Fermi
Gamma-ray
Space Telescope

Application of Classification Trees to Fermi-LAT Data

Alex Drlica-Wagner
on behalf of the Fermi-LAT Collaboration
SLAC-KIPAC-Stanford

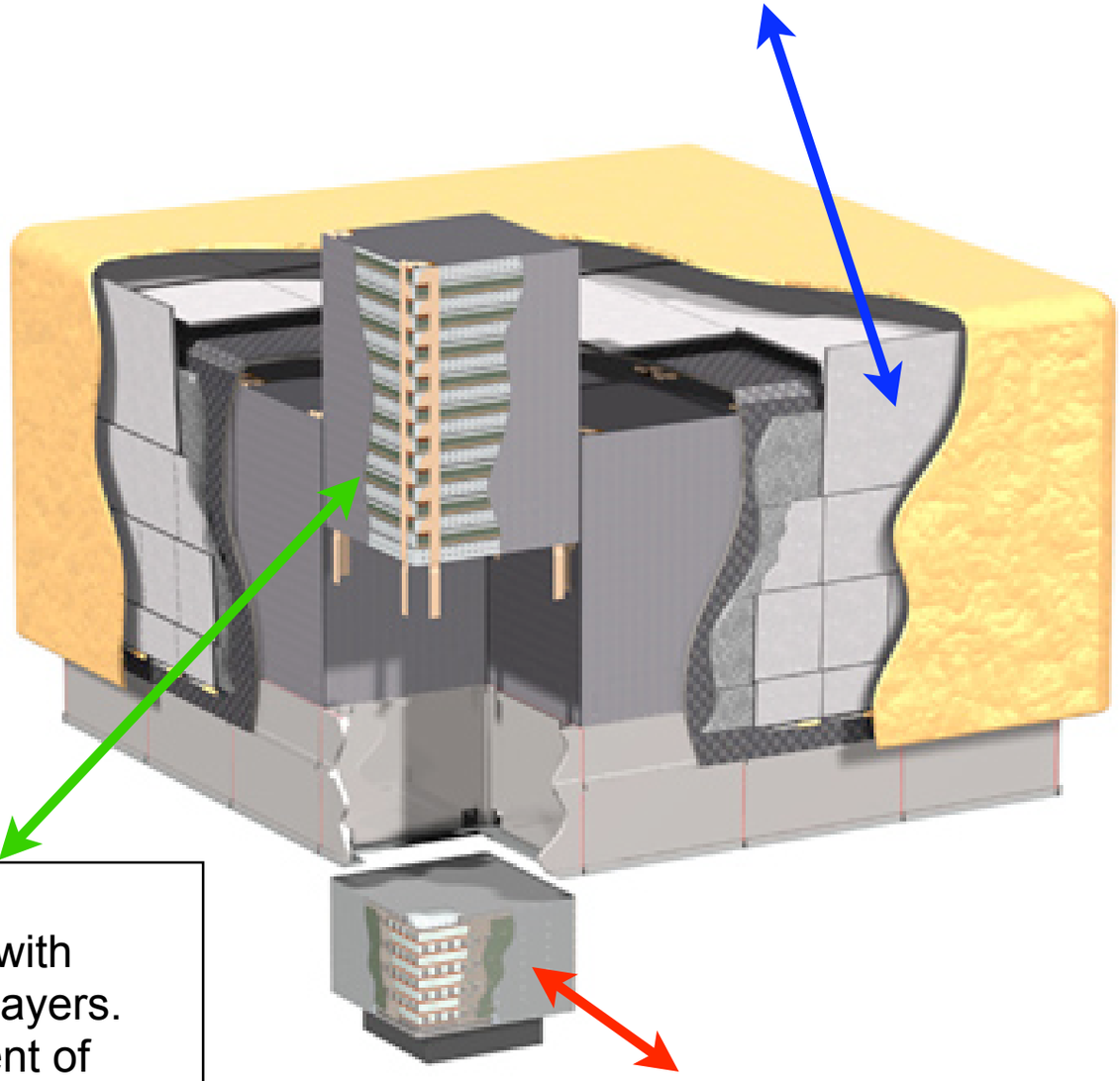
SLAC KIPAC STANFORD UNIVERSITY

Fermi Large Area Telescope (LAT)

- **The Fermi Large Area Telescope (LAT) is an e^+e^- pair-conversion gamma-ray telescope.**

- **The LAT is sensitive to photons from 20 MeV to greater than 300 GeV and accepts events over a 2π sr field of view.**

- **Composed of three detector subsystems:**
  - **Anti-coincidence Detector (ACD)**
  - **Tracker (TKR)**
  - **Calorimeter (CAL)**



Anti-coincidence Detector (ACD)
- Segmented tiles of scintillating fiber.
- Charged particle detection.

Tracker (TKR)
- Silicon strip sensor with tungsten conversion layers.
- Precise measurement of event direction.

Calorimeter (CAL)
- Hodoscopic array of CsI crystals.
- Event energy reconstruction.

Reconstruction of LAT Events

- **Reconstruction algorithms knit together hundreds of readouts into a physical description of each event.**

- **We want to answer these questions about the primary particle in the event:**
  - **Where did the particle come from?**
  - **What energy did the particle have?**
  - **Was the the particle a photon?**

- **The last of these questions is very difficult:**
  - **As a pair-production telescope, the LAT tracks charged particles making it sensitive to cosmic-ray contamination.**
  - **At the LAT orbit cosmic rays can out number gamma-rays by more than a factor of 10^4.**

- **Separating and quantifying a small signal from a large background is a task ripe for machine learning algorithms...**

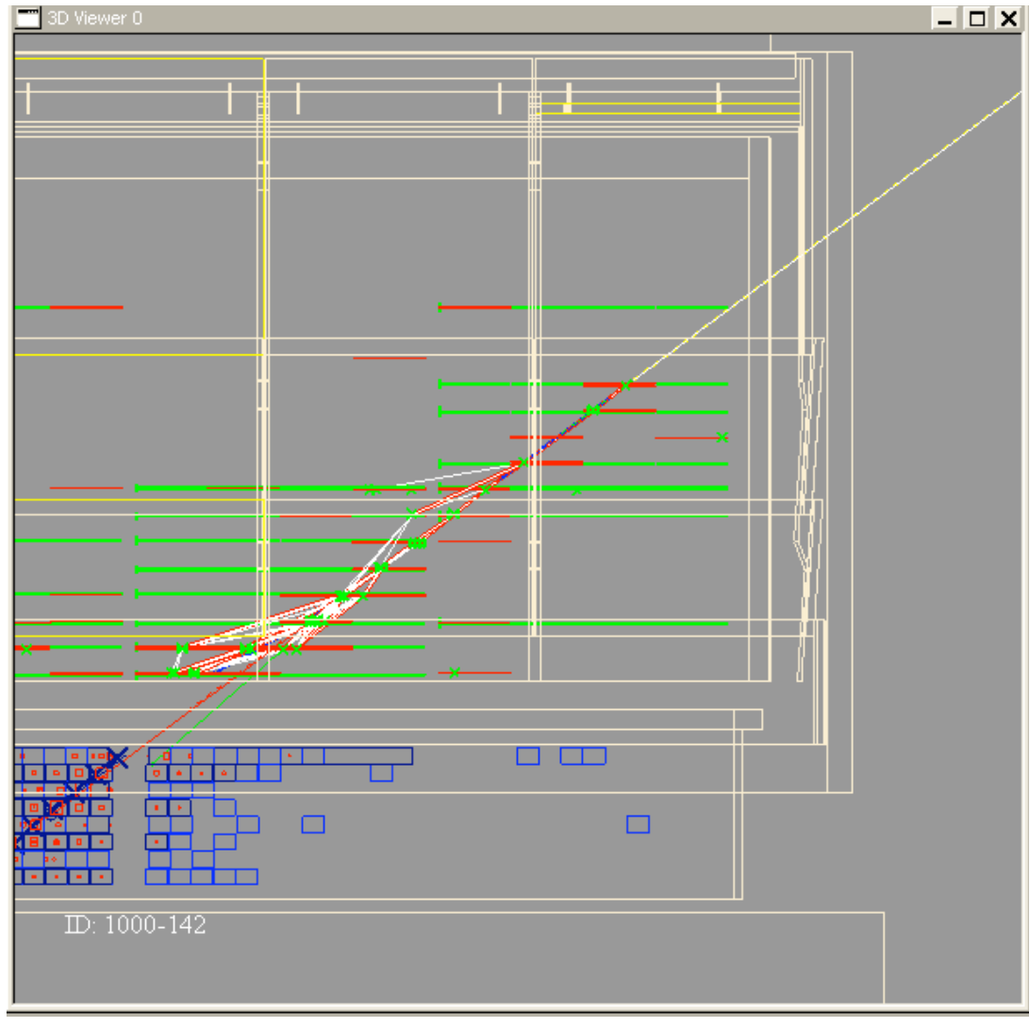
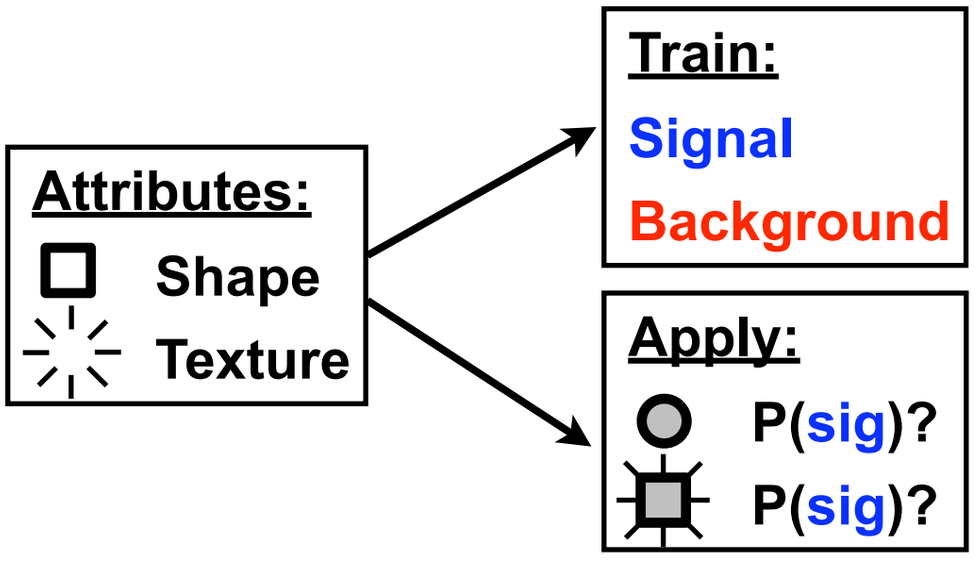
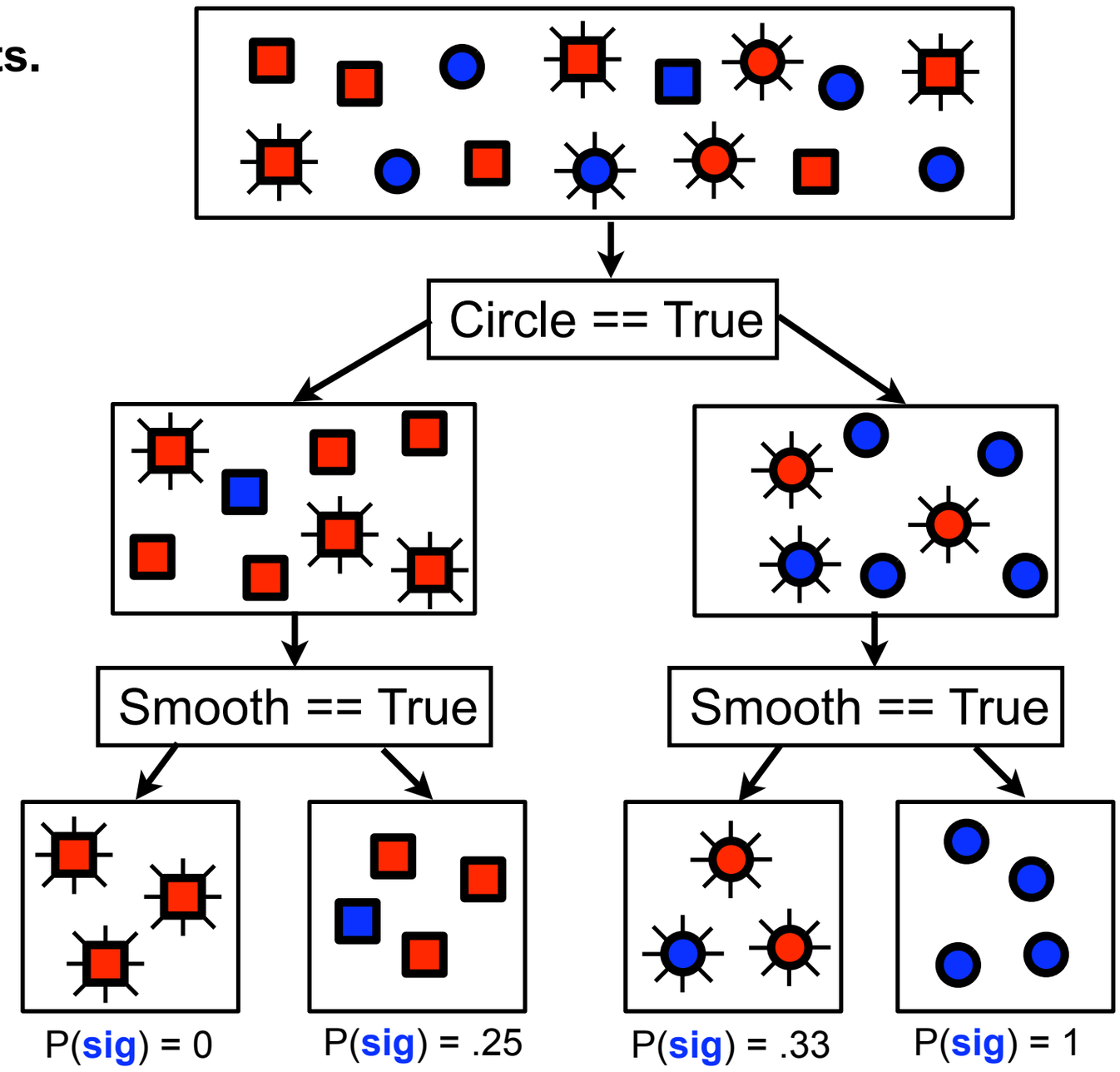


Figure - Reconstruction of an 8 GeV photon event in the LAT.

Classification Trees (CTs)

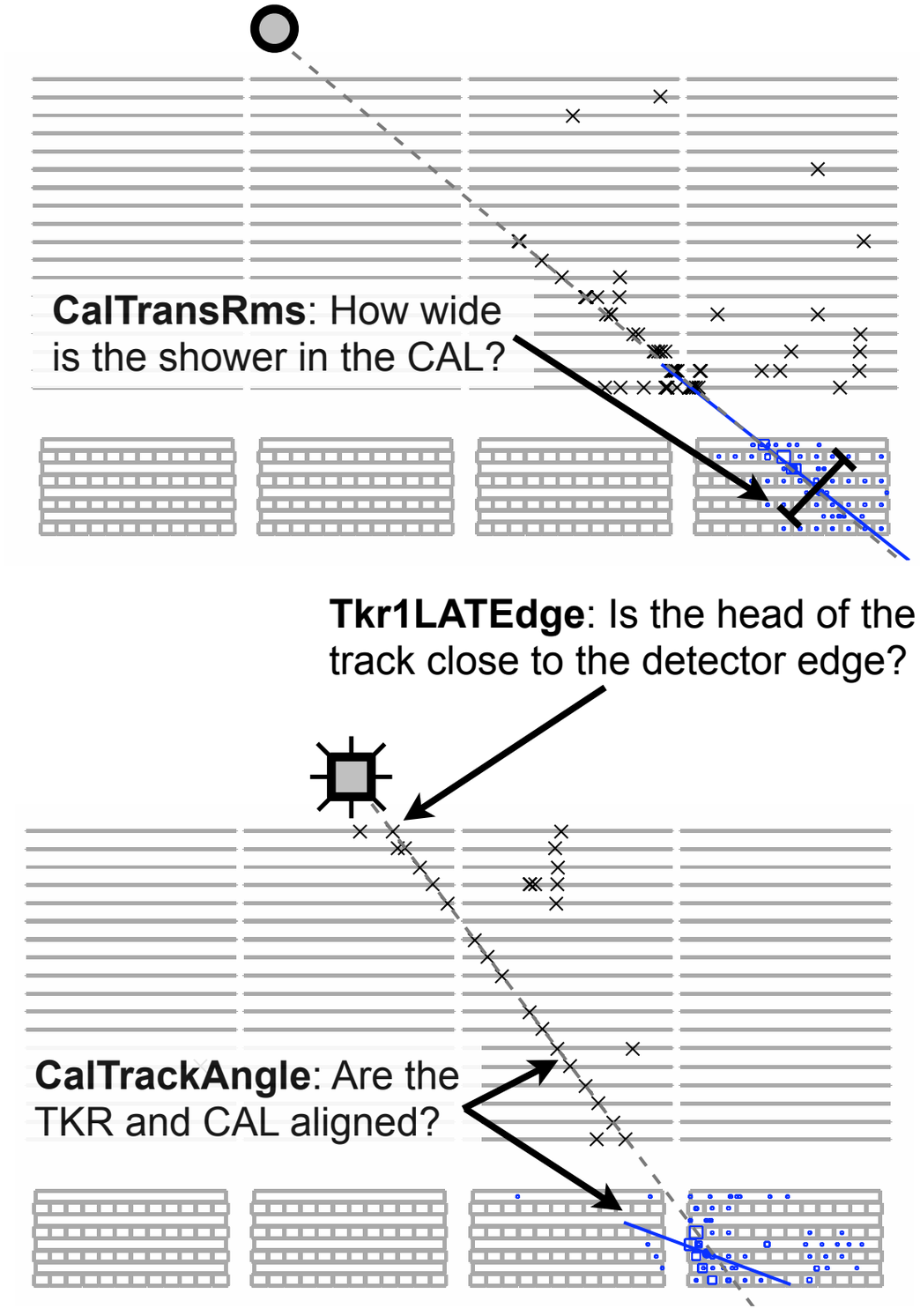
- **Machine learning algorithms use known training data to create criteria that can used to classify new data sets.**
- **Classification Trees (CTs) distinguish signal from background with a set of binary cuts.**
  - **Maximize separation at each cut.**
  - **Terminate at a minimum number of events.**
  - **Find signal probability at end nodes**





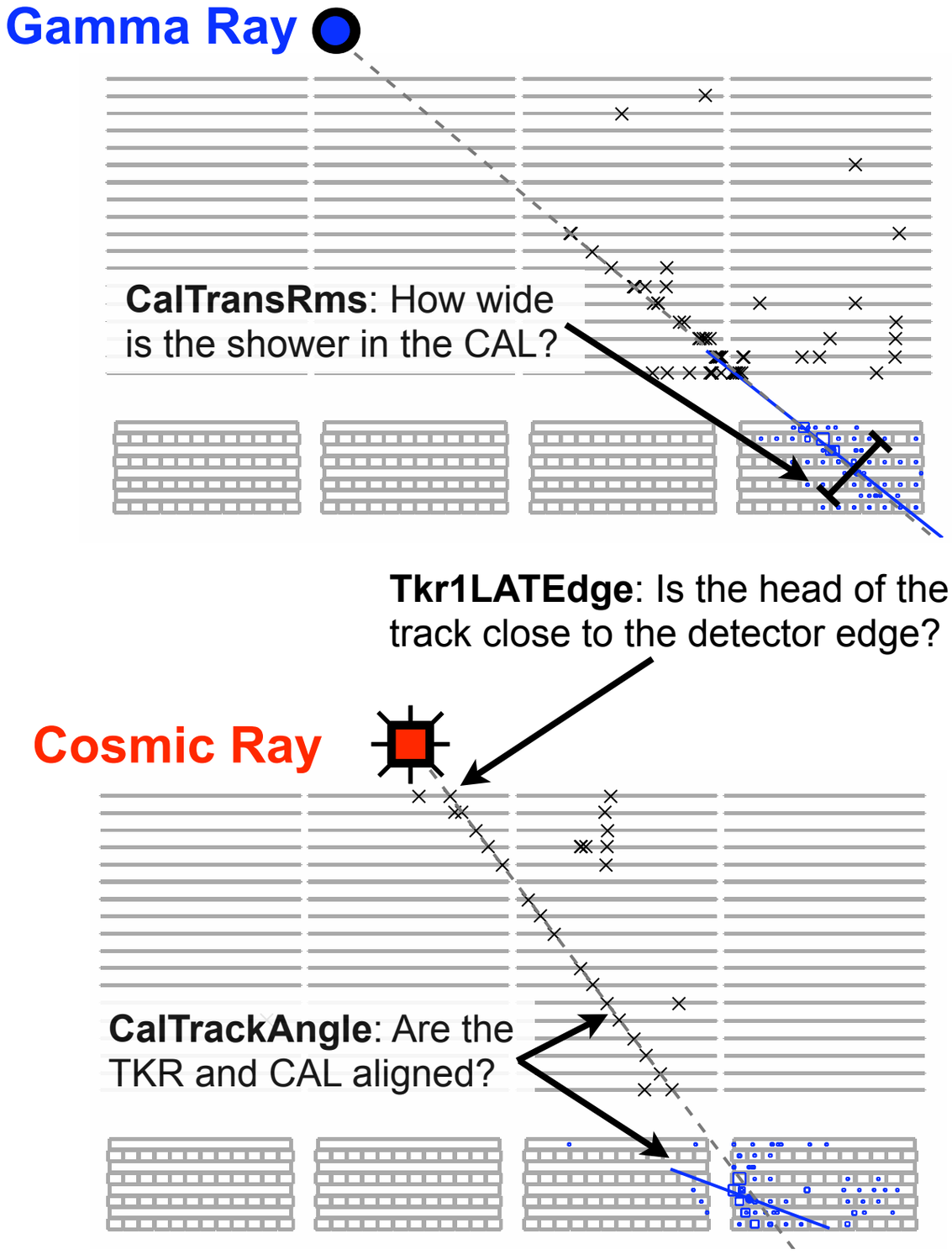
- **In most cases, a "forest" of classification trees is used.**

Using Classification Trees



Which one is the gamma ray?

Using Classification Trees



- **Energy Resolution Knob**
  - **Goal:** Distinguish events with well measured energy from those with poorly measured energy.
  - **Input:** TKR and CAL variables as CT input.
  - **Output:** A "knob" that can be adjusted to trade energy resolution for photon efficiency.

- **Image Resolution Knob**
  - **Goal:** Distinguish events lying within the angular containment radius from those outside.
  - **Input:** Mostly TKR track variables.
  - **Output:** A "knob" that can be adjusted to trade image resolution for photon efficiency.

- **Background Rejection Knob**
  - **Goal:** Distinguish photons from cosmic rays.
  - **Input:** Variables from all subsystems.
  - **Output:** A "knob" that can be adjusted to trade photon purity for photon efficiency.

More on Background Rejection

- **Start with simple cuts**
  - **ACD: Seal gaps, corners and ribbons and reject events that point to hit tiles.**
  - **TKR: Look for vertices (e+ e- pair)**
  - **CAL: Look at transverse shower size.**

- **Train independent CTs for each subsystem.**

- **Global Background Rejection CT:**
  - **Input: Combined information from CAL and TKR (i.e. agreement in direction)**
  - **Input: The probabilities output from the individual subsystem CTs**
  - **Output: A single variable which can be used as a background rejection knob.**

- **Agreement between simulations and data is required. CTs are only as good as the simulated events they are trained on.**

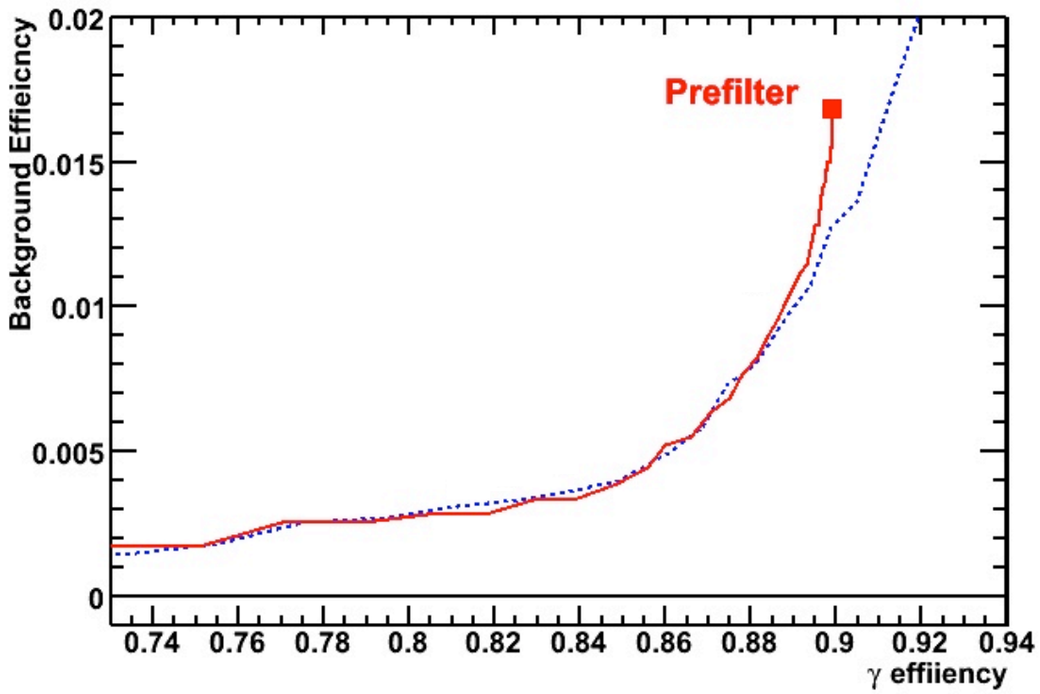


Figure: Example of the background contamination vs. signal efficiency knob derived from classification trees.

TMine Analysis Tool

- **Historically, the Fermi CTs have been generated with Insightful Miner**
  - **Proprietary software designed for the financial industry.**
  - **Uses a visual framework for designing complex analyses.**
  - **Not optimized for large physics data sets.**
  - **Does not interface well with the LAT data processing pipeline.**

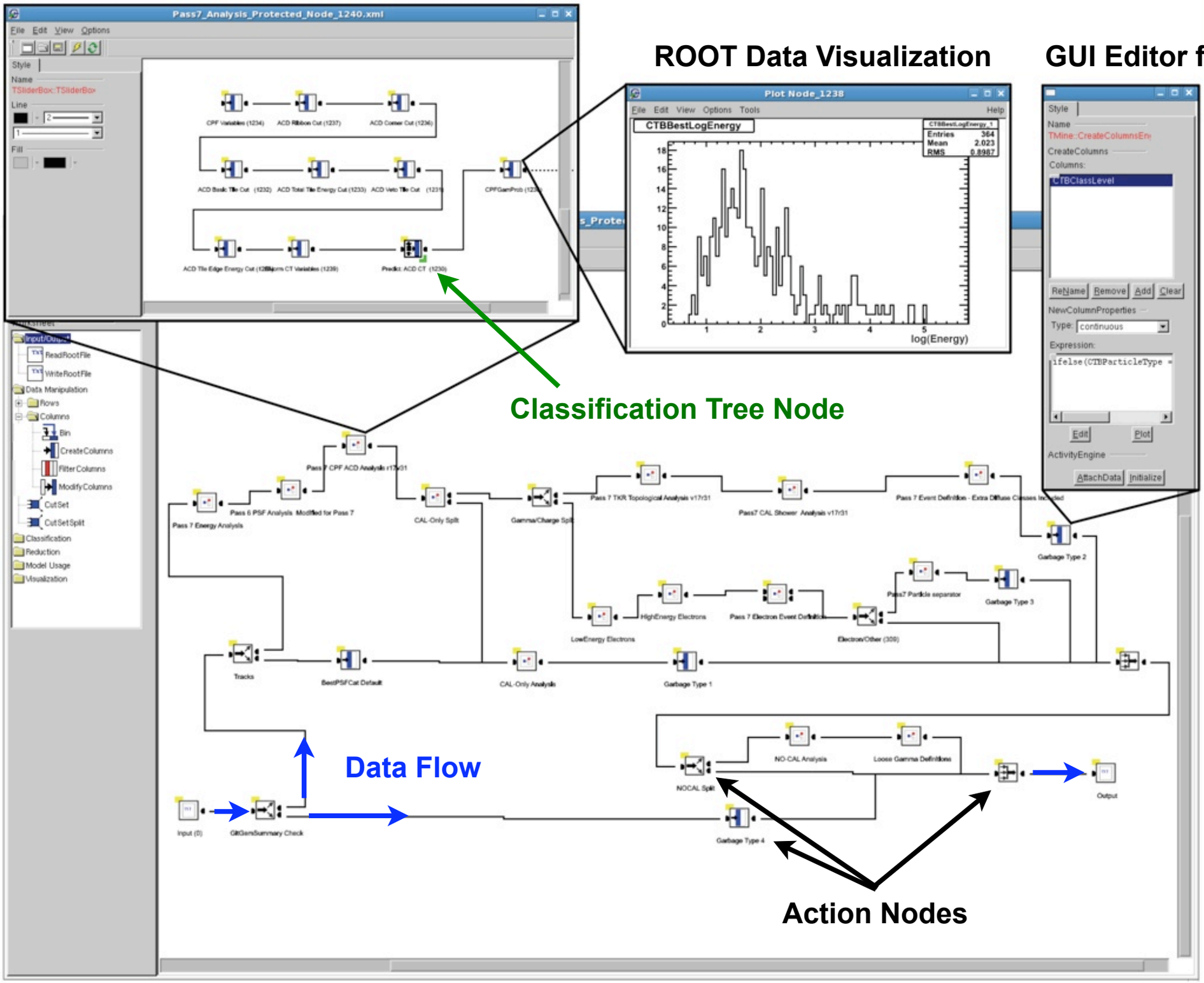
- **TMine is a new open source alternative developed here at SLAC**
  - **Based on ROOT, the de-facto framework for high energy physics experiments.**
  - **Interfaces with the ROOT Toolkit for Multivariate Analysis (TMVA).**
  - **Handles larger data sets with increased speed and efficiency.**
  - **Integrates easily into the LAT pipeline.**

- **TMine goes beyond classification trees**
  - **Provides a GUI framework for designing complex analyses**
  - **Incorporates ROOT data visualization**
  - **Provides command line interface for batch processing.**

- **Rather than telling you about TMine, I'll show you...**

The LAT Event Analysis with TMine

Subsections of the Analysis



Other Applications of TMine

- **Measurement of the Electron Spectrum**
  - The LAT team has published the spectrum of cosmic-ray electrons from 7 GeV to 1 TeV
  - TMine was used to reprocess the large data set used for this analysis

- **Measurements of the Proton Spectrum**
  - Ongoing work here at SLAC
  - TMine used to distinguish protons from leptons and heavy nuclei.

- **Analysis of Unassociated Sources**
  - Ongoing work here at SLAC
  - Could TMine help to identifying some unassociated LAT sources?
  - More from Maria Elena Monzani

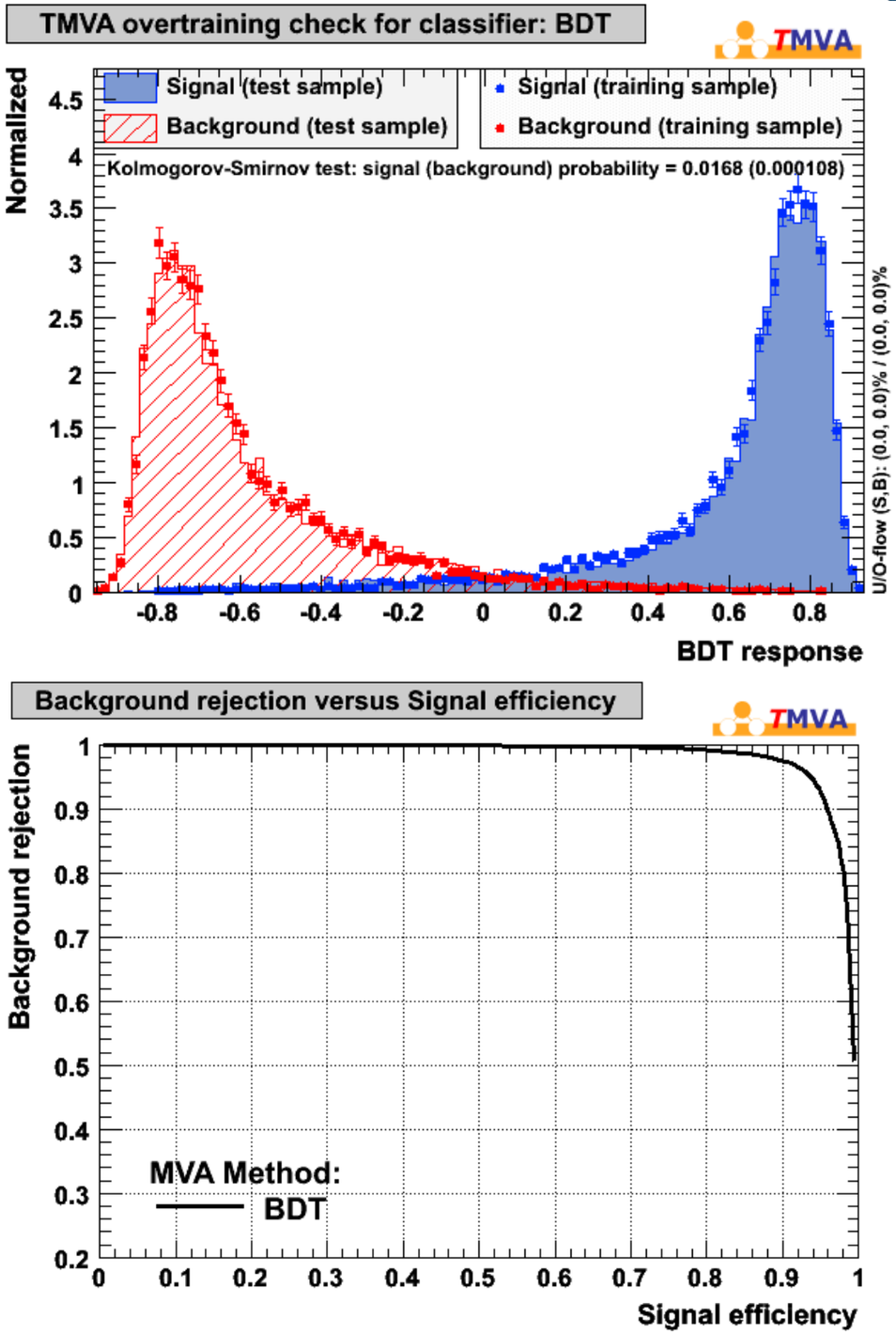


Figure - TMine implementation of the TMVA boosted decision trees to distinguish protons from electrons and positrons.

Summary

- **The problem that the Fermi-LAT faces is difficult**
  - **Cosmic rays out number photons by 10^4 to 1.**
  - **Events must be characterized over a wide range in energy and angle.**

- **Classification trees are crucial for achieving mission science goals.**

- **TMine is a new tool for implementing LAT classification analyses**
  - **Open source and ROOT-based**
  - **Implements powerful TMVA package**
  - **Provides a GUI for designing complex analyses**

- **The future of TMine**
  - **Contribute to the ongoing effort to improve the LAT event analysis.**
  - **Help extend multivariate analyses to high level LAT science.**

References

- W. Atwood. 2011. “The Utilization of Classifications in High Energy Astrophysics Experiments.” Private Communication.
- TMVA Users Guide. 2009. http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf
- R. Brun and F. Rademakers. 1997. Nucl. Inst. & Meth. in Phys. Res. A 389, 81-86.
- A. Hoecker, et al. 2007. PoS ACAT 040
- M. Ackermann, et al. 2010. “Fermi LAT Observations of cosmic-ray electrons from 7 GeV to 1 TeV”. Phys. Rev. D 82, 092003

Extra Slides

Toolkit for Multivariate Analyses

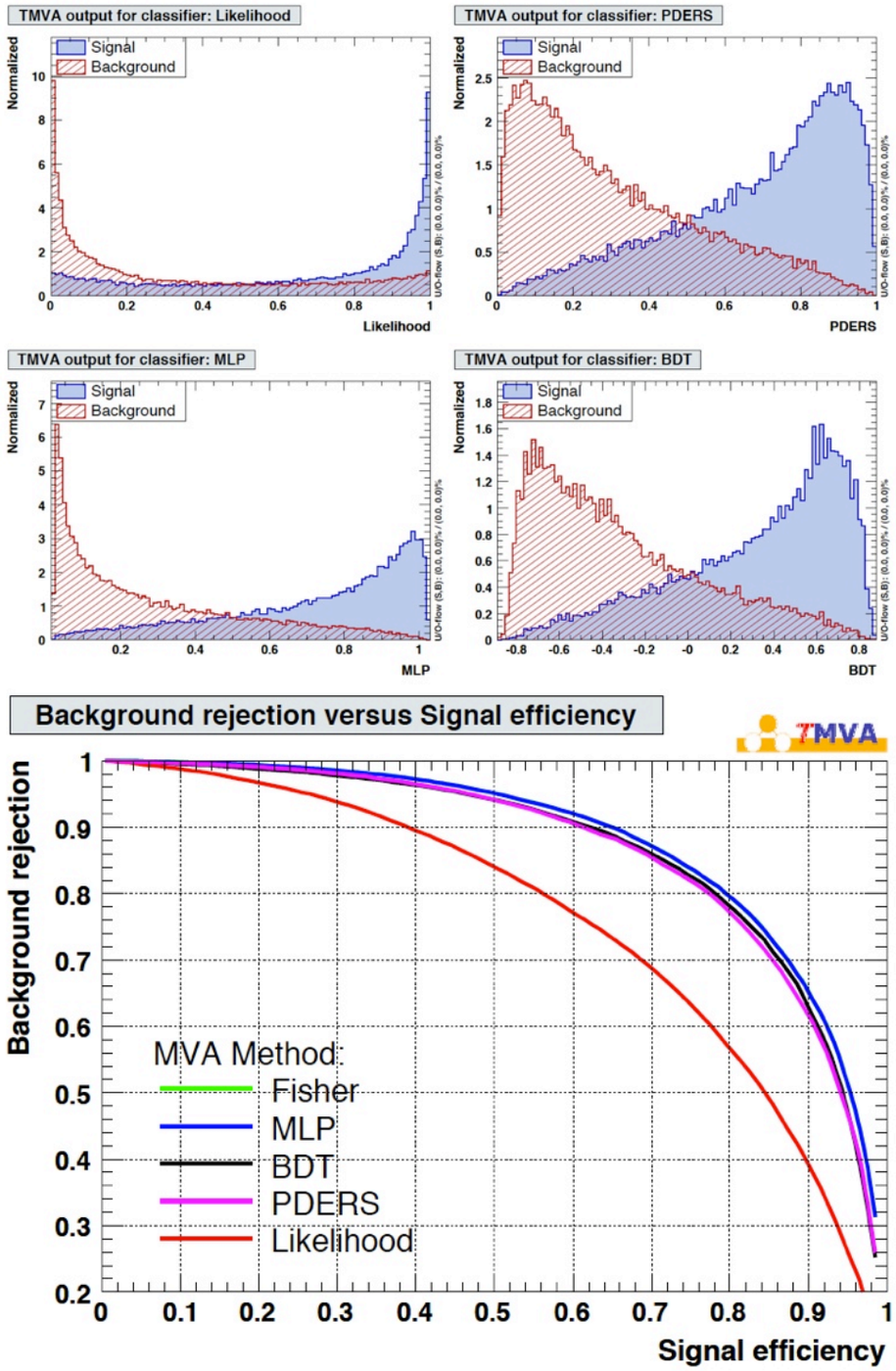


Figure: Toy comparison of some of the TMVA classifiers (TMVA Users Guide)

- **TMVA is the multivariate engine under the TMine hood.**
- **ROOT-integrated environment for processing and evaluating multivariate classification and regression techniques.**
- **Developed at CERN for signal discrimination in large data sets.**
- **Comes standard in current ROOT distributions with multiple classification algorithms fully implemented.**