Background Rejection

S. Ritz

Software Workshop,
January 2000

- Overview of analysis

- Status

- Next steps, needs/improvements

Background Rejection Overview

- Analysis done <u>thus far</u> for two main reasons:

    (1) A reasonable way to quote our effective area.

    (2) A proof of principle and a demonstration of the power of the instrument design.

- Don't expect this to be the final background analysis! Other techniques are available to reduce the backgrounds further with good efficiency. The analysis for the AO response was done in *triage* mode, and there is much to do now.

- Some science topics may require less stringent background rejections than others. Issues of duration, visible energy range, etc.

- Don't expect the simulations of the background to be accurate to this level.

Same points also hold for the event reconstruction we have thus far.

First developer of the background analysis:

Bill Atwood

Lots of work also by Jay, Toby, Heather, Cathie, Sawyer, Jose, Paul ...

straw man

Function: *noun*

Date: 1896

1 : a weak or imaginary opposition (as an argument or adversary) set up only to be easily confuted

2 : a person set up to serve as a cover for a usually questionable transaction

sim·u·la·tion

Function: *noun*

Etymology: Middle English *simulacion,* from Middle French, from Latin *simulation-, simulatio,* from *simulare*

Date: 14th century

1 : the act or process of simulating

2 : a sham object : **COUNTERFEIT**

Ideally (and usually) cut variable distributions are examined several ways, first to check the distribution is sensible and then for implementing the selections:

1) **raw** (after L2T or L3T, depending on tuple)

2) **cumulative** - the distribution of the next cut variable after all previous cuts. Note, this order is **<u>arbitrary</u>** (mostly) and the distributions can be <u>misleading</u>, so….

3) **"all but"**: look at each variable distribution with every cut but this one applied.

4) **niche areas**: check for effects of each cut in different energy ranges and different angles of incidence. (usually done with merit first; in some cases after a phone call from Jose!)

5) **interplay with track quality cuts**: the effects of the track quality cuts and the background rejection cuts are not orthogonal: track quality cuts usually help in background rejection somewhat, and background cuts sometimes help clean up PSF. In one case, an **"all but"** background distribution was empty! Optimize these together.

6) **n-dimensionally** (usually 2 at a time) : look for correlations and domains of well-clustered S/B for like variables.

Note that a neural net very well addresses (3), (5) and (6). These cuts are not orthogonal, and there is a better space in which to make them!

• try to keep the cuts away from steep areas, or right next to individual events (avoid fine-tuning).
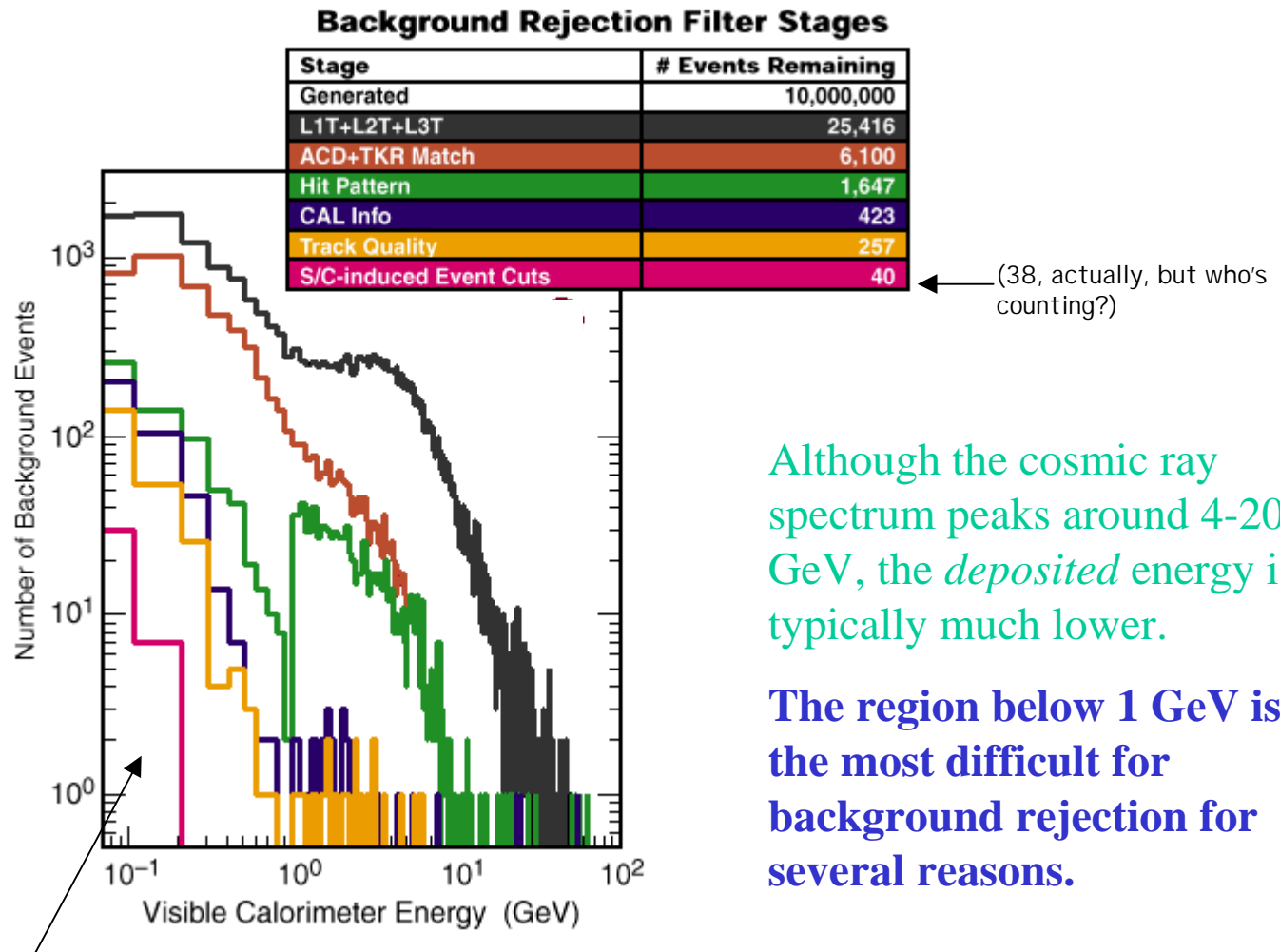
• process is iterative:

With each variable, look at distributions for gammas and background and choose a preliminary cut value.

Scan remaining background events and lost gamma events for adjusting cut and to determine potentially new cut variables.

Check for cut redundancy and correlation. Check impact on instrument performance. Merit is particularly useful here.

As a practical matter, some days are spent mostly improving the rejection and other days are spent mostly improving the gamma efficiency.

Overview: Visible (CAL) Energy Distributions at Various Stages

Background Rejection Filter Stages

Stage	# Events Remaining
Generated	10,000,000
L1T+L2T+L3T	25,416
ACD+TKR Match	6,100
Hit Pattern	1,647
CAL Info	423
Track Quality	257
S/C-induced Event Cuts	40

(38, actually, but who's counting?)

Although the cosmic ray spectrum peaks around 4-20 GeV, the *deposited* energy is typically much lower.

The region below 1 GeV is the most difficult for background rejection for several reasons.

Note, after all selections, no background events remain with visible energy greater than 200 MeV. This wasn't easy.

Some references (beyond meeting presentations):

1) DoE proposal

2) Note of 9 August 1999 (describes cuts and problem areas fairly well, needs distributions included)

3) AO response

however, better documentation is needed. Public version of merit close, but needs some updating/cleanup.

STEPS

• L3T - has been a loose application of a few simple cuts, can be done several different ways, and has not had much attention in the past because it didn't matter much (just a convenient and practical subset of background rejection analysis cuts). <u>NOW IT IS STARTING TO MATTER (e.g., on-board physics analysis), so we should work on defining this more clearly -- more tomorrow.</u>

• The famous VETO_DOCA (only for CsI_Energy_Sum<20) - getting better, but still somewhat *broca*. We have good ideas how to improve this.

STEPS (continued)

- "Hit pattern" - Surplus_Hit_Ratio, with an energy-dependent application.

  ↪ Surplus_Hit_Ratio>2.25 || (CsI_Energy_Sum>1&&fst_X_Lyr>13) || CsI_Energy_Sum>5.

  Hurts off-axis gammas at low energy

- "CAL info" - CsI_Fit_errNrm, CsI_Xtal_Ratio -- **keep events w/little CAL info whenever possible.**

  CsI_Xtal_Ratio>0.25||CsI_No_Xtals<1

  (CsI_Energy_Sum<1.&&CsI_Fit_errNrm<10.)||CsI_Fit_errNrm<4.||CsI_No_Xtals<1

- "Track quality" (from Jose)

- "S/C induced event cuts" - designed to remove cosmics whose primary interaction is in the S/C. This is our single largest residual background!

  No_Vetos_Hit<1.5 || (CsI_Energy_Sum>1. && No_Vetos_Hit<2.5) || CsI_Energy_Sum>50.

  CsI_eLayer8/CsI_Energy_Sum<0.08 || CsI_eLayer1/CsI_Energy_Sum>0.25 ||
  CsI_Energy_Sum>0.35||CsI_No_Xtals<1

  CsI_moment1<15. || CsI_moment1<80.&&CsI_Energy_Sum>0.35||CsI_Energy_Sum>1.||
  CsI_No_Xtals<1

  CsI_Z>-30.||CsI_No_Xtals<1 ← Surprisingly efficient even at high energy

  CsI_No_Xtals_Trunc<20.||CsI_Energy_Sum>75.||fst_X_Lyr<12 ← Only needed in BACK

 Quality_Parm>10 (composite track quality parameter, cut effective against low-energy stubs from splash-up)

(Will work on making this one continuous in energy with a normalized moment)

Next steps:

- Improve low energy Aeff, work on inefficiencies (Surplus_Hit_Ratio, CsI_Fit_errNrm)

- VETO_DOCA: needs work. Mainly a tracking issue. Seed tracks with hit tiles, track quality selections for loop.

- Document, put correct implementation into merit.

- Simplify analysis (make prettier, simpler). Bring in neural net. More sophisticated tracking (downward V) & CAL pattern recognition.

- Further improvements in rejection (currently, integrated background rate is ~ 6% of extragalactic diffuse rate). Also, study background rate differentially (by visible energy bin). More work on upward-going energy events.

- Start validation of simulation of hadronic response.

Needs, infrastructure improvements:

• Simulation event output improvements: better format; embed version & generation info more completely; **save full underlying event**.

• Structured tuple -- store more info about *all* tracks, etc.

• Continue to improve version validation. Schedule & plan for orderly development. Dust-settling periods. (example, TKR recon rewrite)

• More hardware realism: CAL noise, TKR TOT, ACD noise, etc. Review material.

• ID generated events uniquely and implement automated system to extract subsets of events from massive generated files.

• GUI enhancements: interact w/all objects (tracks, cal info). Draw reconstructed photon direction. Interactive track refits. PS file generation.

• *What are the real background fluxes!?!?*

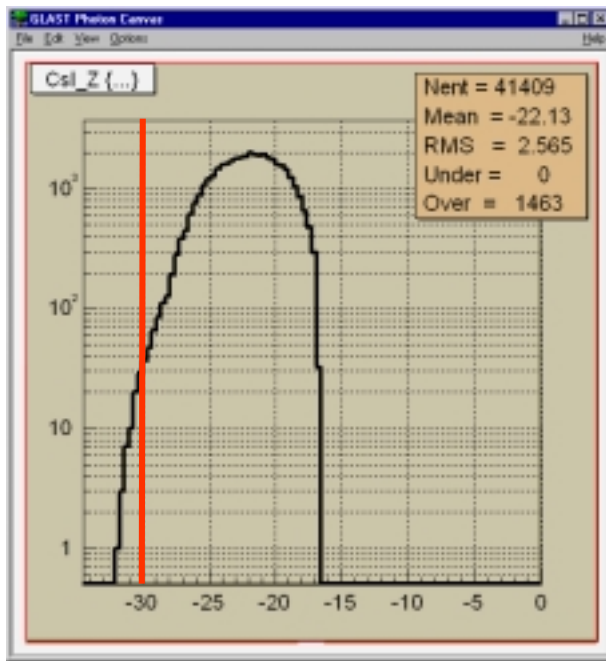• Analysis platform definition to allow tools sharing.
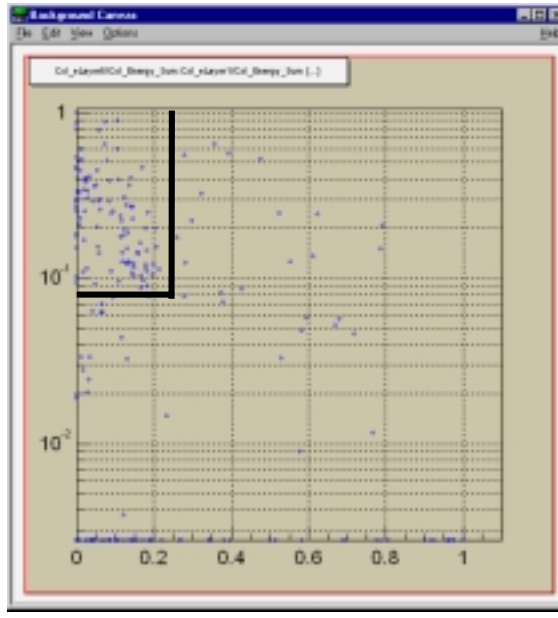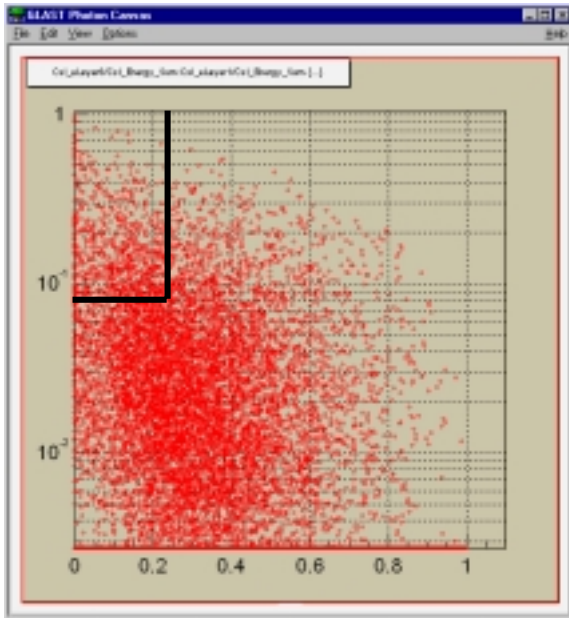
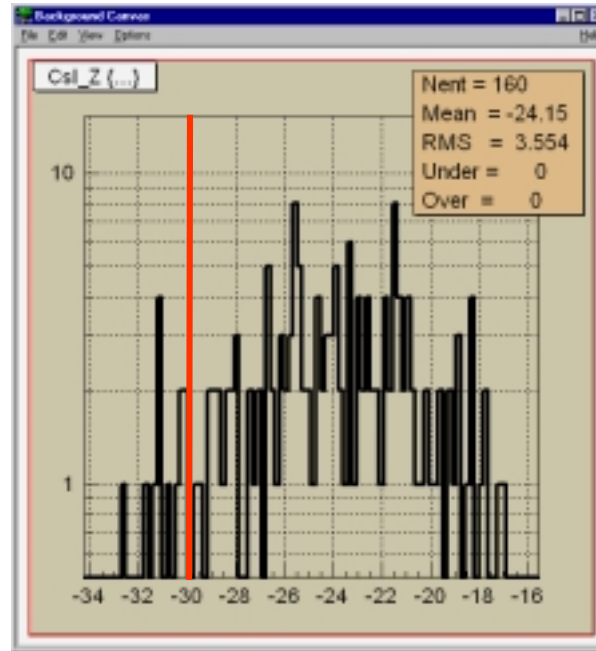Cal_2Dfrac (fraction of energy in front and back layers of CAL)
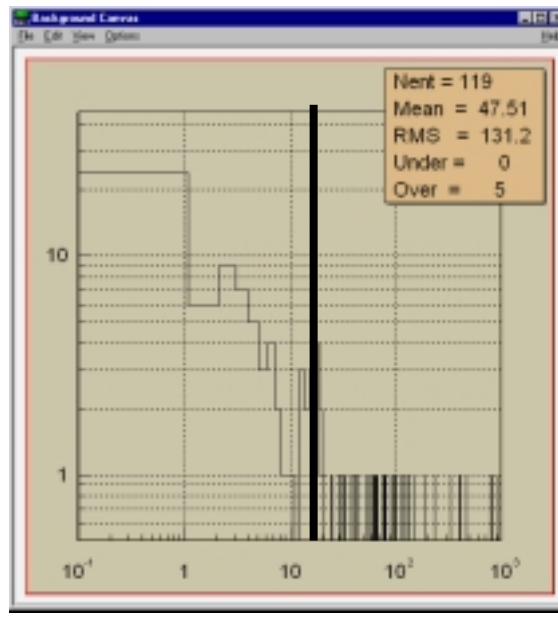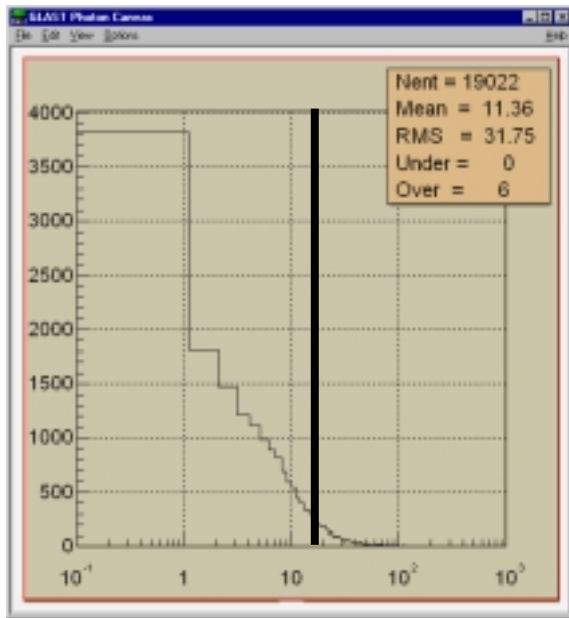
CsI_moment1 for E<350 MeV

All_gamma   Backgroundmix

All_gamma

Backgroundmix