Study of $Z \rightarrow e^+e^-$ Reconstruction with Summer08 Samples

Jeffrey Berryhill
Kalanand Mishra
Fermilab

EWK-electron December 5, 2008

$Z \to e^+e^-$ cross-section

Formula for $Z \to e^+e^-$ cross section can be written as

$$\sigma = \frac{N}{\int \mathcal{L} \cdot A \cdot \epsilon^2}$$

where

N = number of signal events reconstructed
∫ L = integrated luminosity
A = acceptance
ε = efficiency for single electron reconstruction

There are two important caveats in using this formula:

1. Typically one uses tag-and-probe with $Z \to e^+e^-$ events to measure ε. If tag selection overlaps with probe selection, then one needs to take this overlap into account when calculating ε. ⇒ See next slide

2. For a simple one-step efficiency measurement, where tag and passing probe coincide, ε and σ are anti-correlated. ⇒ a couple of slides later

Average efficiency using tag-and-probe - I

If tag selection never overlaps with probe selection, then there are two disjoint categories, passing probes N_{TP} and failing probes N_{TF},

$$N_{TP} = N \times \varepsilon \times \text{tag\_efficiency}$$
$$N_{TF} = N \times (1 - \varepsilon) \times \text{tag\_efficiency}$$

N is the size of the tag-and-probe sample

If tag and passing probe selection are identical, then there are two categories: double tag events N_{TT}, and tag+ failing probe N_{TF}

$$N_{TT} = N \times \varepsilon \times \varepsilon$$
$$N_{TF} = 2N \times \varepsilon \times (1 - \varepsilon)$$

$$\Rightarrow \varepsilon = 2N_{TT} / (2N_{TT} + N_{TF})$$

If tag selection overlaps with probe selection, then there are three disjoint categories: N_{TT}, N_{TP}, and N_{TF}.
One can solve for efficiency in terms of N_{TT},N_{TP},N_{TF}, or perform a multi-category simultaneous fit to N, efficiency, and efficiency(T|P). However, the statistical errors are now all potentially correlated.

(see AN2007-019 appendix A, also talk by J. Berryhill, July 7, 2008)

Average efficiency using tag-and-probe - II

Alternatively, define bins in such a way that events are never reused:

1. Divide the tag-and-probe sample into two halves, using some variable completely uncorrelated with any electron property (e.g., event number modulo 2). Call them sample 1 and sample 2.
2. Categorize all tags by some 2-fold discrete property (charge). Call them sample + and sample -.
3. In sample 1, require all tags to be in category +, probes in category -. In sample 2, require all tags to be in category -, probes in category +.

Measure the binned e^- efficiency in sample 1.
Measure the binned e^+ efficiency in sample 2.

<u>Advantages:</u>

★All opposite sign tag-probe pairs are used, and used exactly once.
★"Double tag" events no longer exist.
★If the efficiencies for the two charges are nearly the same, then one can even think of averaging the two weighted by uncertainties.
★There is no loss of statistics in this procedure.

Uncertainty in the cross section calculation

[Taken from Jeffrey Berryhill's presentation in EWK-electron, July 7, 2008]

For a simple one step efficiency measurement, where tag and passing probe selection coincide, ε and σ are anti-correlated.

For a non-overlapping but identical tag and probe selection, we derived $\varepsilon = 2N_{TT} / (2N_{TT} + N_{TF})$. Let's denote this relation as $\epsilon = \frac{2P}{2P+F}$

$$\begin{aligned}\sigma &= \frac{N}{\int \mathcal{L} \cdot A \cdot \epsilon^2} \\ &= \frac{1}{\int \mathcal{L} \cdot A} \cdot \frac{P}{4P^2/(2P+F)^2} \\ &= \frac{1}{\int \mathcal{L} \cdot A} \cdot P(1+F/2P)^2\end{aligned}$$

$$\begin{aligned}\delta\sigma^2_{corr} &= \frac{1}{\int \mathcal{L}^2 \cdot A^2} \cdot P(1 + \frac{F}{P} + \frac{1}{2}(\frac{F}{P})^2 + \frac{1}{4}(\frac{F}{P})^3 + \frac{1}{16}(\frac{F}{P})^4) \\ \delta\sigma^2_{uncorr} &= \frac{1}{\int \mathcal{L}^2 \cdot A^2} \cdot P(1 + 3\frac{F}{P} + \frac{7}{2}(\frac{F}{P})^2 + \frac{7}{4}(\frac{F}{P})^3 + \frac{5}{16}(\frac{F}{P})^4)\end{aligned}$$

If the correlation is accounted for correctly (top), the uncertainty in the cross section is reduced relative to naively treating them as uncorrelated (bottom).

The correlation can be accounted for correctly via a simultaneous fit to N and ε). After 10 pb^{-1} integrated luminosity, this measurement will not be statistics limited, in which case we can ignore the correlation. We will be overestimating our statistical error slightly.

Our strategy

- ★ Perform the binned efficiency measurements in such a way that double-tag events and binned correlations do not exist.
- ★ Ignore the correlation between N and ε.

Average efficiency calculation

[input from Anil Singh, Punjab university]

- Store e^+ and e^- efficiency as a function of η and p_T into text tables.
- Have a mechanism to read these text tables.
- Distribute the reconstructed $Z \to e^+e^-$ events over the rows (bins) of the efficiency lookup table.
- Find the occupancy of each bin and also the corresponding efficiency and error values.
- Calculate the average efficiency and error in usual way.

key	Efficiency/Error
i	$\epsilon_i{}^-, \delta\epsilon_i{}^-$

+

key	Efficiency/Error
j	$\epsilon_j{}^+, \delta\epsilon_j{}^+$

=

	12		14	
21				
			ij	
		43		
				55

$$\bar{\epsilon} = \sum_{dibins} \frac{N_{ij} * \epsilon_{ij}}{N}$$

$$\delta\epsilon^2 = \sum_{\mathbf{i}} \{\sum_{\mathbf{j}} \mathrm{f_{ij}} \epsilon_{\mathbf{j}}{}^+\}\{\delta\epsilon^-{}_{\mathbf{i}}\}^2 + \sum_{\mathbf{j}} \{\sum_{\mathbf{i}} \mathrm{f_{ij}} \epsilon_{\mathbf{i}}{}^-\}\{\delta\epsilon^+{}_{\mathbf{j}}\}^2$$

$Z \rightarrow e^+e^-$ event selection

Tag electron selection

- GsfElectrons.
- $|\eta| < 1.4442 \,||\, 1.56 < |\eta| < 2.5$
- $p_T > 20$ GeV/c.
- Track isolation:
  - $\Sigma\, p_t^{tracks} / p_t^{el} < 0.2$
  - Inner cone radius 0.02
  - Outer cone radius 0.2
  - p_T of tracks > 1.5 GeV/c
- Electron ID: Robust
- Trigger: HLT_LooseIsoEle15_LW_L1R

Probe electron selection

SuperCluster → GsfElectron → Isolation → ID → HLT

- $|\eta| < 1.4442 \,||\, 1.56 < |\eta| < 2.5$
- $p_T > 20$ GeV/c.
- Fit the tag-probe invariant mass to get the number of signal events.

Data Sample

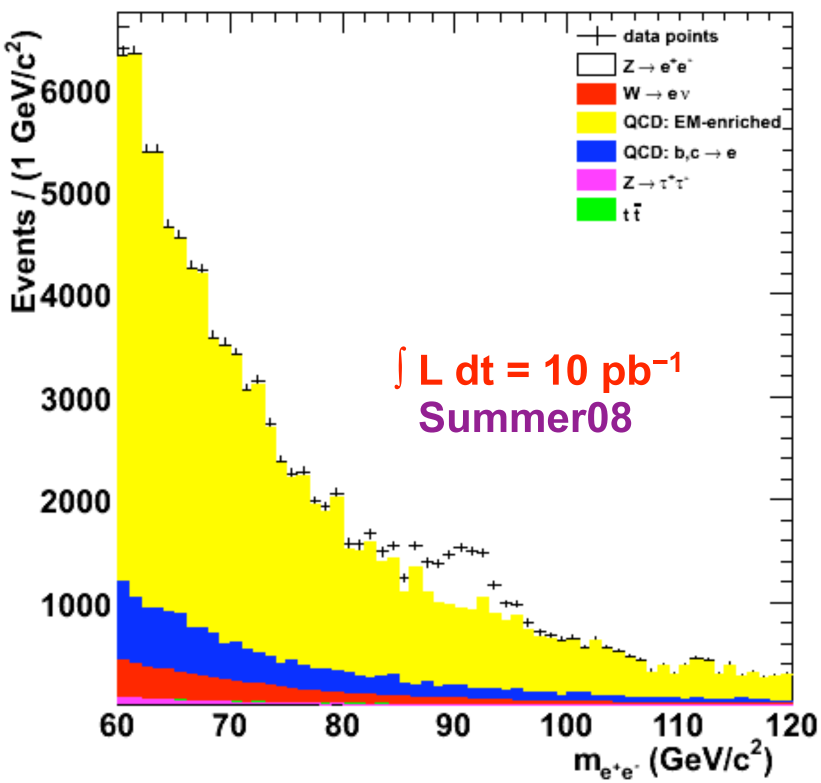
```
/Zee/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/Wenu/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/Ztautau/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_BCtoE_Pt20to30/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_BCtoE_Pt30to80/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_BCtoE_Pt80to170/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_EMenriched_Pt20to30/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_EMenriched_Pt30to80/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/QCD_EMenriched_Pt80to170/Summer08_IDEAL_V9_v1/GEN-SIM-RECO
/TauolaTTbar/Summer08_IDEAL_V9_v1/GEN-SIM-RECO.
```

Sample	# events (in million)	cross section (in pb)	filter efficiency
$Z \to e^+e^-$	1.0	1232	0.701
$W \to e\nu$	1.1	11865	0.738
$Z \to \tau^+\tau^-$	1.1	11840	1.0
QCD $b, c \to e$ with $20 < \hat{p}_T < 30$ GeV/c	2.2	4×10^8	0.00048
QCD $b, c \to e$ with $30 < \hat{p}_T < 80$ GeV/c	2.0	1×10^8	0.0024
QCD $b, c \to e$ with $80 < \hat{p}_T < 170$ GeV/c	0.8	1.9×10^6	0.012
QCD EM enriched with $20 < \hat{p}_T < 30$ GeV/c	5.1	4×10^8	0.008
QCD EM enriched with $30 < \hat{p}_T < 80$ GeV/c	11.9	1×10^8	0.047
QCD EM enriched with $80 < \hat{p}_T < 170$ GeV/c	5.7	1.9×10^6	0.15
$t\bar{t}$	0.15	242	1.0

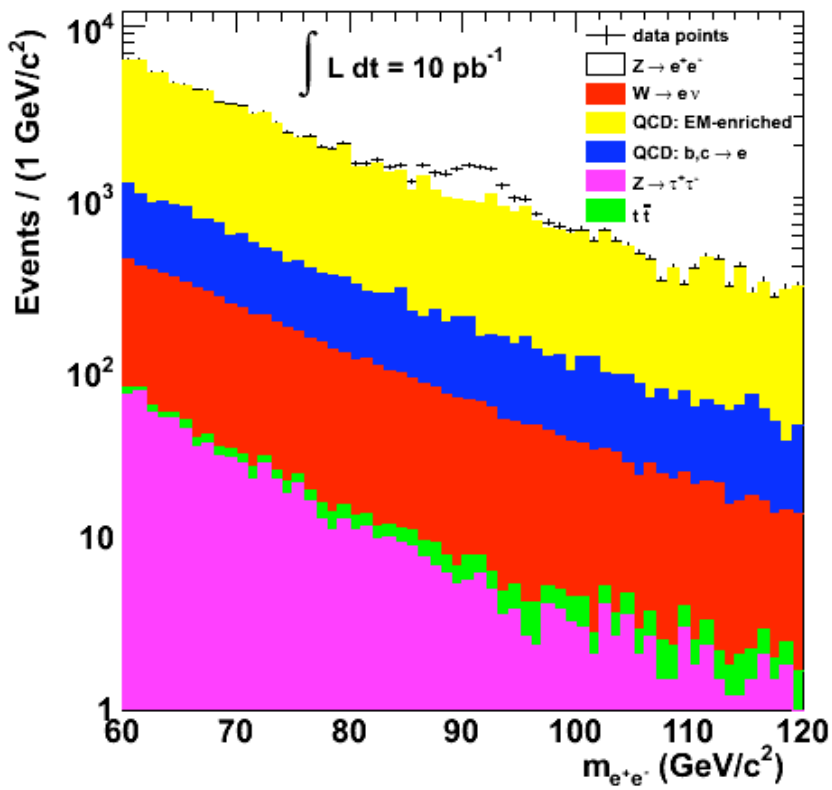
Study of signal purity for $Z \to e^+e^-$ - I

- Dielectron invariant mass for events passing single electron trigger criteria.
- The second electron is just a super cluster.
- No requirements of matching track or isolation for the second electron.

Signal purity = 16.4 % (in the signal region, ±3σ around Z peak)



Linear scale

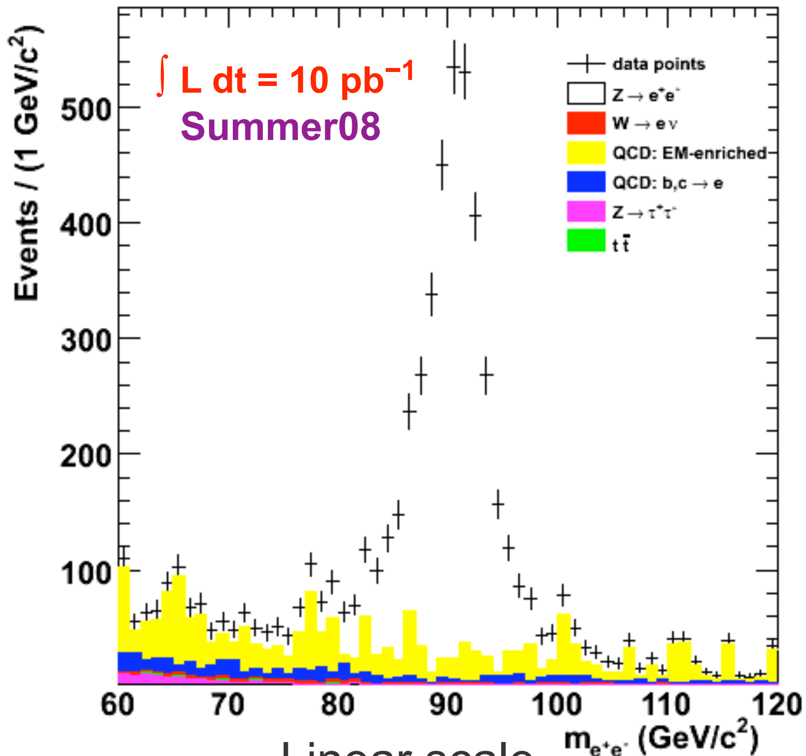


Logarithmic scale

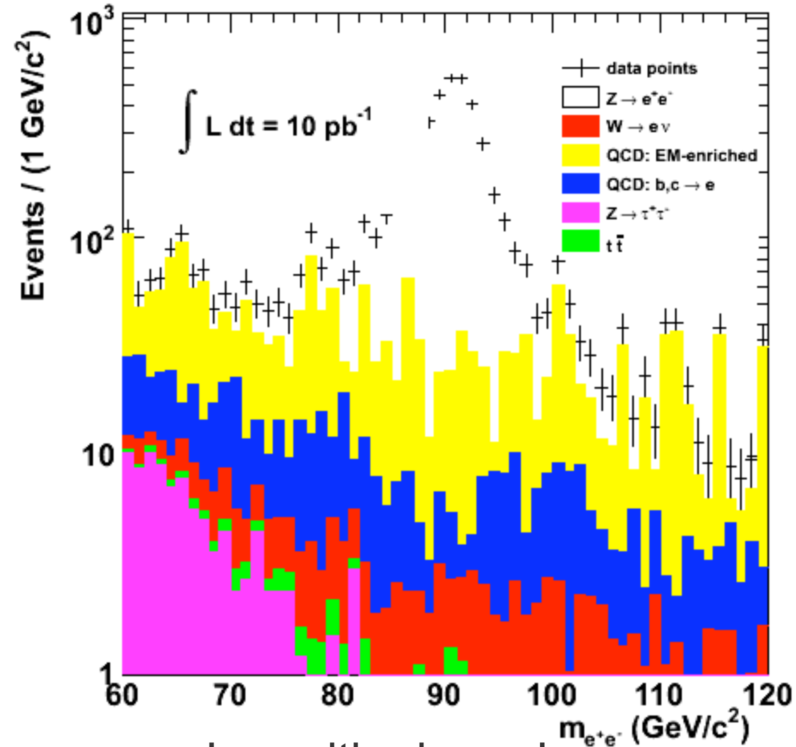
Study of signal purity for $Z \rightarrow e^+e^-$ - II

- The second electron is a super cluster matched to a track.
- No isolation requirement for the second electron.

Signal purity = 84.7 %



Linear scale

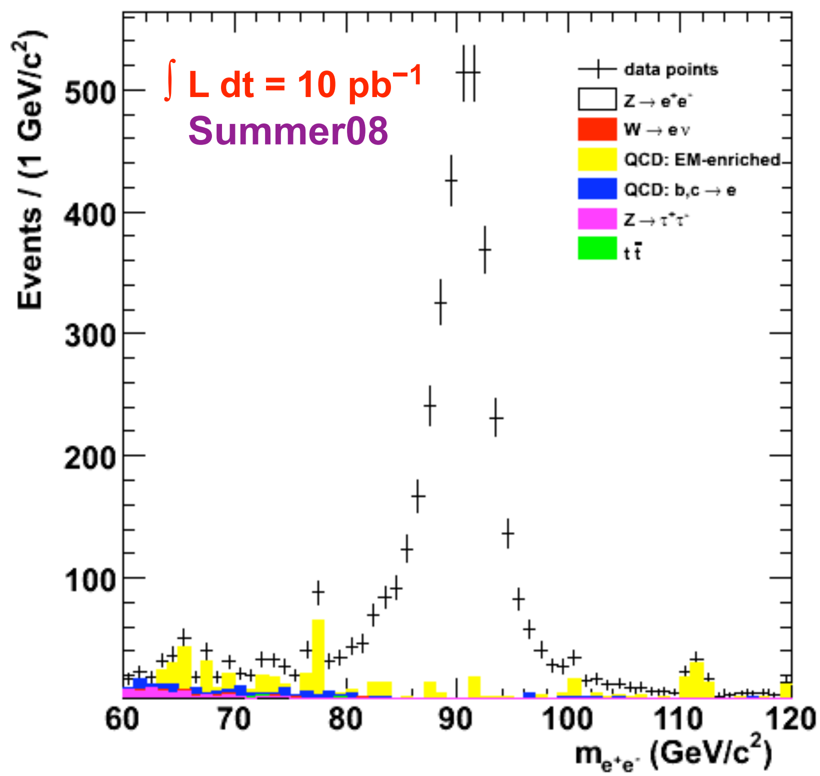


Logarithmic scale

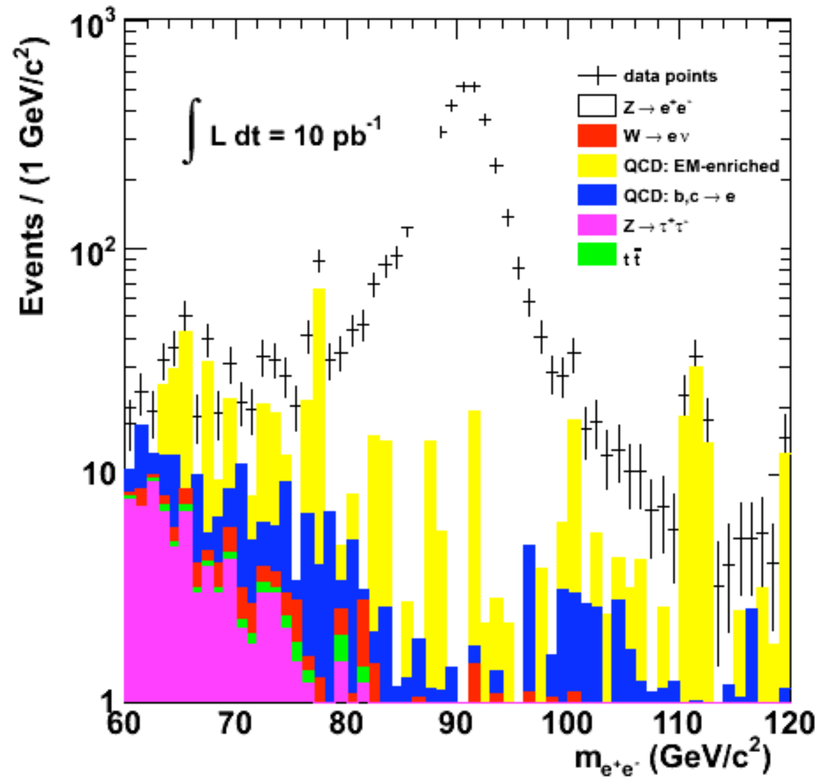
Study of signal purity for $Z \to e^+e^-$ - III

- Additionally, the second electron is isolated.

Signal purity = 96.5 %



Linear scale

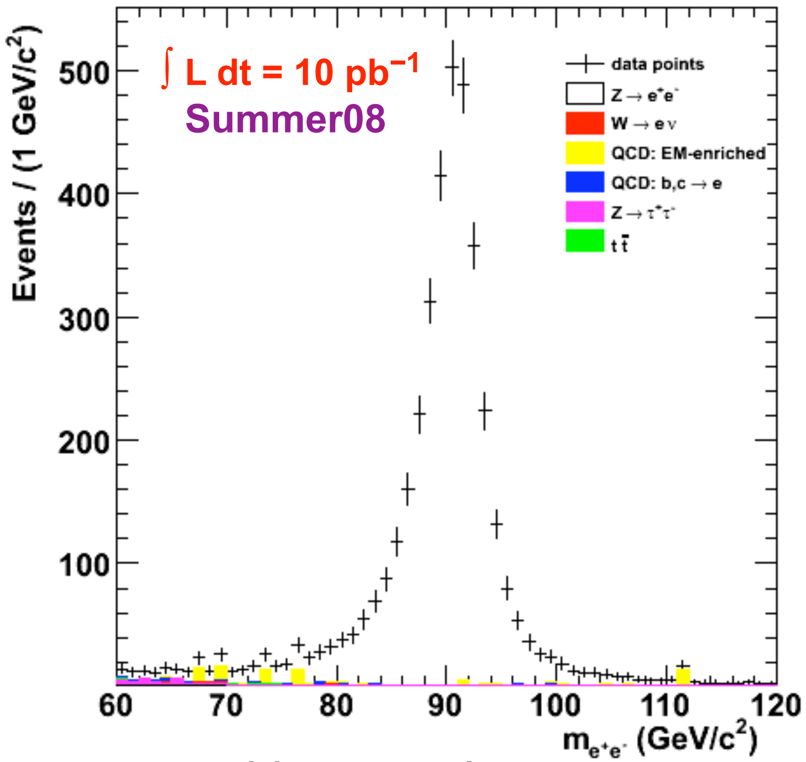


Logarithmic scale

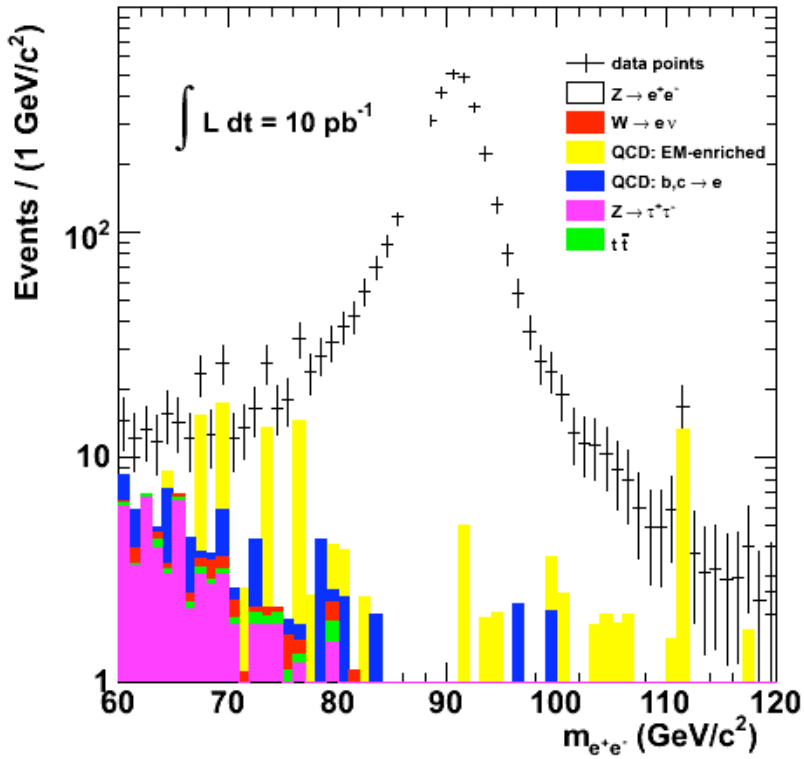
Study of signal purity for $Z \rightarrow e^+e^-$ - IV

- Additionally, the second electron passes "loose" electron identification criteria.

Signal purity = 99.1 %



Linear scale

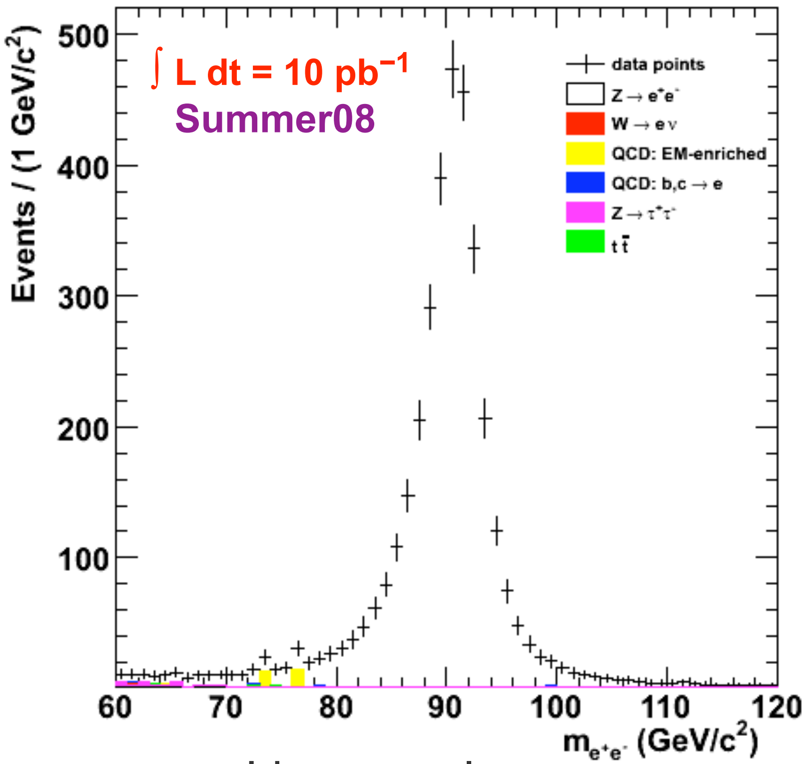


Logarithmic scale

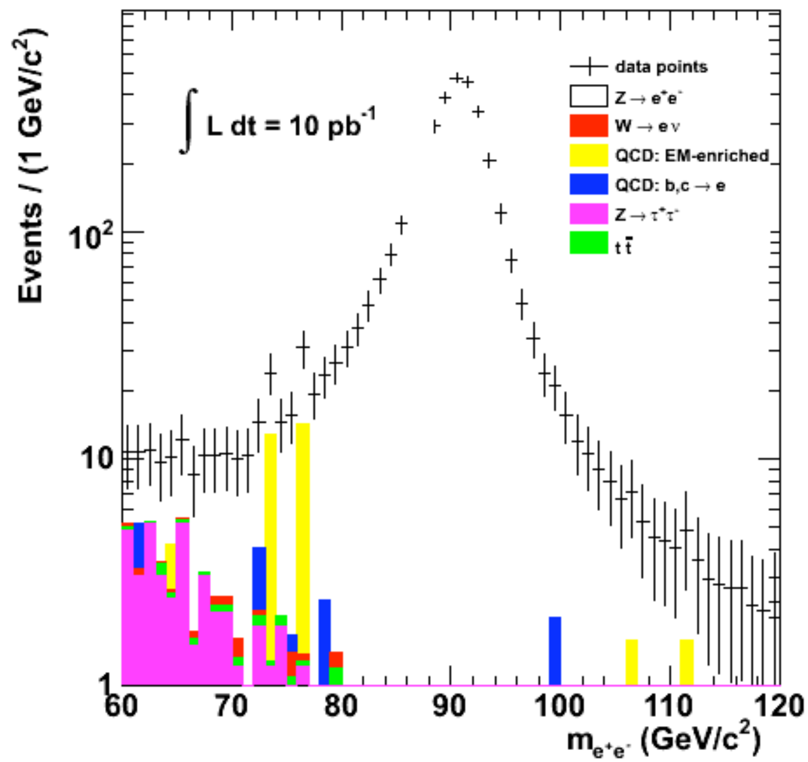
Study of signal purity for $Z \rightarrow e^+e^-$ - V

- Additionally, the second electron passes single electron trigger criteria.

Signal purity = 99.8 %



Linear scale



Logarithmic scale

Summary of $Z \rightarrow e^+e^-$ signal purity

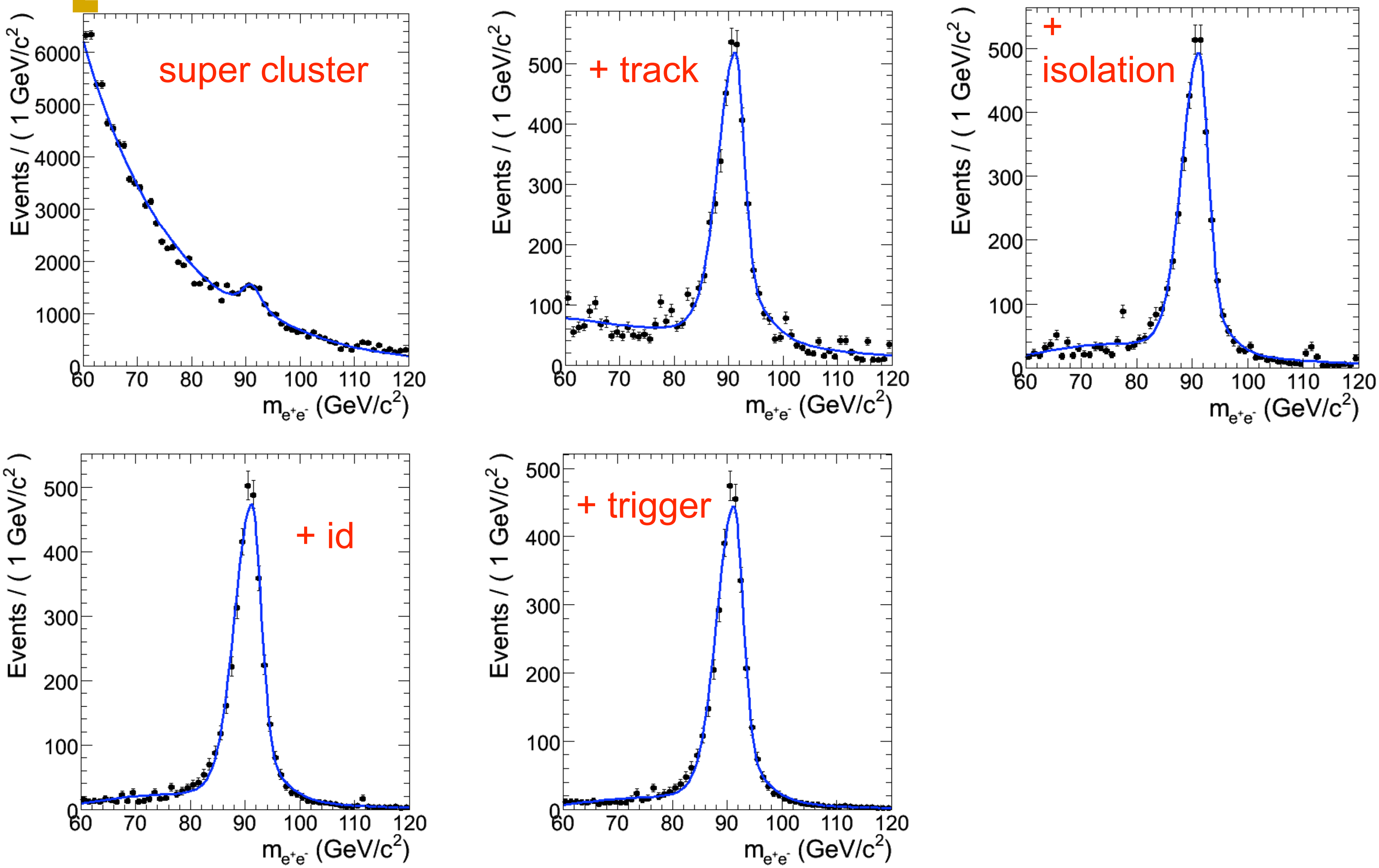
Signal purity for various choices of second electron selection

	super cluster	+ track (Gsf electron)	+ isolation	+ identification	+ passing trigger
$Z \rightarrow e^+e^-$ signal	**16.4%**	**84.7%**	**96.5%**	**99.1%**	**99.8%**
$W \rightarrow e\nu$	4.9%	0.9%	0.4%	0.1%	0.04%
QCD: b,c → e	**9.8%**	**2.5%**	**0.5%**	**0.3%**	**0.09%**
QCD: EM enriched	**68.2%**	**11.4%**	**2.4%**	**0.4%**	**0.0%**
$Z \rightarrow \tau^+\tau^-$	0.5%	0.3%	0.1%	0.03%	0.0%
ttbar	0.2%	0.2%	0.1%	0.09%	0.08%

Understanding the background shape

- ★ As seen on the previous slides, when the background shape is obtained from a large sample and scaled to 10 pb^{-1} statistics, the shape gets distorted and does not reflect the one expected in 10 pb^{-1} data.
- ★ This causes the statistical uncertainty in the background estimation to be quite unreliable.
- ★ One possible way to estimate the statistical uncertainty in background estimation would be to generate several hundred toy Monte Carlo experiments using the fitted background shape (and sampling the signal events for 10 pb^{-1} luminosity from the large signal sample available) and take the spread in the number of background events as statistical uncertainty.
- ★ Another way to study background uncertainty would be to relax the electron selection requirement. This will increase background statistics for 10 pb^{-1} integrated luminosity.
- ★ We are working on both these approaches.

Estimation of $Z \to e^+e^-$ signal yield for 10 pb^{-1}



$Z \rightarrow e^+e^-$ signal yield and efficiency

Number of $Z \rightarrow e^+e^-$ signal events with 10 pb^{-1}

	super cluster	+ track (Gsf electron)	+ isolation	+ identification	+ passing trigger
$Z \rightarrow e^+e^-$ signal	4013 ± 191	3650 ± 95	3164 ± 86	3328 ± 81	3142 ± 73
background	103079 ± 368	2444 ± 88	1344 ± 75	637 ± 68	456 ± 55

The errors are not quite right yet because of the background shape uncertainty mentioned earlier, but the numbers give a rough estimation and show the availability of the machinery to obtain them.

Status & summary

✓ We have performed a study of the $Z \to e^+e^-$ reconstruction using Summer08 samples.

✓ Working on tag-and-probe efficiency tables by defining the bins in such a way that the double-tag events do not exist. Separate tables by charge.

✓ Study of background shape under way.