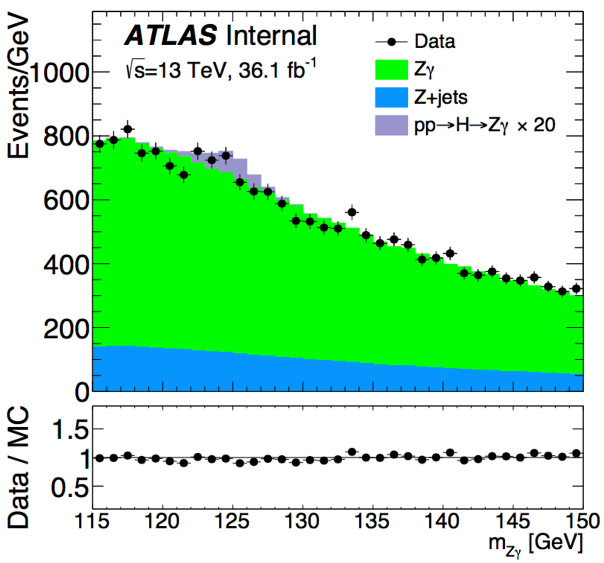
Background decomposition measurement based on photon Identification and isolation

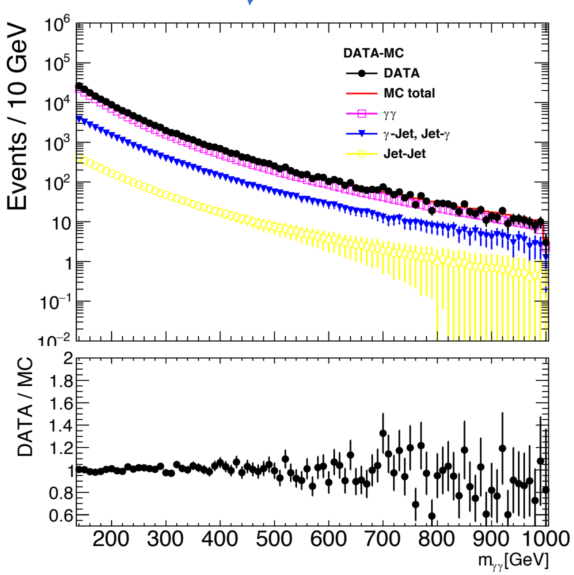
Shuo Han

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Introduction

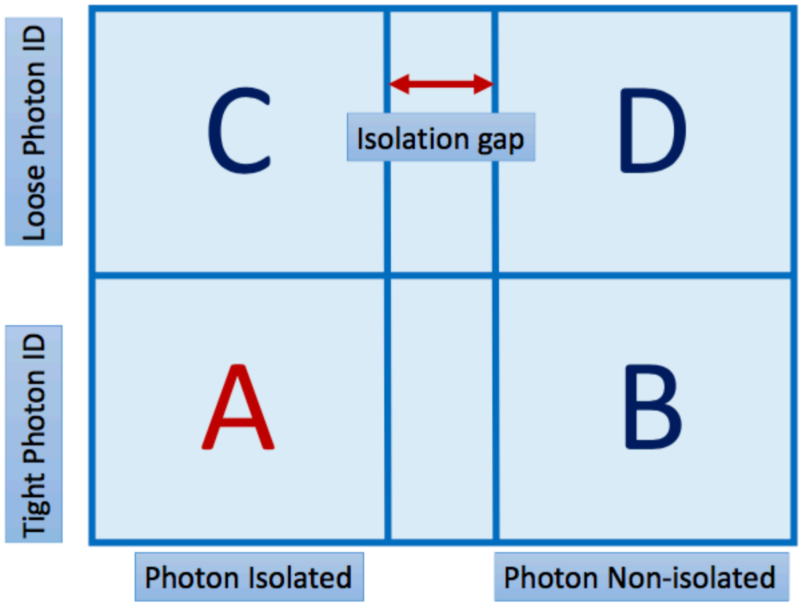
- **Motivation:**
  - Mixture the background for background modeling and background function description study
  - Measure the decomposition of reducible background (with jets) and irreducible background (with photons), by data-driven method
- **Method:**
  - 2-D sideband method (ABCD method)
  - 2x2-D sideband method
  - Dynamic template fit method





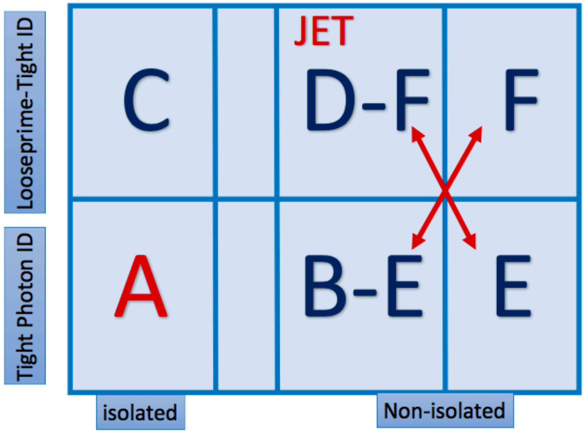
2-D sideband method

- **Definition (use photon final states as example):**
  - Use the number of photon candidates observed in the sidebands of a two-dimensional (e.g. E_T^{iso}, and photon identification) distribution to estimate the amount of background in signal region.
- **Input**
  - DATA
  - MC of irreducible background
  - MC of reducible background (we always don' t have one with enough event number)



2-D sideband method

- **Ideal assumption (use photon ID and Iso as example)**
  - No irreducible component (signal) leakage, irreducible background are all in region A, $N_{B,C,D}^{reducible} = N_{B,C,D}$
  - X and Y axis are not correlated $\frac{N_A^{reducible} \times N_D^{reducible}}{N_B^{reducible} \times N_C^{reducible}} = 1$
  - $Purity = \frac{N_A^{irreducible}}{N_A} = 1 - \frac{N_A^{reducible}}{N_A} = 1 - \frac{N_B N_C}{N_A N_D}$ (totally data-driven)
- **Corrections:**
  - Signal leakage: fraction of irreducible events in B, C, D c_B, c_C, c_D
    - Use irreducible background MC



  - $\frac{N_A^{reducible} \times N_D^{reducible}}{N_B^{reducible} \times N_C^{reducible}} = Rmc$
    - Use reducible background MC like Z+jet
    - Assuming high E_T^{iso} region and low E_T^{iso} region have the same correlation to photon ID, use BDEF regions

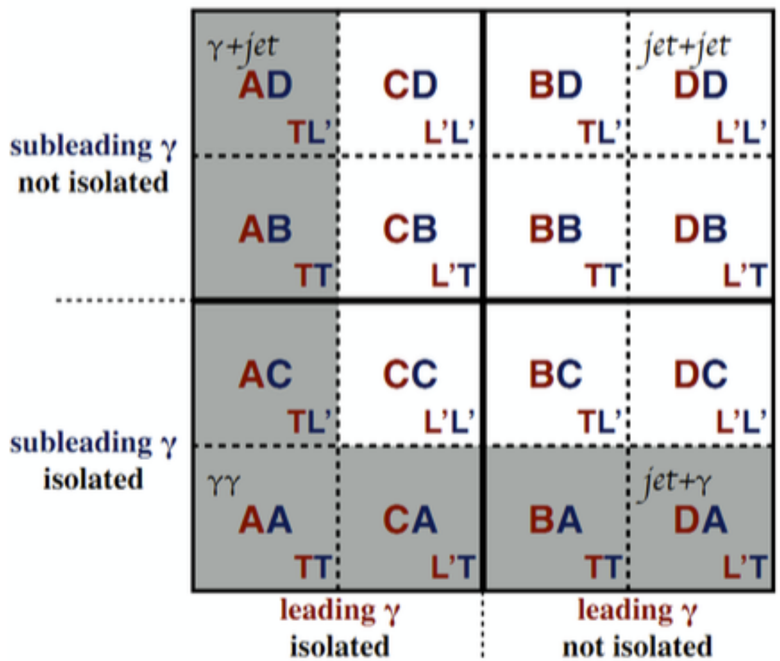
2-D sideband method

- **With corrections we got 5 parameters and 5 equations**
  - $N_A = N_A^{reducible} + N_A^{irreducible}$
  - $N_B = N_B^{reducible} + N_A^{irreducible} \times c_B$
  - $N_C = N_C^{reducible} + N_A^{irreducible} \times c_c$
  - $N_D = N_D^{reducible} + N_A^{irreducible} \times c_D$
  - $\frac{N_A^{reducible} \times N_D^{reducible}}{N_B^{reducible} \times N_C^{reducible}} = Rmc$

- **Solution:**
  - $Purity = \frac{N_A^{irreducible}}{N_A}, N_A^{irreducible} = \frac{b}{2a} \times \left(-1 + \sqrt{1 - \frac{b^2}{4ac}}\right)$
  - $a = c_B \times c_C \times Rmc - c_D$
  - $b = N_D + N_A \times c_D - N_B \times c_C \times Rmc - N_C \times c_B \times Rmc$
  - $c = N_B \times N_C \times Rmc - N_A \times N_D$

2x2-D sideband method

- **Definition:**
  - Extrapolation of ABCD method, considering leading and sub-leading photons in yy background
  - 4 dimensions (ID of leading photon, ID of sub-leading photon, E_T^{iso} of leading photon, E_T^{iso} of sub-leading photon)
  - ABCD -> AA, AB, ⋯ DC, DD (16 regions)



- **Input**
  - DATA
  - MC of irreducible background

2x2-D sideband method

- **Define**
  - Yield of yy, yj, jy, jj components $Y_{yy}, Y_{jy}, Y_{yj}, Y_{jj}$
  - Isolation rate of photon $\varepsilon_1^{iso}, \varepsilon_2^{iso}$
  - ID rate of photon $\varepsilon_1^{ID}, \varepsilon_2^{ID}$
  - Isolation fake rate of jet $f_1^{iso}, f_2^{iso}, f'^{iso}_1, f'^{iso}_2$ $(f - yj, jy, f' - jj)$
  - ID fate rate of jet $f_1^{ID}, f_2^{ID}, f'^{ID}_1, f'^{ID}_2$
  - Correlation between ID and isolation fake rate $C_1^{id-iso}, C_2^{id-iso}$ (set to 1±0.2)
  - Correlation between isolation fake rate in jet-jet events C_{jj}^{iso}

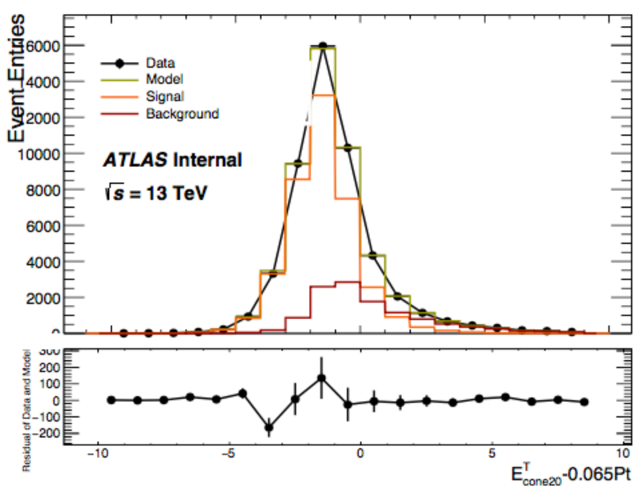
➢ relationship to ABCD method: $\frac{C_i^{id-iso}\,(1-f_i^{iso}-f_i^{ID}-f_i^{iso} f_i^{ID} C_i^{id-iso})}{(1-f_i^{iso} C_i^{id-iso})(1-f_i^{ID} C_i^{id-iso})} = Rmc_i$

- **In each region we have the equation like**
- $N_{AA} = Y_{yy} \times \varepsilon_1^{iso} \varepsilon_2^{iso} \varepsilon_1^{ID} \varepsilon_2^{ID} + Y_{yj} \times \varepsilon_1^{iso} \varepsilon_1^{ID} f_2^{iso} f_2^{ID} C_2^{id-iso} + Y_{jy} \times f_1^{ID} f_1^{iso} C_1^{id-iso} \varepsilon_2^{ID} \varepsilon_2^{iso} + Y_{jj} \times f'^{iso}_1 f'^{iso}_2 f'^{ID}_1, f'^{ID}_2 C_{jj}^{iso} C_1^{id-iso} C_2^{id-iso}$
- 16 equations > 13 unknown parameters -> likelihood fit

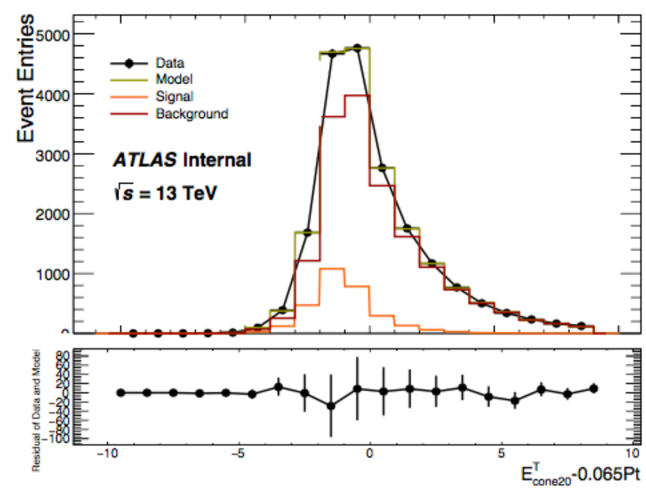
Dynamic template fit method

- **Definition (use photon final states as example):**
  - A binned maximum likelihood fit to the E_T^{iso} distribution of photon candidates selected in data which pass the tight ID. Where the distribution is from simulation of irreducible background and a data template for reducible background
- **Input**
  - DATA
  - MC of irreducible background

Tight – ID (SR)



LoosePrime4 – Tight (CR)



Dynamic template fit method

- **Assumptions**
  - Use signal leakage $Leakage_{ibin}$ from MC
  - Use signal shape $fraction_{ibin}$ from MC
  - Background shape is same in signal and control regions
- **Fit parameters:**

  - signal ($Z+\gamma$) yields in the signal region N_{SR}^{Sig}.
  - background (Z+jet) yields in the signal region N_{SR}^{Bkg}.
  - background (Z+jet) yields in every E_T^{cone20} bins of the control region N_{ibin}^{Bkg}. $\times$ *number of Bins*
- **Fit points:**

$$N_{ibin}^{SR} = N_{SR}^{Sig} * Fraction_{ibin} + N_{SR}^{Bkg} * \frac{N_{ibin}^{Bkg}}{\sum_{ibin=1}^{Nbins} N_{ibin}^{Bkg}}$$

$\times$ *number of Bins*

$$N_{ibin}^{CR} = N_{SR}^{Sig} * Fraction_{ibin} * Leakage_{ibin} + N_{ibin}^{Bkg}$$

- **Equivalent to ABCD method (Rmc = 1) when Nbins = 2, use maximum likelihood fit when Nbins > 2**

backup

Result of Zgamma search

- Zgamma search background decomposition is using
  - ABCD method (Rmc from BDEF region)
  - Dynamic template fit method
- There' s a good consistency in this search

ABCD

	$\ell\ell\gamma$ purity
2015+2016 inc.	$0.813 \pm 0.012^{+0.012}_{-0.019}$
VBF cate	$0.985 \pm 0.022^{+0.000}_{-0.052}$
rel pT cate	$0.865 \pm 0.087^{+0.000}_{-0.222}$
high pTt ee	$0.731 \pm 0.122^{+0.128}_{-0.000}$
low pTt ee	$0.813 \pm 0.019^{+0.013}_{-0.053}$
high pTt $\mu\mu$	$0.863 \pm 0.048^{+0.022}_{-0.000}$
low pTt $\mu\mu$	$0.806 \pm 0.016^{+0.010}_{-0.007}$
ttbar yield inc.	

Template fit

	$\ell\ell\gamma$ purity
2015+2016 inc.	$0.838 \pm 0.005^{+0.000}_{-0.031}$
VBF cate	$0.970 \pm 0.030^{+0.002}_{-0.000}$
rel pT cate	$0.925 \pm 0.018^{+0.010}_{-0.000}$
high pTt ee	$0.899 \pm 0.018^{+0.007}_{-0.026}$
low pTt ee	$0.827 \pm 0.008^{+0.000}_{-0.042}$
high pTt $\mu\mu$	$0.871 \pm 0.018^{+0.000}_{-0.028}$
low pTt $\mu\mu$	$0.824 \pm 0.007^{+0.000}_{-0.041}$

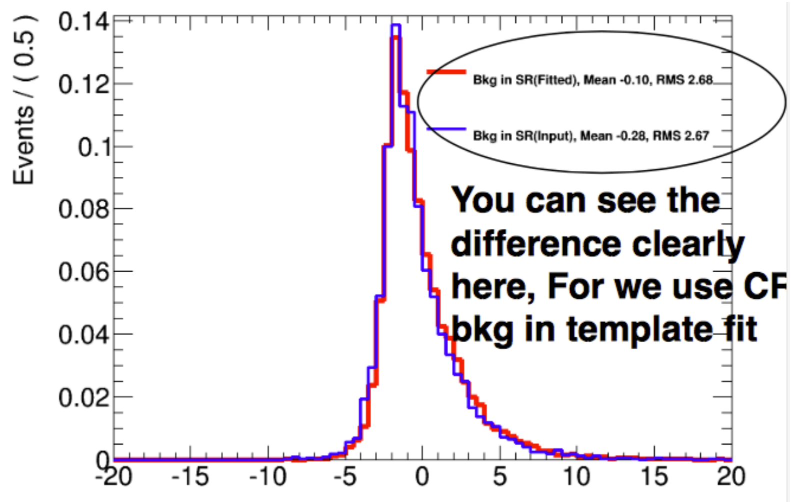
Exist problem

- **ABCD method:** BDEF is not always work (the assumption that high E_T^{iso} region and low E_T^{iso} region have the same correlation factor is not true)
  - Only when we are sure the correlation is very small between X and Y axis, we can simply assume Rmc = 1
- **Mimic data for Jet fake rate study (Purity = 0.262):**
  - Signal MC : single photon;
  - Background MC : di-jet

X-axis	Rmc	Strategy of Rmc	Purity (mimic data)
Ptcone20-0.05pT	1.05±0.04	(B’ D’ EF, mimic data)	0.283±0.026
Ptcone20-0.05pT	1.03±0.04	(B’ D’ EF, MC_jet)	0.296±0.032
Ptcone20-0.05pT	**1.08±0.04**	**(ABCD, MC_jet)**	**0.262±0.031**
Etcone20-0.065pT	1.05±0.05	(B’ D’ EF, mimic data)	0.365±0.031
Etcone20-0.065pT	1.03±0.06	(B’ D’ EF, MC_jet)	0.377±0.039
Etcone20-0.065pT	**1.20±0.05**	**(ABCD, MC_jet)**	**0.262±0.038**

Exist problem

◆ **Template fit method:** the assumption that background shape is same in signal and control regions is not true



Jet fake rate study
Done by Cong

- Template fits with following input shapes:

I. Normal way: S_SR, S_CR, Data_CR to fit Data_SR/Data_CR
II. **Same bkg shape in SR to fit Data_SR/Data_CR**
III. Same signal shape in SR to Data_SR/Data_CR

Signal Region	Signal Yield	Bkg Yield	Total Yield	Purity
Original Inputs	2.90545e+08	8.17705e+08	1.10825e+09	**26.2%**
Input shapes(I)	411622727	696154192	1107776920	37.0%
Input shapes(II)	291408532	817693393	1109101926	26.2%
Input shapes(III)	414300188	693559131	1107859319	37.4%

Summary

- Data driven method for background decomposition is possible, when we have reducible and irreducible background.

- ABCD method is based on the event numbers and can be accurate when we trust background simulations, but it have problem when we don' t have reducible background simulation.

- Template fit method don' t need a reducible background simulation, but it rely on the photon isolation variables shape, where the reducible background shape is not always the same between signal and control regions

- My recommendation:
  - ABCD (Rmc from MC) > Template fit > ABCD (Rmc from BDEF) ~ ABCD (Rmc = 1)