Estimation of the QCD background yield in the $1\tau\,0L$ category

Huiling Hua [1] **Fabio Iemmi** [1]
Hongbo Liao [1] Hideki Okawa [2] Yu Zhang [2]

[1]Institute of High Energy Physics (IHEP), Beijing

[2]Fudan University, Shanghai

July 14, 2021

General idea

- **QCD simulations are not reliable**, especially at high jet multiplicity
  - Large uncertainties on the theoretical cross sections
  - You usually have low selection efficiency $\implies$ few events to use in the analysis
- Find a **QCD-enriched region** (control region, CR) **in data**
- **Important properties** for a good CR:
  - It should be verified that it's indeed **QCD-enriched**
  - It should be **depleted from signal and other backgrounds**
  - It should be as kinematically **close** as possible **to the signal region** (SR)
  - It should be **orthogonal to the SR**

General idea

- QCD is only dominant in $1\tau\,0L$ category
  - $\approx$ 50% of the background yield in $1\tau\,0L$
- All the remaning **major backgrounds** ($t\bar{t}$ and $t\bar{t}$+X) **and signal involve** top quarks, i.e., **bottom quarks** in the final state
- First try: **revert the request on the number of b tagged jets in the event**

	N_{τ_h}	N_ℓ	N_{jets}	N_{bjets}
$1\tau\,0L$	1	0	≥ 8	≥ 2
$1\tau\,0L$ ctl	1	0	≥ 8	0

- $N_{bjets} = 0$ is meant to reject all the top-related processes
- Since setup is ready, check this for each VSjet DeepTau WP

Yields

- Signal region

	VVT	VT	T	M	L	VL	VVL	VVVL
tttt	4	6	7	10	13	20	27	23
tt	2146	3074	4384	6371	9861	17860	29339	25721
QCD	368	2378	4842	7461	15443	32927	61744	57889
tt+X	74	102	140	192	279	460	725	653

- Control region

	VVT	VT	T	M	L	VL	VVL	VVVL
tttt	0	0	0	0	0	0	0	0
tt	96	143	202	294	468	901	1550	1347
QCD	958	2411	4581	8087	15054	46772	89186	81403
tt+X	4	5	6	8	12	25	46	41

Fake rate method

- Inspired by EXO-19-015
- The **large QCD simulated yield** that we get **in CR should come from fake taus** (do we agree on this? Important!)
- Estimate the background completely from data by doing

$$N_{\text{fake-}\tau} = \sum_{p_T,\eta} N_{\text{fake-}\tau}(p_T,\eta) = \sum_{p_T,\eta} \left[N_{F,\overline{T}}(p_T,\eta) \times \frac{\text{FR}(p_T,\eta)}{1-\text{FR}(p_T,\eta)} \right]$$

- Using same (p_T, η) binning of EXO-19-015: $p_T \in [20, 30, 75, 150, \text{Inf}]$; $\eta \in [0, 1.5, 2.3]$

Fake rate method

QCD yield estimation

F. Iemmi

- **Compute FR(p_{T}, η) in my QCD CR**
- Compute $N_{\mathrm{F},\overline{\mathrm{T}}}$ in the so-called application region (**AR**), i.e., **SR with the exception of fakeable-not-tight taus**
- **But before that:** try **closure test on MC QCD**
  - Compute FR in CR, **apply the method in** the same **CR** (AR == CR)
  - **Compare with number** of events in CR **you count from MC**
  - This should close (at least approximately, I think)

```
[fabioiemmi@lxslc713 fake_rate]$ python printQCDYield.py
Fake rate method has been applied in the CR
Number of events by counting: 7978.89217198
Number of events from fake rate method: 8635.58006458
```

Conclusions

- **Computed FR in CR** and **applied** the fake rate **method in the very same CR**
- Expect closure
- **MC and FR prediction differ by ≈8%**
- **Can we claim for closure?**
- If so, shall I move to data?
  - Get FR from CR, apply in AR