Measuring distributions of quark and gluon jets by jet parameters at a hadron collider

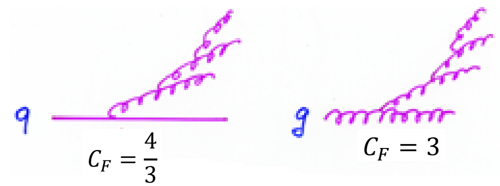
S. Shulga, D. Budkouski

JINR (Dubna, Russia)

XXVIth International Baldin Seminar on High Energy Physics Problems
"Relativistic Nuclear Physics and Quantum Chromodynamics"
September 15-20, 2025
Dubna, Russia

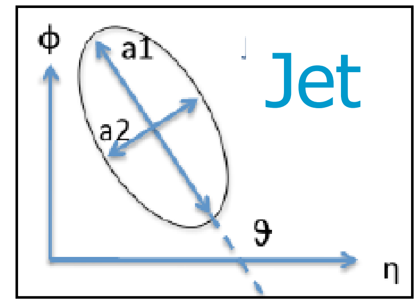
Jet parameters

- A jet is a collection of particles in a narrow cone
- Jet properties depend on kinematics: P_T^{jet} and $\eta^{jet} = -\ln\left(\tan\frac{\Theta^{jet}}{2}\right)$
- q-jets and g-jets have **different properties**
- This difference is determined by different color charge with which quark and gluon are associated



In asymptotic limit $E_{jet} \to \infty$,
$r^{parton} = \frac{\langle n^g \rangle}{\langle n^q \rangle} \to \frac{C_A}{C_F} = \frac{9}{4}$

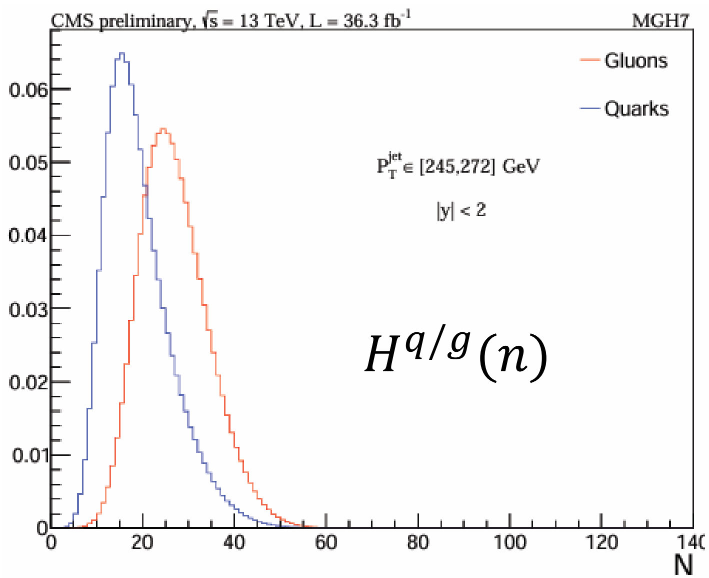
- Tree jet parameters that are sensitive to jet flavour[1]
  - Particle multiplicity inside jet: $mult \quad = V_1$
  - Jet minor axis in (η, φ)-space: $a_2 \to -\log(a_2) \quad = V_2$
  - Fragmentation parameter: $p_T D \equiv \frac{\sqrt{\sum_i p_{T\,i}^2}}{\sum_i p_{T\,i}} \in [0, 1] \quad = V_3$

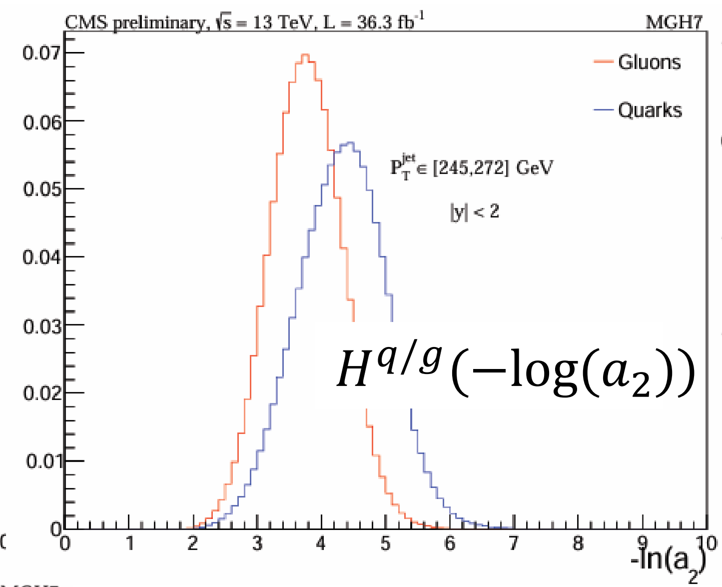


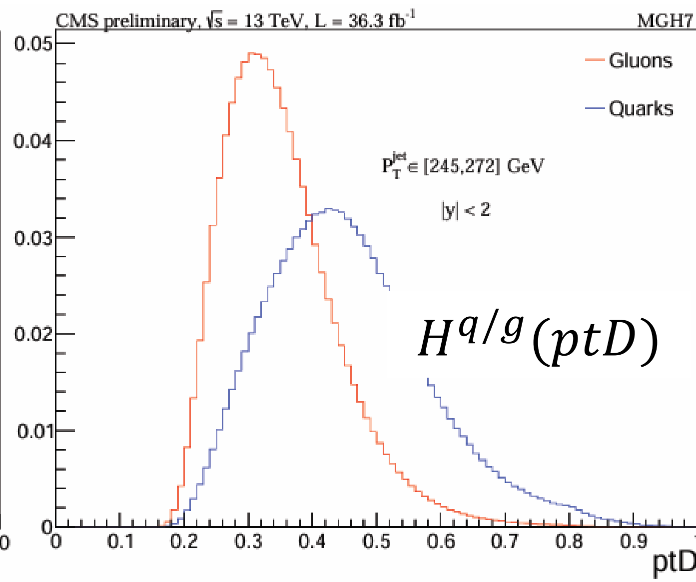
- Combined jet parameter – Quark-Gluon Likelihood discriminator: $QGL \equiv \frac{Q(\vec{V})}{Q(\vec{V})+G(\vec{V})}$

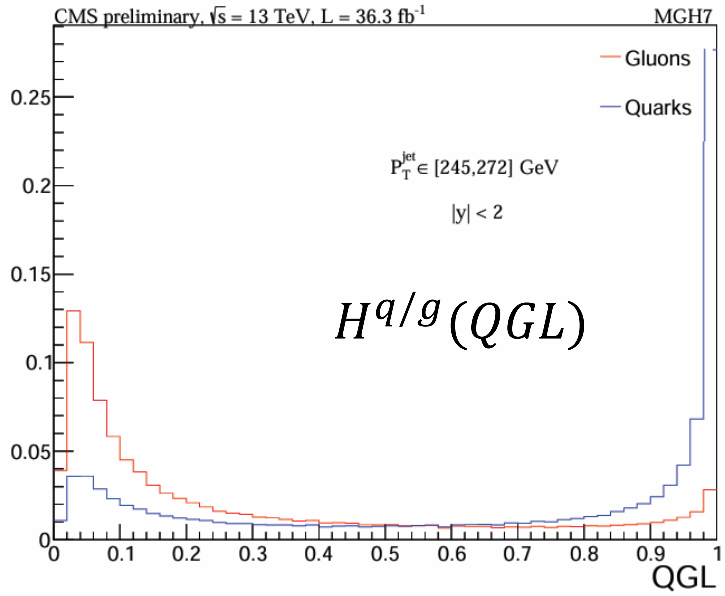
$$Q(\vec{V}) \equiv \prod_{i=1}^{3} H^q(V_i), \qquad G(\vec{V}) \equiv \prod_{i=1}^{3} H^g(V_i) \qquad \vec{V} \equiv (V_1, V_2, V_3)$$

[1] CMS PAS JME-13-002
CMS PAS JME-16-003









- Experimental data shows that q/g-templates in data and MC are different
- How to measure q/g-templates?
- To measure q/g-templates we **need to know q/g-fractions**: α^g and $\alpha^q = 1 - \alpha^g$

$$H(V) = \alpha^g \cdot H^q(V) + (1 - \alpha^g) \cdot H^q(V)$$

- We need two such equations to find $H^g(V)$ and $H^q(V)$

How to measure q/g-templates?

- To find $H^q_{\mathrm{DAT}}(V)$ and $H^g_{\mathrm{DAT}}(V)$ you can use **two samples of jets**:

$$\begin{cases} H_{1,\mathrm{DAT}}(V) = \alpha_1^g \cdot H^g_{\mathrm{DAT}}(V) + \left(1-\alpha_1^g\right) \cdot H^q_{\mathrm{DAT}}(V) \\ H_{2,\mathrm{DAT}}(V) = \alpha_2^g \cdot H^g_{\mathrm{DAT}}(V) + \left(1-\alpha_2^g\right) \cdot H^q_{\mathrm{DAT}}(V) \end{cases}$$

Two samples of jets method

It is assumed here that q/g-templates are **"universal"** =

= independent of jet sample.

$$\Longrightarrow \begin{cases} H^q_{\mathrm{DAT}}(V) = \dfrac{\alpha_2^g \cdot H_{1,\mathrm{DAT}}(V) - \alpha_1^g \cdot H_{2,\mathrm{DAT}}(V)}{\alpha_2^g - \alpha_1^g} \\ H^g_{\mathrm{DAT}}(V) = \dfrac{\left(1-\alpha_1^g\right) \cdot H_{2,\mathrm{DAT}}(V) - \left(1-\alpha_2^g\right) \cdot H_{1,\mathrm{DAT}}(V)}{\alpha_2^g - \alpha_1^g} \end{cases}$$

Data-driven SF for q/g-templates in CMS [4]

Two samples of jets method

[1] CMS PAS JME-13-002
CMS PAS JME-16-003

$$\begin{cases} H^q_{\rm DAT}(V) = \dfrac{\alpha^g_{2,\rm MC} \cdot H_{1,\rm DAT}(V) - \alpha^g_{1,\rm MC} \cdot H_{2,\rm DAT}(V)}{\alpha^g_2 - \alpha^g_1} \\ H^g_{\rm DAT}(V) = \dfrac{(1-\alpha^g_{1,\rm MC}) \cdot H_{2,\rm DAT}(V) - (1-\alpha^g_{2,\rm MC}) \cdot H_{1,\rm DAT}(V)}{\alpha^g_2 - \alpha^g_1} \end{cases} \quad (1)$$

$\alpha^g_{1,2,\rm MC}$ - **MC g-fractions**

$$S_f(V) \equiv \frac{H^f_{\rm DAT}(V)}{H^f_{\rm MC}(V)} \quad (2)$$

- What do we get if we use measured q/g-templates Eqs.(1) to measure $\alpha^g_{\rm DAT}$:

$$H_{\rm DAT}(V) \sim \alpha^g_{\rm DAT} \cdot H^g_{\rm DAT}(V) + (1-\alpha^g_{\rm DAT}) \cdot H^q_{\rm DAT}(V)\ ?$$

- The answer is obvious from Eqs.(1): $\alpha^g_{k,\rm DAT} = \alpha^g_{k,\rm MC}$
- But, measurements show a strong inequality between data and MC: $\alpha^g_{k,\rm DAT} \neq \alpha^g_{k,\rm MC}$
- This means that q/g-templates (1) and **SF** (2) are **incorrect**!
- To extract q/g-templates from data we proposed to use $\alpha^g_{\rm k,DAT}$ in Eqs.(1) [4]

[4] S.S., D.Budkouski Phys. Part. Nucl. Lett. 18(2) 2021

- It is possible to measure g-fraction α^g:

  you just need to set model q/g-**templates** $H^{q,g}_{\mathrm{MC}}(V)$

- Fit equation to find g-fraction[2]:

$$H_{\mathrm{DAT}}(V) \sim \alpha^g_{\mathrm{DAT}} \cdot H^g_{\mathrm{MC}}(V) + (1 - \alpha^g_{\mathrm{DAT}}) \cdot H^q_{\mathrm{MC}}(V)$$

- g-fraction in data jet sample is fundamentally model-dependent

- Model-dependence of measured g-fraction is not reducible within given model

- g-fraction can be measured in one V-bin:

$$\alpha^g_{\mathrm{DAT}}(V) = \frac{H_{\mathrm{DAT}}(V) - H^q_{\mathrm{MC}}(V)}{H^g_{\mathrm{MC}}(V) - H^q_{\mathrm{MC}}(V)}$$

- Measured g-fraction for jet parameter V is average of g-fractions in V-bins:

$$\alpha^g_{\mathrm{DAT}} = \langle \alpha^g_{\mathrm{DAT}}(V) \rangle$$

- Dispersion of $\alpha^g_{\mathrm{DAT}}(V)$ determines "model uncertainty" of measured g-fraction[3]

[2] S.S., D.Budkouski Phys. Phys. Part. Nucl. 55(1) 2024
[3] S.S. Phys. At. Nucl. 87(4) 2024

Correct data-driven SF [4,5]

Two samples of jets method

[5] S.S., e.a. CMS Note 2018/31, 2018
[4] S.S., D.Budkouski Phys. Part. Nucl. Lett. 18(2) 2021

$$\begin{cases} H^{q}_{\mathrm{DAT}} = \dfrac{\alpha^{g}_{2,\mathrm{DAT}} \cdot H_{1,\mathrm{DAT}} - \alpha^{g}_{1,\mathrm{DAT}} \cdot H_{2,\mathrm{DAT}}}{\alpha^{g}_{2,\mathrm{DAT}} - \alpha^{g}_{1,\mathrm{DAT}}} \\ H^{g}_{\mathrm{DAT}} = \dfrac{\left(1-\alpha^{g}_{1,\mathrm{DAT}}\right) \cdot H_{2,\mathrm{DAT}} - \left(1-\alpha^{g}_{2,\mathrm{DAT}}\right) \cdot H_{1,\mathrm{DAT}}}{\alpha^{g}_{2,\mathrm{DAT}} - \alpha^{g}_{1,\mathrm{DAT}}} \end{cases} \quad (1)$$

$\alpha^{g}_{1,2,\mathrm{DAT}}$ - **measured g-fractions**

$$S_f(V) \equiv \frac{H^{f}_{\mathrm{DAT}}(V)}{H^{f}_{\mathrm{MC}}(V)} \quad (2)$$

- What do we get if we use measured q/g-templates Eqs.(1) to measure $\alpha^{g}_{\mathrm{DAT}}$:

$$H_{\mathrm{DAT}}(V) \sim \alpha^{g}_{\mathrm{DAT}} \cdot H^{g}_{\mathrm{DAT}}(V) + (1-\alpha^{g}_{\mathrm{DAT}}) \cdot H^{q}_{\mathrm{DAT}}(V) \; ?$$

- The answer is obvious from Eqs.(1): we get right and the same initial $\alpha^{g}_{k,\mathrm{DAT}}$! So, 2nd iteration to get $\alpha^{g}_{k,\mathrm{DAT}}$ is impossible ⇒ model dependence of measured $\alpha^{g}_{k,\mathrm{DAT}}$ is irreducible

- I remember, that it was assumed that **two jets samples are kinematically equivalent** and have **the same jet environment** in event. So, disadvantage of SF defined by Eqs.(1)-(2) - **small statistics** in the second jet sample

- This disadvantage can be eliminated since the kinematic differences between jet samples can be taken from the simulation ↘

Improved two samples of jets method[6]

- Source Eqs. with "**non-universal**" q/g-templates:

$$(1)\quad \begin{cases} O_1 = \alpha_1^g\, O_1^g + (1-\alpha_1^g) O_1^q \\ O_2 = \alpha_2^g\, O_2^g + (1-\alpha_2^g) O_2^q \\ O_2^q - O_1^q = \Delta O_{\mathrm{MC}}^q \\ O_2^g - O_1^g = \Delta O_{\mathrm{MC}}^g \end{cases}$$

Two "measures of **JFKNU**"

JFKNU – **J**et **F**lavour **K**inematical **N**on-**U**niversality

- It is useful to introduce averaged ("**universal**") q/g-templates:

$$(2)\quad \begin{cases} O^f = A_1^f \cdot O_1^f + A_2^f \cdot O_2^f \\ \Delta O_{\mathrm{MC}}^f = O_2^f - O_1^f \end{cases}$$

$$A_k^f \equiv \frac{N_k \cdot \alpha_k^f}{N_1 \cdot \alpha_1^f + N_2 \cdot \alpha_2^f}$$

N_k - number of jets in kth jet sample

- Inverse expression of "**non-universal**" O_k^f via "**universal**" O^f:

$$\begin{cases} O_2^f = O^f + \Delta O_{\mathrm{MC}}^f \cdot A_1^f \\ O_1^f = O^f - \Delta O_{\mathrm{MC}}^f \cdot A_2^f \end{cases}$$

$O^{q,g} \equiv H_{\mathrm{DAT}}^{q,g}$ | $O_{1,2} \equiv H_{1,2\ \mathrm{DAT}}$ | [6] S.S. Phys. Part. Nucl. **55**(4) 2024

Eqs.(1) (p.8):

$O^{q,g} \equiv H^{q,g}_{\mathrm{DAT}}$ $\quad$ $O_{1,2} \equiv H_{1,2\ \mathrm{DAT}}$

Improved two samples of jets method [6]

$$\begin{cases} \alpha_1^g\, O^g + \alpha_1^q\, O^q = \tilde{O}_1 \equiv O_1 + \Delta O^g_{\mathrm{MC}} \cdot A_2^g \cdot \alpha_1^g + \Delta O^q_{\mathrm{MC}} \cdot A_2^q \cdot \alpha_1^q \\ \alpha_2^g\, O^g + \alpha_2^q\, O^q = \tilde{O}_2 \equiv O_2 - \Delta O^g_{\mathrm{MC}} \cdot A_1^g \cdot \alpha_2^g - \Delta O^q_{\mathrm{MC}} \cdot A_1^q \cdot \alpha_2^q \end{cases}$$

Two "measures of **JFKNU**"

"**Universal**" q/g-templates

$$\begin{cases} O^q = \dfrac{\alpha_2^g\, \tilde{O}_1 - \alpha_1^g\, \tilde{O}_2}{\alpha_2^g - \alpha_1^g} \\ O^q = \dfrac{\alpha_2^g\, \tilde{O}_1 - \alpha_1^g\, \tilde{O}_2}{\alpha_2^g - \alpha_1^g} \end{cases}$$

$\tilde{O}_1$ and $\tilde{O}_2$ are measured q/g-templates with JFKNU corrections

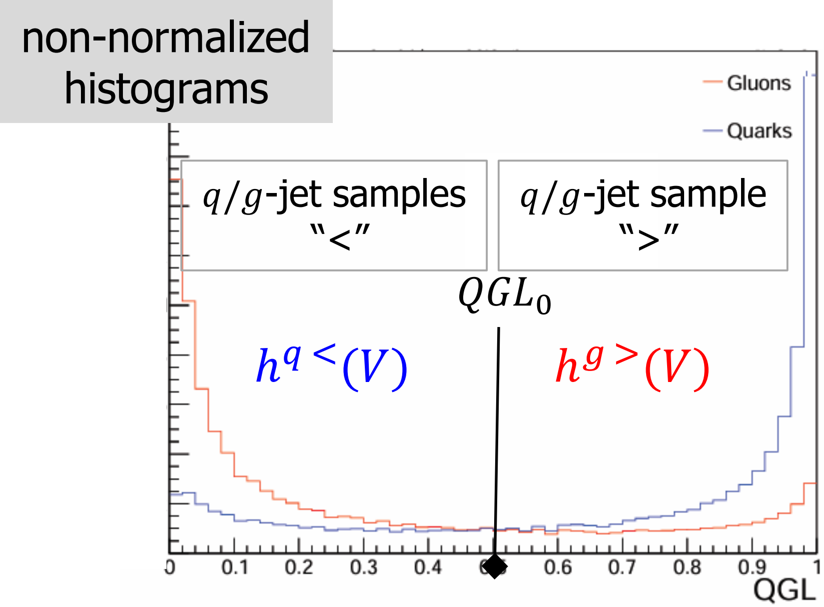
"**Non-universal**" q/g-templates

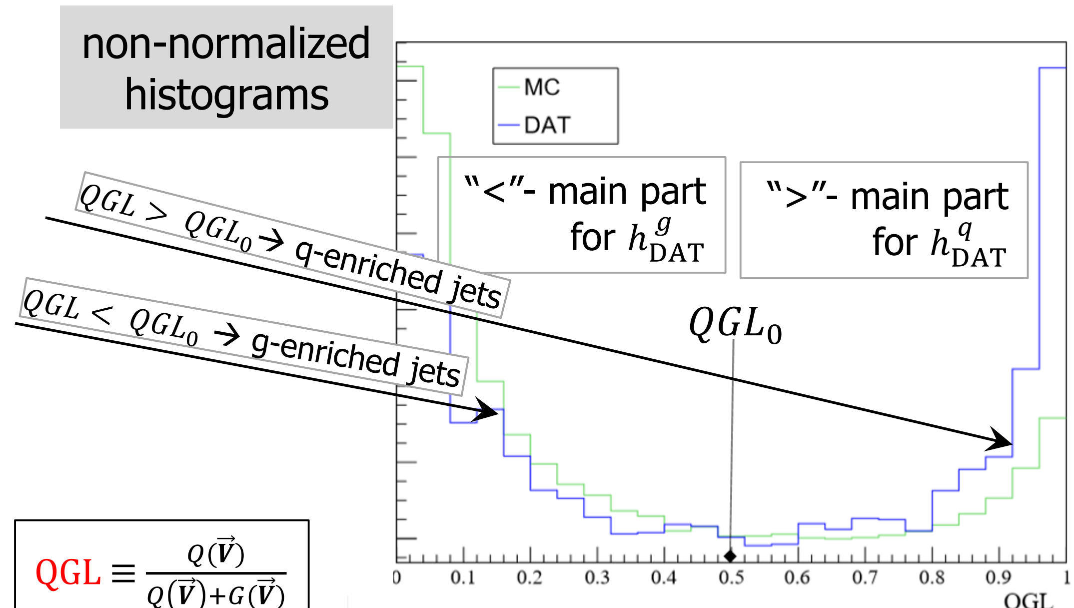
$$\begin{cases} O_2^f = O^f + \Delta O^f_{\mathrm{MC}} \cdot A_1^f \\ O_1^f = O^f - \Delta O^f_{\mathrm{MC}} \cdot A_2^f \end{cases}$$

$$f = q/g$$

- Improved two jet sample method allow us to measure q/g-templates for arbitrary channel
- As a result, we can build QGL discriminator with taking into account JFKNU
- Main uncertainty is uncertainty of the measured $\alpha^g_{1,2}$
- Is it possible to find two q/g-templates **using one jet sample**? Yes, it is!

[6] S.S. Phys. Part. Nucl. **55**(4) 2024





$$\mathrm{QGL} \equiv \frac{Q(\vec{V})}{Q(\vec{V})+G(\vec{V})}$$

V – any jet parameter, not QGL!

New!

Single sample of jets method

- We intend to use QGL to define **main parts** of q/g-templates using w.p. QGL_0:

$$(1) \quad \begin{cases} h^{q}_{\mathrm{DAT}}(V) \approx h^{>}_{\mathrm{DAT}}(V) \\ h^{g}_{\mathrm{DAT}}(V) \approx h^{<}_{\mathrm{DAT}}(V) \end{cases}$$

- How to fix approximate equalities in (1)?

$$\begin{cases} h^{>}_{\mathrm{DAT}}(V) \equiv \text{true } q\text{–jets} - \text{wrong } q\text{–jets} + \text{lost } q\text{–jets} \\ h^{<}_{\mathrm{DAT}}(V) \equiv \text{true } g\text{–jets} - \text{wrong } g\text{–jets} + \text{lost } g\text{–jets} \end{cases}$$

wrong q-jets ≡ lost g-jets ≡ $h^{g\,>}(V)$

lost q-jets ≡ wrong g–jets ≡ $h^{q\,<}(V)$

$$\Rightarrow \begin{cases} h^{q}_{\mathrm{DAT}}(V) \equiv h^{>}_{\mathrm{DAT}}(V) - h^{g\,>}(V) + h^{q\,<}(V) \\ h^{g}_{\mathrm{DAT}}(V) \equiv h^{<}_{\mathrm{DAT}}(V) - h^{q\,<}(V) + h^{g\,>}(V) \end{cases}$$

- We intend to take $h^{q\,<}(V)$ and $h^{g\,>}(V)$ **from the model** !

- We utilize the QGL property: fractions of "wrong q/g-jets" and "lost q/g-jets" much less than fraction of "true q/g-jets"

- Next, we move on to normalized histograms ($h \to H$):

Single sample of jets method

New!

$$\begin{cases} H_{\mathrm{DAT}}^{q} \approx \dfrac{\beta_{\mathrm{DAT}}^{>}}{\alpha_{\mathrm{DAT}}^{q}} \cdot \left[H_{\mathrm{DAT}}^{>} - \alpha_{\mathrm{DAT}}^{g\,>} \cdot H_{\mathrm{MC}}^{g\,>} + \alpha_{\mathrm{DAT}}^{q\,<} \cdot \dfrac{\beta_{\mathrm{DAT}}^{<}}{\beta_{\mathrm{DAT}}^{>}} \cdot H_{\mathrm{MC}}^{q\,<} \right] \\ H_{\mathrm{DAT}}^{g} \approx \dfrac{\beta_{\mathrm{DAT}}^{<}}{\alpha_{\mathrm{DAT}}^{g}} \cdot \left[H_{\mathrm{DAT}}^{<} + \alpha_{\mathrm{DAT}}^{g\,>} \cdot \dfrac{\beta_{\mathrm{DAT}}^{>}}{\beta_{\mathrm{DAT}}^{<}} \cdot H_{\mathrm{MC}}^{g\,>} - \alpha_{\mathrm{DAT}}^{q\,<} \cdot H_{\mathrm{MC}}^{q\,<} \right] \end{cases}$$

- $\alpha_{\mathrm{DAT}}^{g\,>} = f\left(H^{>}, H_{\mathrm{MC}}^{g\,>}, H_{\mathrm{MC}}^{q\,>}\right)$ - measured g-fraction in q-enriched subsample ">"
- $\alpha_{\mathrm{DAT}}^{q\,<} = f\left(H^{<}, H_{\mathrm{MC}}^{g<}, H_{\mathrm{MC}}^{q<}\right)$ - measured q-fraction in g-enriched subsample "<"
- $\alpha_{\mathrm{DAT}}^{g} = f\left(H, H_{\mathrm{MC}}^{g}, H_{\mathrm{MC}}^{q}\right)$ - measured g-fraction in total jet sample
- $\alpha_{\mathrm{DAT}}^{q} = 1 - \alpha_{\mathrm{DAT}}^{g}$
- $\beta_{\mathrm{DAT}}^{>,<} \equiv \dfrac{N_{\mathrm{DAT}}^{>,<}}{N_{\mathrm{DAT}}}$
- $N_{\mathrm{DAT}}^{>} + N_{\mathrm{DAT}}^{<} = N_{\mathrm{DAT}}$

- The main uncertainty of $H_{\mathrm{DAT}}^{q,g}$ is from the model uncertainty of measured q/g-fractions
- The contributions from the model tails are significantly smaller than the main contribution, since

$$\alpha_{\mathrm{DAT}}^{g\,>} \ll \alpha_{\mathrm{DAT}}^{f}$$
$$\alpha_{\mathrm{DAT}}^{q\,<} \ll \alpha_{\mathrm{DAT}}^{f}$$

Conclusion

- A method for measuring q/g-templates using a single sample of jets is proposed
- Single sample of jets method has practical advantages over the method of two samples of jets
- Methods for measuring q/g-templates are being developed to improve the QGL discriminator for CMS

Thank you for your attention!
Have a nice autumn!