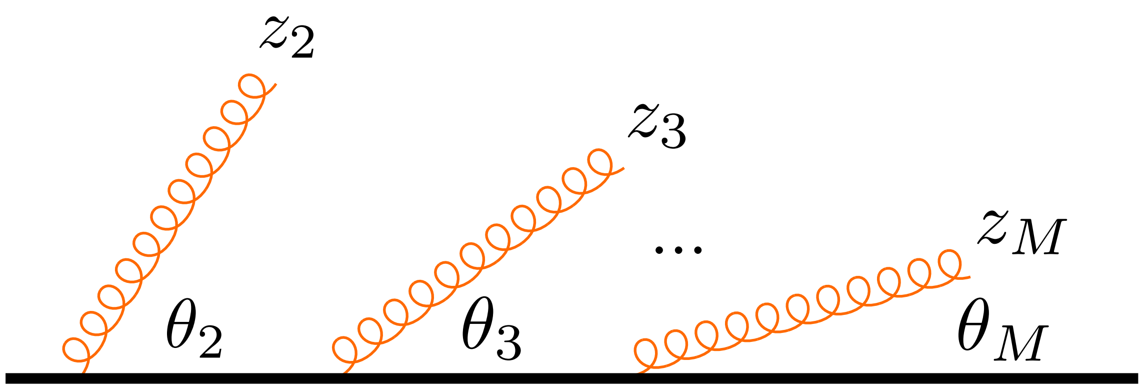
Degree		0	1	2	3	4	5	6
All EFPs	by degree	1	1	3	8	23	66	212
	cumulative	1	2	5	13	36	102	314
SO basis	by degreee	1	1	2	4	7	12	22
	cumulative	1	2	4	8	15	26	49
2-collinear basis	by degree	1	1	2	4	8	16	36
	cumulative	1	2	4	8	16	32	68

Power Counting EFPs

Strongly ordered

- We start our investigation by applying a very restrictive power counting: strong **energy** <u>and</u> **angular** ordering

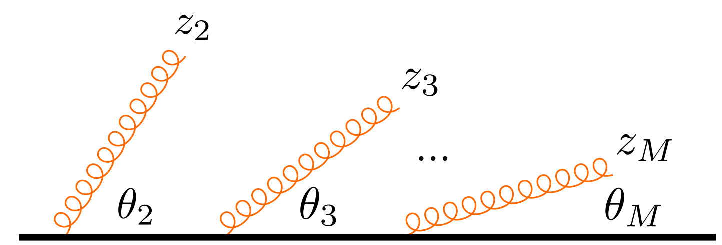
$$z_{i+1} \gg z_i \qquad \text{and} \qquad \theta_{i+1} \gg \theta_i$$



$$z_1 = 1 - \sum_{i=2}^{M} z_i$$

Strongly ordered

- Use this power counting to perform expansion

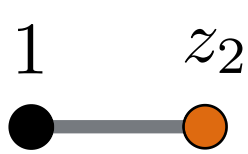
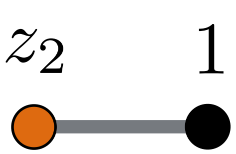


$z_1 = 1 - \sum_{i=2}^{M} z_i$

Full expression:

 $= \sum_{i,j=1}^{M} z_i z_j \theta_{ij}$ → Sum over all particles

Strongly ordered (SO) expansion:

 $\overset{\text{SO}}{=}$  $+$  $+ \mathcal{O}(z_3\theta_3)$

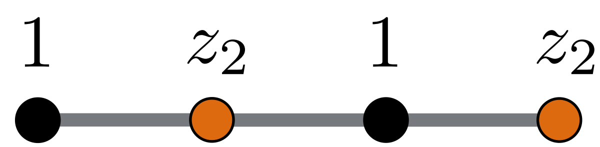
$\overset{\text{SO}}{=} 2z_2\theta_2 + \mathcal{O}(z_3\theta_3)$

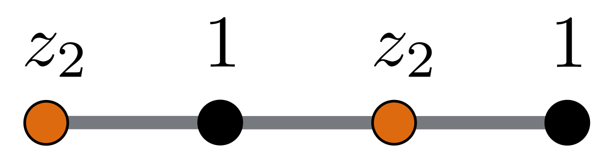
↘ Angle between dominant emission and hard prong

Strongly ordered

 $\overset{\rm SO}{=} \; 2z_2\theta_2 + \mathcal{O}(z_3\theta_3)$

- Let's have a look at a more complicated EFP: 4-dots

 $\overset{\rm SO}{=}$ 

$+$ 

$+ \mathcal{O}(z_2 z_3 \theta_2^2 \theta_3)$

$$\overset{\rm SO}{=} \; 2z_2^2\theta_2^3 + \mathcal{O}(z_2 z_3 \theta_2^2 \theta_3)$$

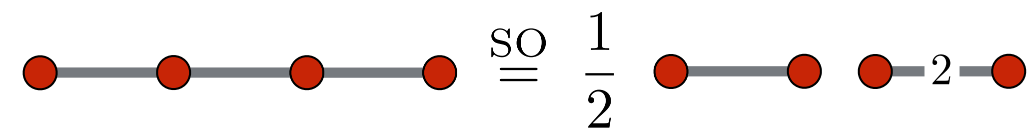
- In the SO limit we can write 4-dots in terms of the dumbbell EFP

 $\overset{\rm SO}{=} \; \frac{1}{2}$  

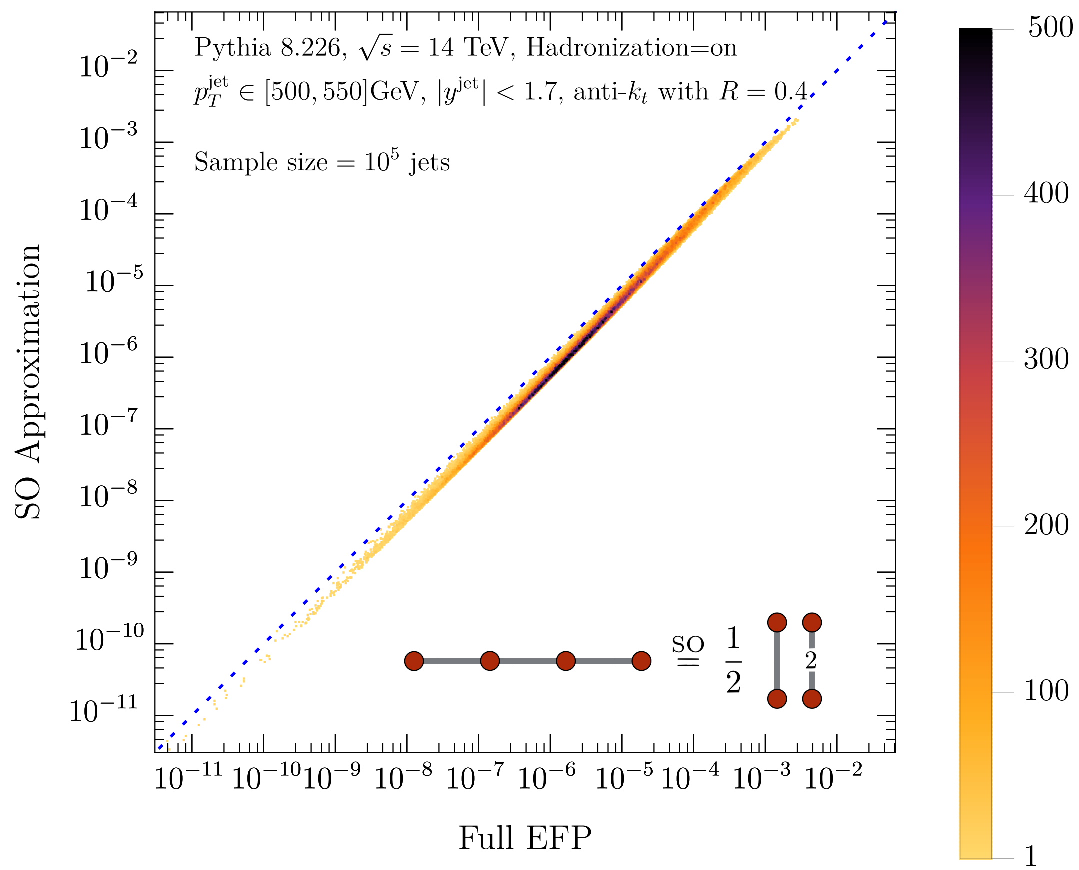
$+ \mathcal{O}(z_2 z_3 \theta_2^2 \theta_3)$

Strongly ordered

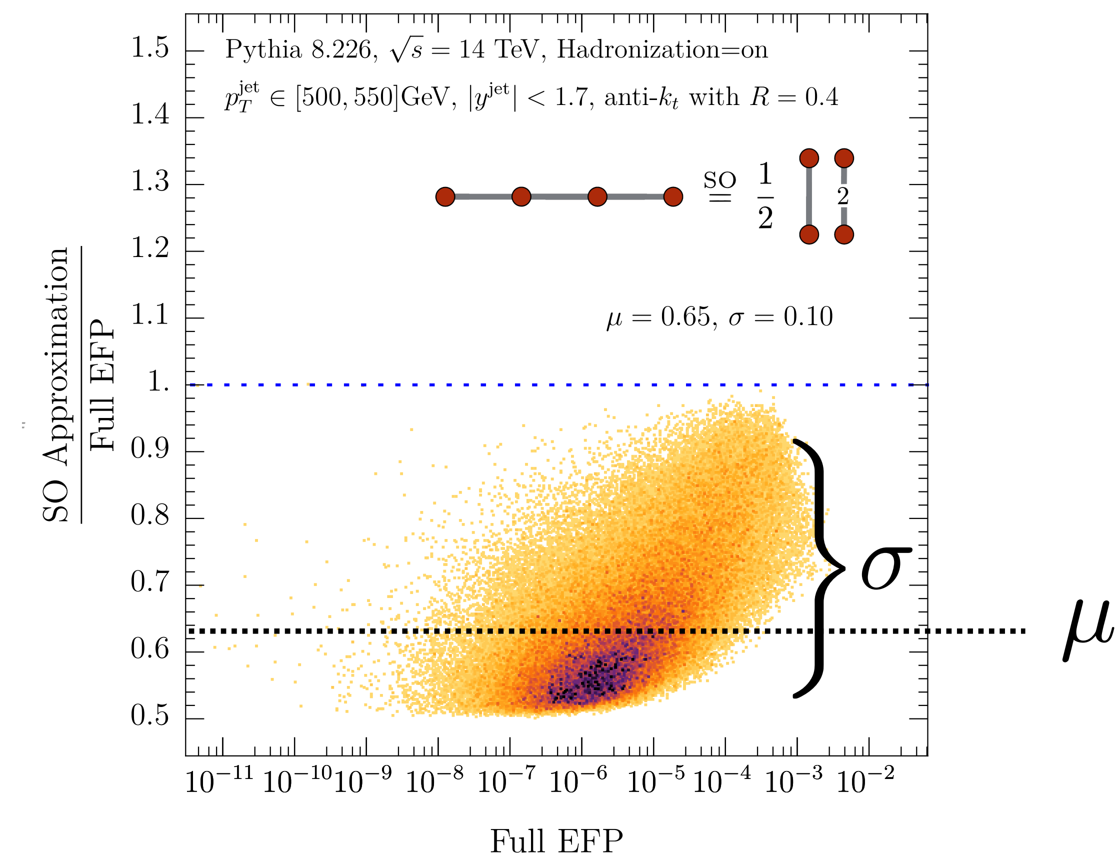
- How does this relation hold in Pythia?



Correlation



Correlation ratio



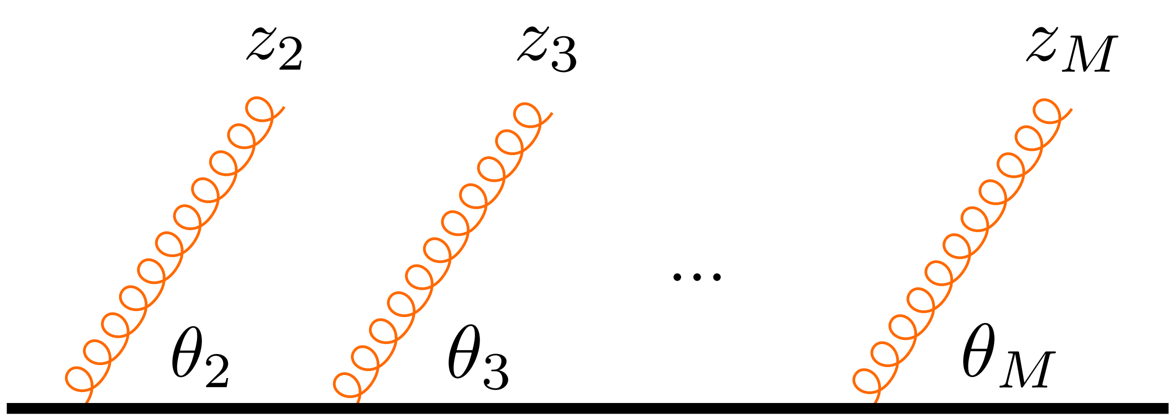
- Correlation is clearly present but expression is off by an overall factor:

$$\mu = 0.65, \quad \sigma = 0.1$$

Cal, Thaler, Waalewijn ’22 (to appear)

1-collinear

- Let's perform a better expansion: the **1-collinear expansion**

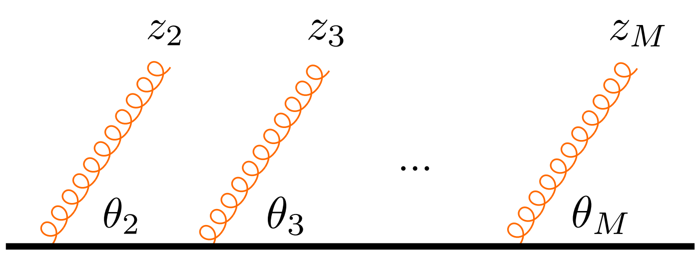


$$z_1 = 1 - \sum_{i=2}^{M} z_i$$

- What do we know in the **1-collinear** approximation?
  - There is **one hard prong**
  - All are emissions are collinear-soft: $z_i \ll 1 \qquad \theta_i \ll 1$
  - We assume no strong ordering

1-collinear

- Use this power counting to perform expansion

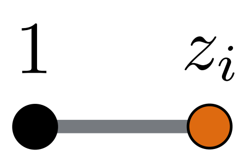
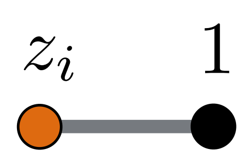


$$z_1 = 1 - \sum_{i=2}^{M} z_i$$

Full expression:

 $= \sum_{i,j=1}^{M} z_i z_j \theta_{ij}$ → Sum over all particles

1-collinear expansion: maximize the number of times the hard prong is assigned

 $\overset{1c}{=}$  $+$  $+\mathcal{O}(z^2)$

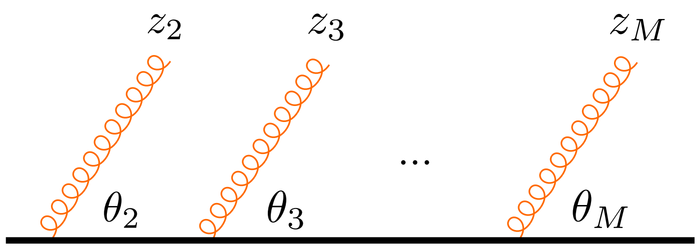
$$\overset{1c}{=} 2\sum_{i=2}^{M} z_i \theta_i + \mathcal{O}(z^2)$$

↓ Sum over collinear-soft particles

→ $\theta_i \equiv$ Angle between particle i and hard prong

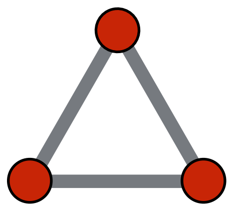
1-collinear

- Use this power counting to perform expansion



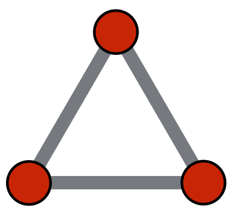
$$z_1 = 1 - \sum_{i=2}^{M} z_i$$

Full expression:

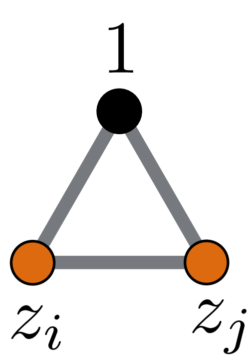


$$= \sum_{i,j,k=1}^{M} z_i z_j z_k \theta_{ij} \theta_{jk} \theta_{ki}$$

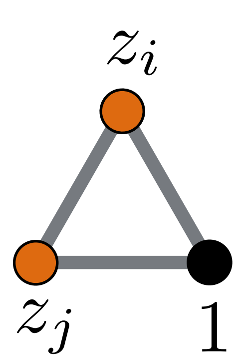
1-collinear expansion:



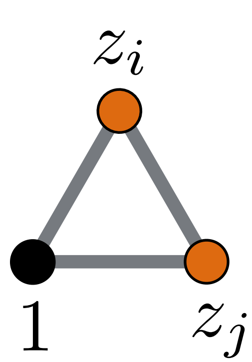
$$\stackrel{1c}{=}$$



$+$



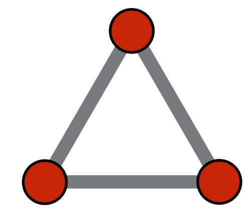
$+$



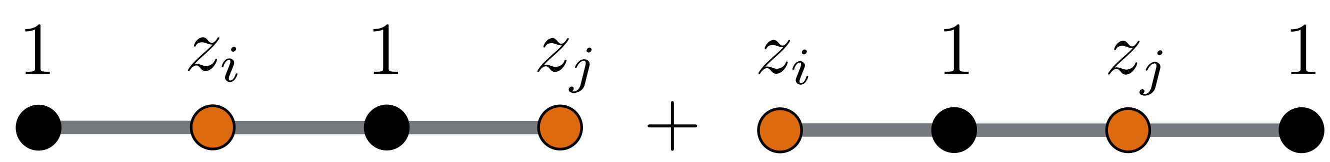
$+\mathcal{O}(z^3)$

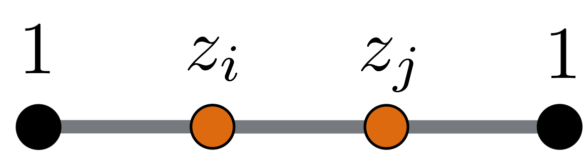
$$\stackrel{1c}{=} 3 \sum_{i,j=2}^{M} z_i z_j \theta_i \theta_j \theta_{ij} + \mathcal{O}(z^3)$$

1-collinear

- So far, at **1c**:  $\stackrel{1c}{=} 2\sum_{i=2}^{M} z_i\theta_i$  $\stackrel{1c}{=} 3\sum_{i,j=2}^{M} z_i z_j \theta_i\theta_j\theta_{ij}$

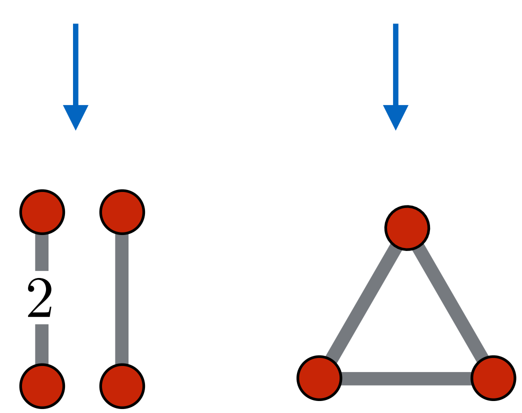
- What happens to 4-dots now?

 $\stackrel{1c}{=}$ 

new ← $+$ 

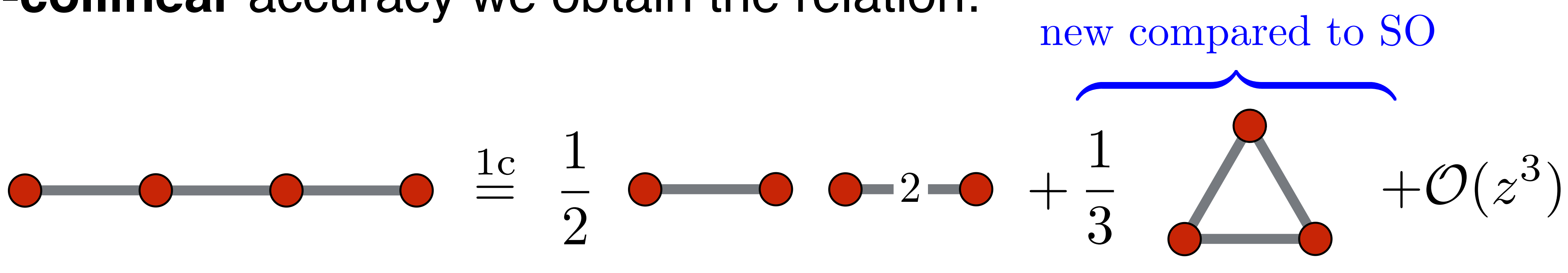
$+\mathcal{O}(z^3)$

$$\stackrel{1c}{=} \sum_{i,j=2}^{M} z_i z_j \left(2\theta_i^2\theta_j + \theta_i\theta_j\theta_{ij}\right) + \mathcal{O}(z^3)$$

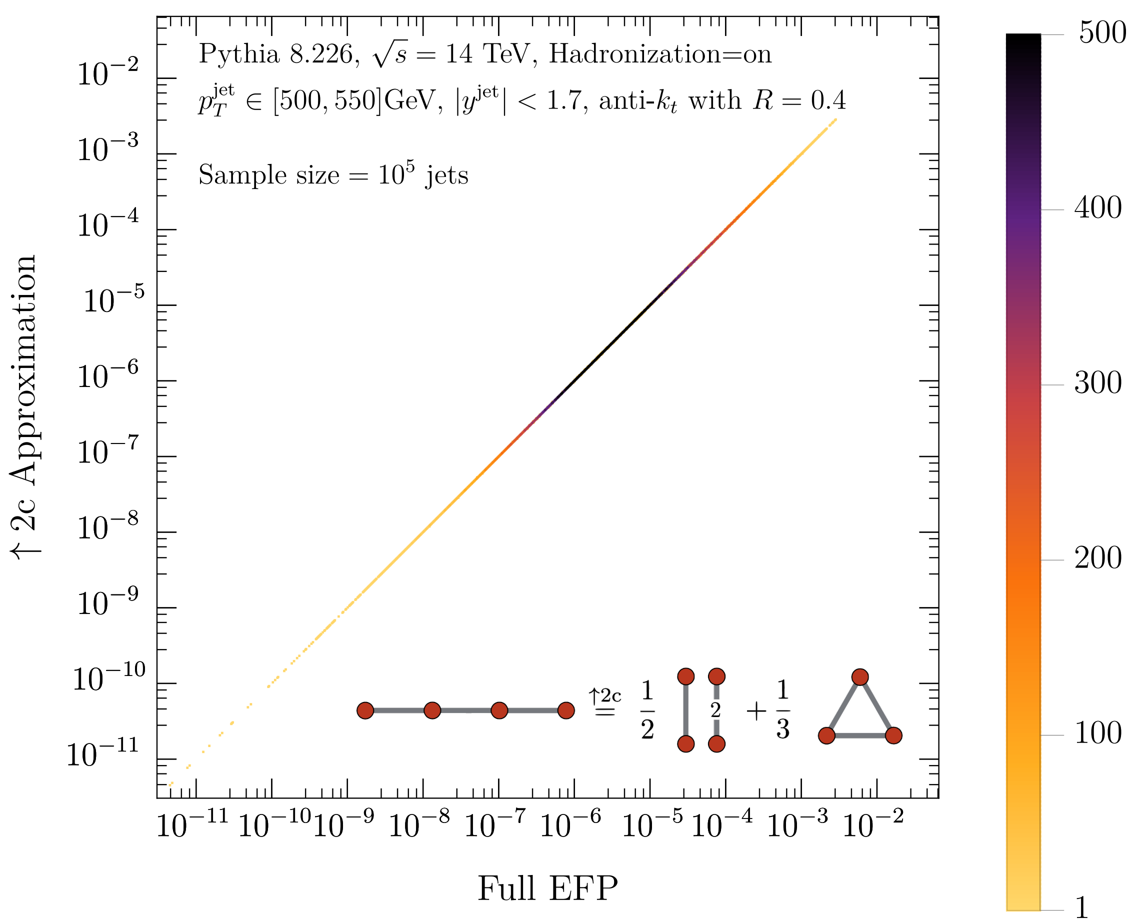


1-collinear

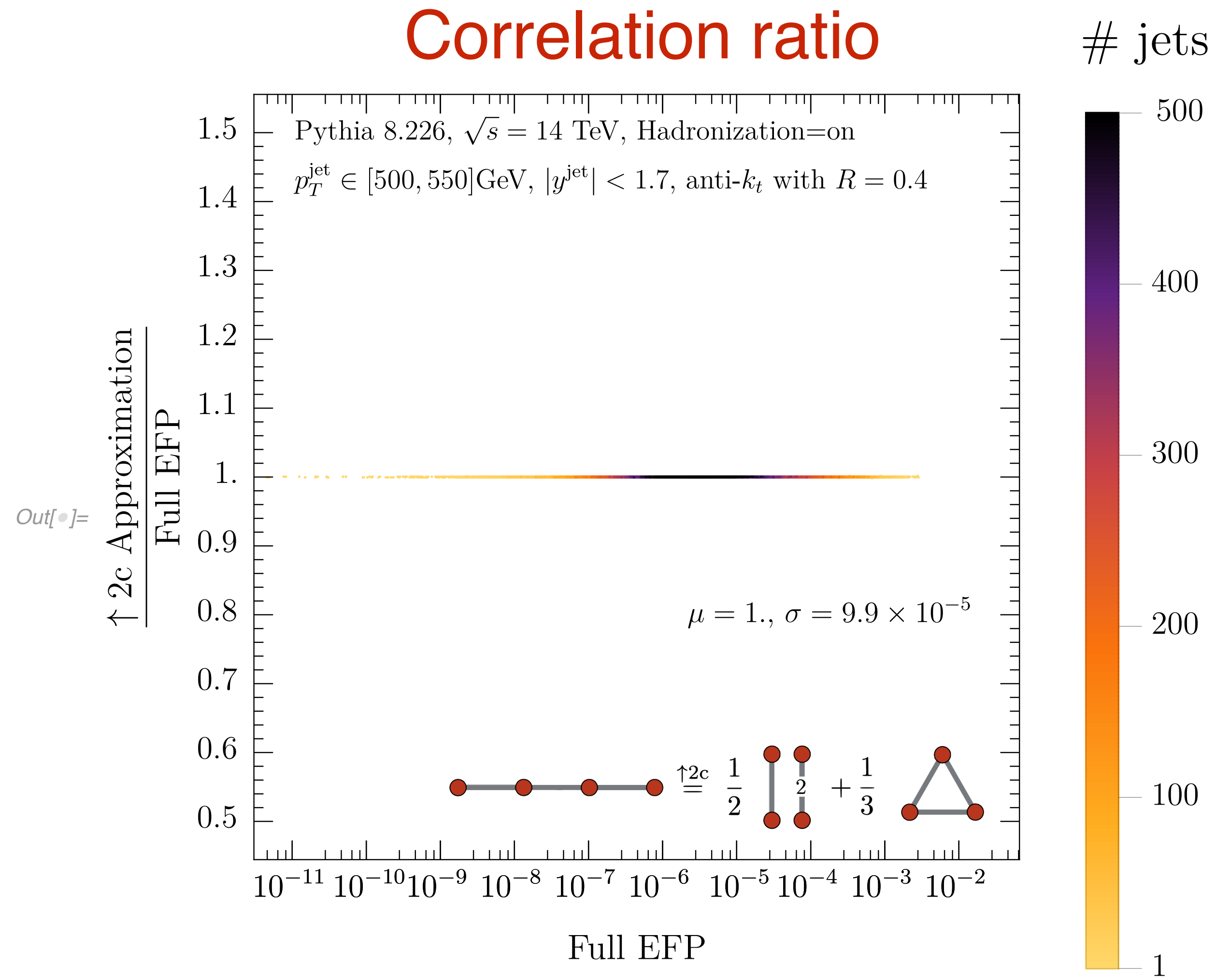
- At **1-collinear** accuracy we obtain the relation:



Correlation



Correlation ratio

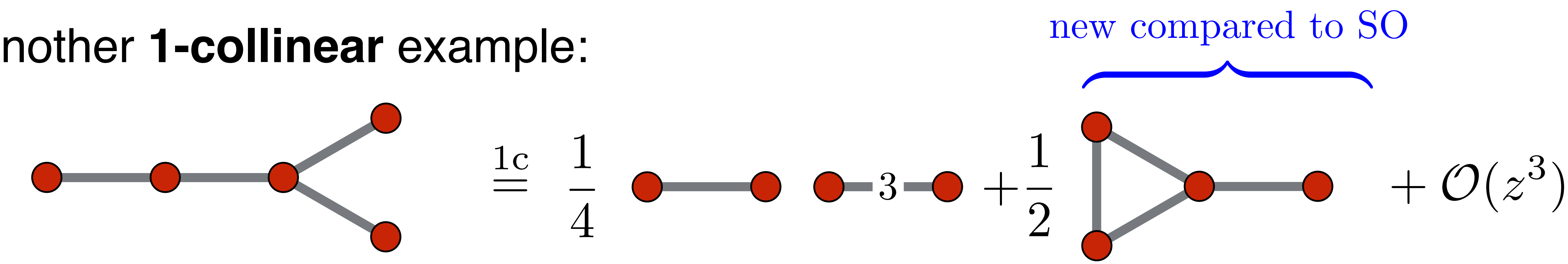


$$\mu = 1.0000, \quad \sigma = 1 \times 10^{-5}$$

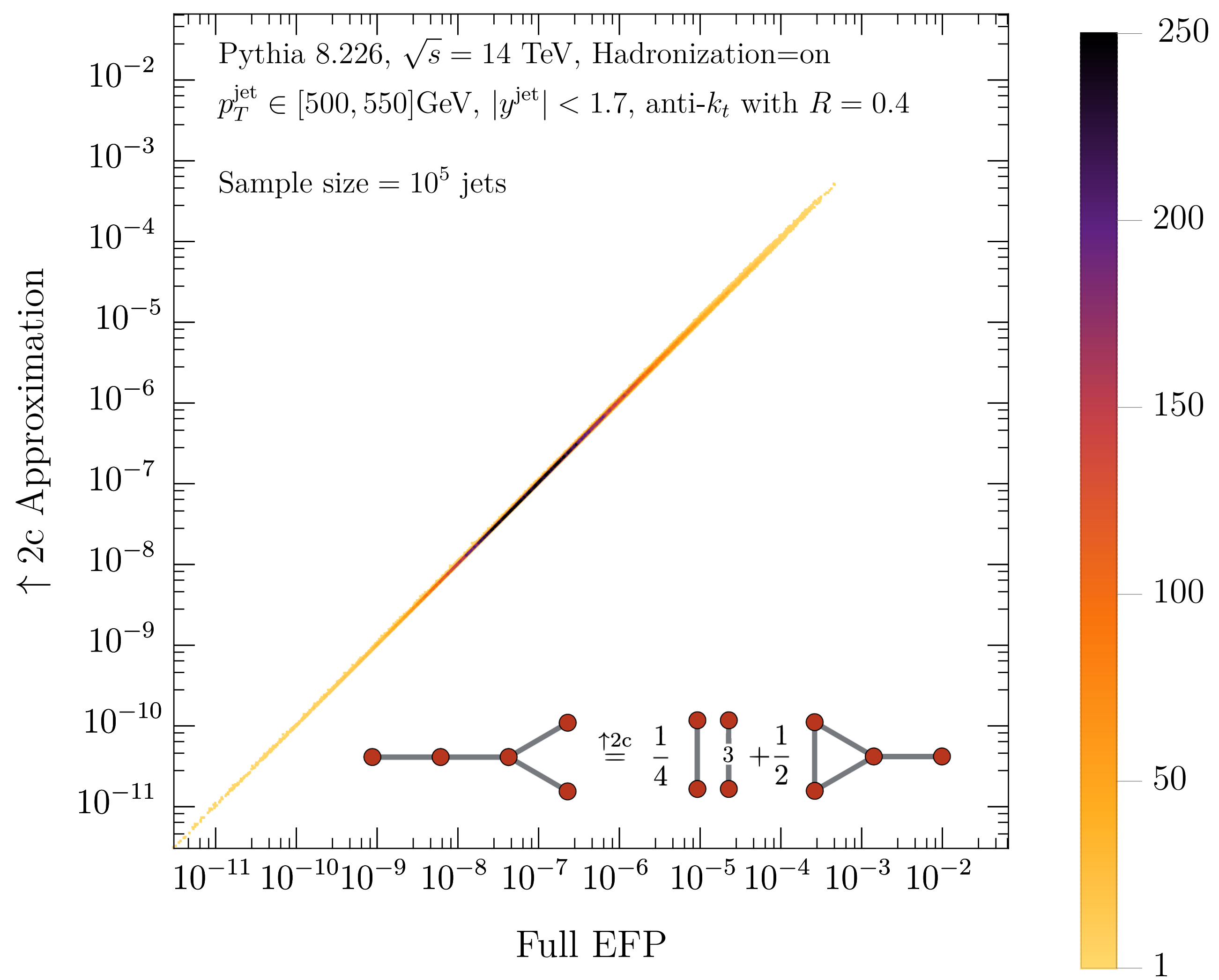
Cal, Thaler, Waalewijn '22 (to appear)

1-collinear

- Another **1-collinear** example:

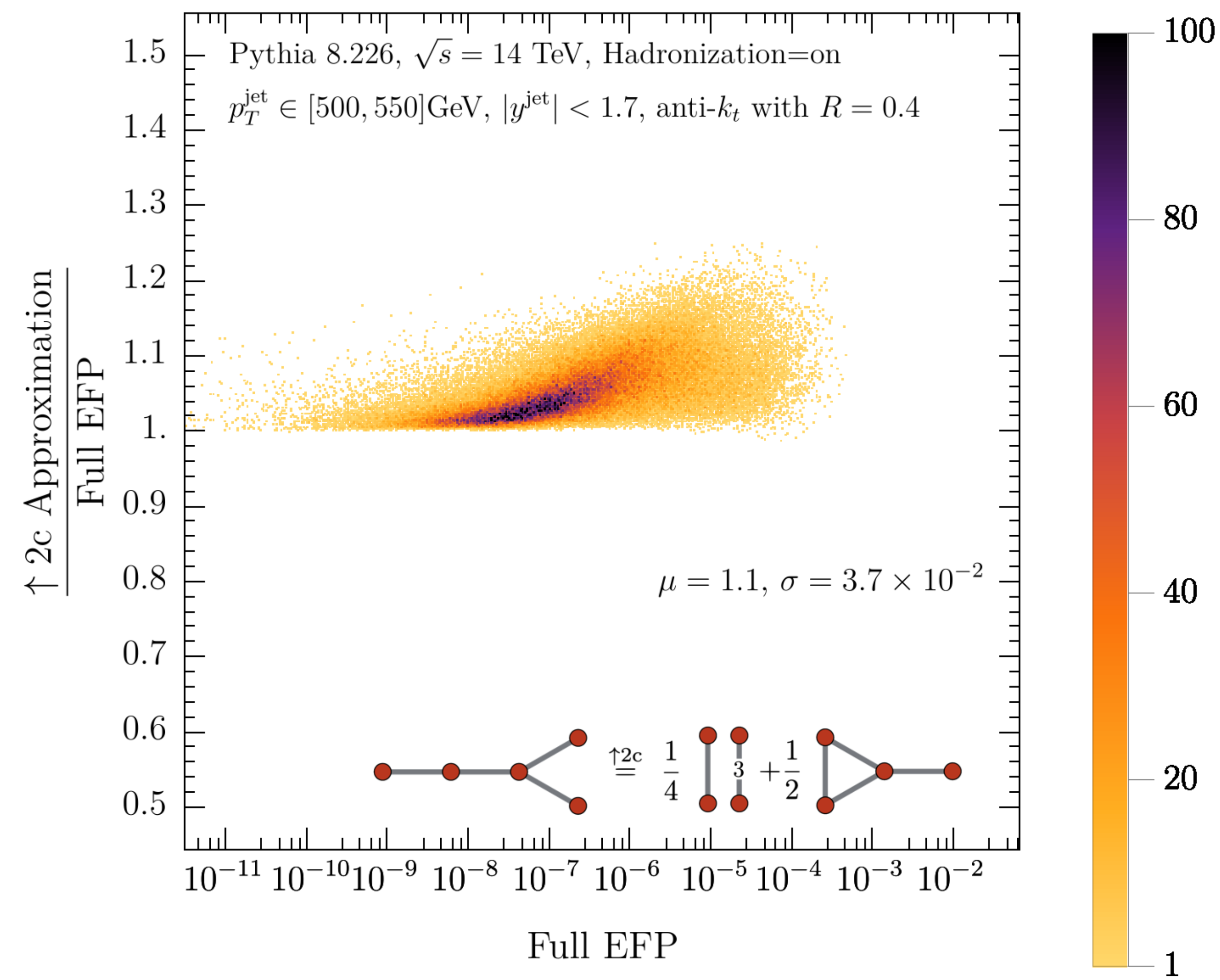


Correlation



$$\mu = 1.1, \quad \sigma = 0.037$$

Correlation ratio

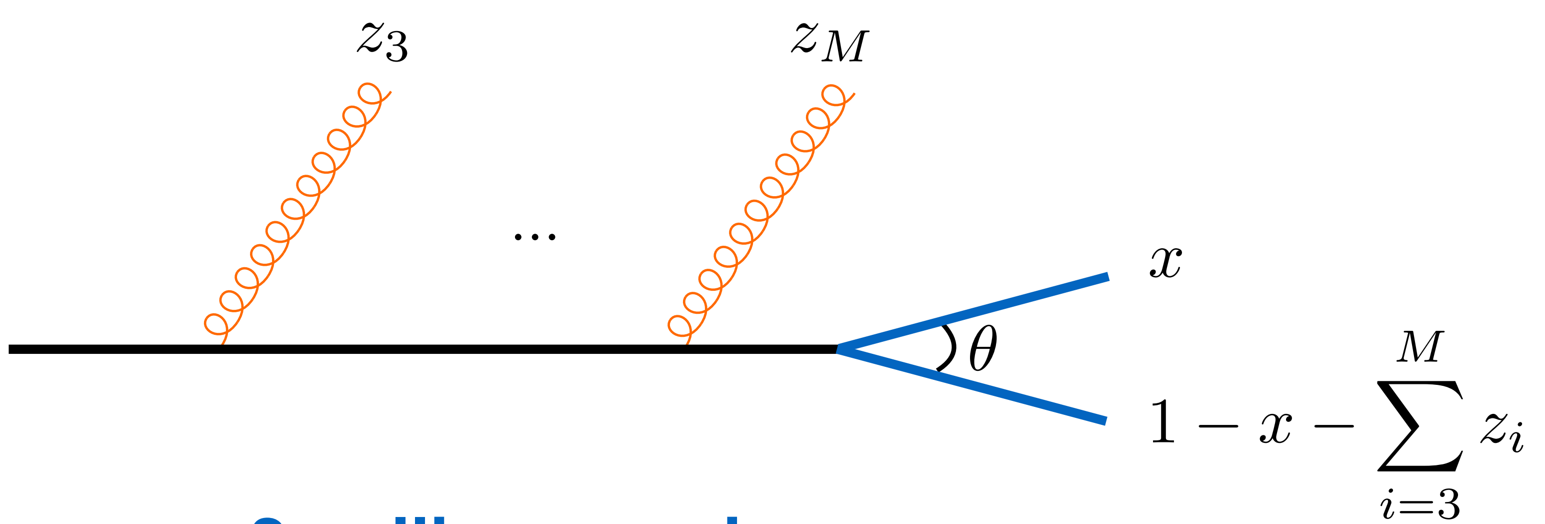


Cal, Thaler, Waalewijn '22 (to appear)

2-collinear

- An even better expansion: **2-collinear expansion**

- **2-collinear approximation**: 2 collinear particles, the remaining are collinear-soft

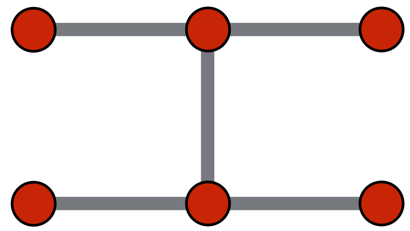


2-collinear regime

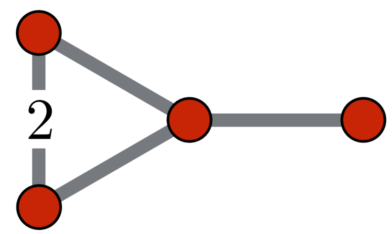
- Note: the general solution must combine the individual 1- and 2-collinear solutions (see how later). We call this the **up-to-2-collinear solution** ($\uparrow 2\mathrm{c}$)

2-collinear

- 1-collinear solution:

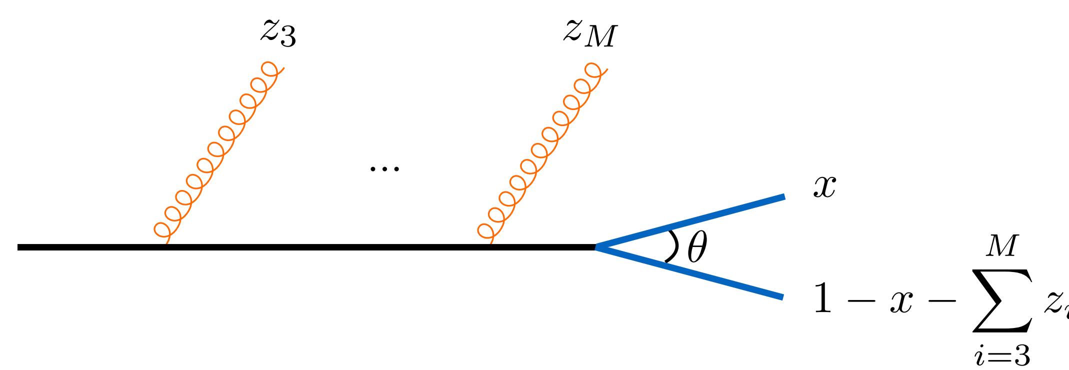


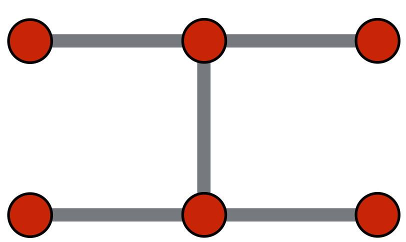
$\overset{1c}{=}$ $\frac{1}{2}$



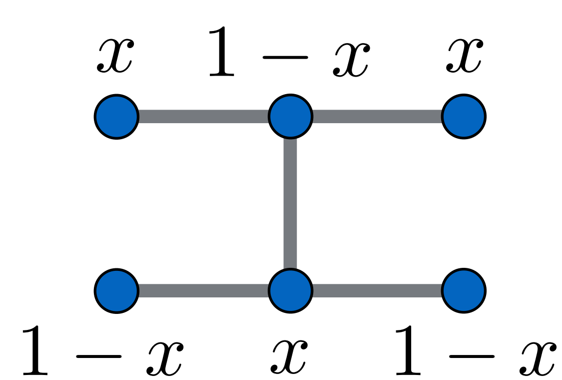
$+ \mathcal{O}(z^3)$

- In the 2-collinear regime:

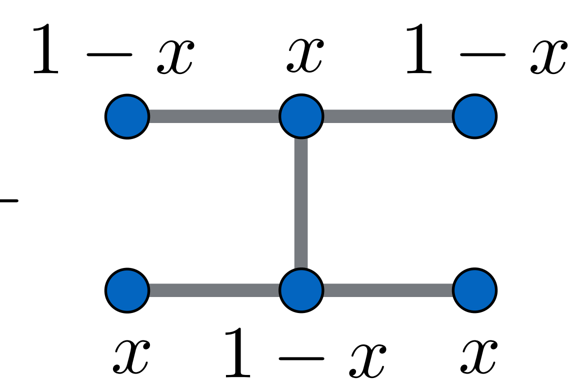




$\overset{2c}{=}$



$+$

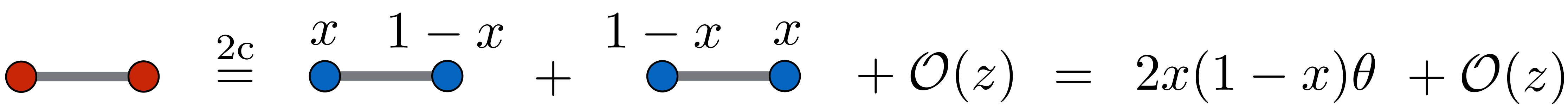


$+ \mathcal{O}(z)$

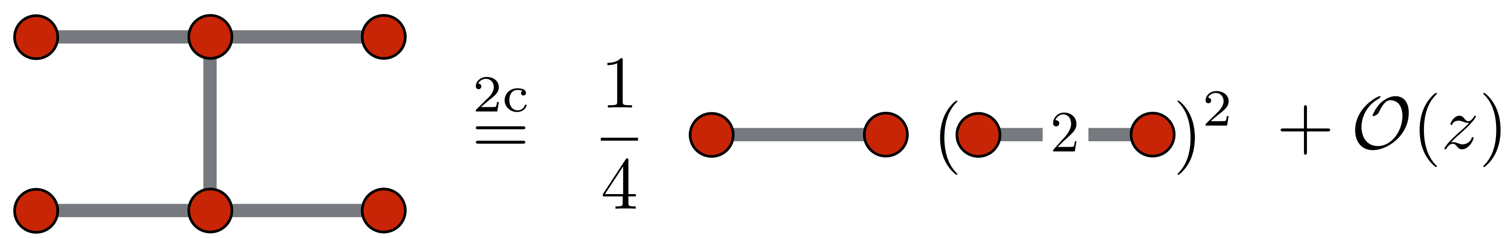
$$\overset{2c}{=} \quad 2x^3(1-x)^3\theta^5 + \mathcal{O}(z)$$

2-collinear

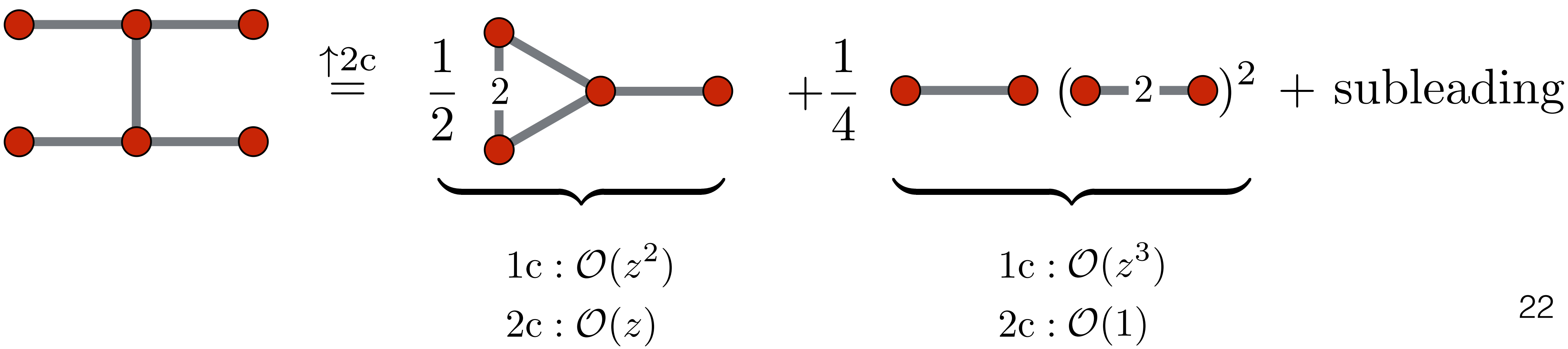
- The dumbbell EFP expansion in the 2-collinear approximation



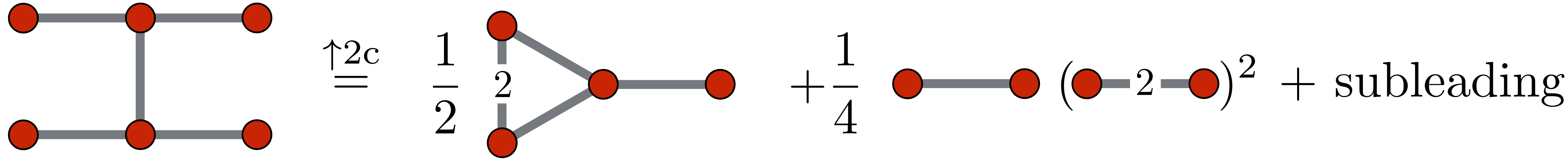
allows us to write:



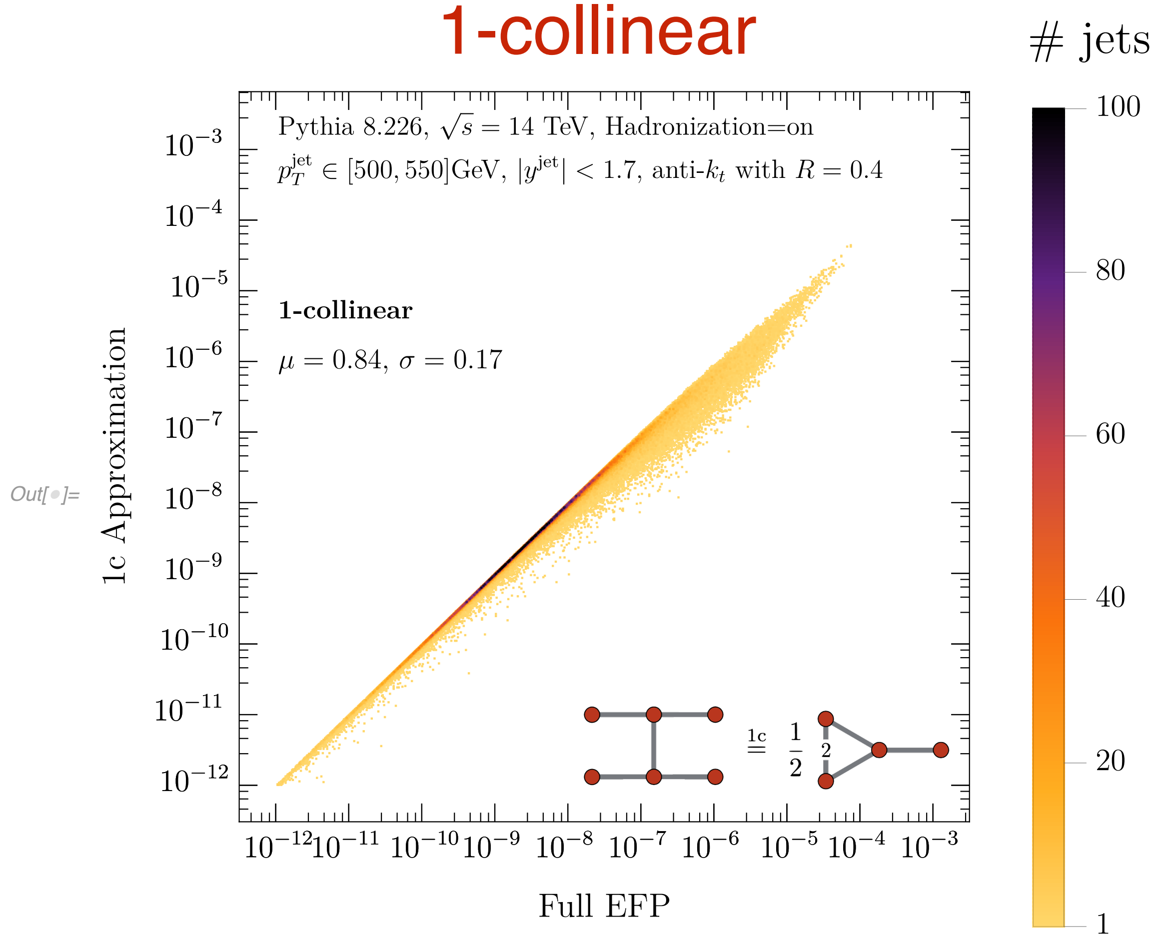
- We obtain the **up-to-2-collinear solution** ($\uparrow 2\mathrm{c}$) by taking the intersection (<u>including degeneracies</u>) of the 1- and 2-collinear solutions



2-collinear

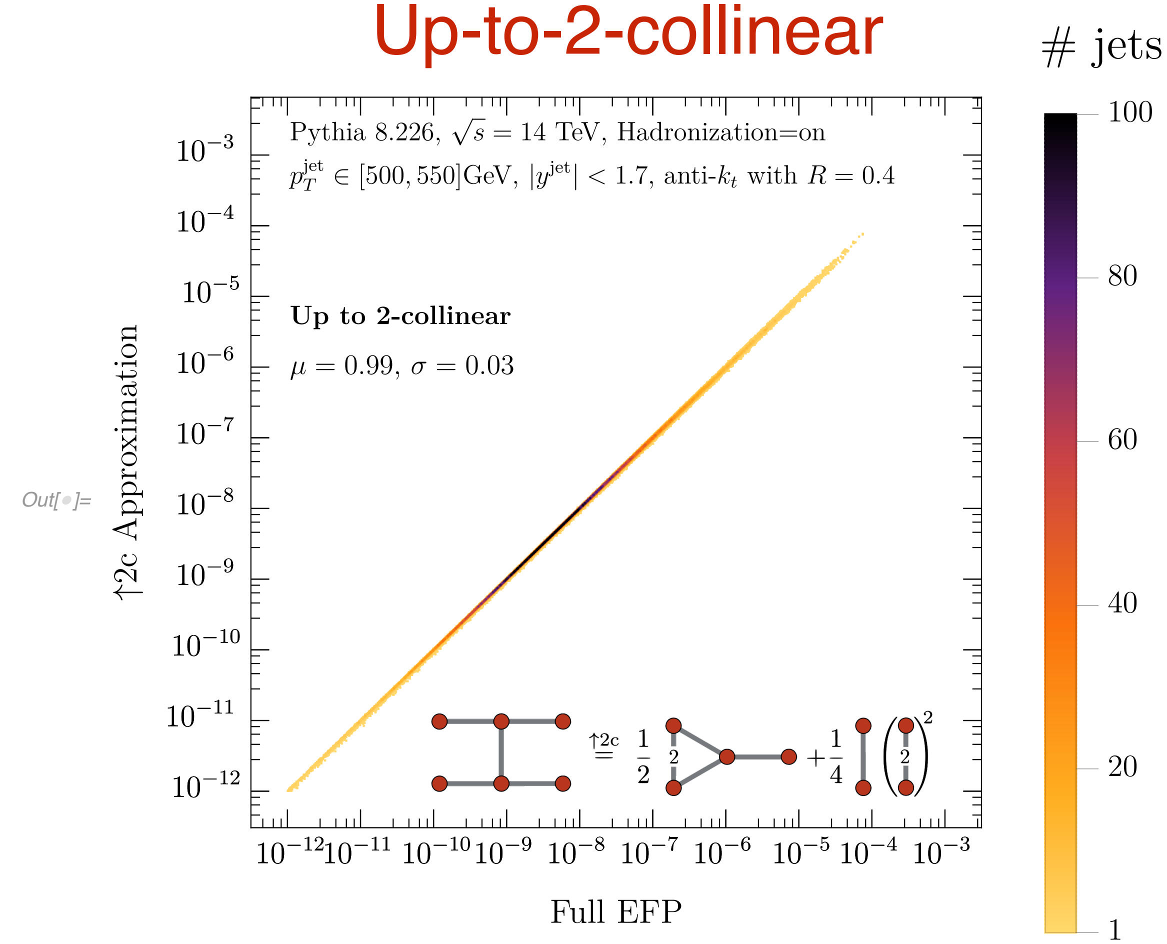


1-collinear



$\mu = 0.84, \sigma = 0.17$

Up-to-2-collinear

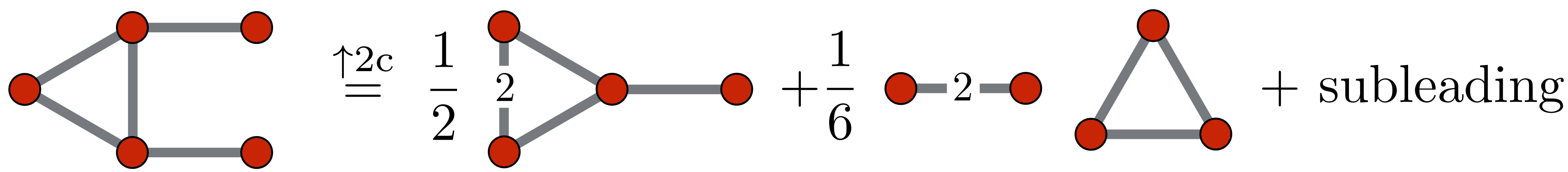


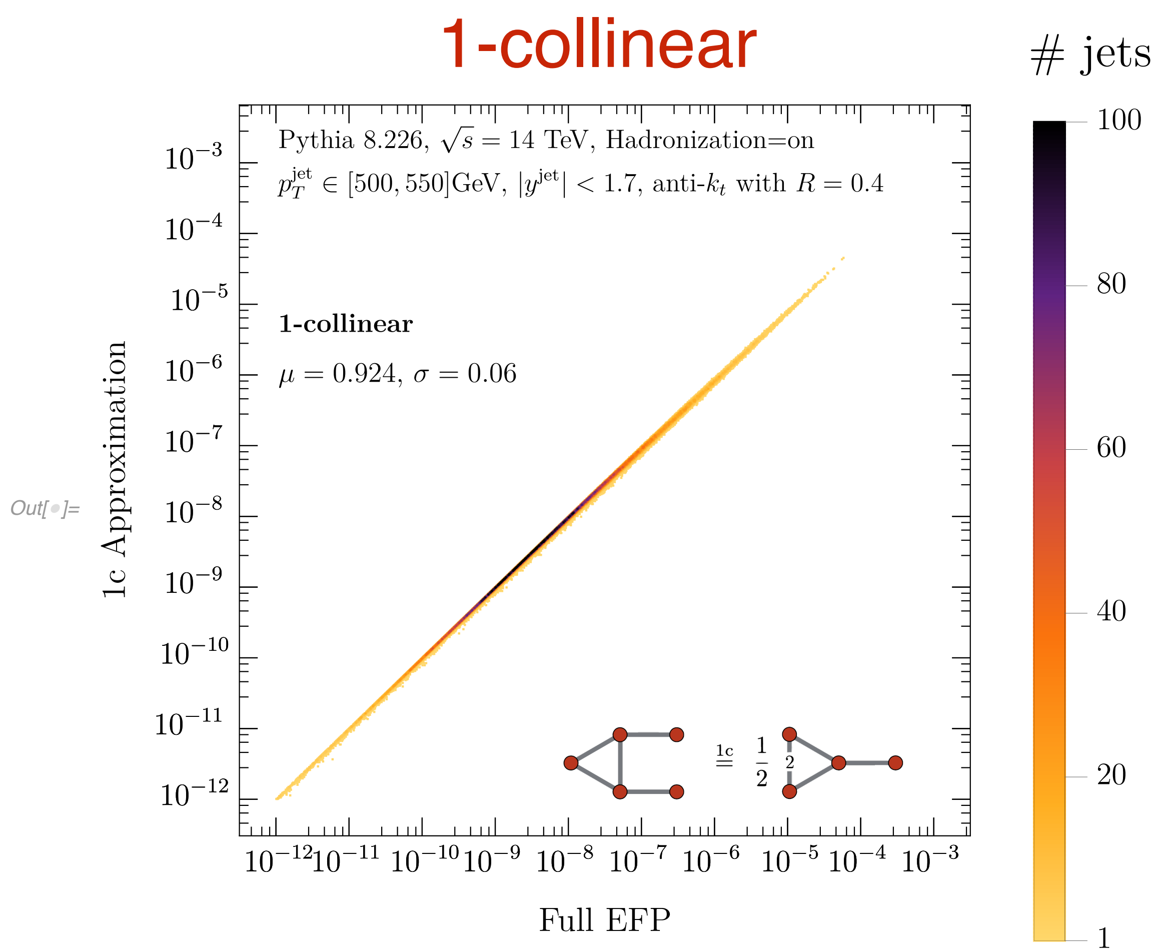
$\mu = 0.99, \sigma = 0.03$

Cal, Thaler, Waalewijn ’22 (to appear)

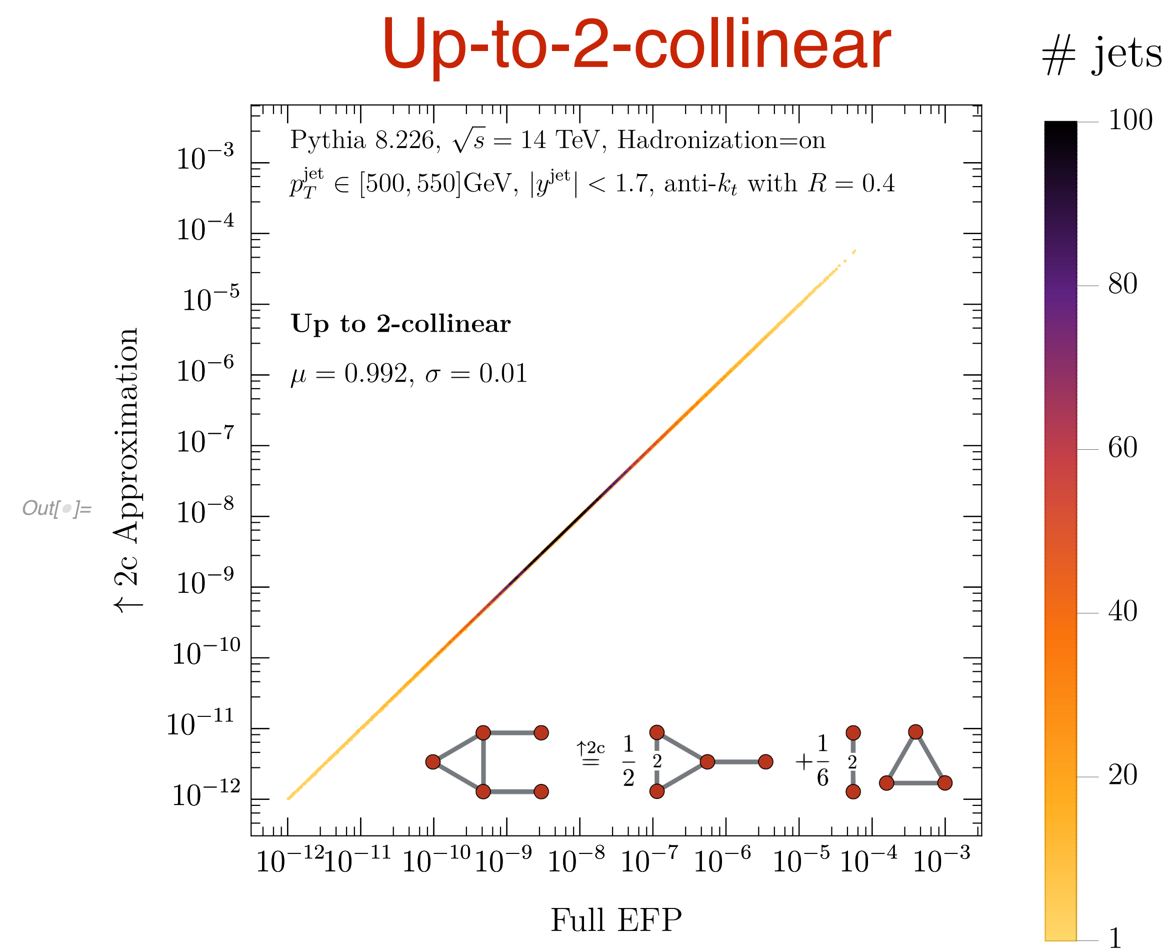
2-collinear

- What about a 3-color EFP? Same reasoning





$\mu = 0.92, \sigma = 0.06$



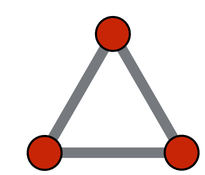
$\mu = 0.99, \sigma = 0.01$

Cal, Thaler, Waalewijn ’22 (to appear)

The collinear basis

- We can use this relations to create a **basis** of linearly independent EFPs at a given accuracy

- We do so by looking at all possible terms that can appear at leading power

	Degree 1	Degree 2	Degree 3	
Expression	$\sum_{i=2}^{M} z_i\theta_i$	$\sum_{i=2}^{M} z_i\theta_i^2$	$\sum_{i=2}^{M} z_i\theta_i^3$	$\sum_{i,j=2}^{M} z_i z_j \theta_i \theta_j \theta_{ij}$
EFP term	$\frac{1}{2}$ 	$\frac{1}{2}$ 	$\frac{1}{2}$ 	$\frac{1}{3}$ 

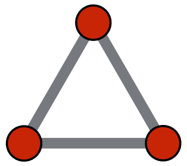
- What about terms like $\sum_{i,j=2}^{M} z_i z_j \theta_{ij}$? → <u>There is no EFP for which it is leading power</u>

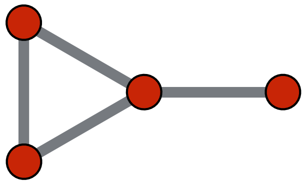
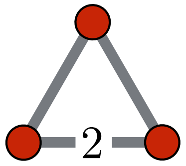
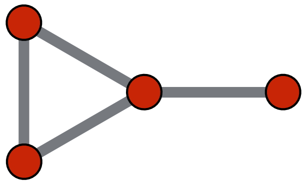
 $\stackrel{1c}{=} 2\sum_{i=2}^{M} z_i\theta_i + \sum_{i,j=2}^{M} z_i z_j \theta_{ij}$

LP **NLP**

The collinear basis

- We can proceed with this exercise to arbitrarily high degree

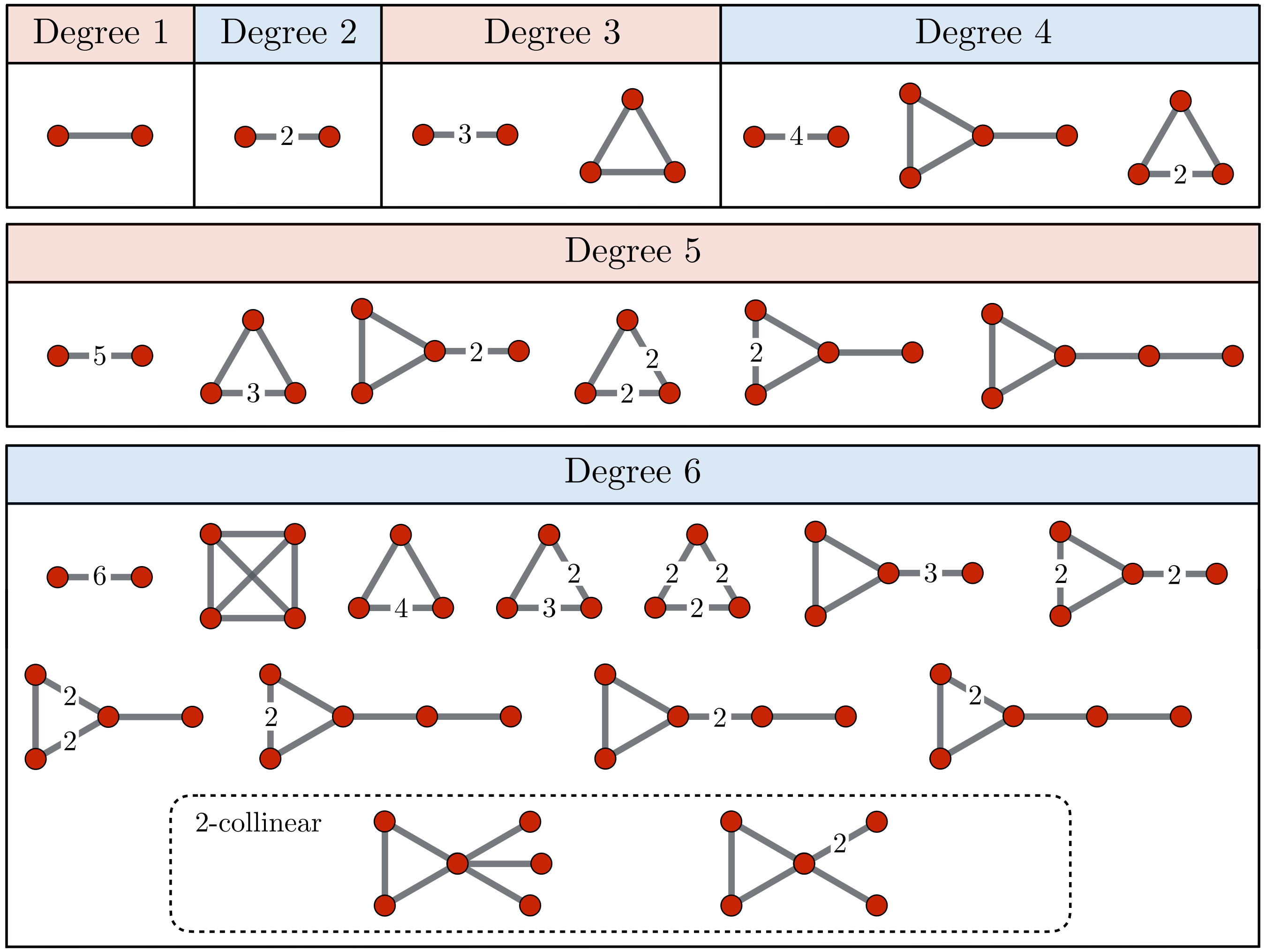
	Degree 1	Degree 2	Degree 3	
Expression	$\sum_{i=2}^{M} z_i \theta_i$	$\sum_{i=2}^{M} z_i \theta_i^2$	$\sum_{i=2}^{M} z_i \theta_i^3$	$\sum_{i,j=2}^{M} z_i z_j \theta_i \theta_j \theta_{ij}$
EFP term	$\frac{1}{2}$ 	$\frac{1}{2}$ 	$\frac{1}{2}$ 	$\frac{1}{3}$ 

	Degree 4		
Expression	$\sum_{i=2}^{M} z_i \theta_i^4$	$\sum_{i,j=2}^{M} z_i z_j \theta_i^2 \theta_j \theta_{ij}$	$\sum_{i,j=2}^{M} z_i z_j \theta_i \theta_j \theta_{ij}^2$
EFP term	$\frac{1}{2}$ 	$\frac{1}{2}$ 	 $-$ 

...

The collinear basis

- We dub this the **2-collinear basis**

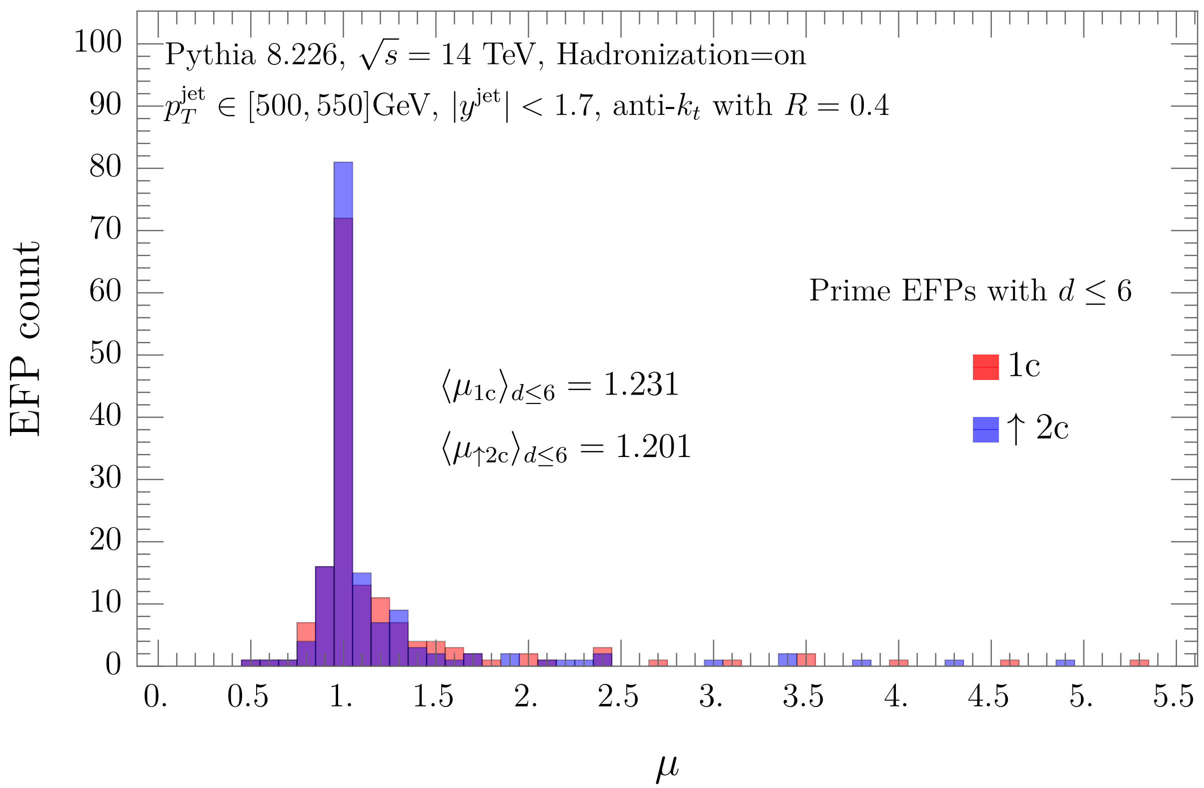


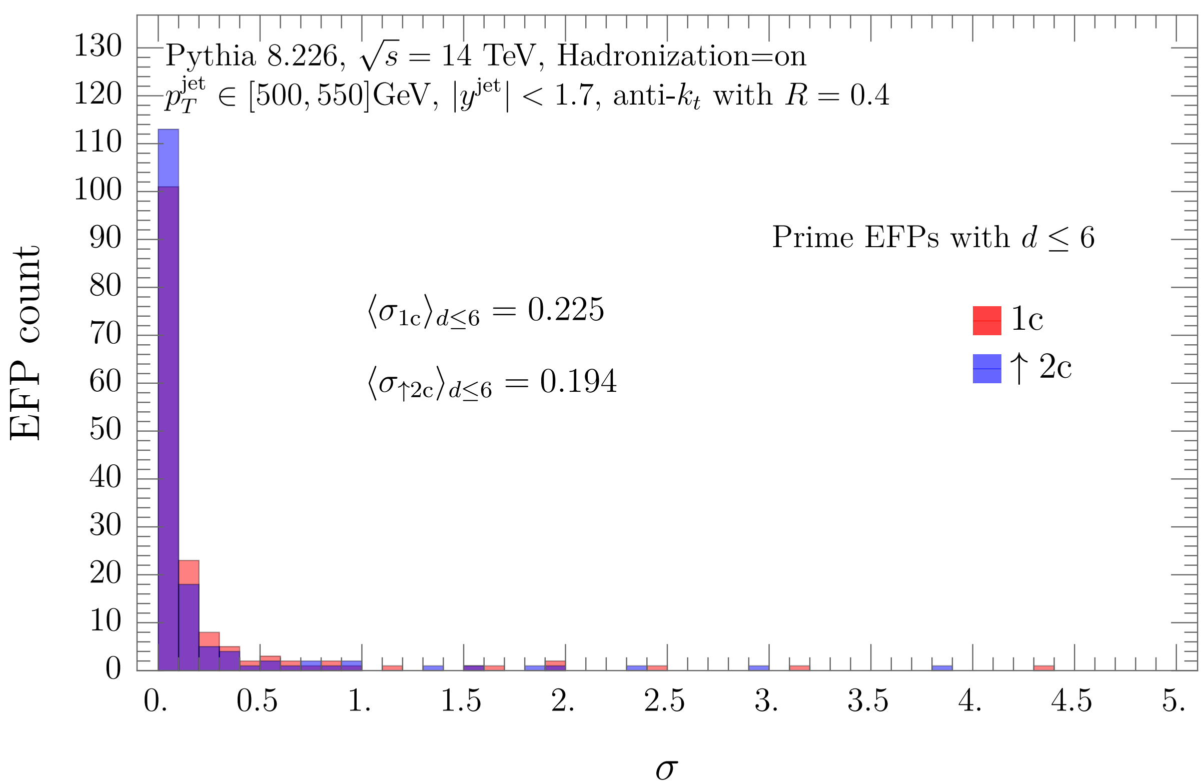
- The 1- and 2-collinear basis are identical apart from two EFPs at degree 6

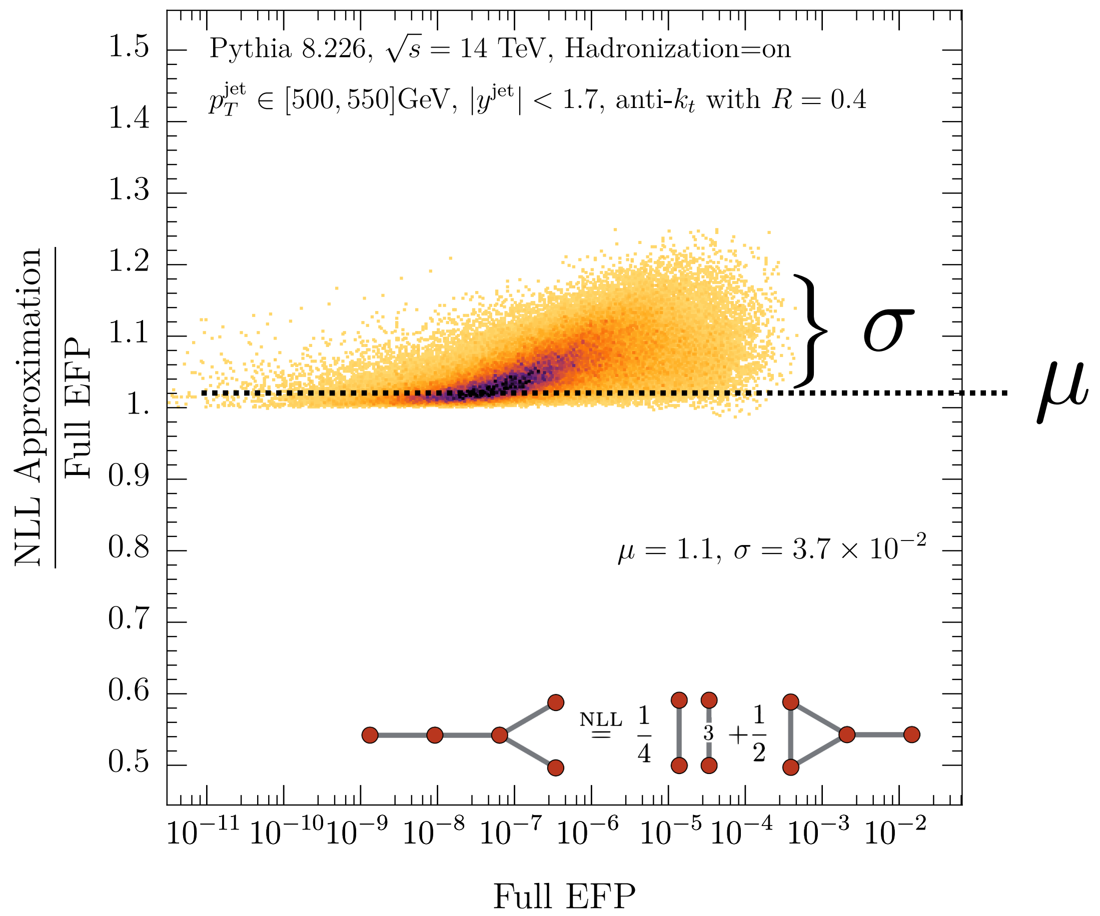
- With these elements we can describe <u>any EFP at up-to-2-collinear accuracy</u>

The collinear basis

- All prime EFPs up to degree 6



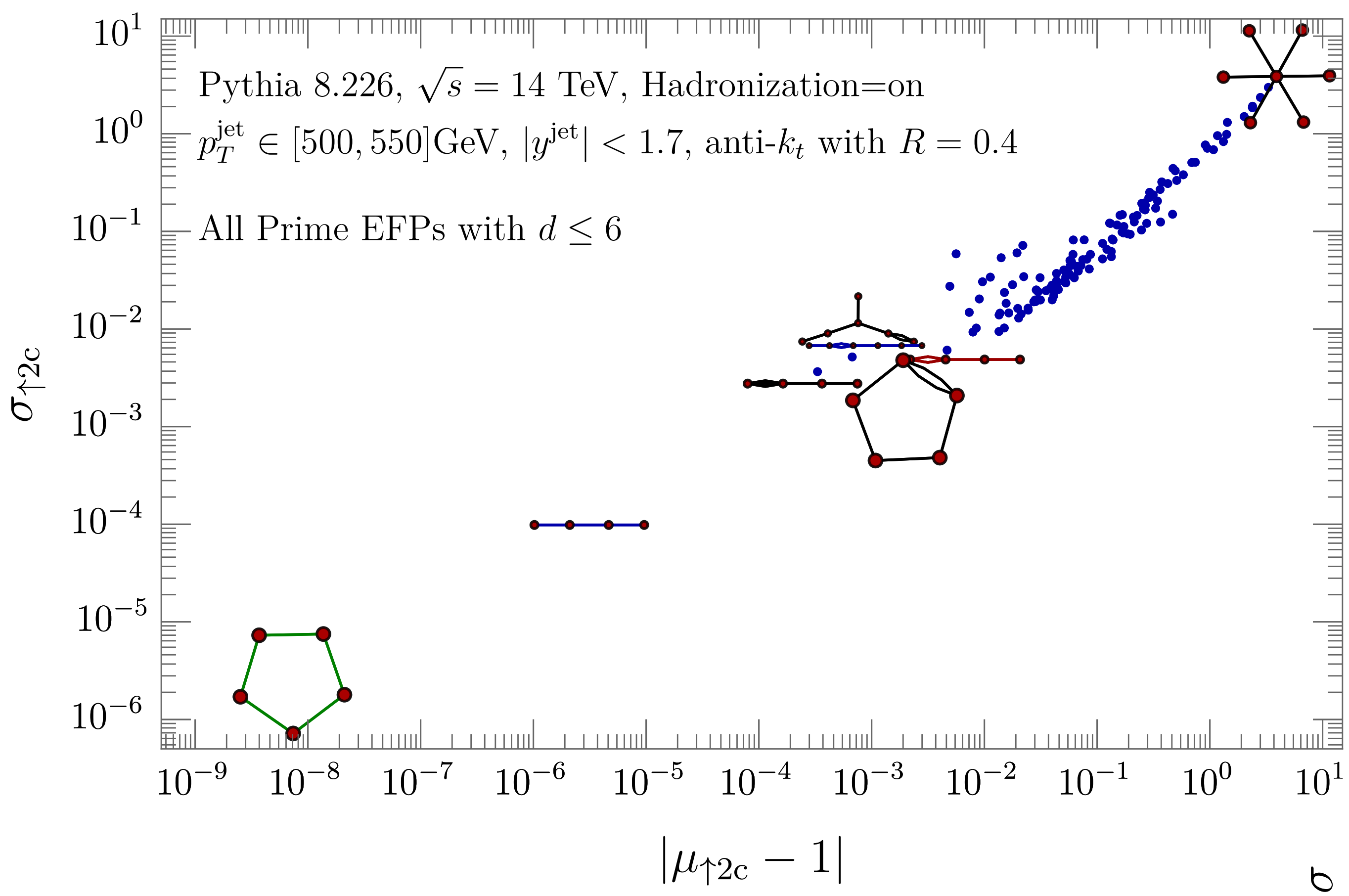




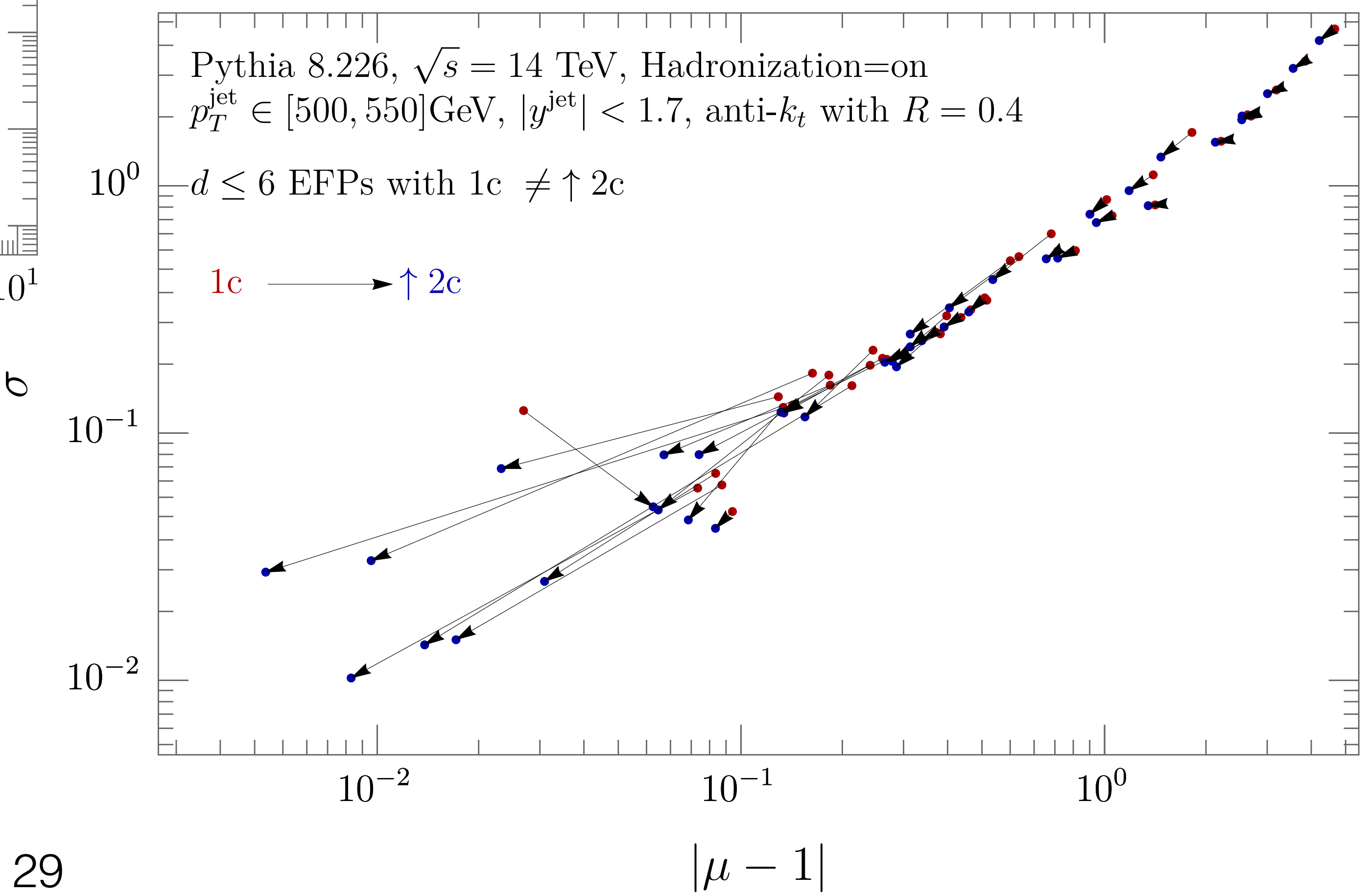
Cal, Thaler, Waalewijn ’22 (to appear)

The collinear basis

- For all EFPs, going from 1-collinear to up-to-2-collinear improves the correlation between the full EFP and its approximation

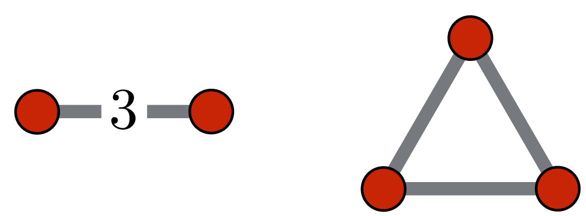
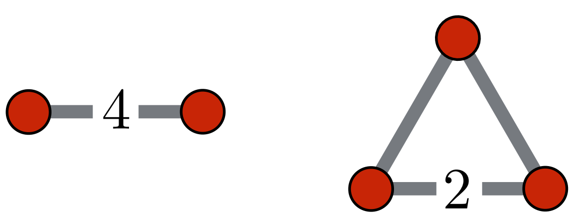


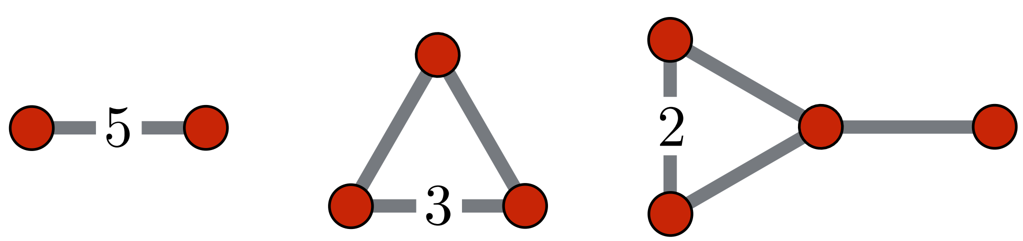
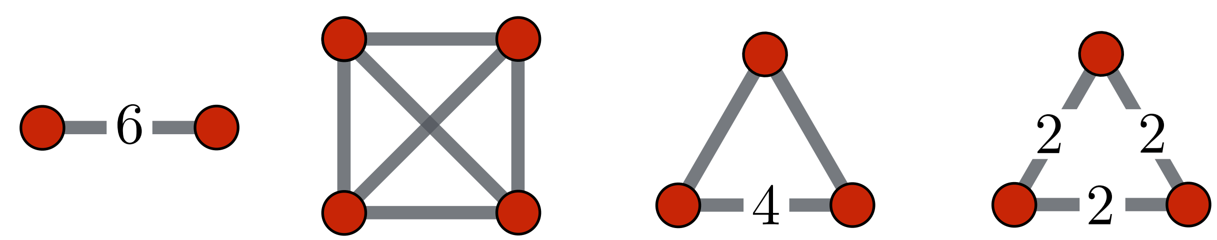
Cal, Thaler, Waalewijn '22 (to appear)



The SO basis

- We can also obtain the **SO basis** by performing the strongly ordered expansion on the 1-collinear basis

Degree 1	Degree 2	Degree 3	Degree 4
			

Degree 5	Degree 6
	

Cal, Thaler, Waalewijn ’22 (to appear)

Bases vs. All EFPs

- The size of the bases is much smaller compared to all possible EFPs

Degree		0	1	2	3	4	5	6
All EFPs	by degree	1	1	3	8	23	66	212
	cumulative	1	2	5	13	36	102	314
SO basis	by degreee	1	1	2	4	7	12	22
	cumulative	1	2	4	8	15	26	49
2-collinear basis	by degree	1	1	2	4	8	16	36
	cumulative	1	2	4	8	16	32	68

- Let's study the differences in quark/gluon tagging performance when using 3 different sets of inputs:

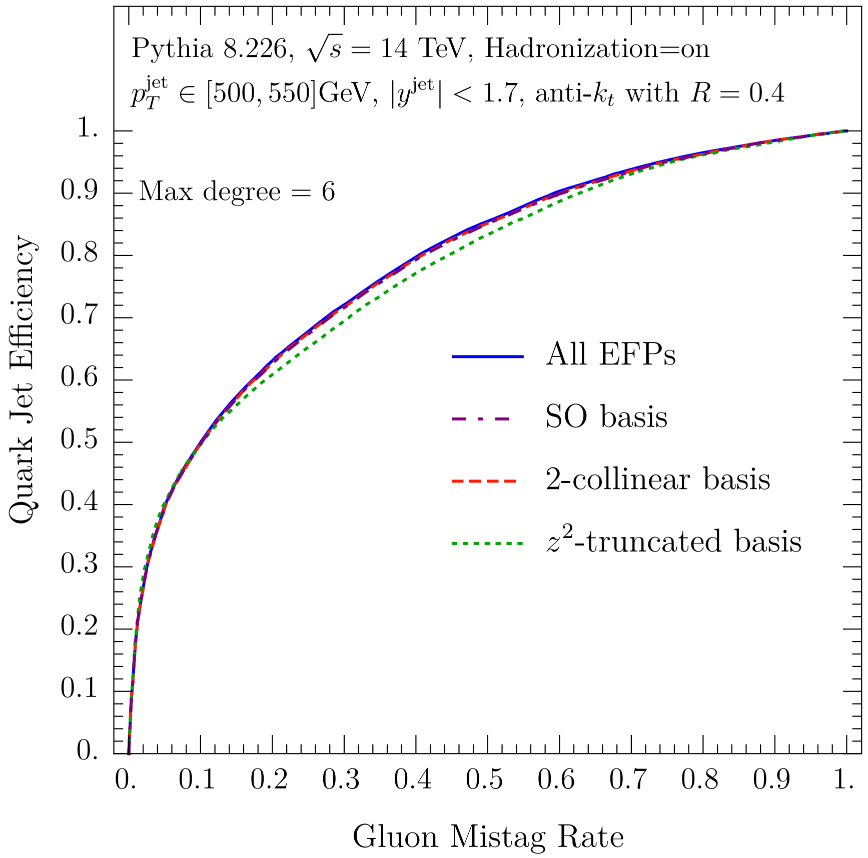
1) All EFPs

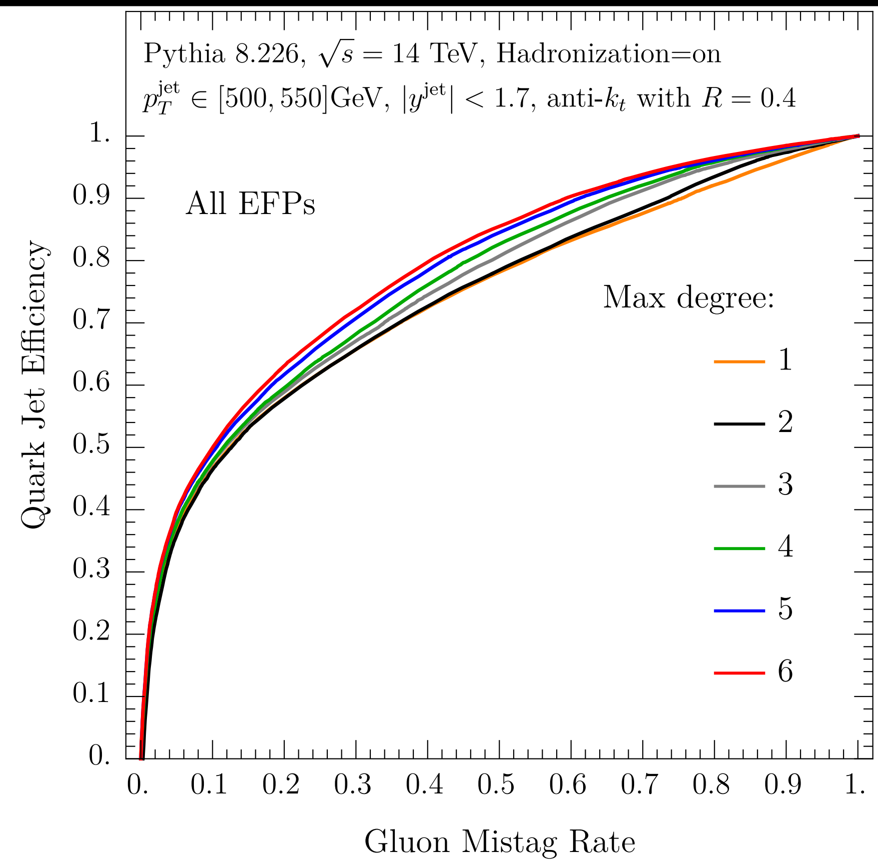
2) The SO basis

3) The 2-collinear basis

Cal, Thaler, Waalewijn '22 (to appear)

Logistic regression





- Area under the curve (AUC) - measure of classification performance

Max degree	All EFPs	SO	2-collinear
1	0.741 ± 0.003	$\mathbf{0.741 \pm 0.003}$	$\mathbf{0.741 \pm 0.003}$
2	0.745 ± 0.003	0.740 ± 0.003	0.741 ± 0.003
3	0.761 ± 0.003	0.755 ± 0.003	0.755 ± 0.003
4	0.770 ± 0.003	0.765 ± 0.003	$\mathbf{0.766 \pm 0.003}$
5	0.784 ± 0.003	0.781 ± 0.003	$\mathbf{0.782 \pm 0.003}$
6	0.792 ± 0.003	$\mathbf{0.789 \pm 0.003}$	$\mathbf{0.789 \pm 0.003}$

Performance nearly identical!

Cal, Thaler, Waalewijn ’22 (to appear)

Conclusions

- We characterize EFPs using power counting, obtaining relations between them that hold at a given accuracy for quark and gluon jets

- These relationships are validated by Pythia, finding excellent agreement

- We can use these relationships to substantially reduce the number of EFPs without affecting tagging performance

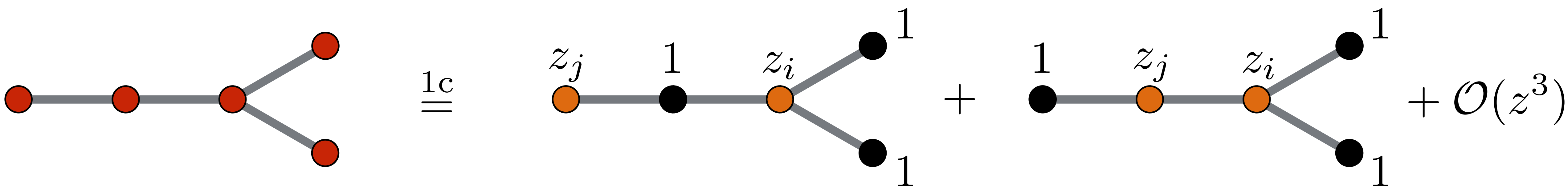
- Future directions:
  - Power counting for 2 and 3 prong jets
  - Analytic resummation of EFPs, looking for patterns in factorization theorems

Dankesfolie

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE)

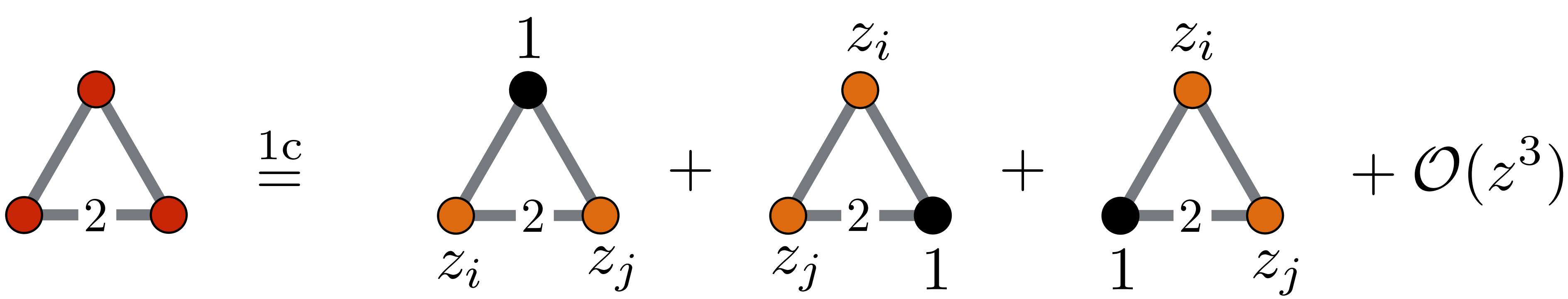
Backup

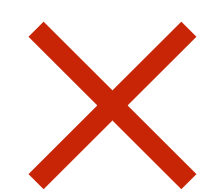
1-collinear



$$\stackrel{1c}{=} \sum_{i,j=2}^{M} z_i z_j (\theta_i^3\theta_j + \theta_i^2\theta_j\theta_{ij}) + \mathcal{O}(z^3)$$

How do we get this term?

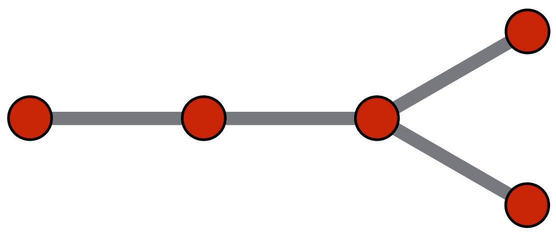




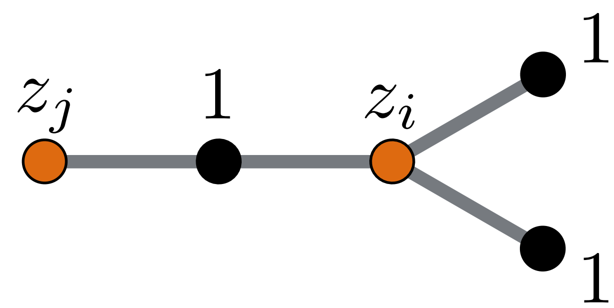
$$\stackrel{1c}{=} \sum_{i,j=2}^{M} z_i z_j (\theta_i\theta_j\theta_{ij}^2 + 2\theta_i^2\theta_j\theta_{ij}) + \mathcal{O}(z^3)$$

Unwanted Wanted

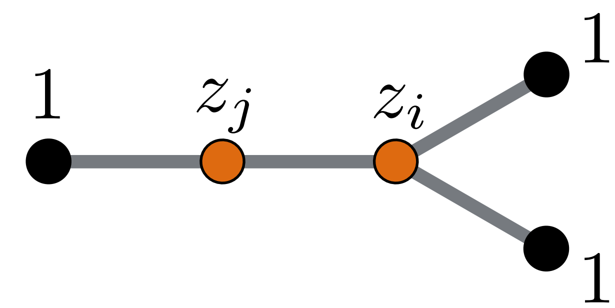
1-collinear



$\overset{1c}{=}$



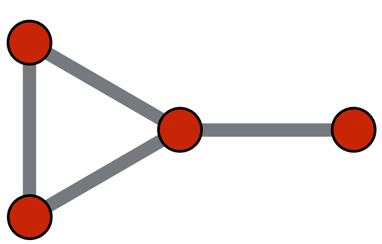
$+$



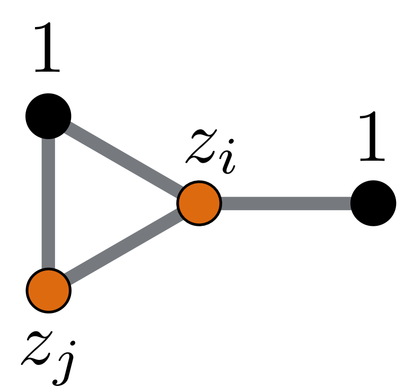
$+ \mathcal{O}(z^3)$

$$\overset{1c}{=} \sum_{i,j=2}^{M} z_i z_j (\theta_i^3 \theta_j + \theta_i^2 \theta_j \theta_{ij}) + \mathcal{O}(z^3)$$

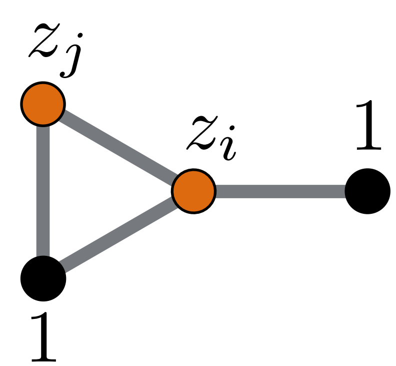
How do we get this term?



$\overset{1c}{=}$



$+$

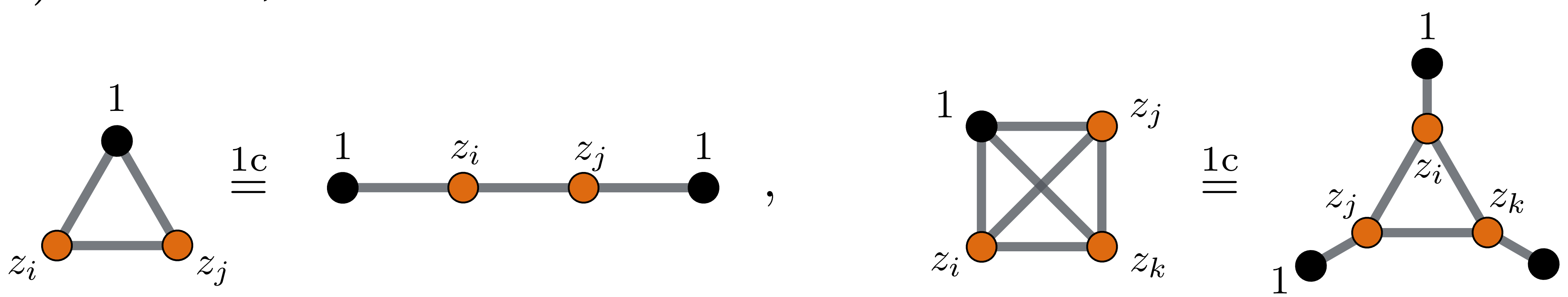


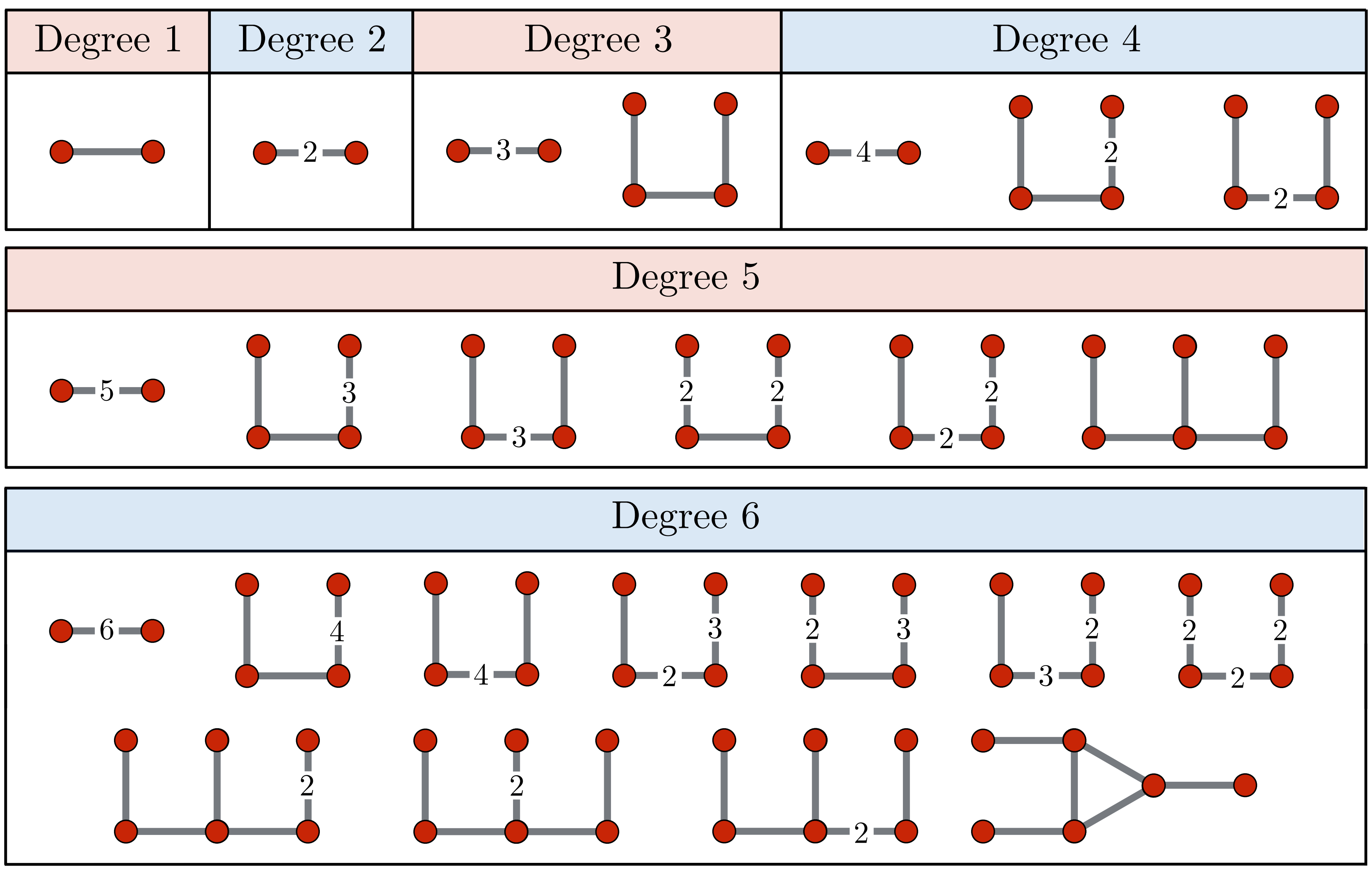
$+ \mathcal{O}(z^3)$

$$\overset{1c}{=} 2 \sum_{i,j=2}^{M} z_i z_j \theta_i^2 \theta_j \theta_{ij} + \mathcal{O}(z^3)$$

Color reduction for 1-collinear

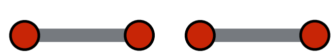
- 1-collinear basis can be "color reduced", i.e. n-color graphs can be traded for $(n-1)$ colored, for $n \geq 3$



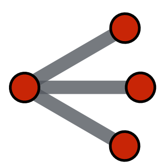
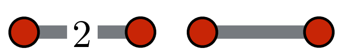


z^2-truncated basis

- Another way to define leading power: **Lowest power of z across all EFPs**

Degree	EFP	$\mathcal{O}(z)$	$\mathcal{O}(z^2)$
1		$2\sum_i z_i\theta_i$	$-2z_{\text{cs}}\sum_i z_i\theta_i + \sum_{i,j} z_i z_j \theta_{ij}$
		$2\sum_i z_i\theta_i^2$	$-2z_{\text{cs}}\sum_i z_i\theta_i^2 + \sum_{i,j} z_i z_j \theta_{ij}^2$
2		$\sum_i z_i\theta_i^2$	$-2z_{\text{cs}}\sum_i z_i\theta_i^2 + \sum_{i,j} z_i z_j (2\theta_i\theta_{ij} + \theta_i\theta_j)$
		0	$4\sum_{i,j} z_i z_j \theta_i\theta_j$

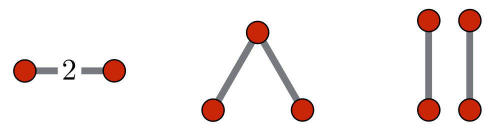
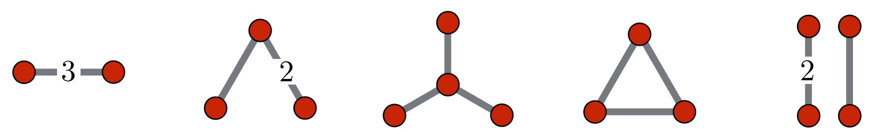
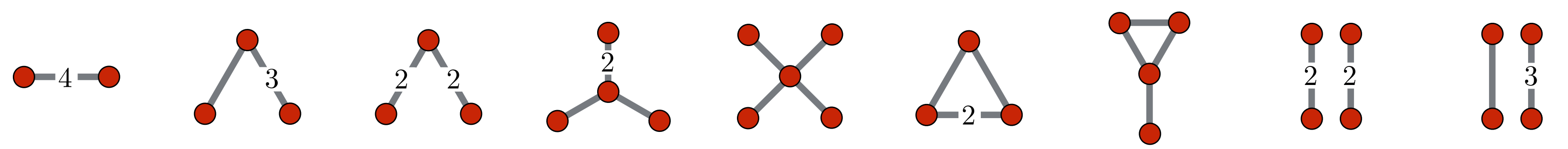
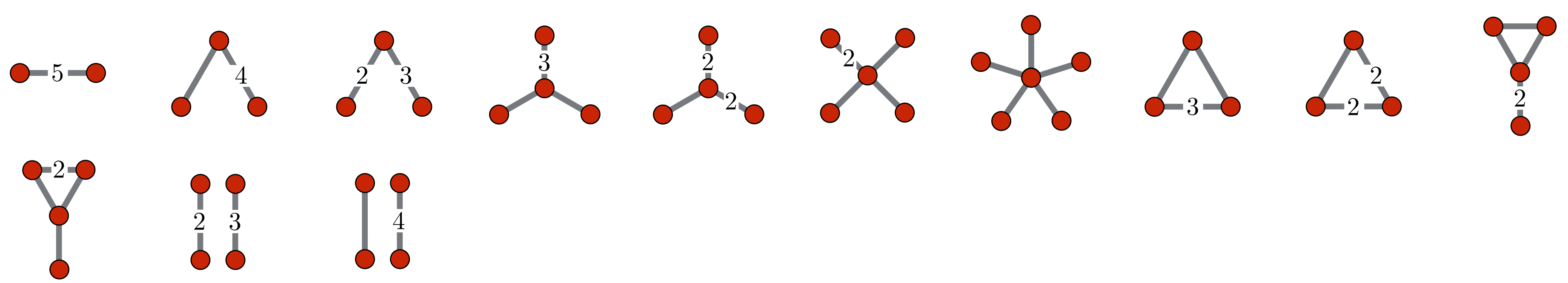
z^2-truncated basis

Degree	EFP	$\mathcal{O}(z)$	$\mathcal{O}(z^2)$
3		$2\sum_i z_i\theta_i^3$	$-2z_{\rm cs}\sum_i z_i\theta_i^3 + \sum_{i,j} z_i z_j \theta_{ij}^3$
		$\sum_i z_i\theta_i^3$	$-2z_{\rm cs}\sum_i z_i\theta_i^3 + \sum_{i,j} z_i z_j(\theta_i^2\theta_{ij} + \theta_i^2\theta_j + \theta_i\theta_{ij}^2)$
		$\sum_i z_i\theta_i^3$	$-3z_{\rm cs}\sum_i z_i\theta_i^3 + 3\sum_{i,j} z_i z_j\theta_i^2\theta_{ij}$
		0	$4\sum_{i,j} z_i z_j\theta_i^2\theta_j$
		0	$\sum_{i,j} z_i z_j(2\theta_i^2\theta_j + \theta_i\theta_j\theta_{ij})$
	…	…	…

- If we keep only $\mathcal{O}(z)$ terms, we end up with only Dumbbells in our basis
- Let's therefore find a basis that can recover all $\mathcal{O}(z^2)$

z^2-truncated basis

- We dub this the **z^2-truncated basis**

Degree	z^2 monomial basis
1	
2	
3	
4	
5	
6