HEP DATA ANALYSIS: GROUP 3

Q. Dong, H. Guo, Y. Huang, F. Li, J. Liu,
Q. Man, C. Que, C. Xi,
K. Zhang, L. Zhang, Y. Zhang, H. Zhou, Y. Zou

TASK 1: EVENT SELECTION

- Select the signal events by putting a cut on the Data tree, do the following:

1. Find a cut variable "ann" using the signal MC tree, so as you can keep 90% of your signal.

2. Given the cut you find above, using the MC background tree, evaluate the corresponding background rejection rate, i.e. the "power" to reject the background (see our lecture note for the definition of the "power").

3. Report the cut you find, and the "power" you get.

TASK 1: PROCEDURES

- How to do event selection? CUT!!
  - Monte Carlo Simulation: Sig & Bkg ◀····· Determined Cut Condition
  - Real Data: Sig? Bkg? ◀····· Apply Cut Condition
  - How to qualify the cut condition?
    1. Efficiency(Purity)
    2. Significance

TASK 1: KEEP 90% EFFICIENCY

- **Efficiency:** ratio between # of signals before and after
- Binary Search.

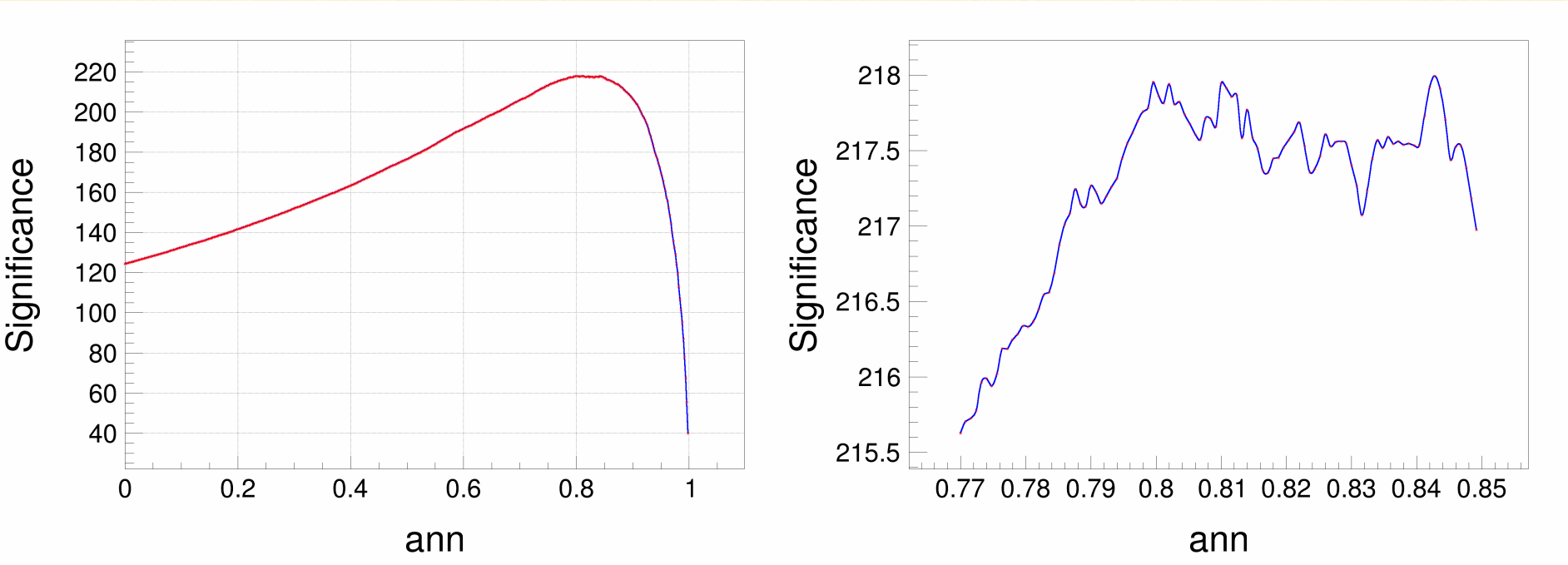
```
// binary search to find the certain point
Int_t bin_max = nbinsx;
Int_t bin_min = 0;
Int_t bin_mid = 0.5*(bin_max+bin_min);
while( bin_min < bin_max ) {
    sum = h1->Integral(bin_mid,nbinsx);
    if( abs(sum/(n_entries+0.0) - ratio) <= eps ) {
        cout<<"==> Cut Point founded!..."<<endl;
        break;
    }
    else if( (sum/(n_entries+0.0) < ratio - eps) ) {
        bin_max = bin_mid;
    }
    else if( (sum/(n_entries+0.0) > ratio + eps) ) {
        bin_min = bin_mid;
    }
    bin_mid = ( (Int_t)(0.5*(bin_max+bin_min)) == bin_mid ) ? bin_mid+1 : (Int_t)(0.5*(bin_max+bin_min)) ;
}
cut = h1->GetBinCenter(bin_mid);
```

Integral(bin_{low}, bin_{up})

TASK 1: RESULTS

Cut	ann >= 0.772875		
	Total	After Cut	Difference
Signal	20,000	18,001	1,999
Background	20,000	2,893	17,107
Significance		215.916	
		H_0 is true	H_0 is false
	Fail to Reject	0.90005	0.14465
	Reject	0.09995	$1 - \beta = 0.85535$

TASK 1: EXTRA



$$S = \sqrt{2(s+b)\log(1+\frac{s}{b}) - 2s}$$

TASK 1: EXTRA RESULTS

Cut	ann >= 0.80		
	Total	After Cut	Difference
Signal	20,000	17339	2661
Background	20,000	2506	17494
Significance		217.833	
		H_0 is true	H_0 is false
	Fail to Reject	0.86695	0.1253
	Reject	0.13305	$1 - \beta = 0.8747$

TASK 2: FIT TO DATA

Fit to the data mass distribution, and extract the mass mean value (μ) and the width (σ) using a composed function.

- Apply the cut you have found in Task 1 to the data tree on the "ann" variable, and draw the "mass" into a histogram.
- Define a composed function: $f(x) = k \cdot f_S(x) + (1 - k) \cdot f_B(x)$ using Gaussian as your signal function and exponential function as your background function.
- Fit the composed function to the data mass histogram, to extract the mean value (μ) and the width (σ) of the signal mass peak.

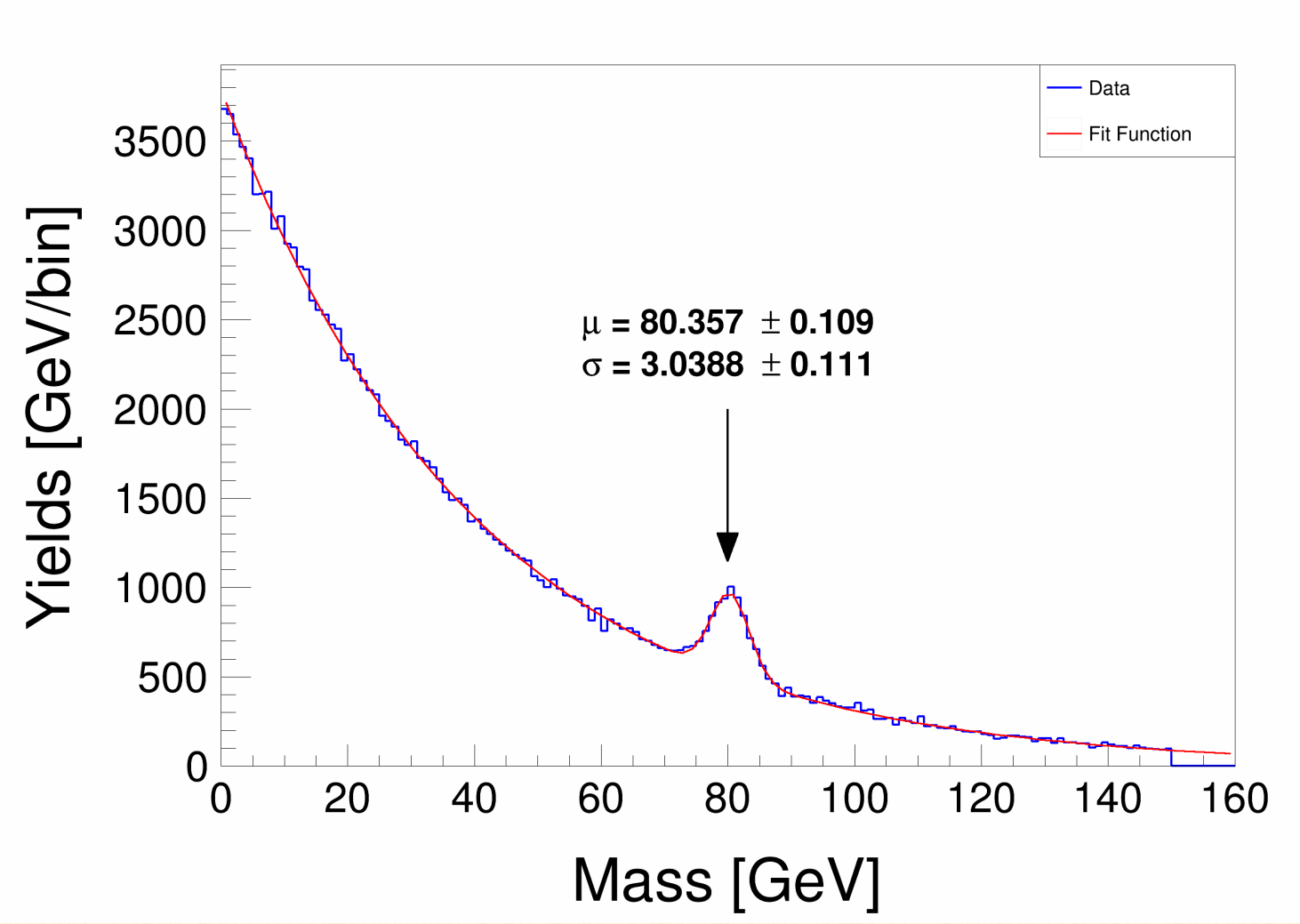
TASK 2: FIT PROCEDURE

1. Fit data in multiple sub ranges
2. Sub ranges fitting results as initial value for whole fitting

```
Double_t par[6];
TF1* f1 = new TF1("f1","gaus",75,85);
TF1* f2 = new TF1("f2","expo",0,160);
TF1* fitfunc = new TF1("fitfunc","[5]*gaus(0)+(1-[5])*expo(3)",0,160);
massh->Fit("f1","QR");
massh->Fit("f2","QR+");
f1->GetParameters(&par[0]);
f2->GetParameters(&par[3]);
par[5] = 0.5;
fitfunc->SetParameters(par);

// Fitting data!!
massh->Fit("fitfunc","");
```

TASK 2: FIT RESULTS



TASK 2: FIT TO DATA

$$f(x) = kAe^{-0.5[\frac{(x-\sigma)}{\mu}]^2} + (1-k)e^{C+\lambda x}$$

Para.	*Value*	*Error*
Amp.	992.23	32.278
Mean	80.357	0.10924
Width	3.0388	0.11061
Constant	8.8722	0.00633
Lambda	-0.0250	0.00008
K	0.4688	0.00276

SUMMARY

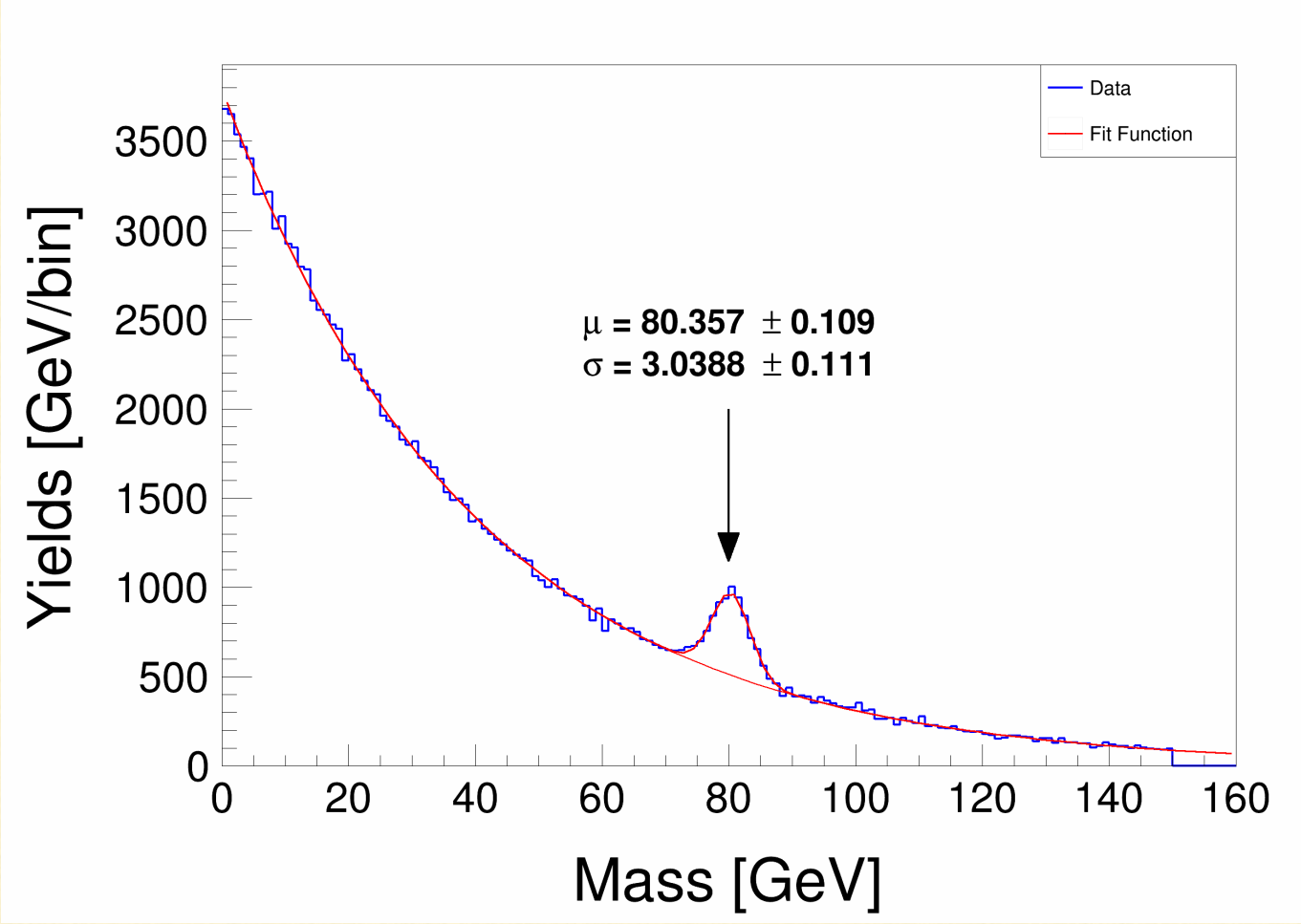
		ann >= 0.772875	ann >= 0.80
MC	Efficiency	0.90005	0.86695
	Significance	215.916	217.833
Data	Mean	80.357 (0.109)	80.384 (0.106)
	Width	3.0388 (0.111)	2.9896 (0.104)
	Local-S	44.401	45.892
	Global-S	9.2136	9.4766

$$\int \mathscr{L} dt \times \sigma \times \epsilon_{eff} = N$$

"We're walking along a rugged path, never knowing what you're going to get."

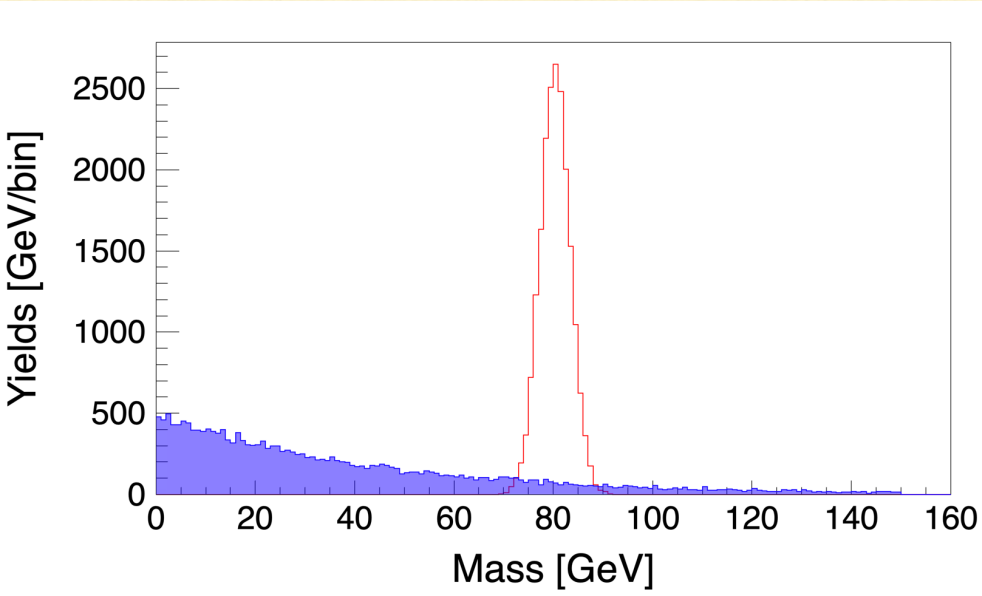
–Yulei Zhang

BACKUP

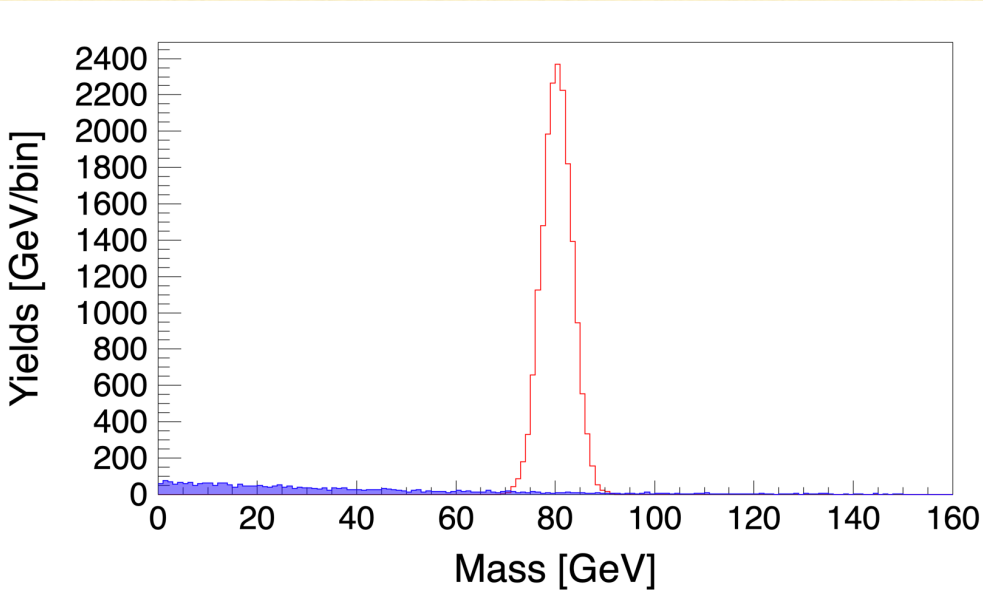

Data
Fit Function
μ = 80.357 ± 0.109
σ = 3.0388 ± 0.111
Yields [GeV/bin]
Mass [GeV]
0
500
1000
1500
2000
2500
3000
3500
0
20
40
60
80
100
120
140
160

BACKUP

Before Cut

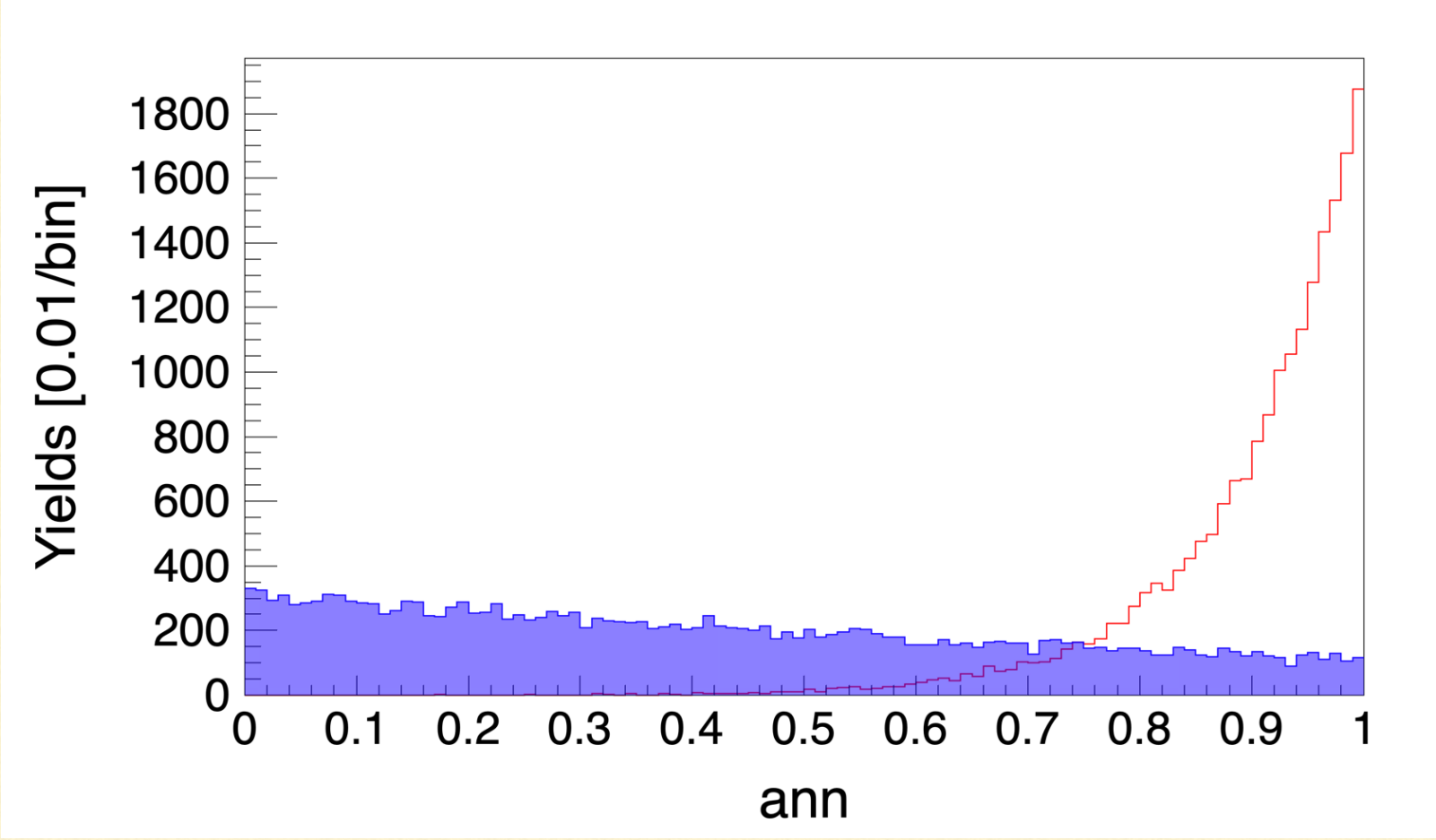


Cut: ann>=0.77

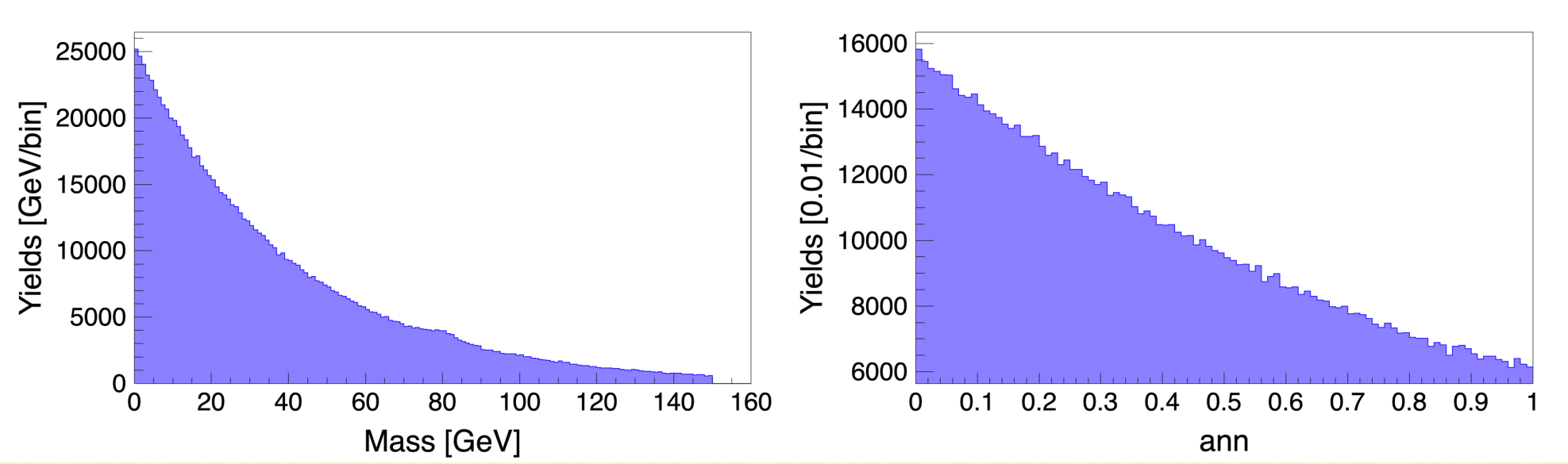


MC Simulation

BACKUP


Yields [0.01/bin]
1800
1600
1400
1200
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ann

BACKUP


Yields [GeV/bin]
25000
20000
15000
10000
5000
0
0
20
40
60
80
100
120
140
160
Mass [GeV]
Yields [0.01/bin]
16000
14000
12000
10000
8000
6000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ann

Data

BACKUP

```
// Apply ann cut on both Sig and Bkg
char temp[50];
sprintf(temp,"ann>=%f",cut_anna);
TCut cut1 = temp;
Long64_t cut_Sig = Sig->Draw("ann",cut1);
Long64_t cut_Bkg = Bkg->Draw("ann",cut1);
Double_t Significance = sqrt( 2*((cut_Sig+cut_Bkg)*log(1.0+(cut_Sig+0.0)/cut_Bkg)-cut_Sig) );
```