CDF/ANA/TOP/CDFR/CDF9894
November , 2008

# Dalitz and Goldstein Method for Measuring the Top Quark Mass In The Di-Lepton Channel

Matthew Hare and Krzysztof Sliwa

Tufts University

Department of Physics and Astronomy
Medford, Massachusetts 02155 USA

ABSTRACT

In this paper, I will describe how the mechanisms by which the Dalitz and Goldstein method for measuring the Top Quark mass work and the techniques used in this analysis.

## Geometrical Construction

The Dalitz and Goldstein[2] method for measuring the Top Quark mass in the Dilepton channel employs a geometrical interpretation of the equations of constraint.

Those equations are:

$$
\begin{aligned}
(l^+ + \nu)^2 &= M_W^2 && (1)\\
(l^- + \bar{\nu})^2 &= M_W^2 && (2)\\
(t - l^+ - b)^2 &= M_\nu^2 = 0 && (3)\\
(\bar{t} - l^- - \bar{b})^2 &= M_\nu^2 = 0 && (4)\\
t^2 = M_t^2 &= M_{\bar{t}}^2 = \bar{t}^2 && (5)\\
-P_x^t &\sim P_x^{\bar{t}} && (6)\\
-P_y^t &\sim P_y^{\bar{t}} && (7)
\end{aligned}
$$

where $t$, $l$, $b$, and $\nu$ are the top quark, lepton, bottom quark, and neutrino 4-momenta. $M_W$, $M_t$, and $M_\nu$ are the masses of the W boson, Top quark, and neutrino. $P_x^t$ and $P_y^t$ are the x and y components of the Top and anti-Top quarks' transverse momenta. Equations 6 and 7 are only approximate and are considered "weak" constraints while the rest are "hard" constraints. Equations 6 and 7 are "weak" due to the possibility that the partons inside of the protons may have some transverse momentum. If this were not the case, then both equations 6 and 7 would be "hard" constraints.

The geometrical construction begins by rewriting equations 1 - 4 in terms of the Top and Bottom quark kinematics.

$$
\begin{aligned}
(\vec{P}_t - \vec{P}_b)^2 &= (E_t - E_b)^2 - M_W^2 \equiv R_W^2 && (8)\\
(\vec{P}_t - \vec{P}_b - \vec{P}_{l^+})^2 &= (E_t - E_b - E_{l^+})^2 \equiv R_\nu^2 && (9)
\end{aligned}
$$

$\vec{P}_t$, $\vec{P}_b$, and $\vec{P}_{l^+}$ are the 3-momenta for the Top, Bottom, and charged lepton; and $E_t$, $E_b$, and $E_{l^+}$ are their corresponding energies. A second pair of equations can be written for the anti-Top quark and its decay products in exactly the same way. These equations have a similar form to the equation for a sphere, in this case in 3-dimensional momentum space.

$$X^2 + Y^2 = R^2 \qquad (10)$$

This suggests that the Top 3-momentum vector $\vec{P}_t$ must lie on the intersection of two spheres of radii $R_W$ and $R_\nu$. The centers of these two spheres are separated by the charged lepton's 3-momentum $\vec{P}_{l^+}$.(Figure 1) The intersection of these two spheres is a circle with radius $r$. If we neglect the lepton masses, the Top Quark energy will be constant on this circle, where $E_0$ is the lowest possible energy for the Top Quark given the kinematic of the event.

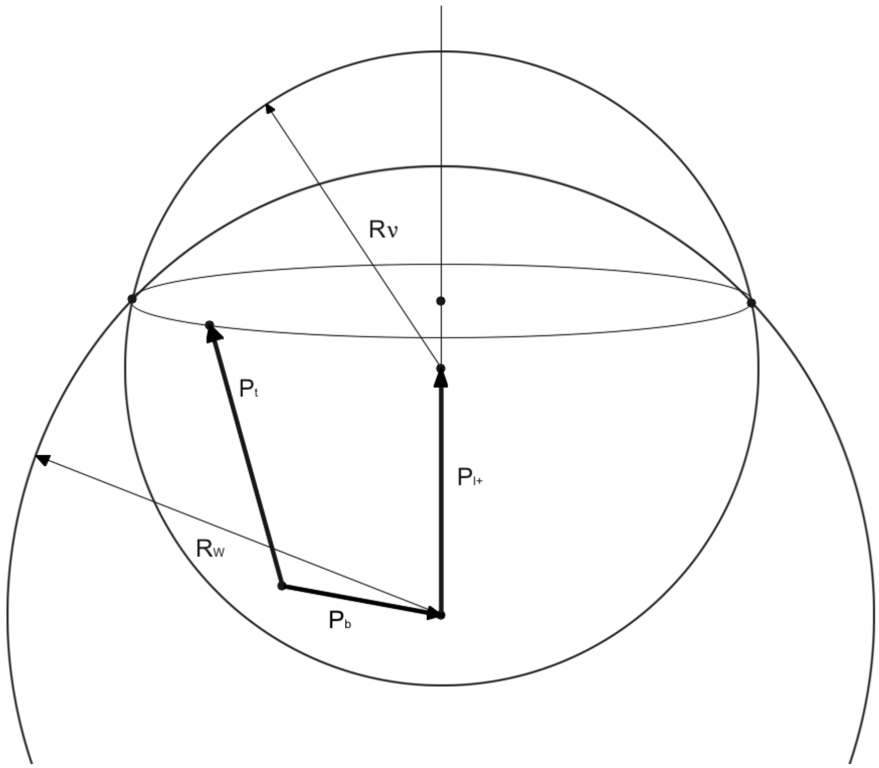


Figure 1: The two spheres of radii $R_W$ and $R_\nu$, whose centers are separated by $\vec{P_t}$, and intersect in a circle.

$$r^2 = \frac{M_W^2}{\mid \vec{P_{l^+}} \mid}(E_t - E_o) \tag{11}$$

$$E_o = E_b - E_l + \frac{M_W^2}{4E_l} \tag{12}$$

Given a different Top Quark 3-momentum $\vec{P_t}$, a different pair of spheres can be constructed. It is possible to choose a $\vec{P_t}$ where the two spheres intersect only at a point, which would correspond to a Top Quark energy equal to $E_0$. For each pair of spheres that intersect, a new circle is created which will correspond to a different Top Quark energy $E_t$. The Top Quark energy will increase in the direction of $\vec{P_{l^+}}$. These circles form the surface of a paraboloid. While the Top Quark energy is constant on these circles, the Top Quark mass , $M_t$, is not; however, if the mass is fixed the Top Quark 3-momentum vector will be confined to a conic section of the paraboloid which is, by definition, an ellipse. The orientation and eccentricity of the ellipse will depend on the assumed mass and the 4-momenta of the leptons and Bottom Quarks. A similar ellipse can be constructed in the same way for the anti-Top Quark.

Once both ellipses are constructed, they are projected into the x-y plane of the detector, one of which is reflected about the origin. Using the "weak" constraints, the problem becomes fully

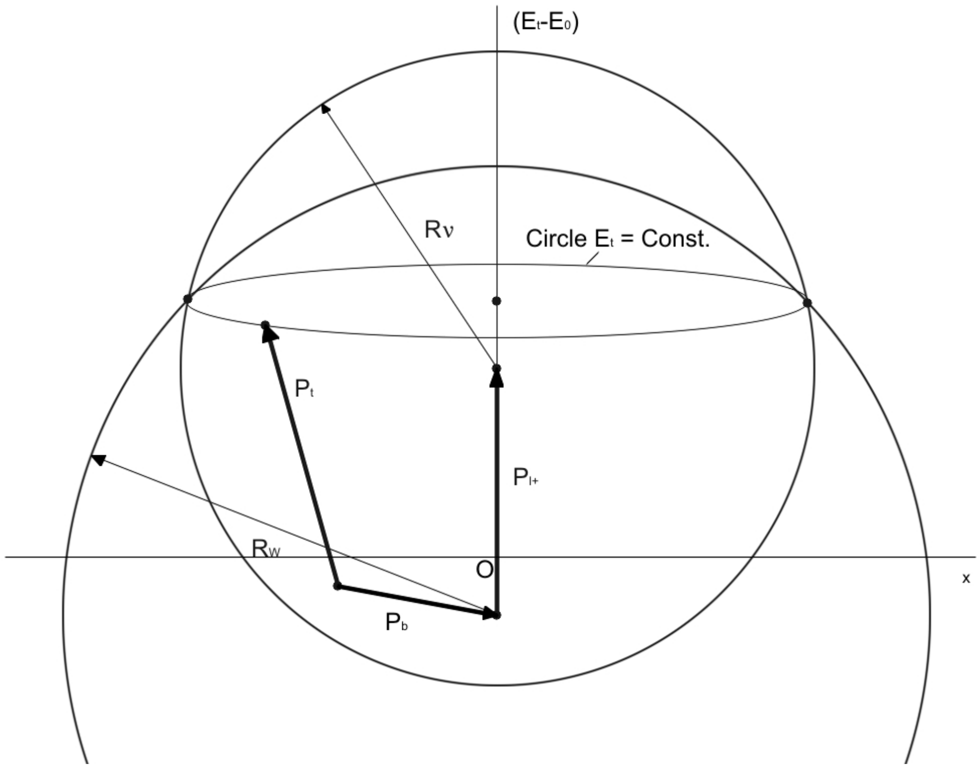


Figure 2: The two spheres of radii $R_W$ and $R_\nu$, showing a circle of intersection with constant energy $E_t$. O is the point where $E_t - E_0 = 0$, the minimum energy allowed for the Top Quark as determined by the kinematics of the event.

constrained where each pair of points, one from each ellipse, corresponds to a possible solution consistent with the assumed top mass and the measured lepton and jet momenta. If these final constraints were "hard", then only the points of intersection of the two ellipses would need to be considered as possible solutions. It is because they are "weak" that every pair of points needs to be considered. Given a pair of points from the projection of the two ellipses, the transverse momentum, $Pt_{t\bar{t}}$, of the Top-anti-Top system can be calculated. Each pair of points is weighted by a likelihood factor, $P(Pt_{t\bar{t}})$, from the $Pt_{t\bar{t}}$ spectrum. The expected shape of the $Pt_{t\bar{t}}$ distribution is determined from Monte Carlo simulation. If a different top mass is assumed, a different pair of ellipses will be created whose projections into the x-y plane will give another set of possible solutions.

### The Likelihood

To determine the most likely mass of the Top Quark a probability distribution is constructed for each combination of leptons and jets in an event. Given an assumed top quark mass, a likelihood value $L_i$ is projected onto the $M_t$-axis of this distribution. The most likely top quark mass corresponds to the peak of this distribution. The likelihood values, $L_i$, are a product of six probability factors.

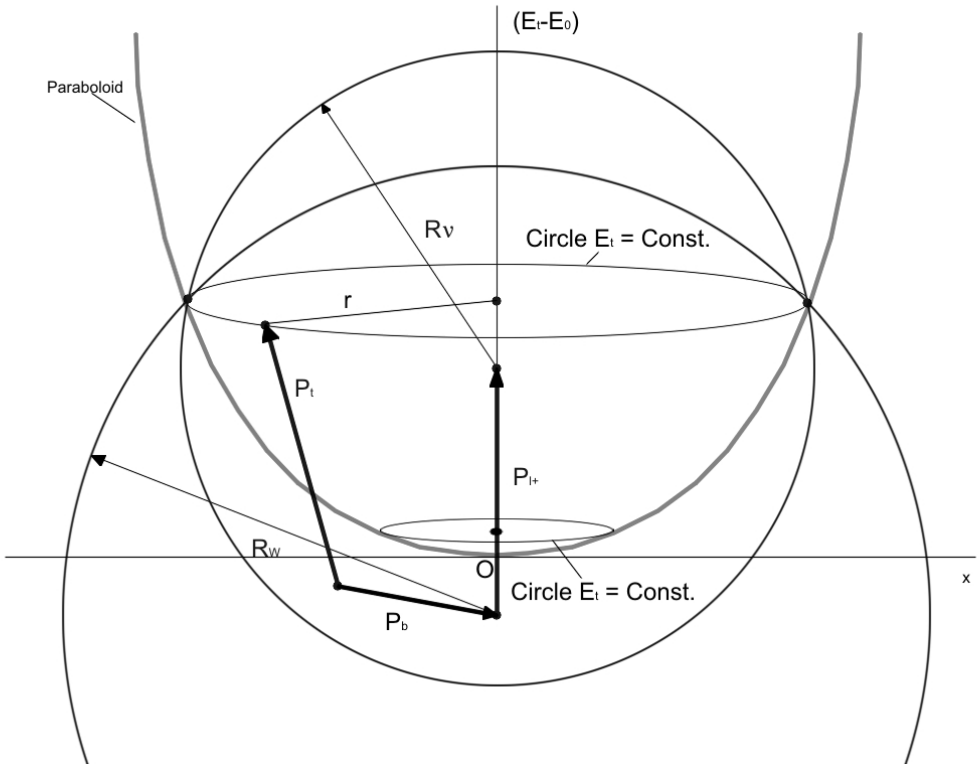


Figure 3: The two spheres of radii $R_W$ and $R_\nu$, showing a circle of intersection with constant energy $E_t$. O is the point where $E_t - E_0 = 0$, the minimum energy allowed for the Top Quark as determined by the kinematics of the event. Also seen here is a second circle of constant $E_t$ corresponding to a different $P_t$ (not drawn) resulting from the intersection of two other spheres (also not drawn). Both circles begin to form the paraboloid surface.

$$L_i = P(P_{t\bar{t}}) \times G(b) \times G(\bar{b}) \times P(x_1, x_2) \times P(l^+) \times P(l^-) \quad (13)$$

$P(Pt_{t\bar{t}})$ is the factor related to the transverse momentum of the Top-anti-Top system

$G(b)$ and $G(\bar{b})$ are the factors related to Jet Energy Smearing

$P(x_1, x_2)$ is the factor related to the Structure Functions

$P(l^+)$ and $P(l^-)$ are the factors related to V-A Calculations

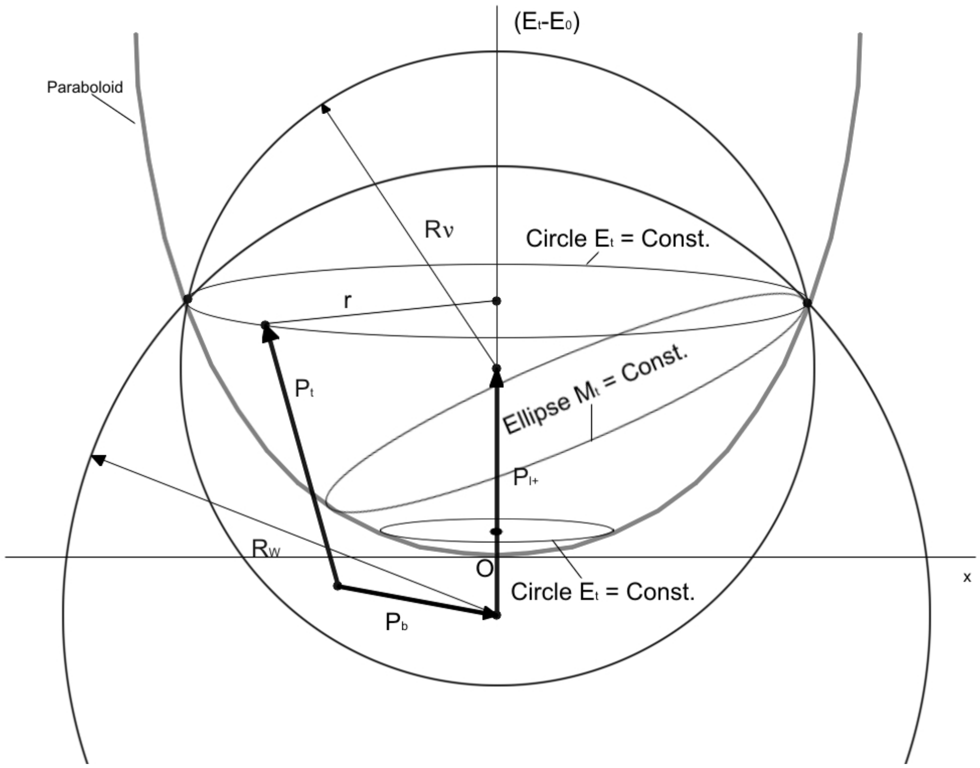


Figure 4: The two spheres of radii $R_W$ and $R_\nu$ showing a circle of intersection with constant energy $E_t$. O is the point where $E_t - E_0 = 0$, the minimum energy allowed for the Top Quark as determined by the kinematics of the event. Also seen here is a second circle of constant $E_t$ corresponding to a different $P_t$ (not drawn) resulting from the intersection of two other spheres (also not drawn). Both circles begin to form the paraboloid surface. Assuming a constant $M_t$ confines $P_t$ to a conic section of the paraboloid which is an ellipse.

**i.** $P(Pt_{t\bar{t}})$

See Geometrical Construction.

**ii.** $P(x_1, x_2)$

A relative likelihood factor is assigned that describes the level of agreement between the Feynman-x values, $x_1$ and $x_2$ , that are calculated from the event and those predicted by theory, i.e. the structure functions for the event,

$$P_{x_1,x_2} = \frac{\sum_{i=qq,gg} F_i(x_1)F_i(x_2)\frac{d\sigma}{d\hat{t}}(\hat{s},\hat{t})_i}{\sum_{i=qq,gg}\frac{d\sigma}{d\hat{t}}(\hat{s},\hat{t})_i} \quad (14)$$

$$x_{1,2} = (E_t + E_{\bar{t}} \pm (t_L + \bar{t}_L))/2P \quad (15)$$

$$\hat{s} = x_1 x_2 s \quad (16)$$

$$\hat{t} = M_t^2 - x_1\sqrt{s}(E_t - t_L) \quad (17)$$

where $i$ labels the $q\bar{q}$ and $gg$ processes; $F_i$ are the structure functions; $\hat{s}$ is the center-of-mass energy; $\hat{t}$ is the momentum transfer of the Top-anti-Top quark production subprocess; $P$ is the proton momentum; $s$ is the square of the proton-anti-proton system in the center-of-mass frame; and $t_L$ is the longitudinal momentum of the Top Quark in the lab frame of the proton-anti-proton system.

**iii.** $P(l^+)$ and $P(l^-)$

For each point a likelihood factor is calculated to describe the agreement of the charged lepton energies calculated in the Top Quark rest frame with the values predicted by V-A calculations.

$$dP(E_l) = (24/M_t^2)E_l(1 - 2E_l/M_t)dE_l \tag{18}$$

**iv.** $G(b)$ and $G(\bar{b})$

Since the b-jet and b-quark measured energies have large errors, the true energies can differ from the measured values. A range of energies is defined which is centered about the measured energy of each of the jets in an event. The range is chosen to be $3\sigma$, where $\sigma$ is the width of jet energy resolution distribution. A probability, G(b), is assigned for each point within the $3\sigma$ range. Points that correspond to b-jet energies that deviate from the measured value will be downgraded by a Gaussian probability factor giving them a lower probability than ones closer to the measured value. This smearing of the jet energies will give a family of ellipses for each jet.

Pairing together all the combinations of ellipses from each family will create a 2-dimensional grid where the axes are indexed by the smeared jet energies. Each point on the grid will have a value which is the $L_i$ that corresponds to smeared jets momenta and the given lepton momenta and assumed top mass. There will be a different grid for each assumed top mass. All together, this will create a 3-dimensional space where the third axis is the assumed top mass. Summing over the assumed top mass axis of this space will reduce it back to a 2-dimensional grid with the smeared jet energies as the remaining axes, however, now the value at each point is a total likelihood. The point on the grid with the greatest total likelihood will be chosen as the most probable solution given this combination of leptons and jets.

It should be noted that this is an important difference in methodology between this analysis and the Run 1 analysis done by Kristo Karr[1]. In that analysis the 2-D smeared jet energy grid was summed over to get the total likelihood for the event instead of picking the best smeared jet combination.

This total likelihood and its corresponding probability distribution will be compared to other combinations of leptons and jets (Bottom Quarks) from the same event where the greater one is favored.

### MET Probability

There is an additional probability calculated for each combination based on the measured missing transverse energy. During the analysis of each combination, the energy and momentum of the neutrinos are obtained. The difference between the two missing neutrinos' x and y momenta in the transverse plane and the missing transverse energy's x and y components are compared and assigned a probability based on a Gaussian shape.

### The Joint Likelihood

Once a combination from an event is chosen via the method described above, and all events have been analyzed, a joint likelihood is created. The joint likelihood is the product of the probability distributions from the chosen combinations from each event. The shape of the probability distribution for each event will not necessarily be the same. Some will be asymmetrical and some may have more than one peak. Choosing the mean or the peak value as the top mass for an individual event may include biases from the shape of the distribution. By taking the product of the event distributions, the joint distribution becomes more Gaussian in shape. Multiple peaks and asymmetries in the individual events are eliminated in the joint distribution as are the possible biases that these characteristics may produce. Since the true top mass for all of the events is the same, the joint probability distribution reveals where all the events are consistent with a given $M_t$, which should point toward the true mass. The arithmetic mean of the joint distribution is the Top Quark Mass.

### The Construction of the Pseudo Experiments

The real data and the Monte Carlo data are treated in exactly the same way in every detail of this analysis except for how the joint likelihood is constructed. When analyzing real data, all of the candidate events take part in the product that forms the Joint Likelihood. When analyzing the Monte Carlo data, all of the events that pass cuts and have a solution are put into a pool. Next, events are randomly selected from the pool and multiplied together to produce the joint likelihood, which forms one pseudo experiment. The number of events selected from the pool to form one pseudo experiment depends on the number of candidate events expected in the real data. The number of events is randomly generated based on a Poisson distribution whose mean is the number of candidate events. Care is taken to not allow an event to appear more than once in the same pseudo experiment, however, all events in the pool are available to every pseudo experiment. The mean and RMS from each pseudo experiment are used to calculate the delta and the pull. The mean, delta, pull, and RMS are each put into histograms to produce their own distributions.

### The RMS Correction

When building the pseudo experiments using the joint likelihood method described above, it becomes necessary to make a correction to the RMS in order to relate it to the errors. If a simple distribution populated by the means of the events in a pseudo experiment was used to make the measurement of the mass, the error on that measurement would be related to the RMS of the distribution by $1/\sqrt{N}$. This is, however, not the case for a joint likelihood distribution. A joint likelihood distribution is too narrow and a correction factor is needed.

In order to explain how the correction factor is calculated, a simple model is needed. The model is a box filled with cards, and on each card is a number. The number of cards in the box is $N$. The sum of all the numbers on the card is $S_{box}$, the average of all the numbers is $A_{box}$ and the standard deviation of the numbers in the box is $\sigma_{box}$. Now, $n$ cards are draw from the box at random without replacement. The expectation value for the sum of the draws, $S_{draw}^{EV}$, is $n * A_{box}$ and the standard error on the sum of the draws from the box is $\sqrt{n} * \sigma_{box}$. The expectation value for the average of the draws, $A_{draw}^{EV}$, is $S_{draw}^{EV}/n$ which is simply $A_{box}$ and the standard error for the average of the draws is $\sqrt{n} * \sigma_{box}/n$ which simplifies to $\sigma_{box}/\sqrt{n}$.

The pool of events in this analysis is very similar to the box of cards. Each event is a card and the number on the card is the mean of the distribution that corresponds to the event. All of the same statistics that applied to the box of cards will apply to the pool of events in exactly the same way. However, in this analysis, the average is not taken, but instead a joint distribution is made. While at first glance this seems to be very different, making a joint distribution out of the individual mass distributions for each event is like taking the average of the numbers on the cards. So, the standard error for a joint distribution should be similar to that of the random draws from a box and the RMS of a joint distribution is related to that error.

$$RMS_{JD} \propto \sigma_{pool}/\sqrt{n} \tag{19}$$

If all possible combinations of $n$ events were drawn from the pool and the mean of each sample was put into a histogram, that histogram would be a total probability distribution. It would be normal in shape even if the parent distribution is not. Its mean would be similar to the mean of the parent distribution and its width would be the error on the mean of the sample distributions. Since the number of possible combinations is an extremely large number, it is not practical to compute every possible combination. Instead, it is sufficient to draw $n$ events at random from the pool $X$ number of times as long as $X$ is large. When this is done, the resulting histogram will not be a complete total probability histogram, however its mean and width, $\sigma_{XPE}$, will be the same as if all possible combinations were sampled. Since the widths of these distributions are the same, $\sigma_{XPE}$ describes the error on the PEs.

$$\sigma_{XPE} \approx \sigma_{pool}/\sqrt{n} \tag{20}$$

It follows that

$$RMS_{JD} \propto \sigma_{XPE} \tag{21}$$

This gives the correction factor, $C$ to the RMS of the joint distributions.

$$C = \sigma_{XPE}/RMS_{JD} \tag{22}$$

and

$$RMS_{corr} = C \times RMS_{raw} \tag{23}$$

where $RMS_{corr}$ and $RMS_{raw}$ are the corrected and uncorrected widths of the individual joint distributions of each PE.

### The Mapping Function

Analyses of a range of Top Mass MC samples are used to study the correlation between the generated and reconstructed mass. Each mass MC sample has its own mean, delta, pull, and RMS distribution. Given the mean distribution for each mass MC sample, a plot of reconstructed mass as a function of MC mass is made and to it a line is fit. It is possible to fit other shapes to the points, however a line does the job very well. That fit is the mapping function that will be used to make the final correction to the measurement.

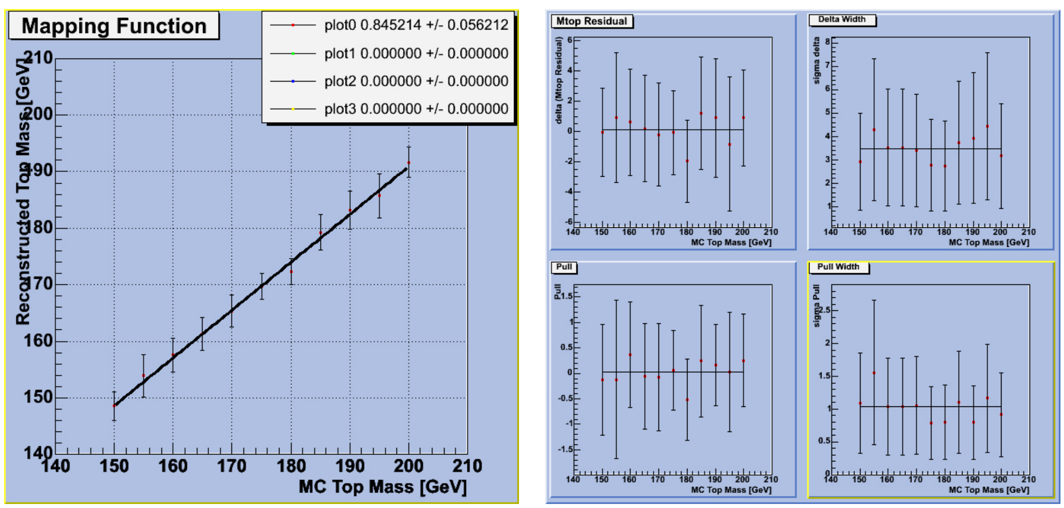


Figure 5: Mapping function from signal only MC to be used as a correction on the real data (Left). Delta, Delta Width, Pull, and Pull Widths for the signal MC data samples after corrections has been applied from the mapping function (Right).

### Additional Types of Combination Selection

In addition to selecting a combination of leptons and jets according to their relative probabilities, three other methods are also studied. The first method is to combine both the most likely combination with its "twin". The "twin" is simply the combination where the jets-leptons assignment is switched. Both probability distributions are added together for each event before the pseudo experiments are constructed. The second method is to select only the combinations in an event that have both of the two leading jets, the jets with the highest transverse energy. After the two combinations are found, the one with the highest probability is chosen. The final method differs from the second in that the two combinations that contain the two leading jets are added together, just like in the first additional method.

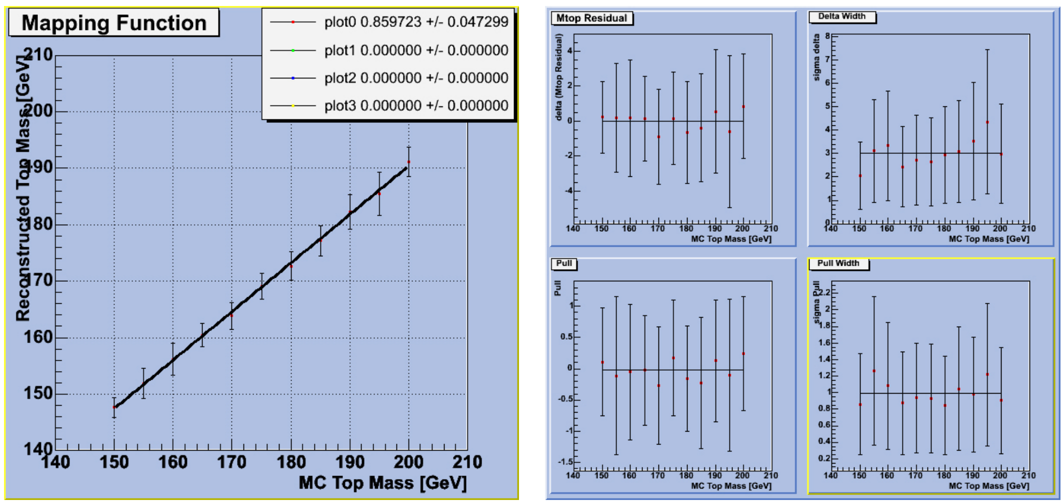


Figure 6: Mapping functions, Deltas, Delta Widths, Pulls, and Pull Widths for the signal MC data samples where the most probable combination is added to its "twin".

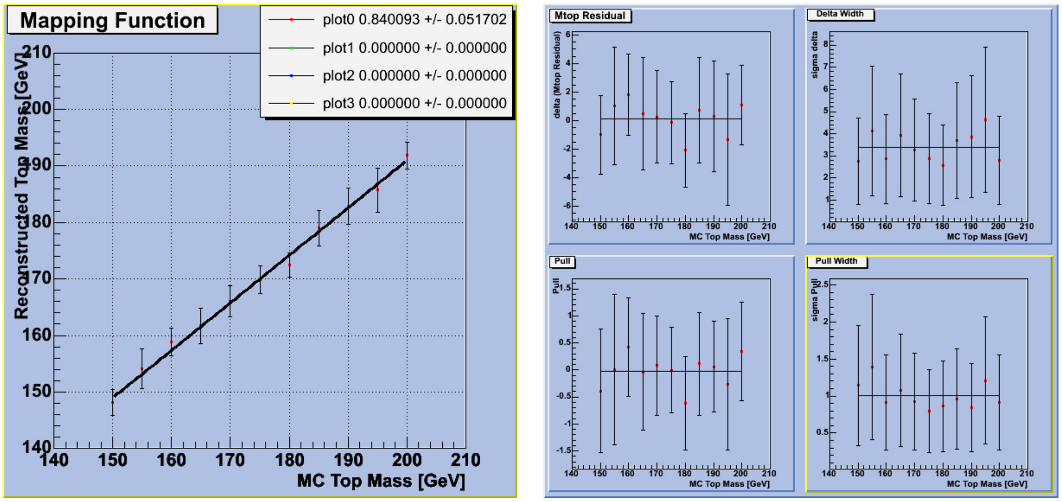


Figure 7: Mapping functions, Deltas, Delta Widths, Pulls, and Pull Widths for the signal MC data samples where the most probable combination is selected out of the combination which contain both of the leading jets.

# References


[1] Kristo M. Karr. *Measurement of the Top Quark Mass by Application of the Dalitz-Goldstein Method to Dilepton Events*. PhD thesis, Tufts University, 1999.

[2] R.H.Dalitz and G. Goldstein. Decay and Polarization Properties of the Top Quark. *Phys. Rev. D*, 45:1531, 1992.

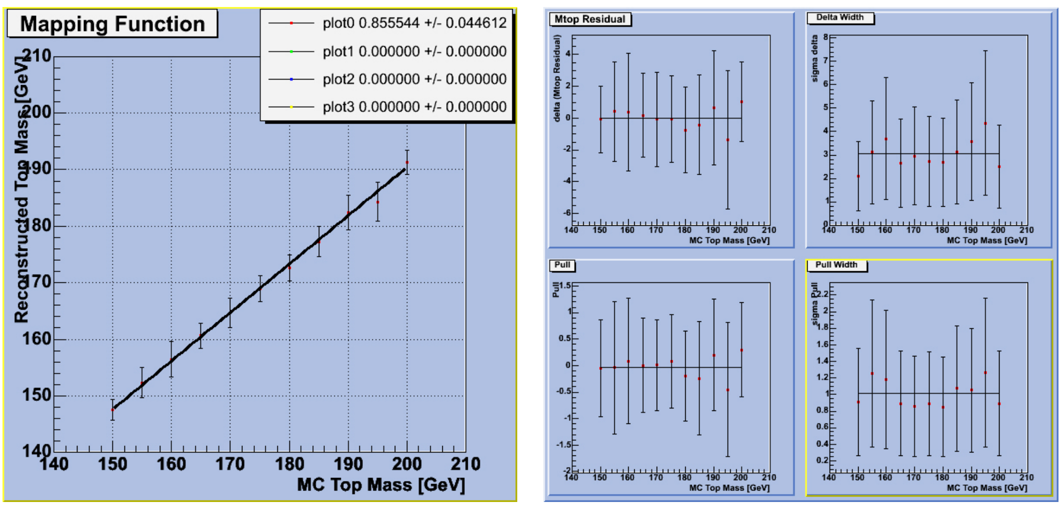


Figure 8: Mapping functions, Deltas, Delta Widths, Pulls, and Pull Widths for the signal MC data samples where the most probable combination is selected out of the combination which contain both of the leading jets and is added to its "twin".