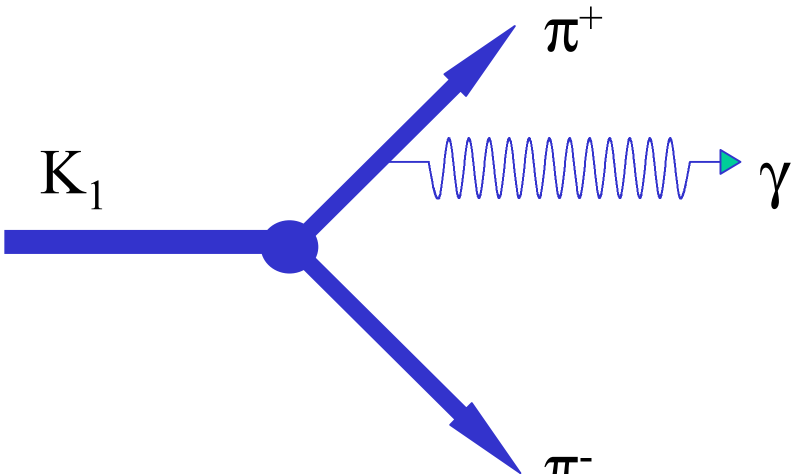
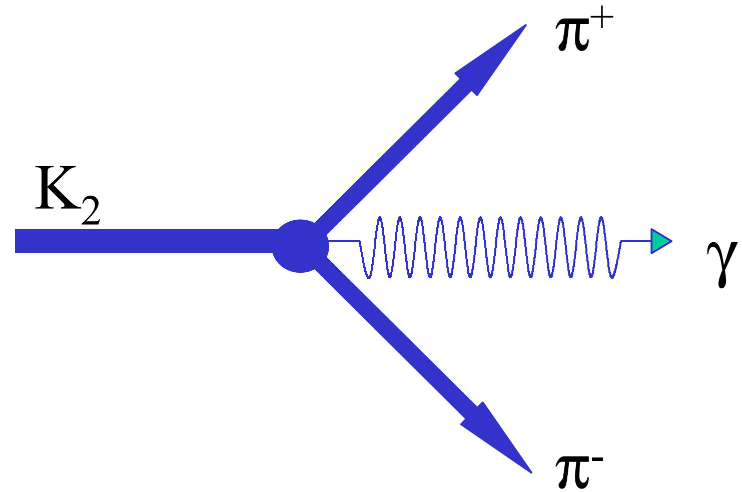
Progress on the $K_L \rightarrow \pi^+\pi^-\gamma$ Analysis

John Shields
University of Virginia
KTeV Collaboration
August 2, 2003

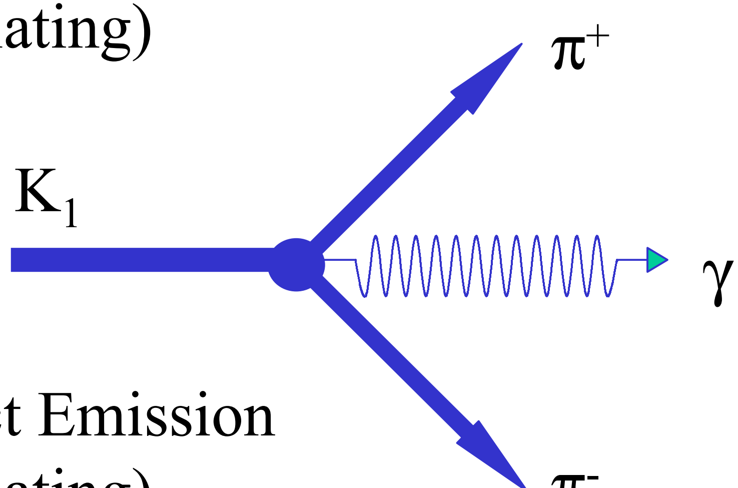
$K_L \rightarrow \pi^+\pi^-\gamma$ Decay



Inner Bremsstrahlung
(CP Violating)



M1 Direct Emission
(CP conserving)



E1 Direct Emission
(CP Violating)

$K_L \to \pi^+\pi^-\gamma$ Differential Decay Rate, Sehgal Model

$$\frac{d\Gamma}{d\omega d\cos\theta} = (const)\left[|\xi_E|^2 + |\xi_M|^2\right]\omega^3\beta^3\left(1 - \frac{2\omega}{M_K}\right)\sin^2\theta$$

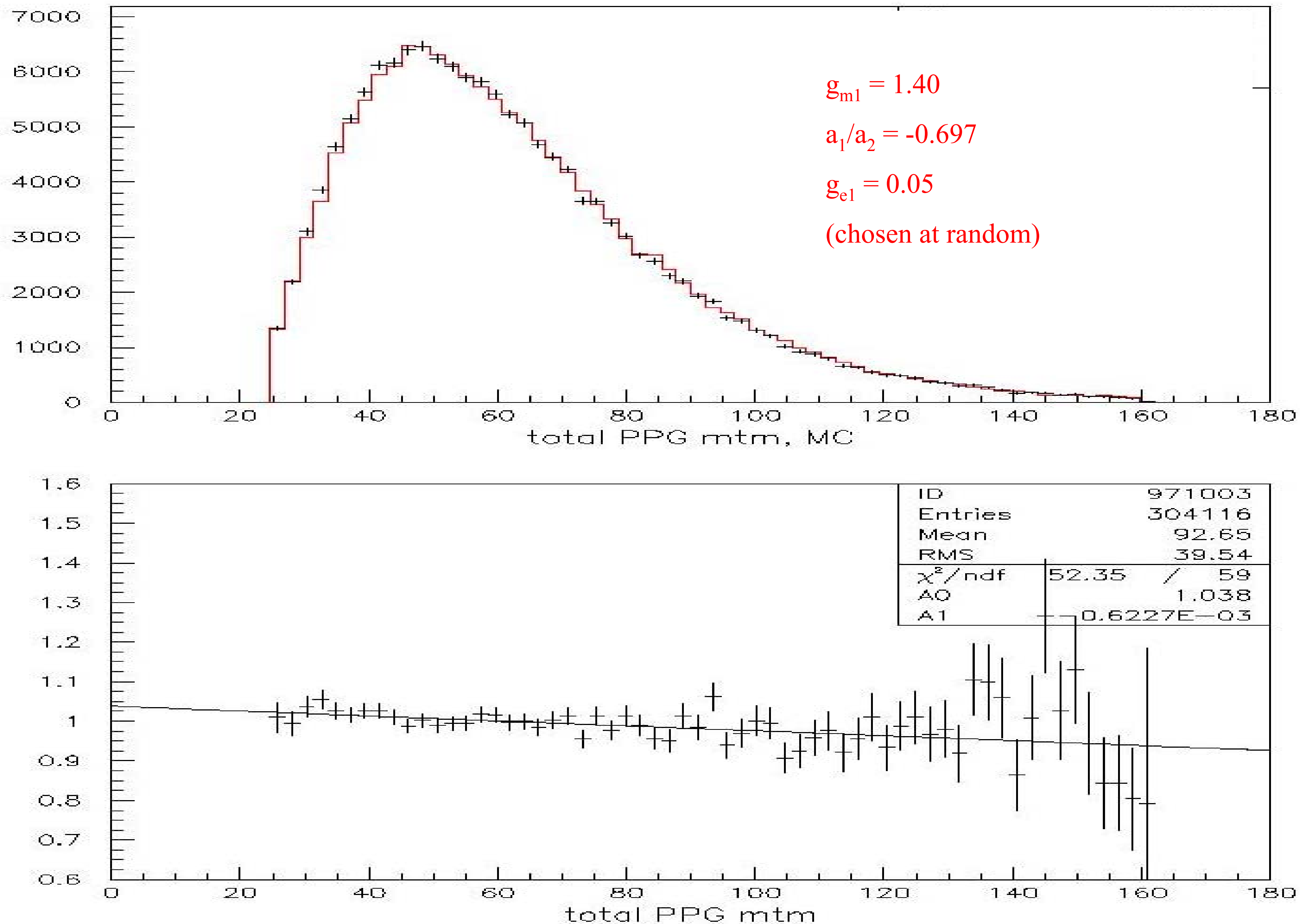
where:

$$|\xi_E| = C_2\left|\frac{\eta_{+-}}{\omega^2\left(1-\beta^2\cos^2\theta\right)}\right| + g_{e1}$$

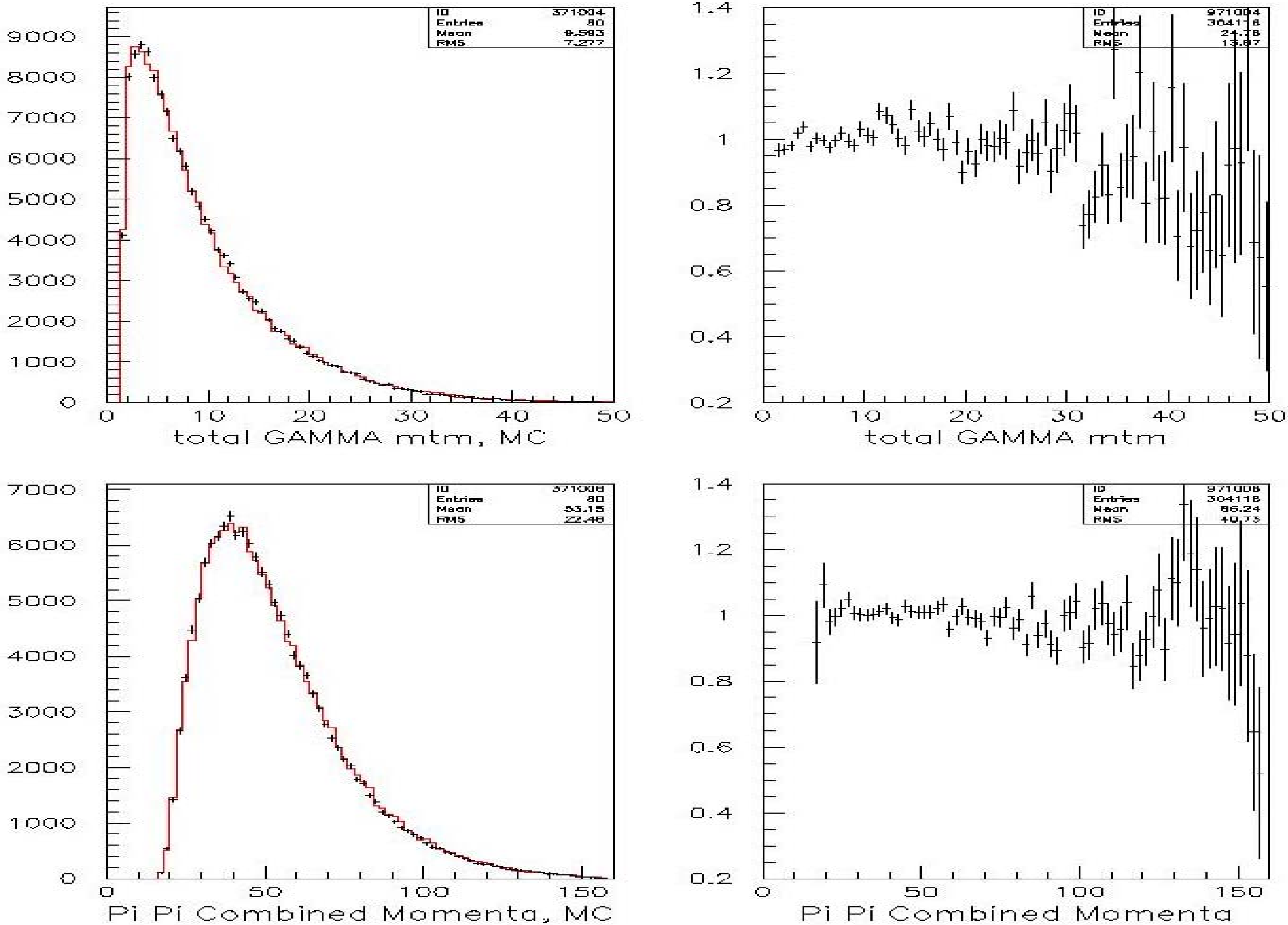
$$|\xi_M| = \tilde{g}_{m1}\left[1 + \frac{a_1/a_2}{\left(M^2{}_\rho - M^2{}_K\right) + 2M_K E^*{}_\gamma}\right]$$

Tasks Completed since Jan, 2003 meeting

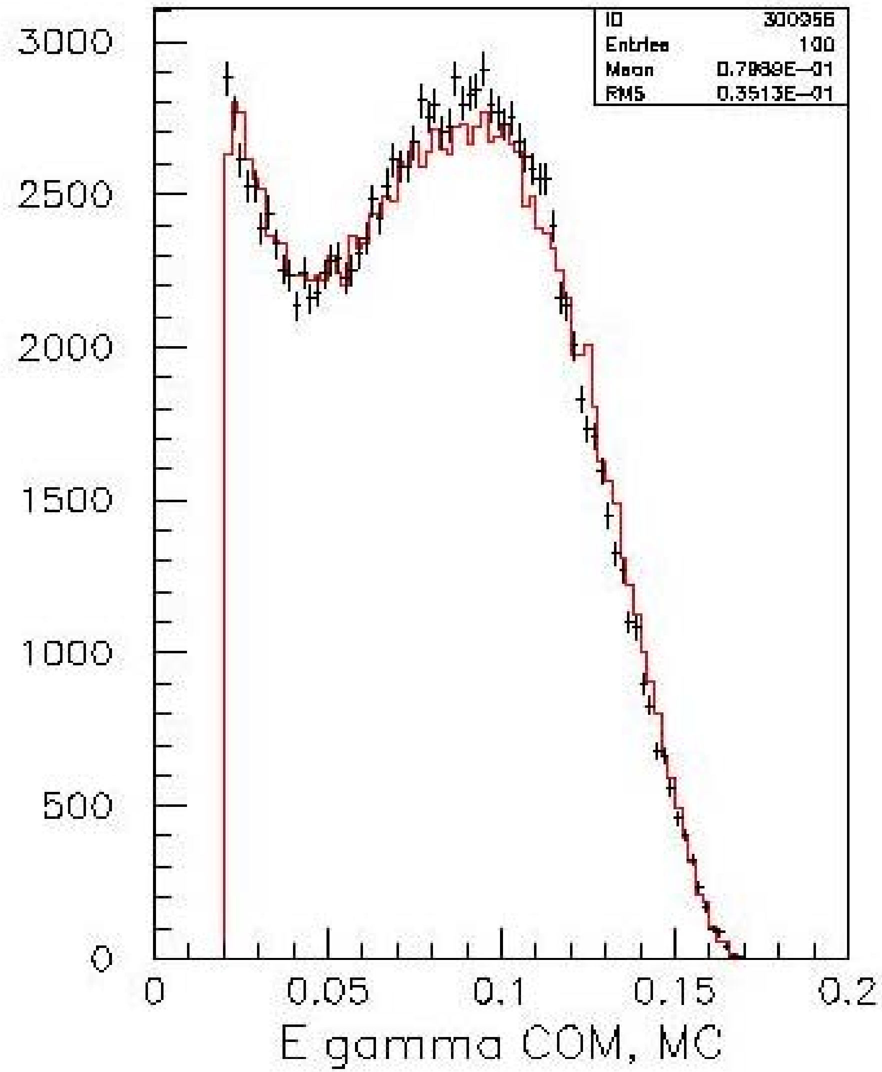
- Upgraded MC and analysis code from Ktevana version v6_00 to v6_01
- Applied RK's L3 filter code correction to the Monte Carlo
  - (accounts for a change in B02 trigger PP0KINE cut during 97 run)
  - (see RK's Jan 3, 2003 write-up for details)
- Added E1 contribution to Monte Carlo matrix element
- Finished implementing a new, more CPU-friendly version of likelihood fitting code **and** tested it extensively on data-sized Monte Carlos with known parameter values



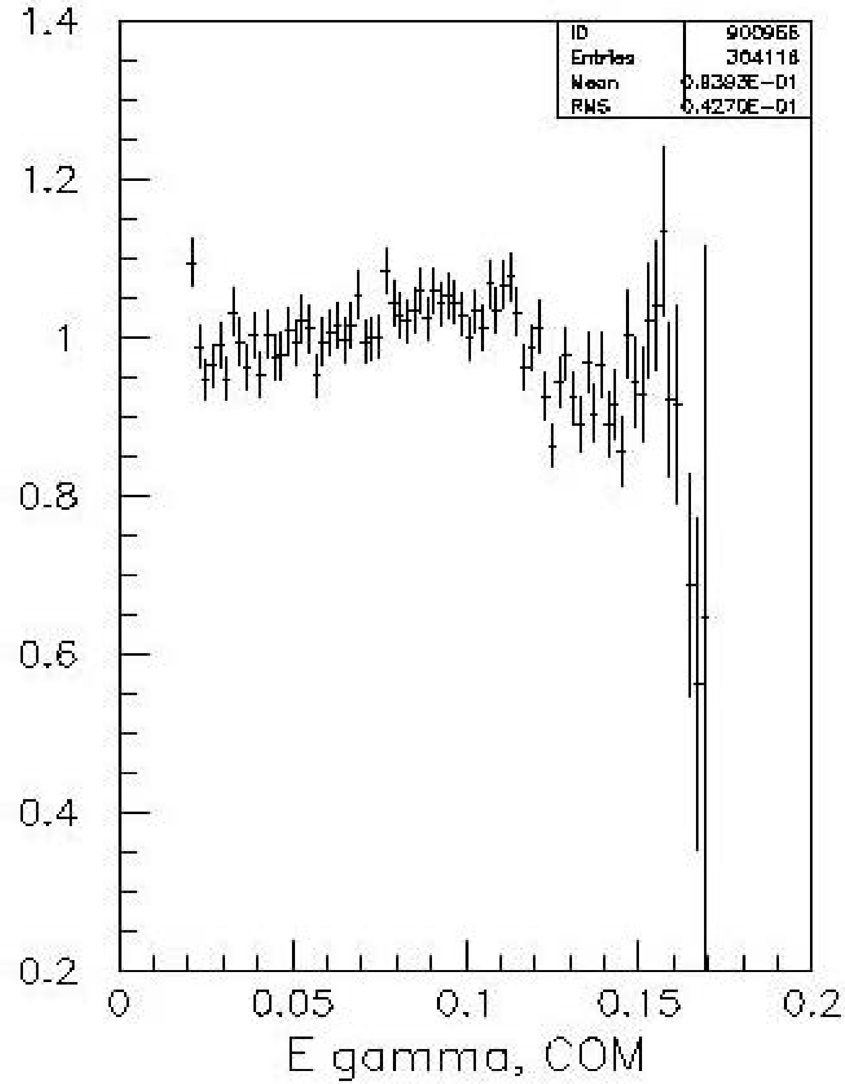
7000
6000
5000
4000
3000
2000
1000
0
0 20 40 60 80 100 120 140 160 180
total PPG mtm, MC
g_{m1} = 1.40
a_1/a_2 = -0.697
g_{e1} = 0.05
(chosen at random)
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0 20 40 60 80 100 120 140 160 180
total PPG mtm
ID 971003
Entries 304116
Mean 92.65
RMS 39.54
χ^2/ndf 52.35 / 59
A0 1.038
A1 -0.6227E-03



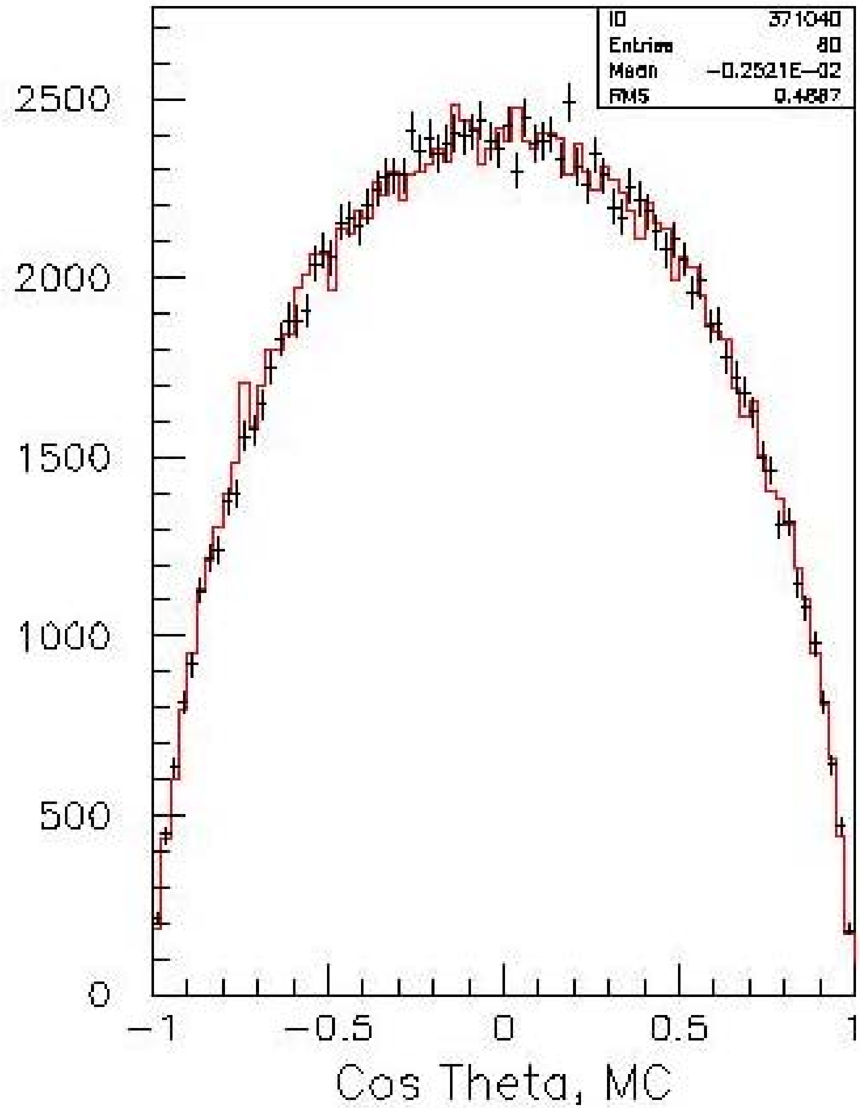
ID 371004
Entries 80
Mean 9.583
RMS 7.277
total GAMMA mtm, MC
ID 971004
total GAMMA mtm
ID 371008
Entries 80
Mean 53.15
RMS 22.48
Pi Pi Combined Momenta, MC
ID 971008
Pi Pi Combined Momenta

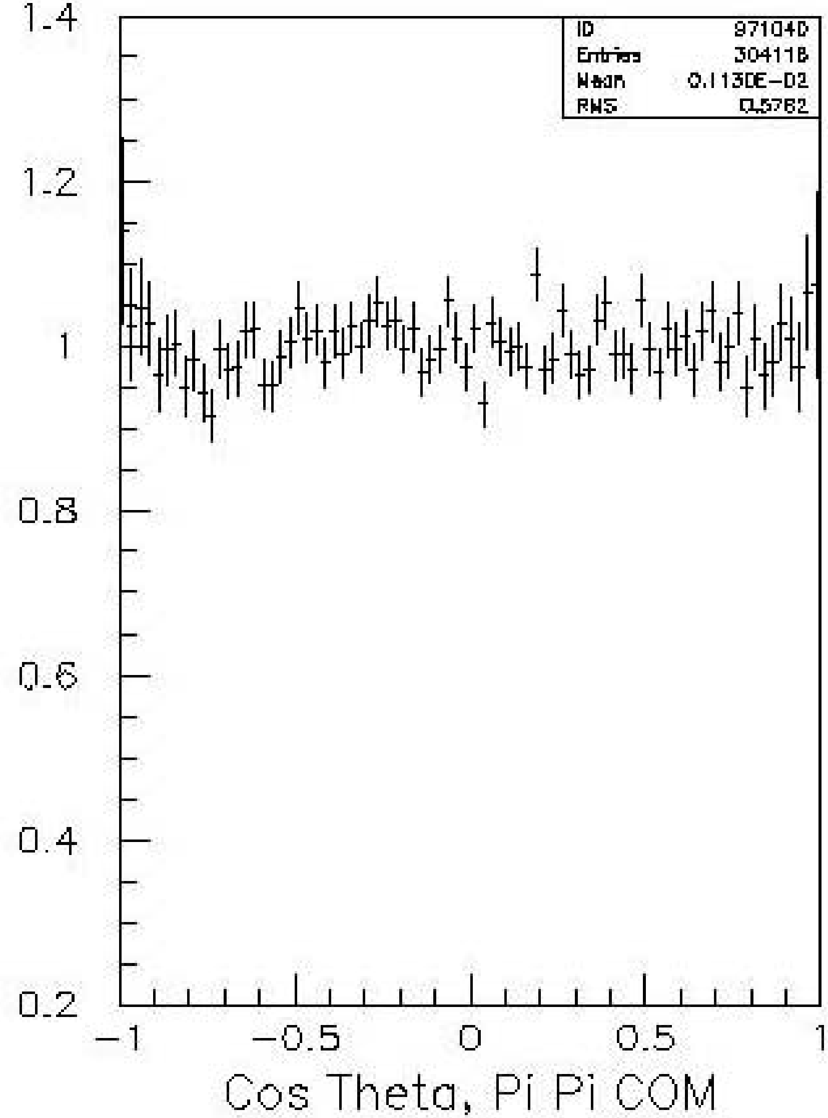


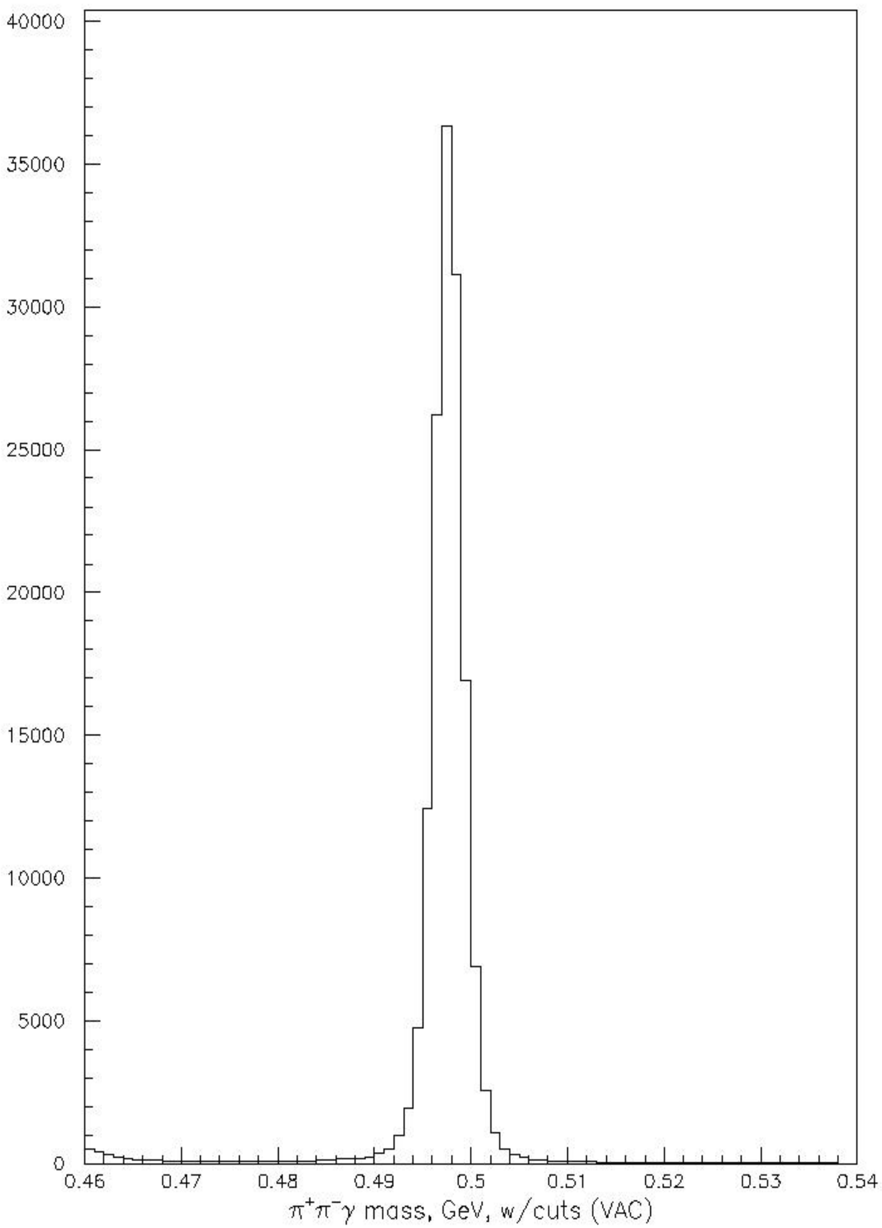
ID 300356
Entries 100
Mean 0.7989E-01
RMS 0.3513E-01
3000
2500
2000
1500
1000
500
0
0
0.05
0.1
0.15
0.2
E gamma COM, MC

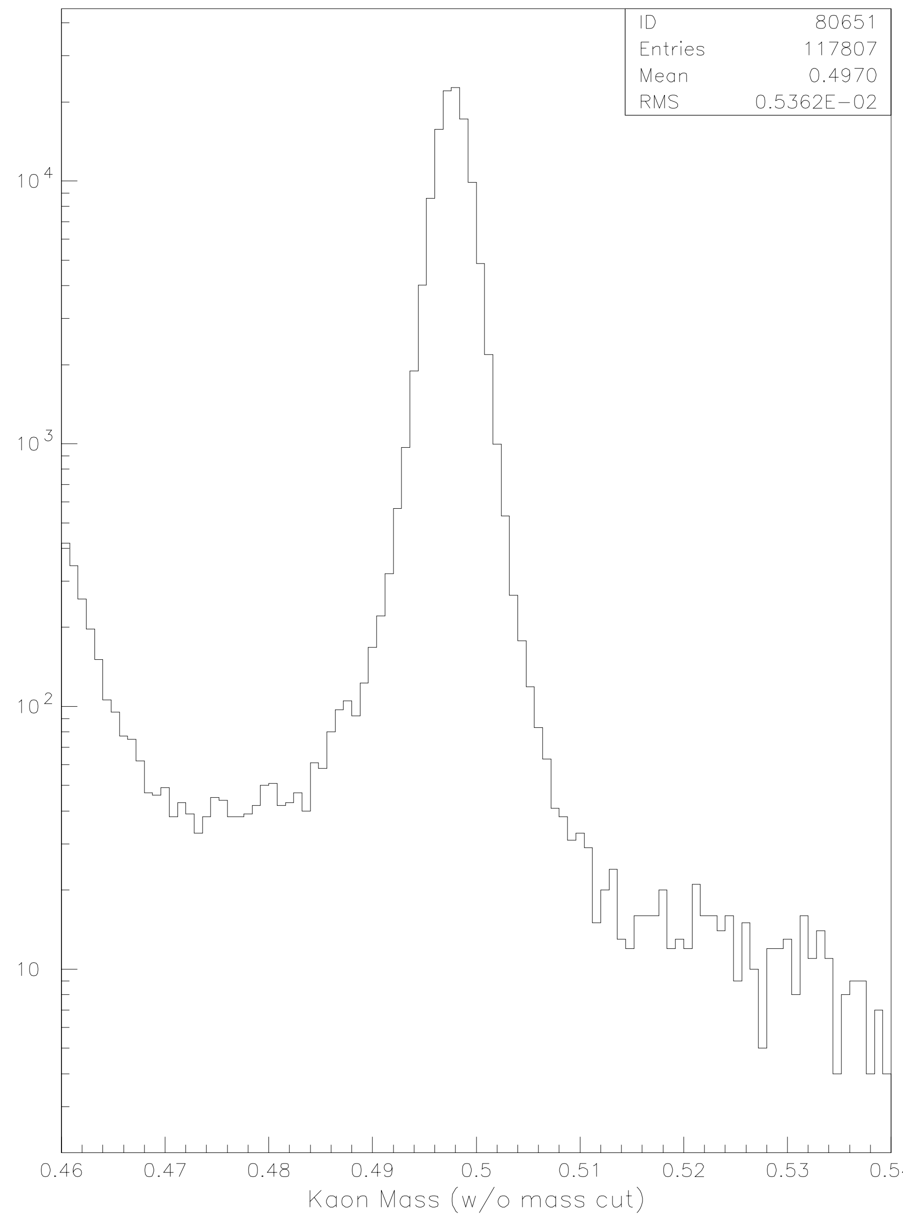


ID 900966
Entries 304118
Mean 0.8393E-01
RMS 0.4270E-01
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.05
0.1
0.15
0.2
E gamma, COM


ID 371040
Entries 80
Mean -0.2521E-02
RMS 0.4887
Cos Theta, MC


ID 971040
Entries 304118
Mean 0.1130E-02
RMS 0.5762
Cos Theta, Pi Pi COM


40000
35000
30000
25000
20000
15000
10000
5000
0
0.46 0.47 0.48 0.49 0.5 0.51 0.52 0.53 0.54
$\pi^+\pi^-\gamma$ mass, GeV, w/cuts (VAC)


ID 80651
Entries 117807
Mean 0.4970
RMS 0.5362E-02
10^4
10^3
10^2
10
0.46 0.47 0.48 0.49 0.5 0.51 0.52 0.53 0.54
Kaon Mass (w/o mass cut)

Current Projects

- Statistical Error Analysis:

  → determining the size of "Big Monte Carlo" (BMC) sample required to make $\sigma_{BMC} \ll \sigma_{Ndata}$ where:

  BMC ≡ Big MC used by the likelihood fitting code for reweighting

  σ_{BMC} ≡ STAT error due to BMC sample size (only)

  σ_{Ndata} ≡ STAT error due to DATA size (only)

  – To do this, σ_{BMC} must be determined as a function of "Q",

    where $Q \equiv N_{BMC} / N_{DATA}$

Current Projects (cont.)

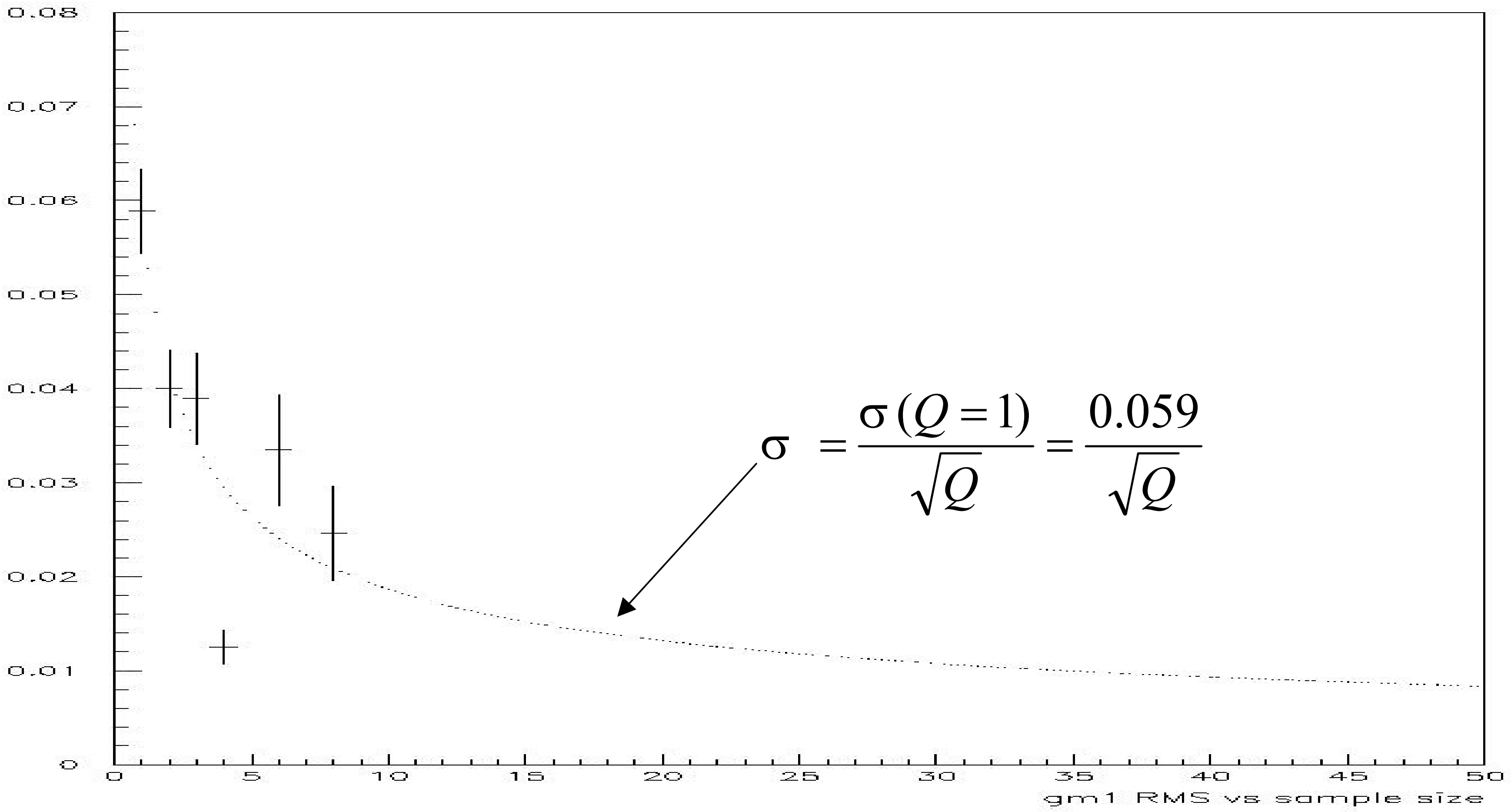
- So, in other words, we need to know what value of $Q = N_{data}/N_{BMC}$ results in $\sigma_{BMC} << \sigma_{Ndata}$? (Q =10?, Q=100?, Q =1000??)

- In Theory:

$$\sigma_{BMC}(Q) = \frac{\sigma_{BMC}(Q=1)}{\sqrt{Q}}$$

- In Practice: preliminary studies do indeed very *roughly* indicate $\frac{1}{\sqrt{Q}}$ dependence
  - Caveats:
    - Studies conducted with very limited number of MC samples (and hence over a very small range of "Q")
    - Further studies with a MUCH greater number of data-sized samples is required to reliably extend the curve for Q >> 1

σ_{BMC} vs. Q (for g_{m1} parameter and limited samples)



PLOT: Distribution of likelihood fits of the SAME "data" sample by different Q=2 "Big Monte Carlos"

•Why are we so concerned about the **exact** Shape of the σ_{BMC} curve?

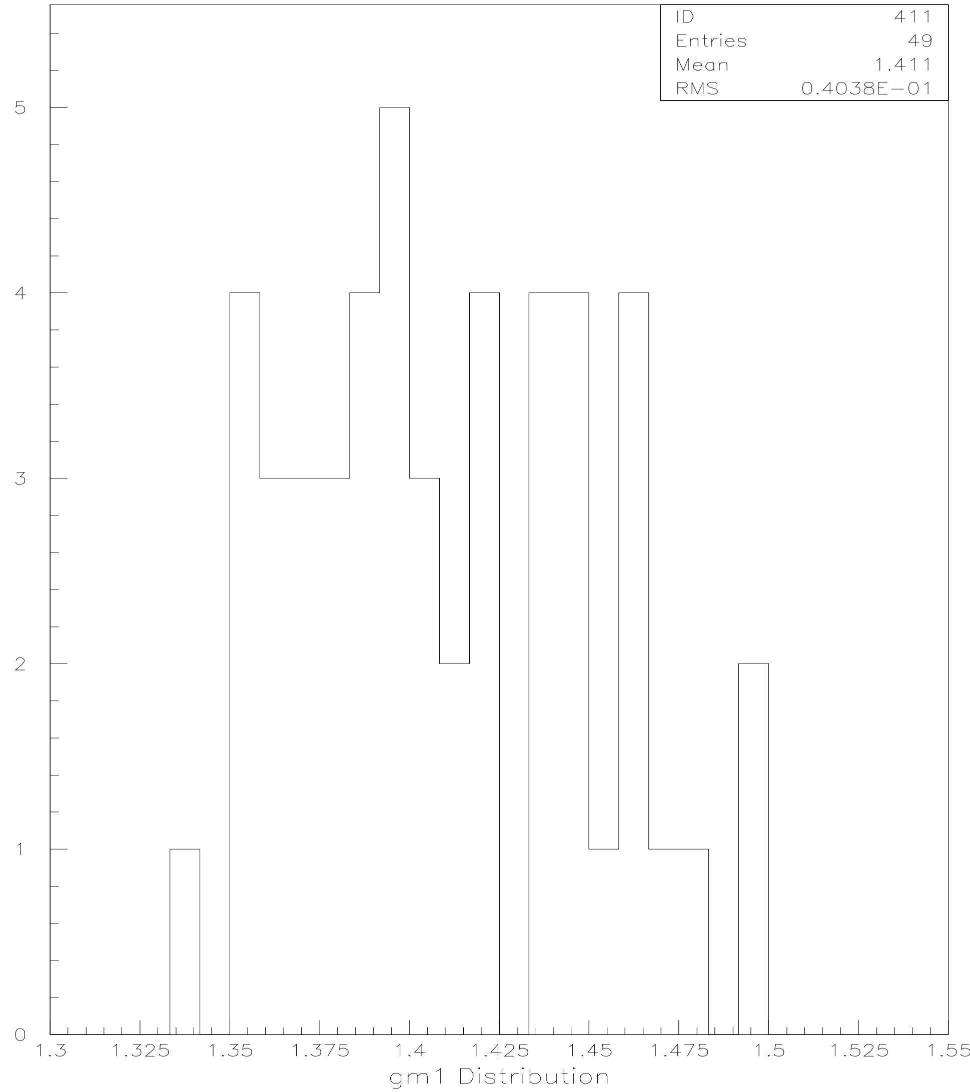
- •likelihood fit process is **massively** CPU-intensive (since Ndata ~120K) and scales linearly in Q

  (~0.5 hr per Q !!)

• Hence, currently:

- •One Q=50 job takes ~1 day of CPU
- • One Q=100 job takes ~2 days of CPU

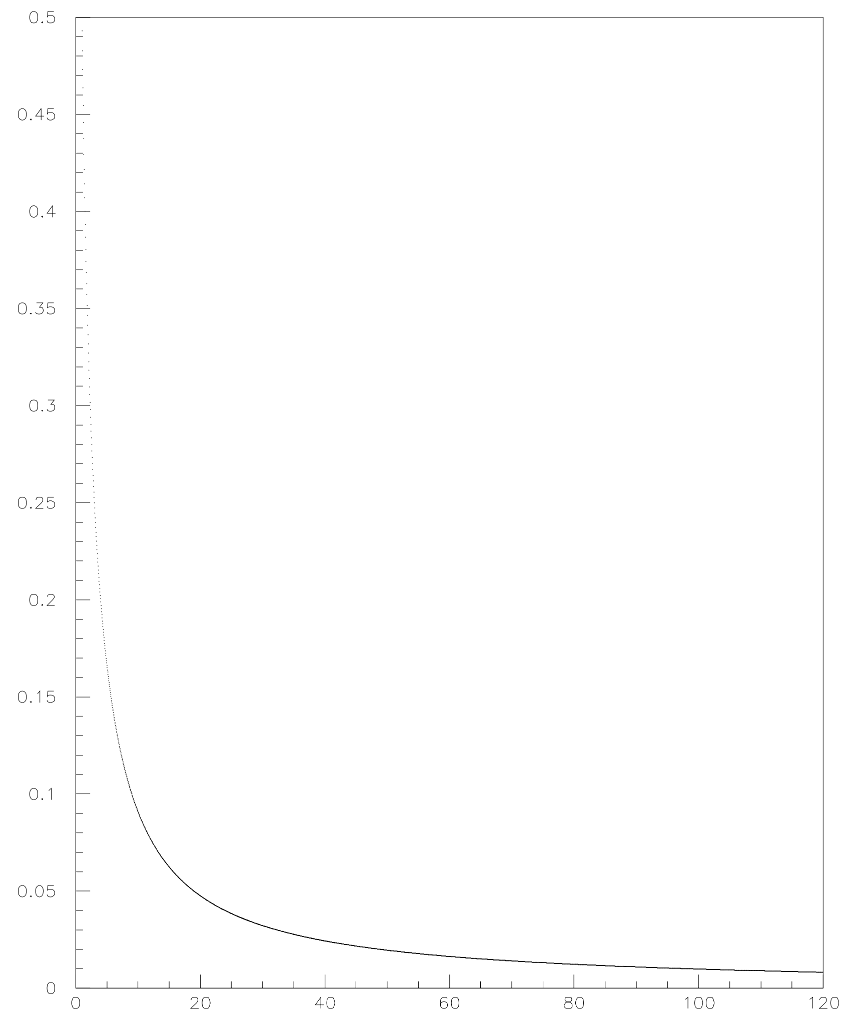
→ Implies the need to carefully select a balance between a small amount of BMC error & the realities of CPU limitations



Theoretical Curve: $\sigma^2_{BMC}/\sigma^2_{DATA}$ vs. Q

- Assuming:
  - $\sigma_{BMC}(Q) = \dfrac{\sigma_{BMC}(Q=1)}{\sqrt{Q}}$
  - $\sigma_{DATA}(Q) = \sigma_{BMC}(Q=1)$
  - $\sigma^2_{TOT}(Q) = \sigma^2_{BMC}(Q) + \sigma^2_{DATA}$
- Then (with a little algebra):

$$\frac{\sigma^2_{BMC}}{\sigma^2_{TOT}} = \frac{1}{Q+1}$$



Current Projects (cont.)

- Getting all my v6_01 MC generation and analysis code running correctly on UVa's shiny new Linux cluster
  - → Should provide ample CPU and disk space to generate sufficient MC to complete my statistical studies

- Optimization studies for analysis cuts

Upcoming Projects

- Complete statistical error analysis
  - (i.e. determine final values for σ_{BMC} and σ_{Ndata})
- Complete final background studies
  - Preliminary studies (with slightly different cuts) indicate that background contributions are small:
    - $K_L \rightarrow \pi^+ \pi^- \pi^0$ ~ 0.3%
    - $K_L \rightarrow \pi^+ \pi^-$ ~ 0.1%

    ** appears to be sufficient to account for virtually ALL observed background
- Extract fit parameter (g_{m1}, a_1/a_2, g_{e1}) values from data, with statistical errors
  - Process is complicated by the large CPU requirements
  - Ultimately, approach may need to be slightly different than the one used for the $K_L \rightarrow \pi^+ \pi^- e^+ e^-$ analysis (which only has ~5000 events)
  - Final choice of "Q" for BMC will have major impact on time/difficulty of process
- Determine g_{e1} statistical sensitivity
  - (i.e. will my result be a value or an upper limit?)
- Begin study of systematic errors
  - Again: likelihood CPU requirements are likely to complicate this!