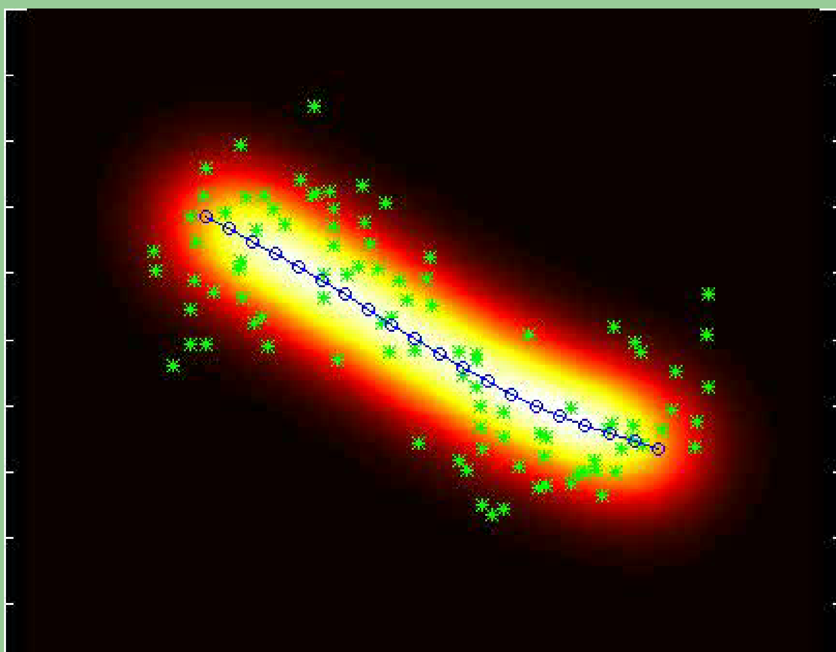
Finding New Physics in Run 2

Greg Landsberg

BROWN UNIVERSITY

JUNE 1, 2002



Advanced Analysis Tools Workshop at Fermilab

Outline

- New Physics' Many Faces
- Search Strategies
- Multivariate Techniques
  - Particle ID
  - Triggers
  - Signal/Background Discrimination
  - Combining Various Channels and Experiments
- Novel Approaches
- A Word of Caution
- Conclusions

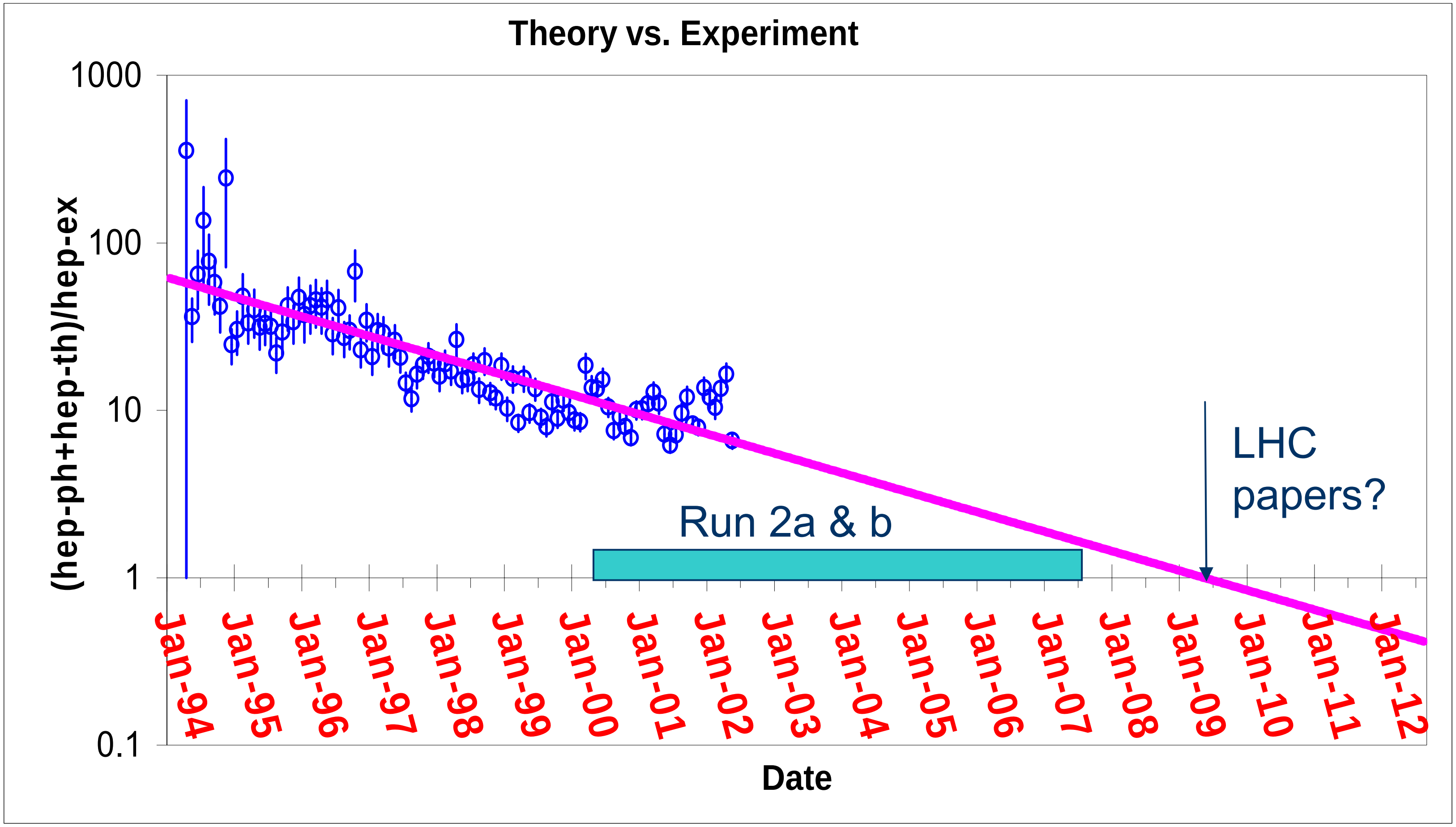
New Physics' Many Faces

- Electroweak/Heavy Flavor/QCD Physics:
  - Most properties are well known
  - Mainly precision measurements
- Higgs physics:
  - Most signatures are known
  - A few free parameters
  - Well-defined search strategy
- Searches for New Physics:
  - Do not even know what it is and where it is
    - Little doubt exists that it must be there, just around the corner
  - Only vague ideas about signatures
    - Many channels, many possibilities of statistical fluctuations
    - Might not be possible to use 'sister channels' as a discovery proof (e.g., RPV SUSY)
  - An ultimate prize and an ultimate challenge

New Physics' Many Faces, cont'd

- Supersymmetry
  - mSUGRA like
  - GMSB
  - $\tilde{G}$MSB
  - AMSB
- Strong Dynamics
  - Technicolor
  - Top See-Saw
  - Compositeness
- Extra Dimensions
  - Large Extra Dimensions
  - Infinite Extra Dimensions
  - Universal Extra Dimensions
- Exotics
  - Leptoquarks
  - Extra gauge bosons
- Something COMPLETELY unexpected

Pick Your Model! – or Maybe Not?



- Do not expect to test EVERY model - we will work with CHANNELS, not MODELS!

Hunting Ground

l+jets+$\not\!E_T$

ll+jets+$\not\!E_T$

jets+$\not\!E_T$

bb+γ

taus+$\not\!E_T$

bb+$\not\!E_T$

Massive Stable Particles

ll+bb+$\not\!E_T$

ll+jets

γ+jets+$\not\!E_T$

Undetectable

$llll$+$\not\!E_T$

Kinks

monojets

γ+ll+$\not\!E_T$

lll+jets+$\not\!E_T$

Non-prompt photons
or Z's

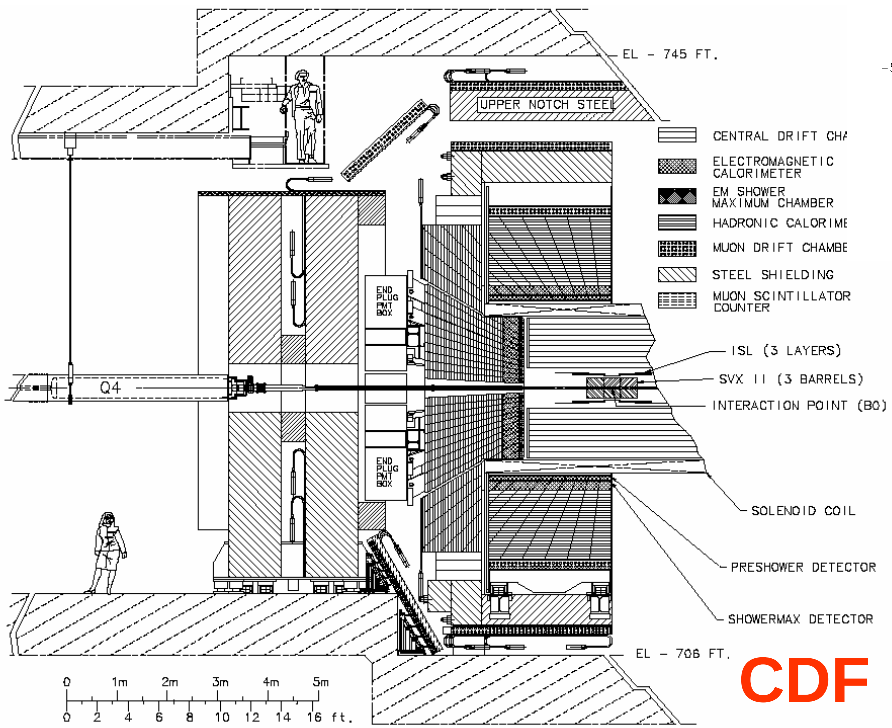
$\gamma\gamma+E_T$

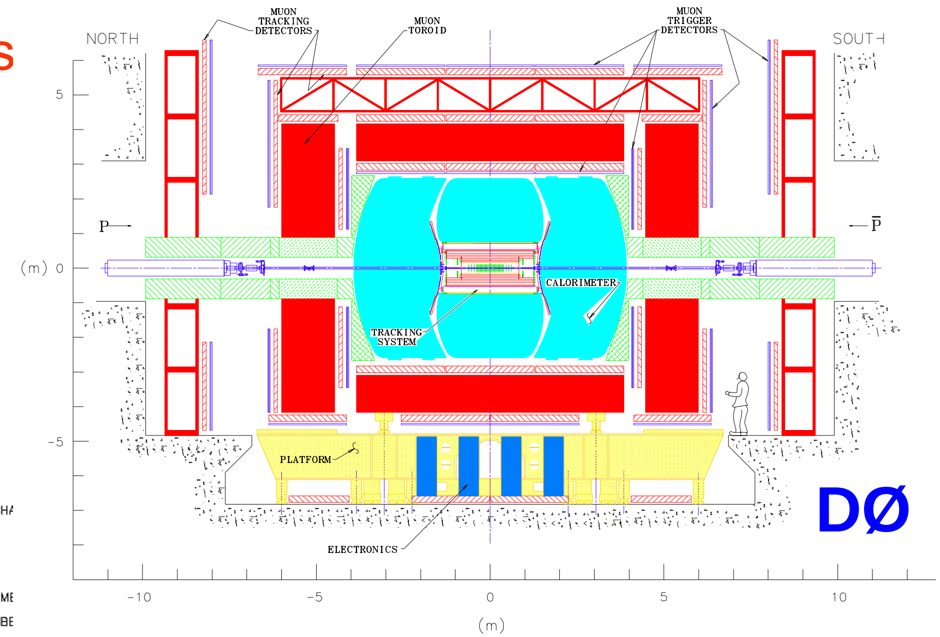
It's a MESS!

Run 2 Detectors

Significantly upgraded experiments with three major subdetectors:

- Hermetic Calorimeters
- Central Tracker
- Muon Detectors





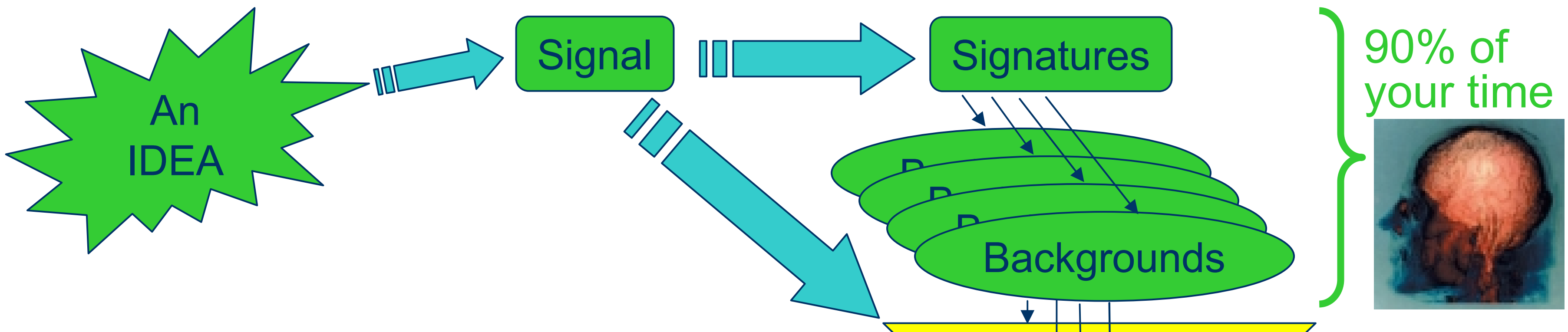
Multiple detectors – multiple variables:

- EM: calorimetry, preshowers, (tracking, ionization)
- Muons: central and outer tracks, ionization, calorimetry
- τ's: tracking, calorimetery
- Jets: tracking, calorimetry

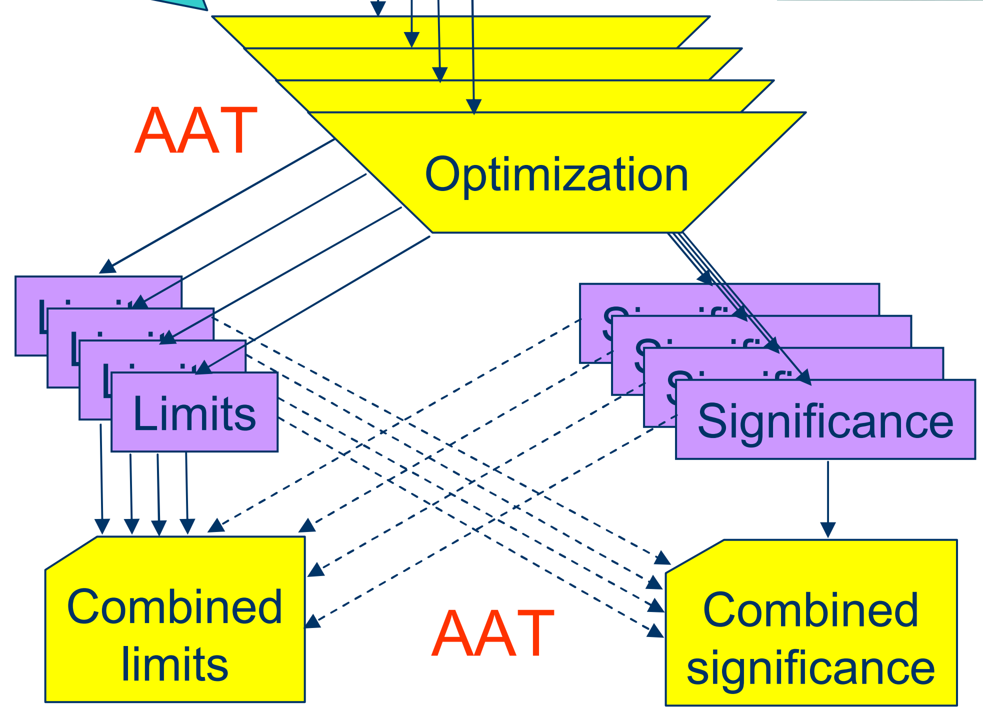
Importance of Advanced Methods

- It's likely that NP will manifest itself as a marginal event excess in a number of related channels
  - Need for an accurate combination of various channels to reach the discovery significance (cf. Higgs)
- It's likely that NP will be overwhelmed by the SM backgrounds in the channels of interest
  - Need for advanced methods of background suppression, while retaining high sensitivity for the signal
- It's possible that NP searches would require special triggers
  - Need for fast triggering methods in a complicated environment
- It's given that with the complexity of modern detectors, particle ID is based on a number of variables
  - Need for advanced multivariate particle ID techniques

A Typical Search for New Physics

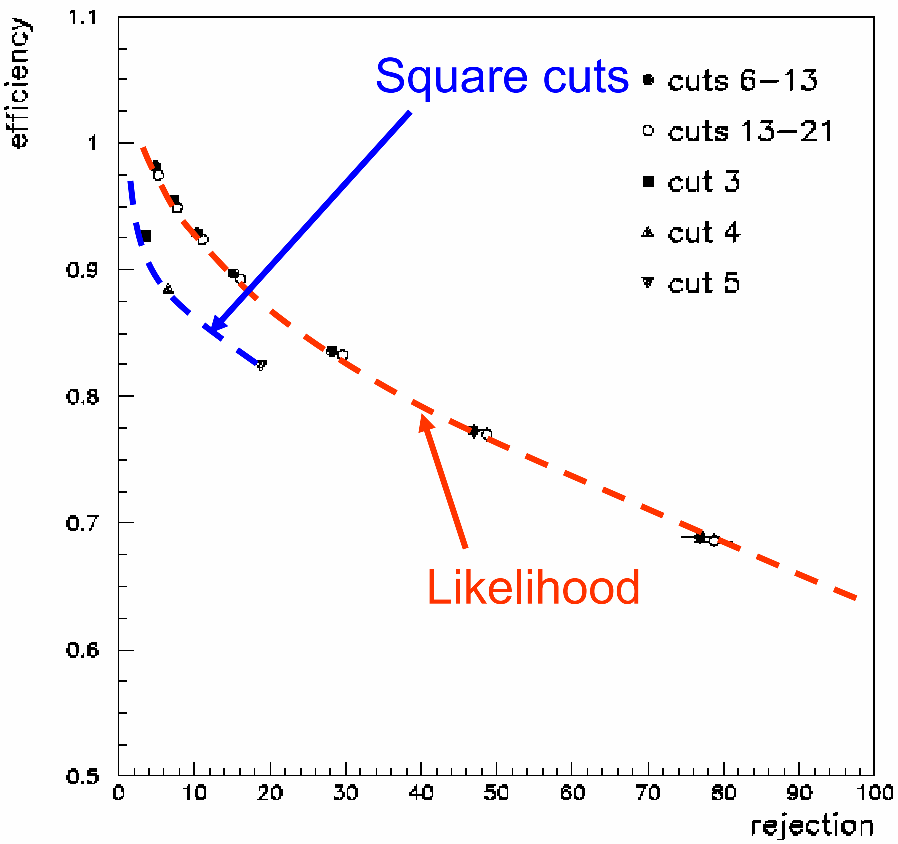


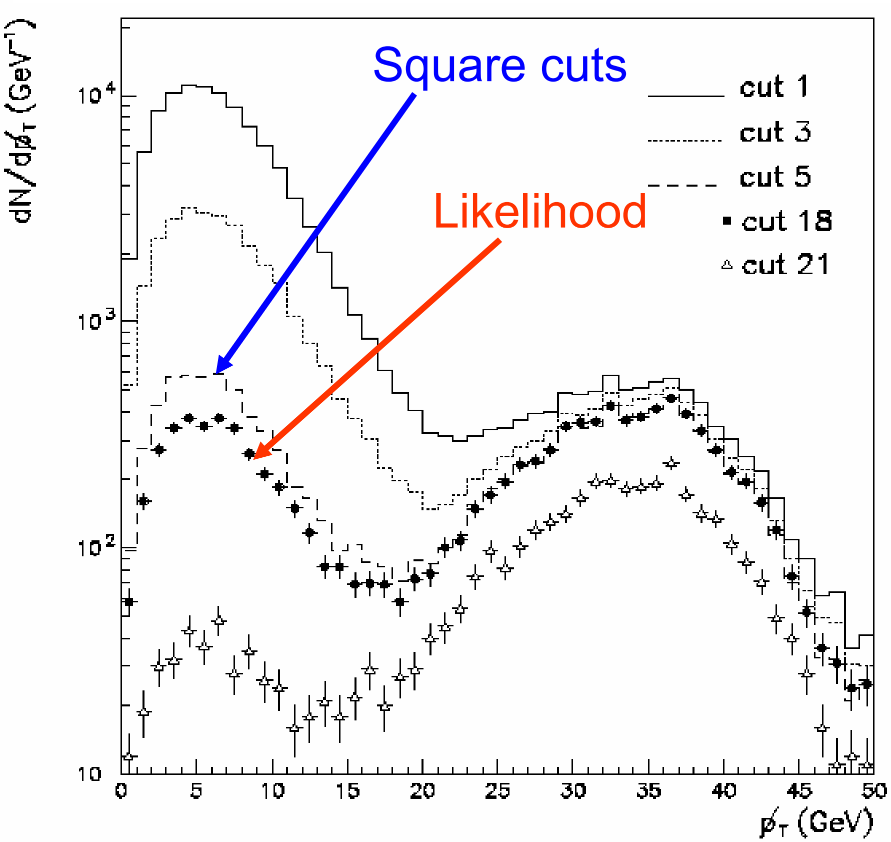
- From a (bright) idea to a limit (discovery):
  - 90% of the time is spent in picking the right signatures and understanding of the backgrounds; AAT are but useless here
  - Only 10% of your time is spent on optimizations, where AAT help
- Don't overuse AAT!



Multivariate Particle ID: Electrons

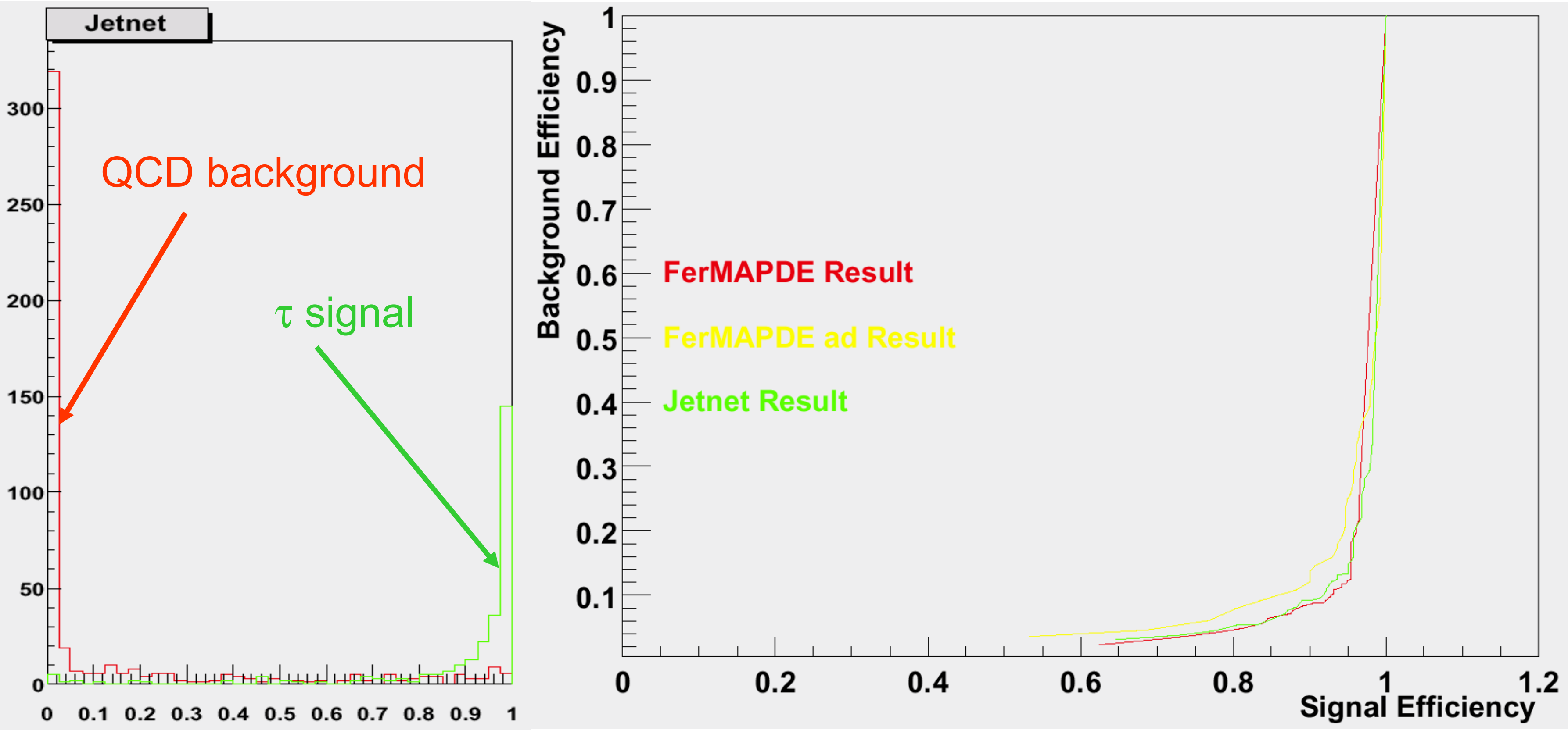
- Example: DØ Run 1, electron likelihood, based on four variables: track match, cluster shape, ionization, and EM fraction
- Gives twice the QCD background rejection per constant efficiency, compared to rectangular cuts





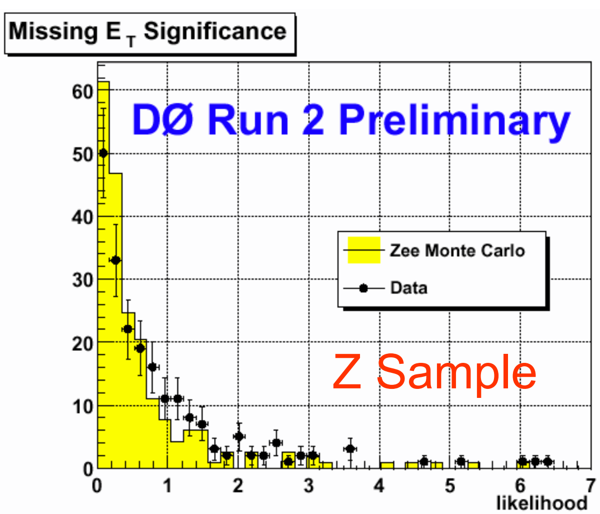
Multivariate Particle ID: Taus

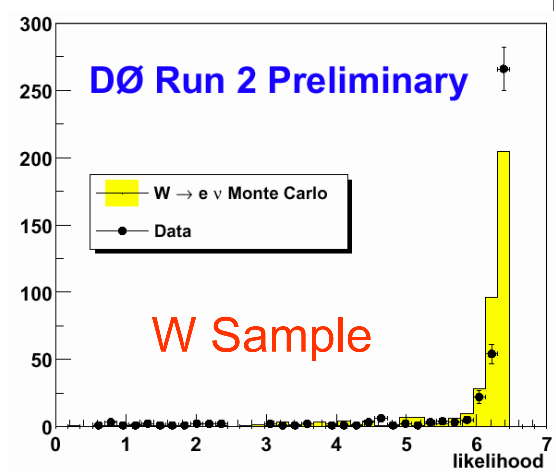
- Example: DØ Run 2, one-prong tau identification via a neural net or PDE techniques
- Gives 30-40% better QCD background rejection per constant efficiency, compared to rectangular cuts

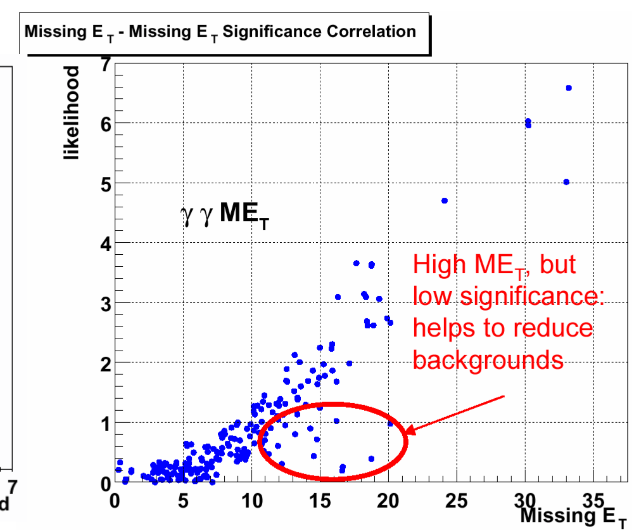


Missing E_T Significance

- Another example of using advanced techniques for particle ID is missing E_T significance (prime SUSY signature):
  - Takes into account PDE for jet energy uncertainties, misvertexing probability, and the probability of a hot cell occurrence to calculate the significance of missing transverse energy on event-by-event basis (i.e., by taking into account the event topology)
  - Gives twice the QCD background rejection for a given efficiency, compared to MET isolation cuts
  - Already being used in DØ Run 2 analyses!

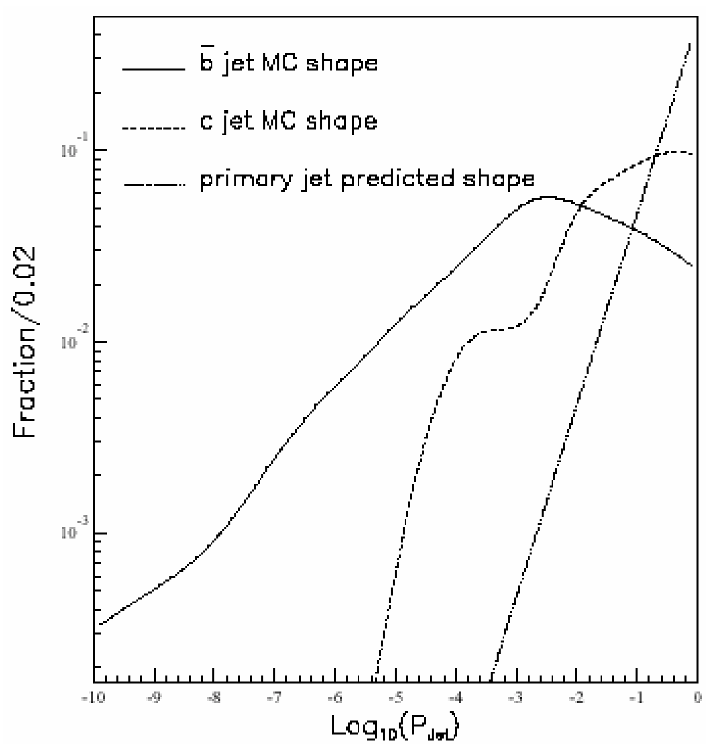


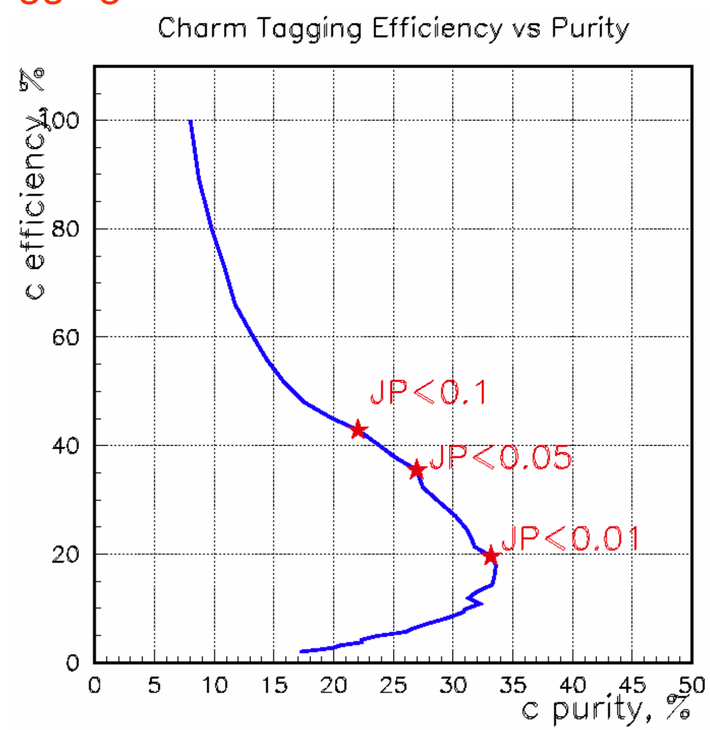




Bottom and Charm tagging

- Multivariate techniques can be used to tag jets coming from charm and bottom quarks
- Example: jet probability used in CDF Run 1 analyses (stop, sbottom, LQ2, LQ3 searches)
  - Define track probability as a measure of track consistency with the primary vertex
  - Define jet probability as combination of track probabilities for all tracks
  - Use this variable for bottom and charm tagging

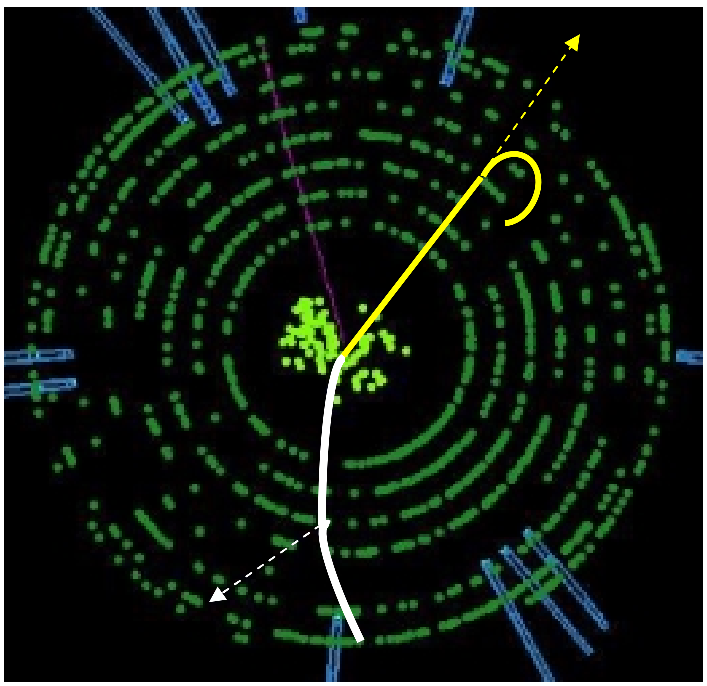


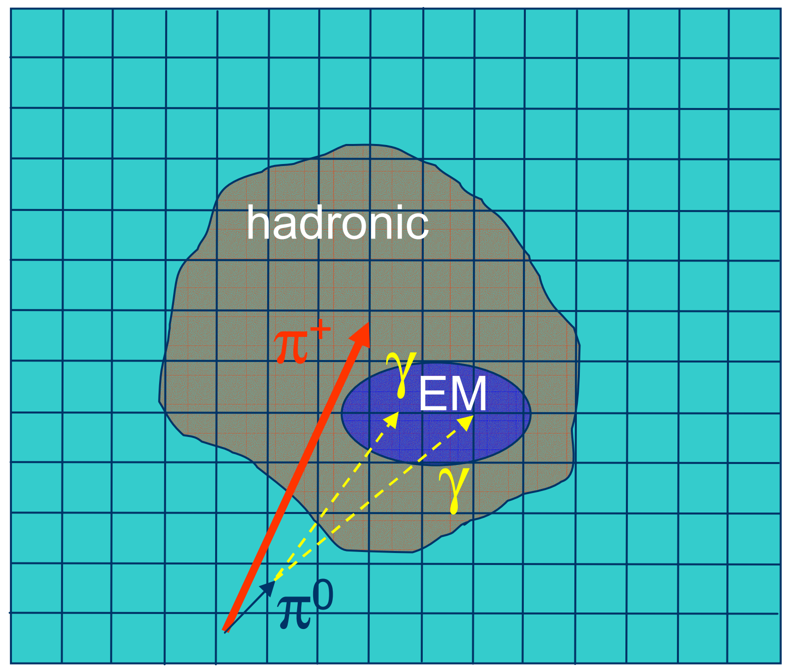


AAT Workshop, June 1, 2002

Triggering and AAT

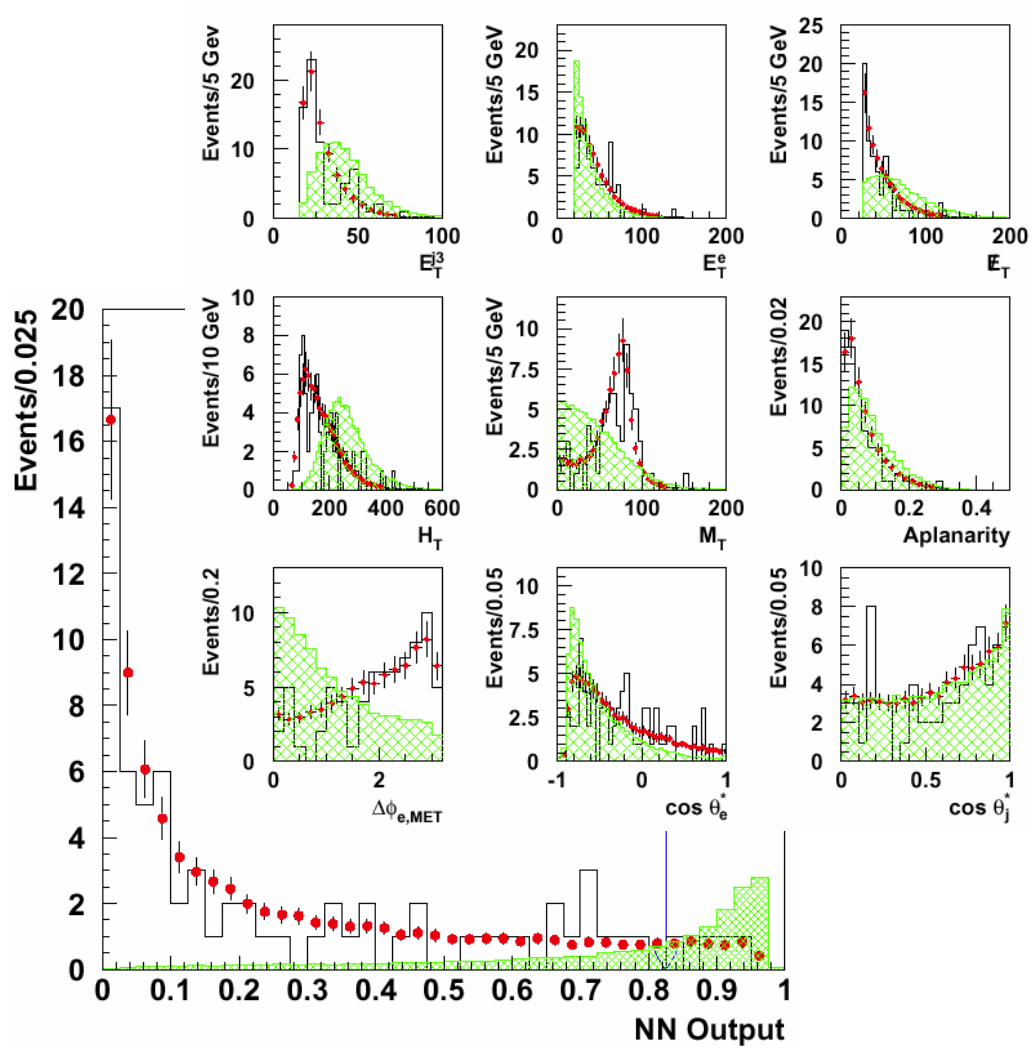
- Although we do not use AAT in the triggers yet, it could be very beneficial, both at the hardware and software trigger levels
- Examples:
  - Looking for kinks or stopped tracks due to long-lived charged particles (e.g., AMSB models with highly-degenerate chargino-neutralino)
  - Identifying $\tau^{\pm} \to \rho^{\pm}\nu_{\tau} \to \pi^{\pm}\pi^{0}\nu_{\tau}$ decays by looking for jet substructure





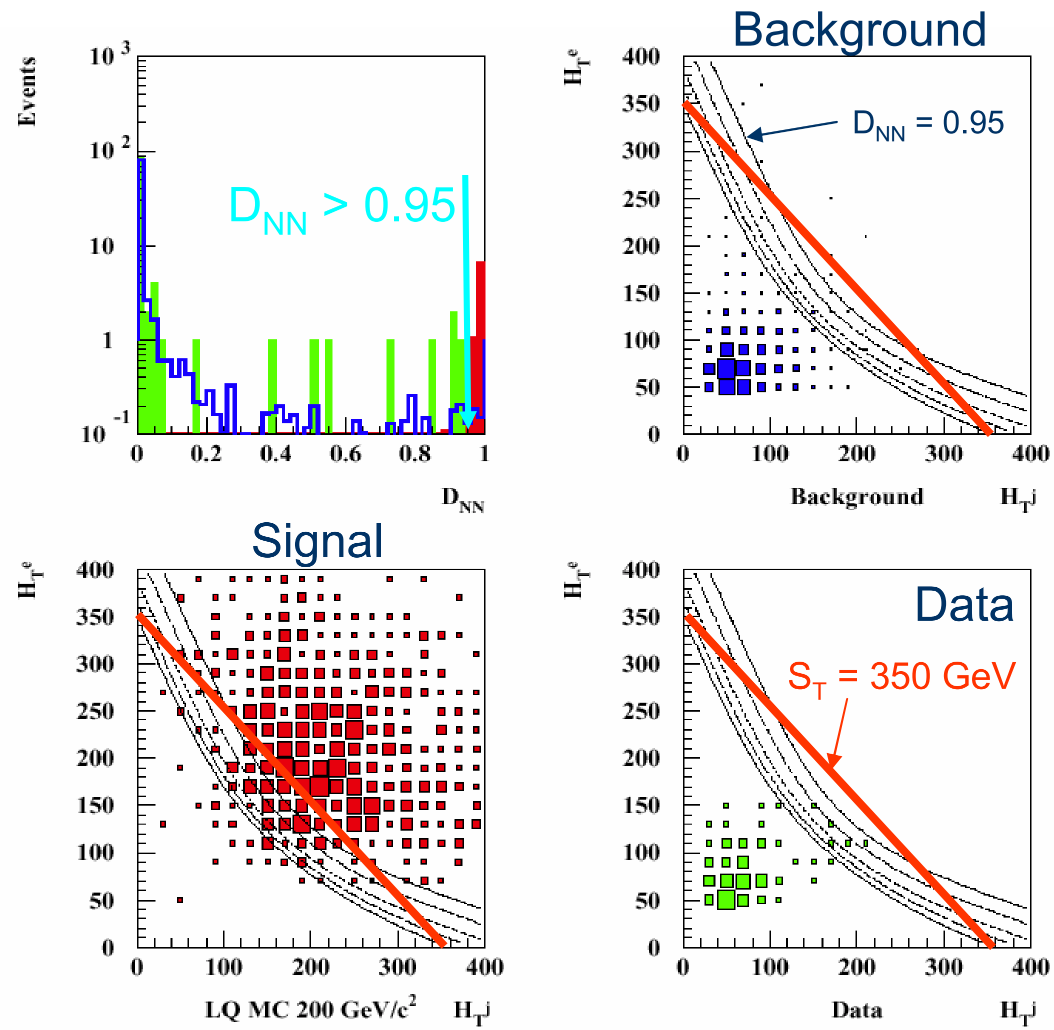
Optimization: Here AAT Help…

- Multivariate techniques are excellent help when:
  - Backgrounds are large and so are signals (e.g., SUSY in jetty channels)
  - Several backgrounds sources with significantly different kinematics
  - There are no mass peaks to look for, i.e. an excess due to new physics is not easily localized in phase space
- An example: DØ Run 1 search for SUSY in the e+jets+ME$_T$ channel
  - 9 variable network
  - No "killer" variable
  - W+jets, top, QCD backgrounds
- Many jetty SUSY/Technicolor channels in Run 2 would benefit from NN



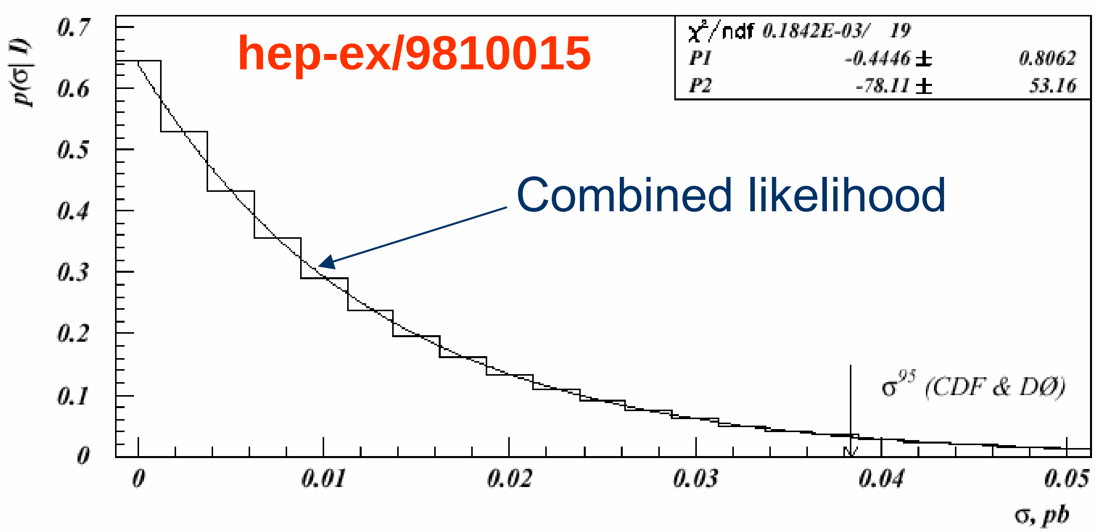
And Here They Do Not…

- One does not need multivariate methods if:
  - Backgrounds are low (SUSY in trileptons)
  - There is a single dominant background with the kinematics significantly different from that for signal
  - There is a single physics background, similar to that for signal (QCD $\gamma\gamma$ background for large extra dimensions search)
  - Signal is localized in the phase space (e.g., a mass peak)
- Example: DØ Run 1 LQ1 search in the eejj channel:
  - Only a marginal improvement compared to a simple one-variable cut



Combination of Various Channels

- Often have to combine several channels and experiments to improve significance
  - Various channels typically have quite different signal-to-background ratios, hence cannot combine spectra
  - Have to take into account correlated errors
  - For combination of several experiments its desirable to “hide” experiment-specific information
- Solution: use Bayesian-likelihood-based combination, with correlated and uncorrelated errors treated separately



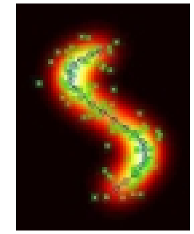
EXAMPLE:

Combined CDF and DØ LQ1 cross section limit:

σ^{95}(CDF) = 0.066 pb

σ^{95}(DØ) = 0.099 pb

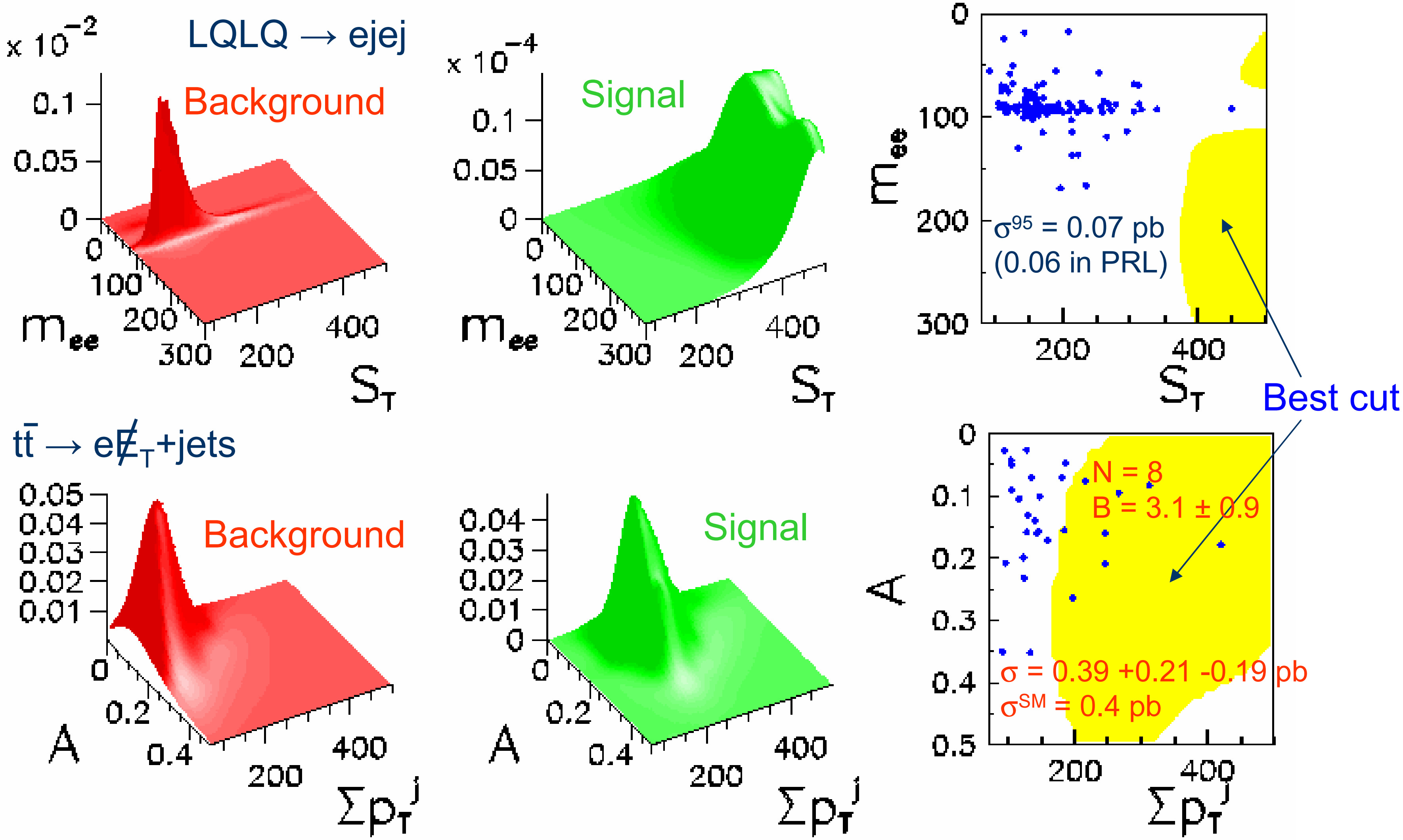
σ^{95}(TeV) = 0.038 pb



QUAERO: Quick-and-Dirty Model Tester

- QUAERO [DØ, PRL **87**, 231801 (2001)] is an attempt to automate the final step of a physics analysis: testing the data against a particular physics model
  - Based on PDE's for signal and background in a multivariate user-defined space
  - Finds the area in the multivariate space that maximizes signal significance; similar to NN
  - Capable of fast generation of user-specified process via PYTHIA with a built-in fast detector simulation
  - Allows to set quick-and-dirty limits on a given process in a matter of minutes and test different variables to discriminate between signal and background
  - Largely ignores systematic errors when setting limits
  - An interesting way of making experimental data public

QUAERO Examples

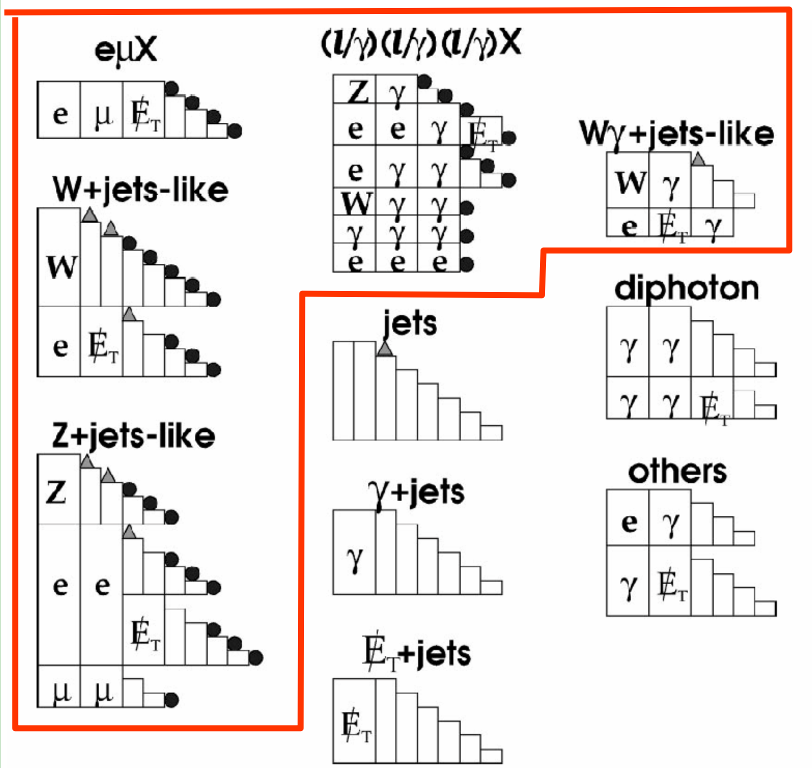


Grand Finale: Have We Missed Anything?

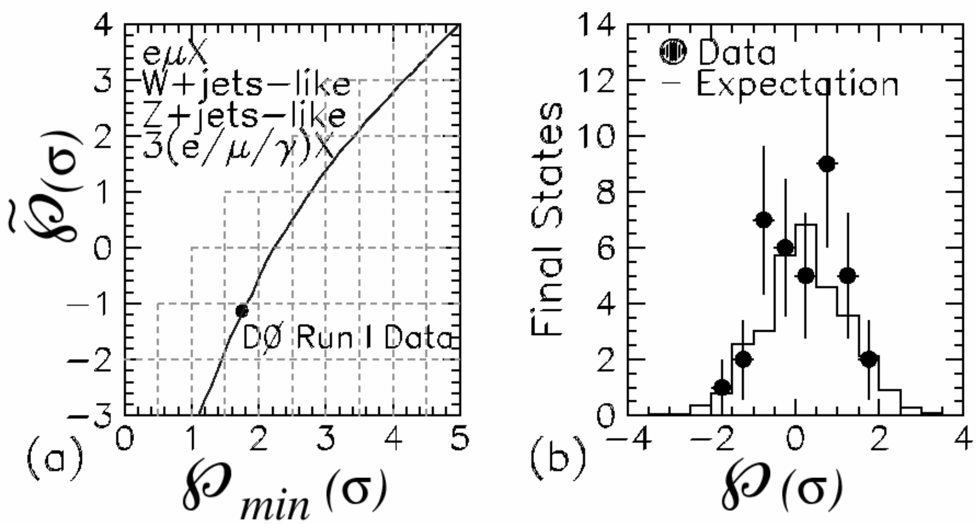
- SLEUTH (formerly known as SHERLOCK) [DØ, PRD **62**, 92004 (2000); PRL **86**, 3712 (2001); PRD **64**, 012004 (2001)] is an attempt to use multivariate techniques to answer the above question
  - Generalizes search for high-p_T physics by identifying global variables, correlated with the p_T of a process, and then looking for “reasonable” contiguous areas in the resulting multivariate space yielding maximum excess of data above the SM background
  - Unusual, data-driven approach; an interesting method to answer the question of whether there was an evidence for new physics in the data *posteriori*
  - Takes into account the fact that search for excesses was performed in many different variables and channels by adjusting the probability accordingly; consequently, trades decreased significance for increased generality
  - Used by DØ to reanalyze Run 1 data in a semi-model-independent way and quantify the degree of the agreement between the data and the SM predictions in some 30 previously studied channels

DØ Run 1 Sleuth Results

- Sleuth is a nice tool to use before closing the door and shutting off the lights: have we forgotten or missed anything
- It crucially relies on background understanding and won't help at this (most time-consuming) step
- Not very competitive with direct search for a known signature (e.g. mass peak)

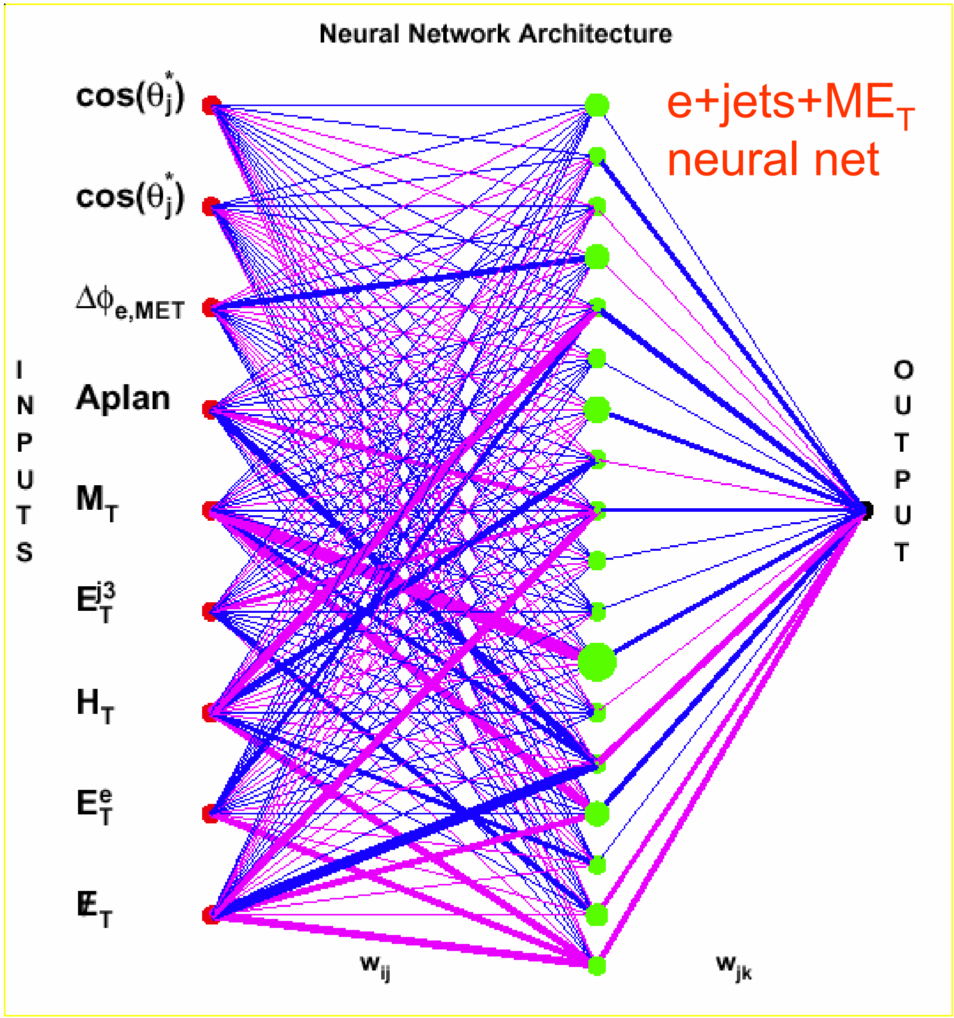


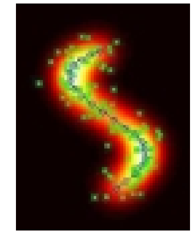
$\tilde{P}(\sigma) = -1.23\sigma \rightarrow \tilde{P} = 0.89$,
i.e. 89% of hypothetical experiments would have seen more "interesting" results in these 32 channels than DØ



A Word of Caution

- It's very dangerous to use NN (or other multivariate techniques) as black boxes
  - GIGO principle
  - Cf. Israeli Army example
- Representative and large training samples are crucial
- Even complicated neural nets can be "opened" up
- Use of two-dimensional projections and other graphic tools to make sure that the NN acts reasonably





Conclusions

- Advanced Analysis Techniques will be important in Run 2 at various levels:
  - Fast pattern recognition at the trigger level
  - Advanced particle ID
  - Maximization of the signal significance via optimal variables and set of cuts
- However, AAT are not a panacea; they are just useful tools:
  - They are only as smart as you are
  - They won’t help to understand the backgrounds
  - If used as a black box, they could be dangerous
- The best neural net is the one between your ears – use it, and you shall succeed!