

# Modern Statistical Methods for GLAST Event Analysis

Robin D. Morris  
USRA-RIACS  
444, Castro St, Suite 320  
Mountain View, CA 94041

Johann Cohen-Tanugi  
SLAC/KIPAC  
Stanford University, Stanford, CA

**Abstract**— We describe a statistical reconstruction methodology for the GLAST LAT. The methodology incorporates in detail the statistics of the interactions of photons and charged particles with the tungsten layers in the LAT, and uses the scattering distributions to compute the full probability distribution over the energy and direction of the incident photons. It uses model selection methods to estimate the probabilities of the possible geometrical configurations of the particles produced in the detector, and numerical marginalization over the energy loss and scattering angles at each layer. Preliminary results show that it can improve on the tracker-only energy estimates for muons and electrons incident on the LAT.

## I. INTRODUCTION

The Large Area Telescope (LAT) [1] is the primary instrument on GLAST, and so it is of utmost importance to extract as much information as possible from the response of the LAT to incident photons and particles. While the quantities of primary interest for each event are few (namely the azimuth, elevation and energy of the incident photon/particle), the rich physics of the interactions of the particles/photons with the LAT makes a principled reconstruction algorithm complex. Whilst the objects of primary interest are photons, the interaction of a photon with the LAT is essentially that of an electron-positron pair. Therefore we concentrate on the basic building blocks of event reconstruction, the analysis of the interaction of charged particles with the LAT.

Figure 1 shows the schematic of an interaction between a charged particle and the LAT. Visible in the figure are 1) multiple Coulomb scattering in the tungsten foils; 2) the production of secondary photons; and 3) the production of secondary charged particles. The GEANT4 toolkit [2] is designed to simulate these physics processes in the *forward* direction. The task in event reconstruction is the *inverse* problem – estimating, from the data of the microstrip responses, which physics processes actually occurred in a particular event. The result is an estimate not only for the original particle and its properties, but also of all secondary particles and photons. To accurately estimate the primary particle, it is necessary to estimate accurately all secondaries.

Figure 2 shows the tree of hypotheses for the physical processes at the first two layers of interaction of a charged particle with the LAT. The final leaves describe the *structure* of the hypothesized event reconstructions. Clearly as we descend the layers the number of branches in this tree explodes. But

only a small number of their leaves will be consistent with the instrument data - with the pattern of microstrips that fired - and, further, this consideration can be used to exclude branches as the tree is descended, further limiting the number of leaves that must be considered. A full event reconstruction consists of two stages:

- 1) The enumeration of the possible event structures consistent with the microstrip data.
- 2) The computation of the parameters of each event structure and their relative probabilities.

Computation of the relative probabilities allows particle identification (by considering the trees formed by different hypothesized incident particles), and, for a particular particle type, the final event reconstruction to be a weighted average of event structures, weighted according to the probability that the microstrip data actually came from that structure.

## II. METHODOLOGY

We concentrate first on the simplest type of event. If the only physical process that actually occurred was multiple Coulomb scattering, then there is only a single (x-y) pair of microstrips at each layer (excluding noisy strips). This is the case for muons of moderate energy (up to a few hundred MeV), where the probability of producing secondary electrons or photons is extremely small, and hence can be neglected. We parameterize the trajectory of the particle by 1) its origin (a point outside the LAT); 2) the position at which it traverses each conversion layer; and 3) its endpoint (also outside the LAT); and from these we derive the incident directions ( $\theta, \phi$ ) and the scattering angles,  $\theta_i$ , at each layer. Finally, we add the incident energy,  $E$ , and the energy deposited in each conversion layer  $\delta E_i$ . Denoting by  $s_i$  the microstrips at each layer, we can write

$$p(\theta, \phi, E, \theta_1, \delta E_1, \dots, \theta_n, \delta E_n | s_1 \dots s_n) \propto p(s_1 \dots s_n | \theta, \phi, E, \theta_1, \delta E_1, \dots, \theta_n, \delta E_n) \times p(\theta, \phi, E, \theta_1, \delta E_1, \dots, \theta_n, \delta E_n)$$

The first term on the right hand side is the likelihood. It takes one of two values – one if the trajectory described by  $\theta, \phi, \theta_1, \dots, \theta_n$  intersects all the microstrips that fired, and zero otherwise. It serves to limit the region of the state space that is of interest. The second term contains all the physics of the interactions of the particle with the LAT. We use conditional

$$\begin{aligned}
& p(\theta, \phi, E) && \text{Priors on azimuth, elevation and energy.} \\
& \times p(\theta_1, \delta E_1 | E, t_1) && \text{Distribution of scattering angle and energy loss for a particle of energy } E \\
& && \text{in a foil of effective thickness } t_1. \\
& \times p(\theta_2, \delta E_2 | E, \delta E_1, t_2) && \text{Same for the particle at layer 2, which has energy } E - \delta E_1 \\
& && \text{and sees a foil of effective thickness } t_2 \\
& && \times \dots \\
& \times p(\theta_n, \delta E_n | E, \delta E_1, \dots, \delta E_{n-1}, t_n) && \text{At layer } n \text{ the particle has energy } E - \delta E_1 - \dots - \delta E_{n-1}.
\end{aligned} \tag{1}$$

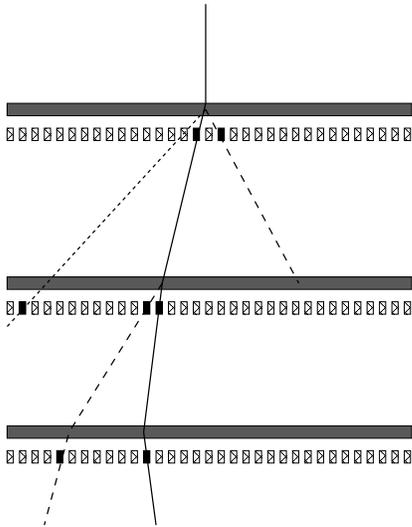


Fig. 1. A schematic of a charged particle interacting with the LAT. The solid line is the incident charged particle. Long-dashed lines indicate secondary charged particles. The short-dashed line is a secondary photon. The solid blobs indicate microstrips that fired.

independence and decompose it as shown in equation 1 at the top of this page.

These scattering distributions are the known distributions for particles of a specified energy incident on a LAT foil [3]. They are also parameterized by the effective thickness,  $t_i$  due to non-normal incidence. The non-Gaussian tails of the scattering angle distributions were modeled by a second Gaussian component. For muons, the energy loss was parameterized as a Landau distribution, with the distribution's parameters being functions of energy.

The parameters of primary interest, however, are the azimuth, elevation and energy, and so the distribution of primary interest is  $p(\theta, \phi, E)$ . This is obtained from (1) by marginalization. This is performed numerically using Markov chain Monte Carlo (MCMC). Originally developed in physics [4], it has been extensively developed in statistics in the past 20 years, and is now a standard tool for use in the analysis of complex, high-dimensional probability distributions [5]. It works by simulating a Markov chain whose equilibrium distribution is constructed to be the distribution of interest (in this case, the

distribution in (1)). Averages over the distribution can then be made by forming averages over the states of the simulated Markov chain. For example, the mean energy is estimated by forming the mean of the energy variables over a length of the simulated chain, while ignoring all the other variables. Collecting all the variables into  $\mathbf{x}$ , and initializing  $\mathbf{x} \leftarrow \mathbf{x}_0$  the MCMC algorithm is iteration of

- 1) propose a change,  $\mathbf{x} \leftarrow \mathbf{x}'$  with some proposal distribution  $\pi(\mathbf{x}'; \mathbf{x})$
- 2) accept the change with  $p_a = \frac{p(\mathbf{x}')\pi(\mathbf{x}; \mathbf{x}')}{p(\mathbf{x})\pi(\mathbf{x}'; \mathbf{x})}$ , and set  $\mathbf{x} \leftarrow \mathbf{x}'$ , else retain  $\mathbf{x}$

The proposal distribution at each stage may be chosen to only change some of the elements of  $\mathbf{x}$ . For this work, we use a cycle of proposals that successively proposes changes to  $\theta, \phi, E, \theta_1, \delta E_1, \dots, \theta_n, \delta E_n$ .

For electrons incident on the LAT foils, as well as multiple Coulomb scattering, there is appreciable probability of producing a secondary photon, and a small probability of producing a secondary electron. (At 100MeV these probabilities are  $\simeq 0.25$  and  $\simeq 0.01$  respectively.) We restrict the discussion here to events which contain at most secondary photons. Typically, for electrons of a few hundred MeV the secondary photons are not detected. They carry energy away from the electron which is “lost” to the tracker.

In the forward direction this is modeled by a mixture distribution. With probability  $p_{ns}(E)$  no secondary is produced, and the energy loss follows a Landau distribution. With probability  $p_s(E)$ , a photon is produced and the energy lost has two components, a Landau distributed component from multiple Coulomb scattering, plus a component distributed as  $1/E$  representing the energy carried away by the photon. The energy loss distribution in this case is a convolution of the Landau distribution with the  $1/E$  distribution.

The samples generated by the MCMC algorithm represent the distribution over the trajectories' parameters. To compute the probability for an event structure it is necessary to compute the normalizing factor that was omitted from equation (1). This can be done by using the MCMC output to construct an importance sampling distribution, and using samples from that distribution to compute the normalizing factor. This will be discussed elsewhere.

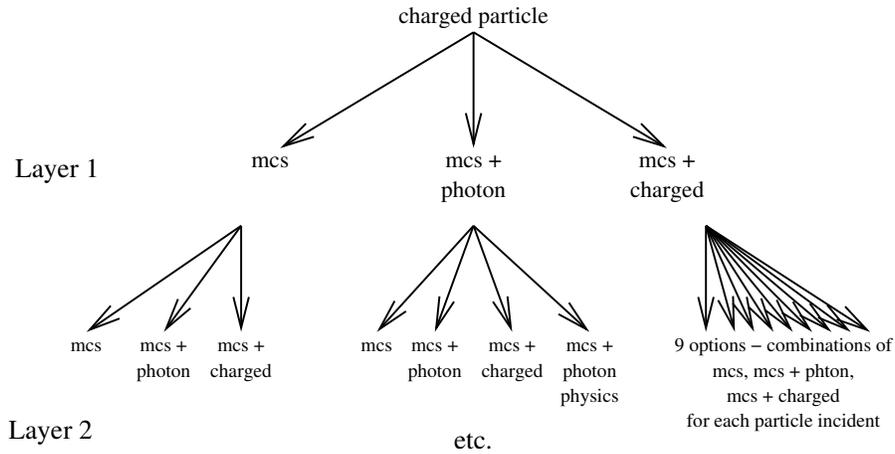


Fig. 2. The first two layers of the tree of possible event structures

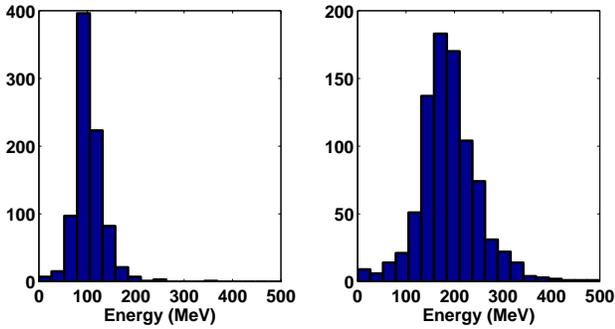


Fig. 3. Energy estimates for muons incident in a  $45^\circ$  cone

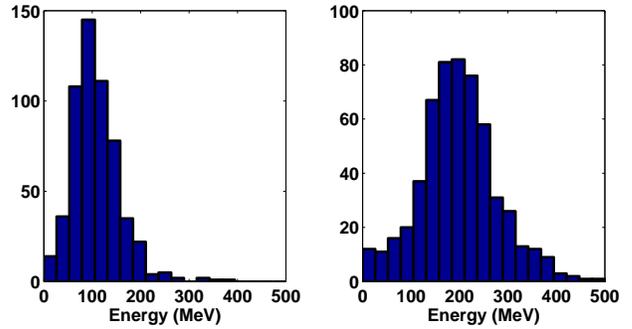


Fig. 4. Energy estimates for electrons incident in a  $45^\circ$  cone

### III. RESULTS

One thousand events were simulated [6] for each of four sources – muons and electrons of 100 and 200 MeV. The incidence directions were chosen randomly within a 45 degree cone. Figures 3 and 4 show the energy estimates of the reconstruction. We do not show here direction estimates because, for charged particles, the accuracy of the direction estimate is determined almost entirely by multiple Coulomb scattering in the top foil. For electron events, those events where two charged tracks were detected by the reconstruction algorithm were not analyzed. In all four cases the estimates are unbiased; the histograms are centered accurately on either 100MeV or 200MeV. The histograms for electrons show more dispersion than those for muons, due to the effect of energy being transferred into photons which are not detected. Note that these estimates were made using only the first 12 regular GLAST layers of the tracker and did not use any information from the calorimeter.

To illustrate the sensitivity of this approach to event reconstruction, consider the event shown in figure 5. This is the simulation of a 100MeV electron incident on the LAT. The cyan lines indicate the microstrips that fired, and that are considered as part of the event; the grey lines are microstrips

that fired due to “noise”. The near-vertical red line is the trajectory of the incident electron. The white lines are low-energy photons that are produced by the primary electron. The short, highly curved, red lines are low-energy electrons (“delta rays”) that typically do not trigger microstrips. The highlighted line (in purple) is a high-energy secondary photon, and is the object of interest here.

Figure 6 shows the marginal posterior distribution for the energy of this electron, given the LAT response. It is peaked around 65MeV. Because of the stochastic nature of the physics of the detector, the response for each individual electron may not be peaked at the true value. However, figures 3 and 4 show that the reconstruction methodology as a whole is unbiased.

Figure 7 shows the marginal posterior distribution for the energy lost by the electron in layer 9 of the LAT. Recall that the physics of energy loss by electrons in tungsten foils is from multiple Coulomb scattering and, potentially, from secondary production. For layer 9, figure 7 follows closely the Landau distribution for energy loss due to multiple Coulomb scattering. Note that there is no direct measurement of the energy loss, so in this case the estimated distribution of the energy lost is essentially the prior distribution from the physics (plus an effect due to the uncertainty of the energy of the incident

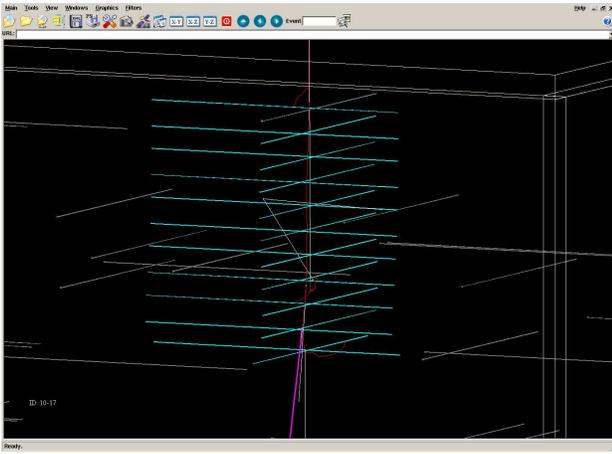


Fig. 5. A simulation (“Monte Carlo”) of a 100MeV electron normally incident on the GLAST LAT

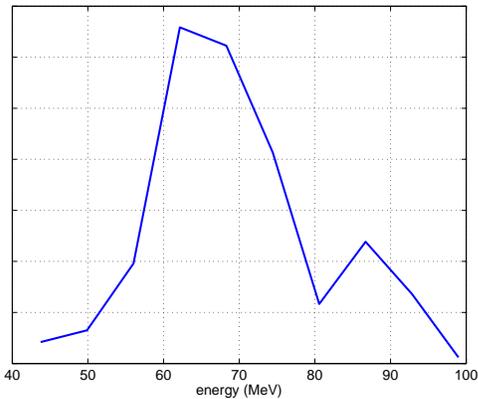


Fig. 6. One electron event: marginal distribution for the electron’s energy

electron at layer 9).

Figure 8 shows the estimated distribution of energy loss at layer 10. In this case the distribution does not resemble the Landau distribution. Instead, it shows that significant energy (approximately 38MeV) was lost from the electron as it traversed this layer. Because this event only had a single set of hits at each layer, this result indicates that the energy was lost into a secondary photon, a photon that did not interact with the LAT after production. The only information in the data about the change in energy of the incident electron in this layer is that the scattering angles become (on average) larger in layers below the layer at which the photon was produced. Close examination of the simulated event (figure 5, the “Monte Carlo”) shows that indeed a photon was produced at one of the lower layers of the LAT, indicated in purple in figure 5.

The new methodology, from a detailed analysis of the behavior of the primary electron, revealed the production of a secondary photon.

#### ACKNOWLEDGMENTS

Our analysis methodology is built upon the software provided by the LAT Science Analysis Software Team. RDM is

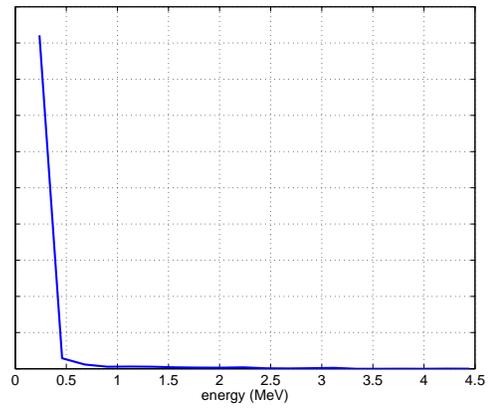


Fig. 7. Marginal distribution for energy loss at layer 9

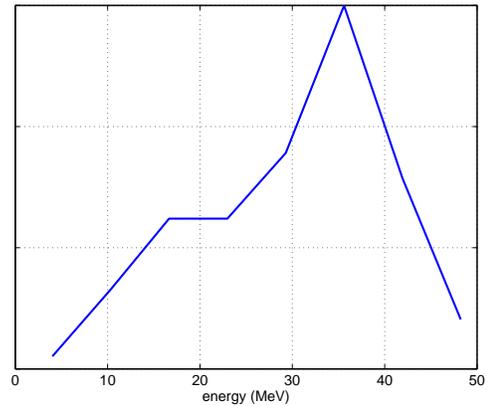


Fig. 8. Marginal distribution for energy loss at layer 10

supported by a grant from the NASA AISR program.

#### REFERENCES

- [1] <http://www-glast.slac.stanford.edu>
- [2] S. Agostinelli et al., “GEANT4 - a simulation toolkit”, *NIM-A* **506** 3, pp 250-303 (2003)
- [3] W.-M Yao et al., *Journal of Physics G*, **33**, 1 (2006)
- [4] N. Metropolis et al. “Equation of state calculations by fast computing machines”. *J. Chem. Phys.* **21**, 1087-1092 (1953).
- [5] W.R. Gilks et al (eds), *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC (1995)
- [6] W. Atwood et al. “The full simulation of the GLAST LAT high energy gamma ray telescope”. In *Perugia 2004, Calorimetry in particle physics*, pp 329-336 (2004)