Advanced Event Analysis Methods at the Energy Frontier

Reinhard Schwienhorst

Boston University Particles & Fields Seminar, 12/7/2005

Outline

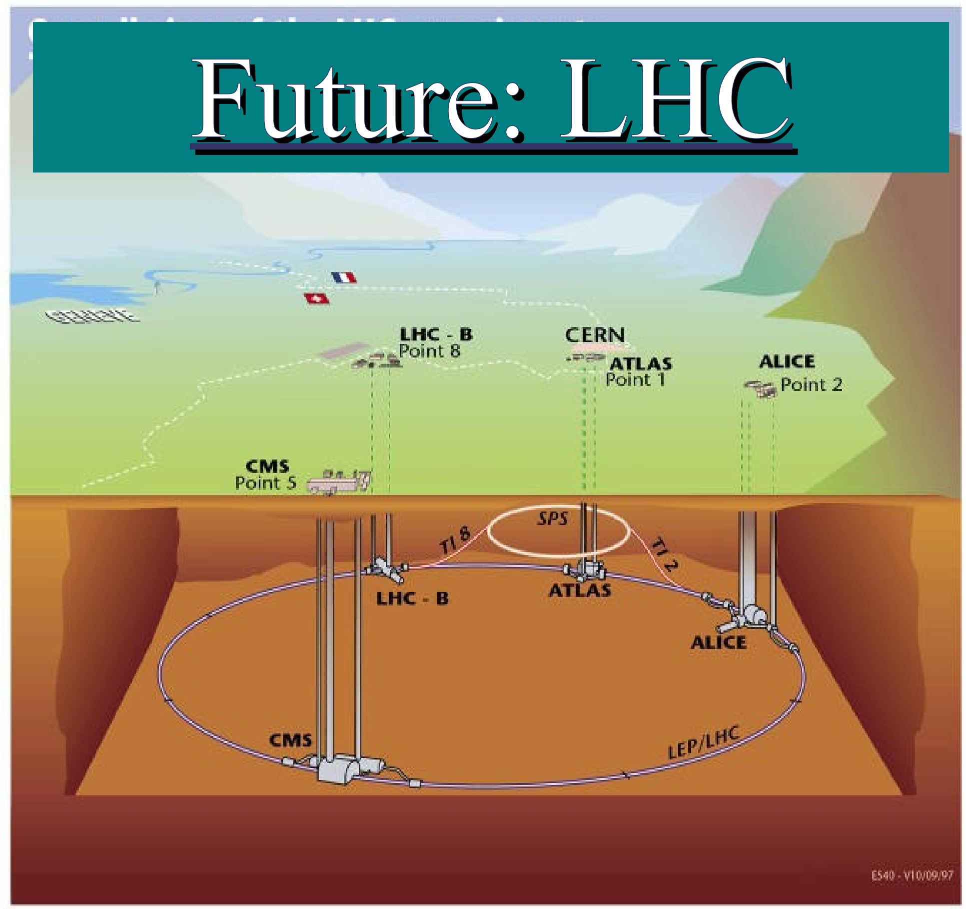
- Introduction
  - Physics at the energy frontier
- Analysis procedure
- Event Analysis methods
  - Neural Network
  - Decision Tree
  - Boosting
  - Bayesian Limit
- Conclusions

Disclaimer

- Focus on the energy frontier
  - Problems of low statistics
- Example: single top quark search at the Tevatron
  - My area of expertise
- Bayesian statistics where applicable
  - I am a statistical philosophy agnostic
  - Methods presented here are independent of which philosophy is followed
- Focus on general principles and guiding ideas
  - Intuitive procedure, not necessarily mathematically rigorous

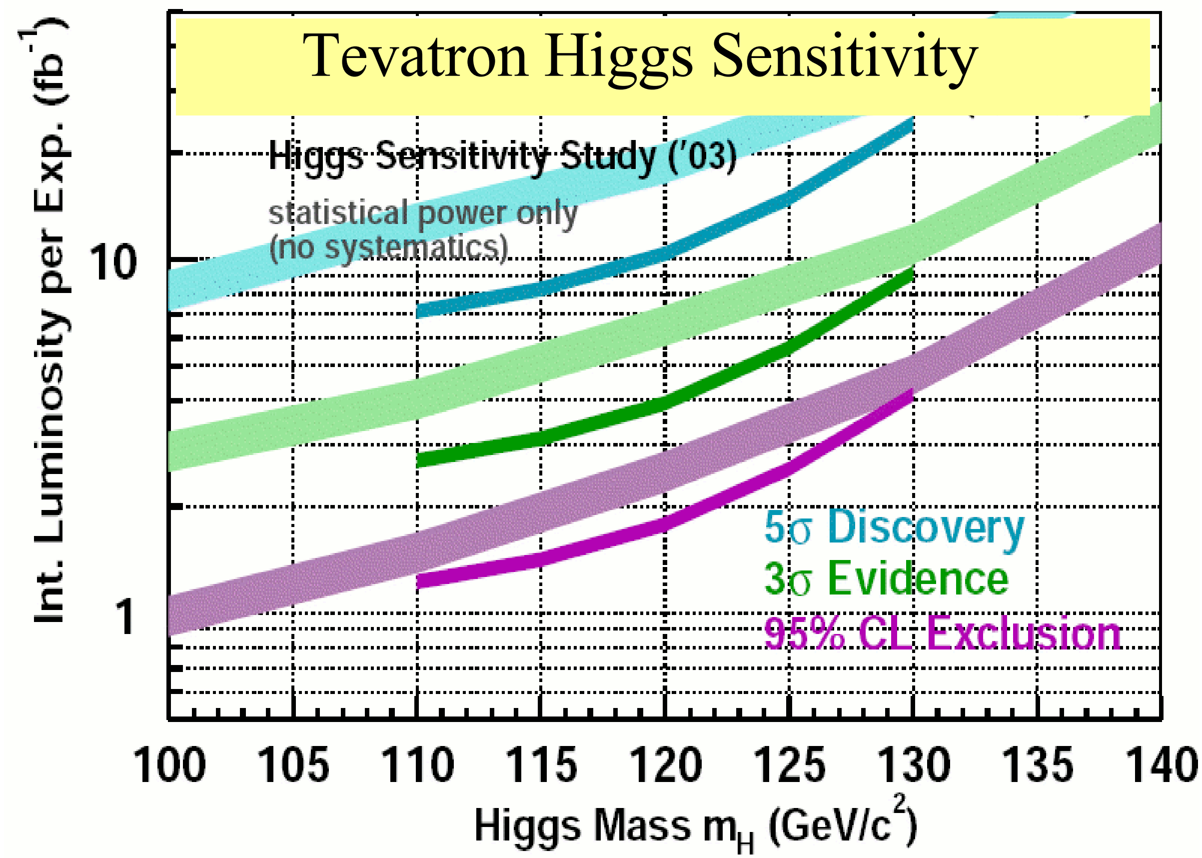
The Energy Frontier

- Colliding particles at the highest available energies
    - Probe structure of matter at the most fundamental level
  - Observe interactions at the smallest possible distances
  - Produce never-before-seen particles



Searches at the Energy Frontier

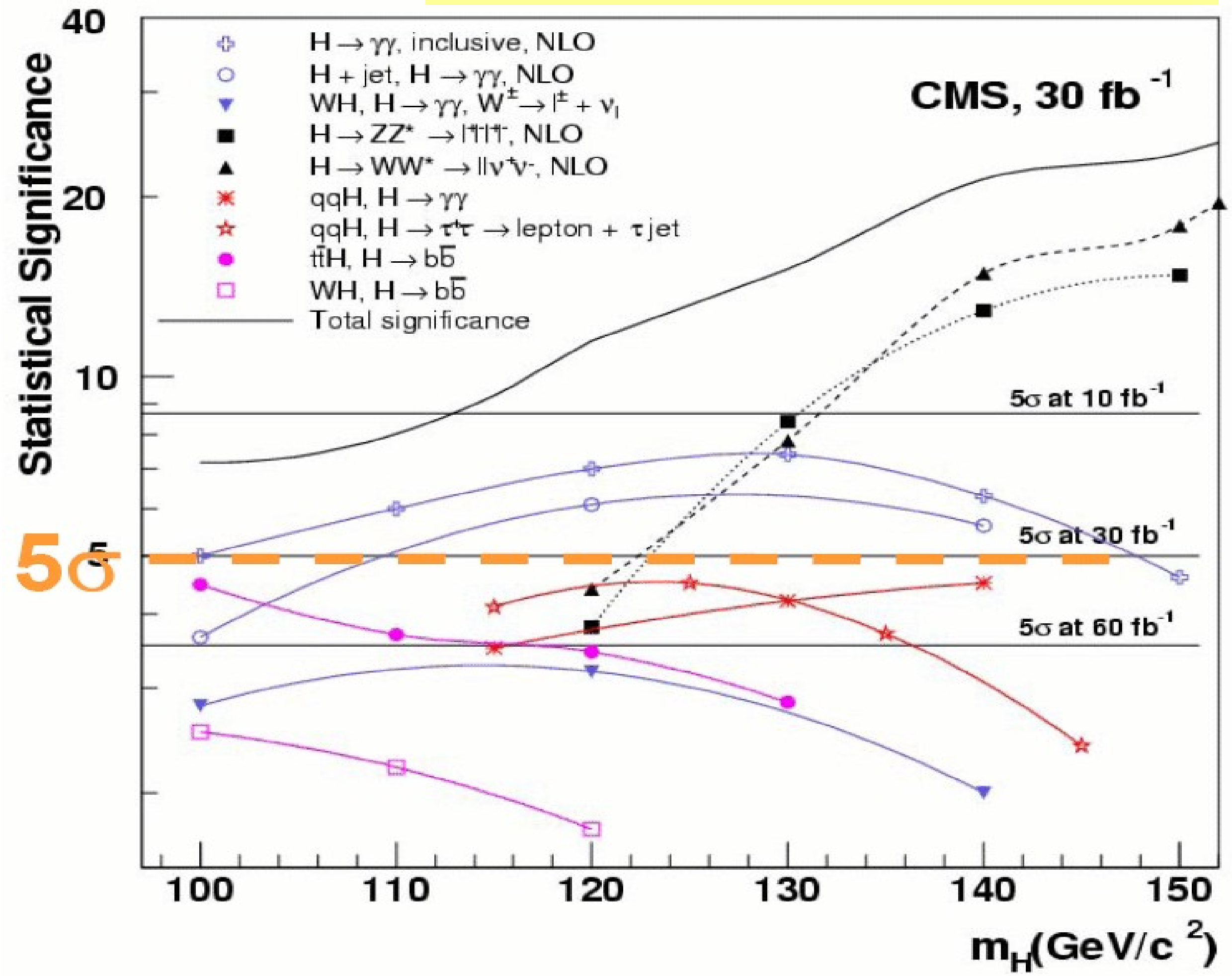
- Searches for new particles, phenomena, couplings
  - Tevatron:
    - Single top quark production
    - Higgs boson search
    - SUSY
    - Extra dim
    - …



Searches at the Energy Frontier

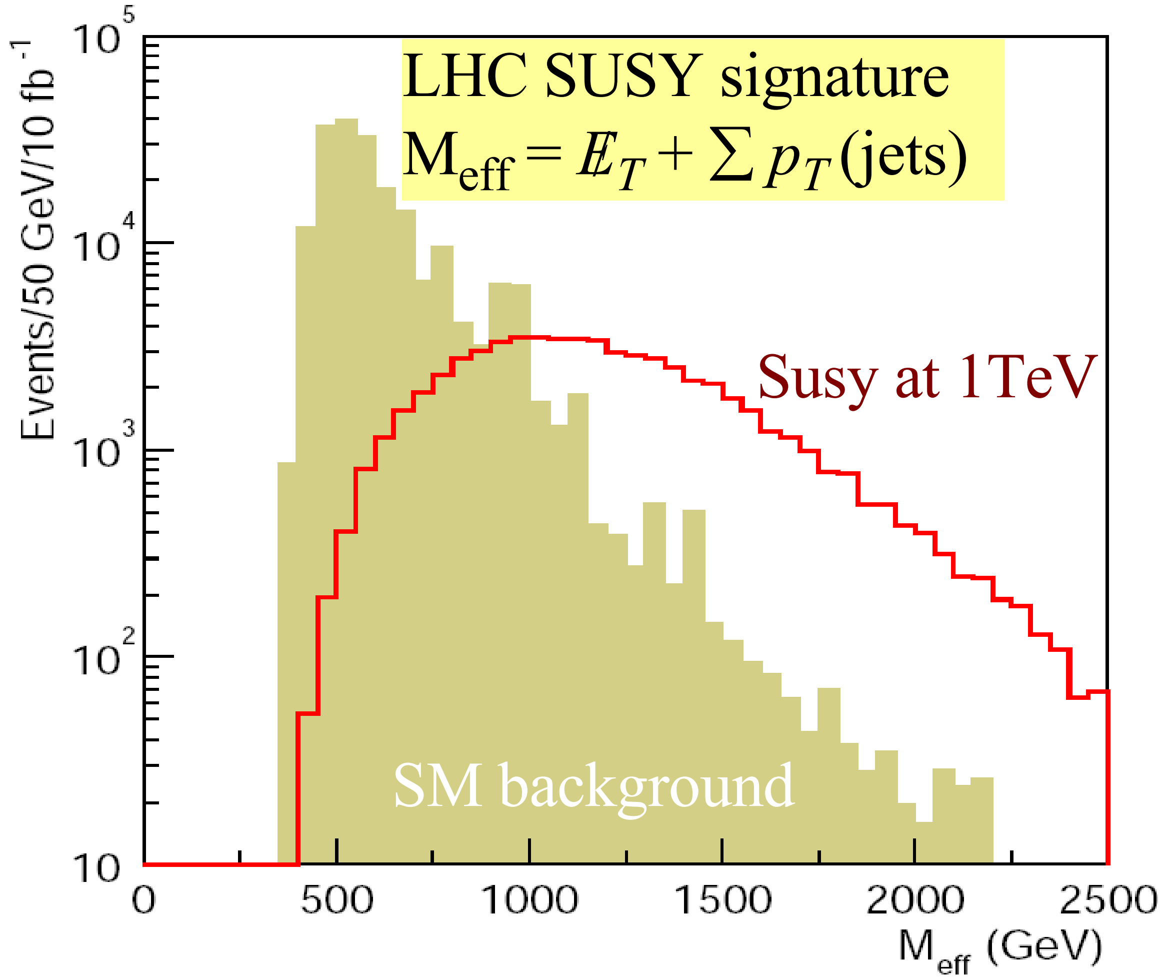
- Searches for new particles, phenomena, couplings
  - Tevatron:
    - Single top quark production
    - Higgs boson search
    - SUSY
    - Extra dim
    - …
  - LHC:
    - Higgs search

LHC Higgs Sensitivity



Searches at the Energy Frontier

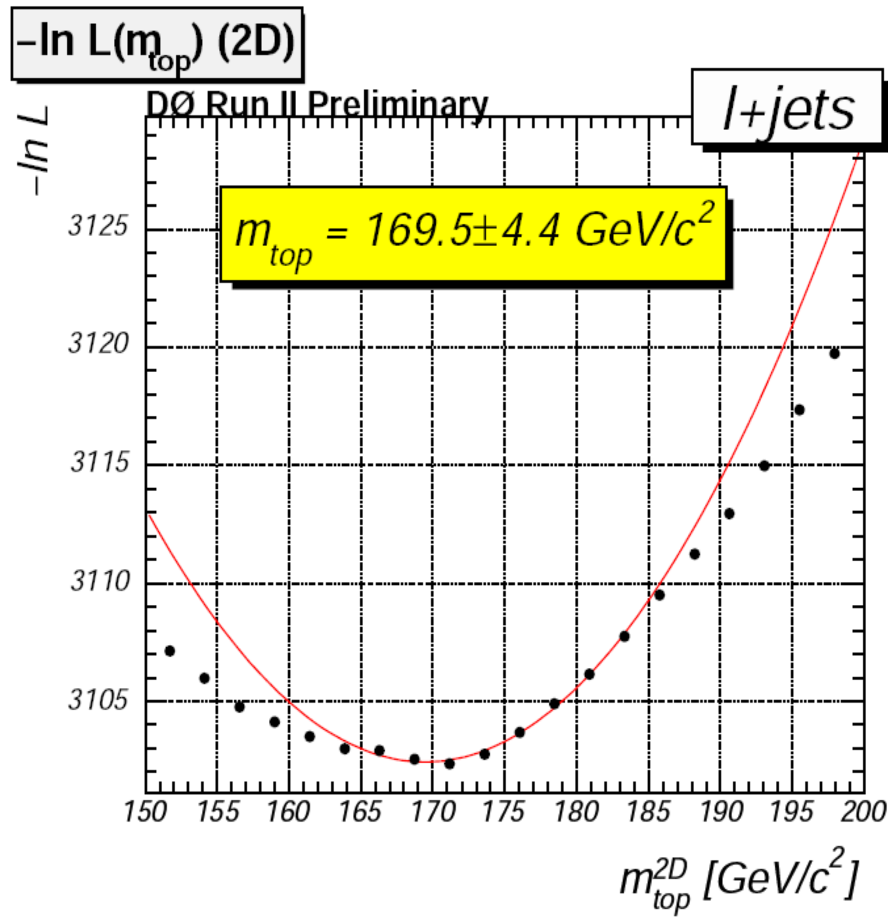
- Searches for new particles, phenomena, couplings
  - Tevatron:
    - Single top quark production
    - Higgs boson search
    - SUSY
    - Extra dim
    - …
  - LHC:
    - Higgs boson search
    - SUSY
    - Extra dim
    - …



Measurements at the Energy Frontier

- First measurements of properties, couplings
  - With samples of limited size
  - Example:
    Tevatron top quark mass

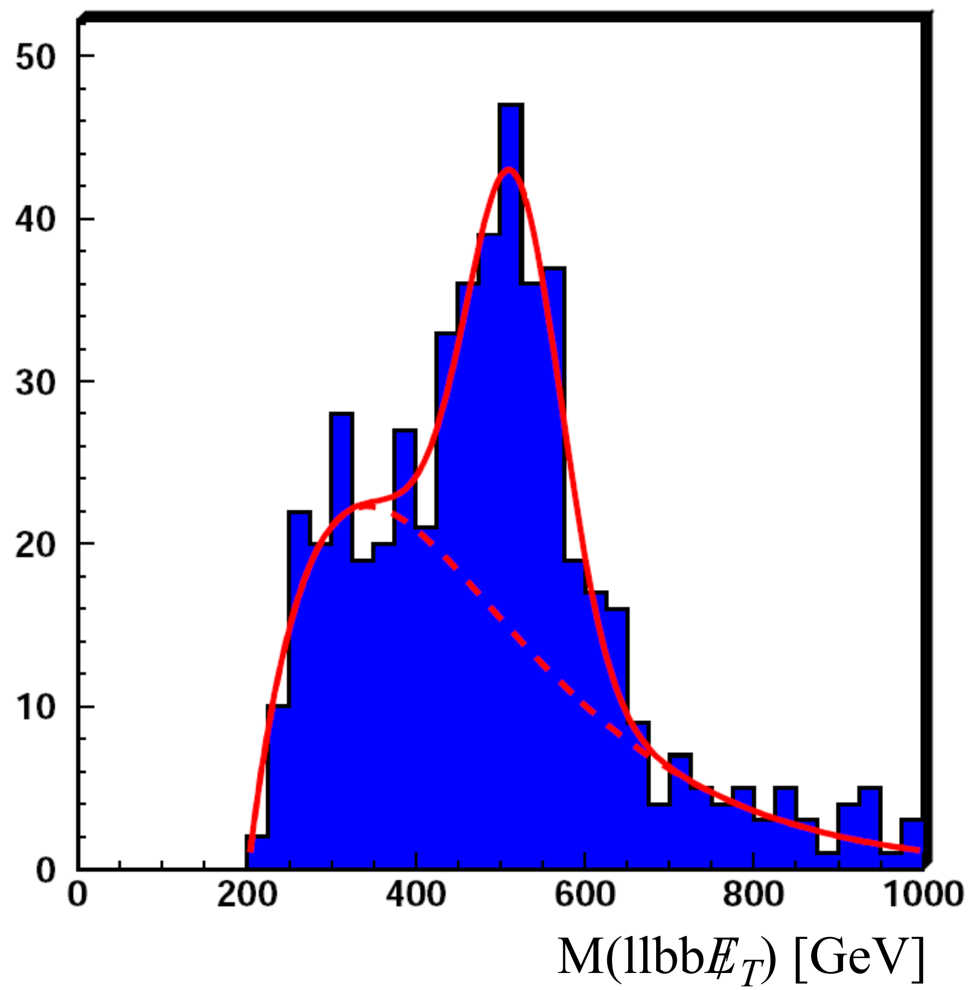
DØ top mass based on 55 top quark pair events in 0.3 fb^{-1}



Measurements at the Energy Frontier

- First measurements of properties, couplings
  - With samples of limited size
  - Example:
    LHC Susy particle masses

LHC $\tilde{b}$ mass:
100 signal
events in 30 fb^{-1}

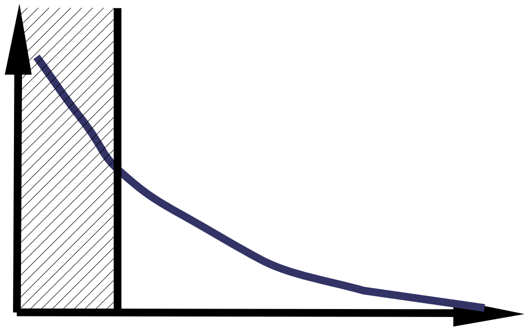


Physics at the Energy Frontier

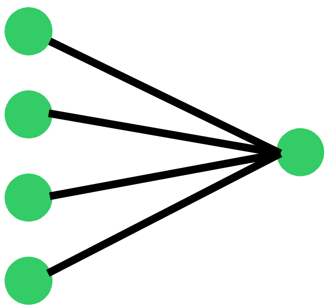
- Searches for new particles, phenomena, couplings
- First measurements of properties, couplings

Making the most out of small samples of events

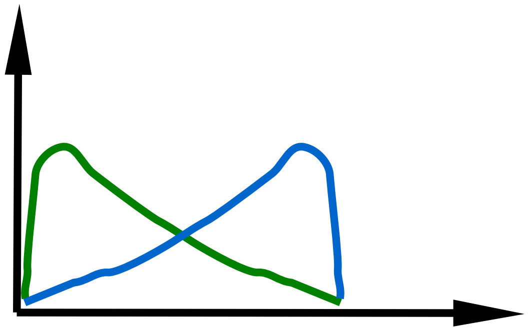
Analysis outline



1. Event selection
   – Object identification
   – Background modeling

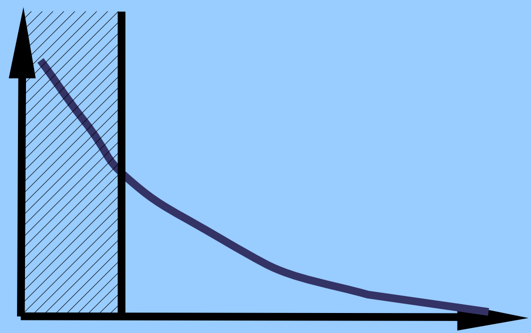


2. Event analysis
   – Discriminating variables
   – Cut/combine in multivariate analysis

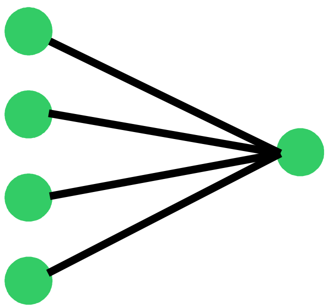


3. Statistical analysis
   – Measurement with uncertainty
   – Confidence limit

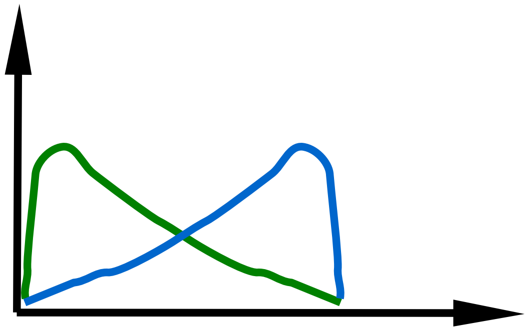
Analysis outline



1. Event selection
   - Object identification
   - Background modeling



2. Event analysis
   - Discriminating variables
   - Cut/combine in multivariate analysis

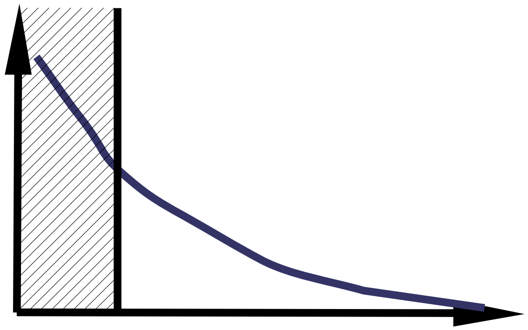


3. Statistical analysis
   - Measurement with uncertainty
   - Upper/Lower confidence limit

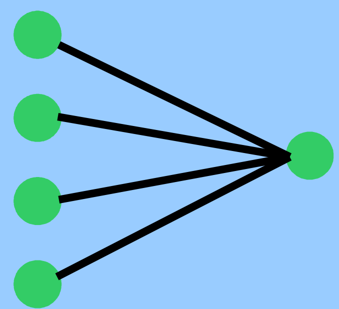
Event Selection

- Select events with specific final state objects
  - Object ID, remove mis-reconstructed events
  - Cuts on p_T and η for leptons, jets, MET
  - Specific numbers of leptons, jets
  - Possibly b-quark tagging
- Figure out possible SM backgrounds
  - Any process resulting in same final state
- Compare background sum to observed data
  - In background-dominated sample
  - In additional samples where no signal is expected
  - Compare total event counts
  - Compare shapes of important distributions

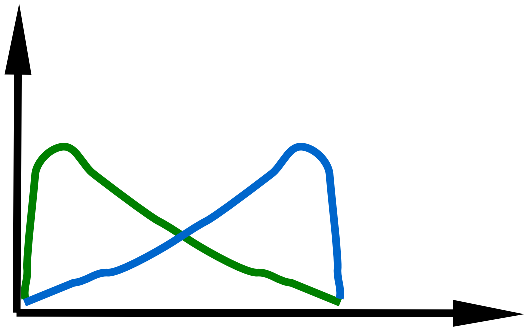
Analysis outline



1. Event selection
   – Object identification
   – Background modeling



2. Event analysis
   – Discriminating variables
   – Cut/combine in multivariate analysis



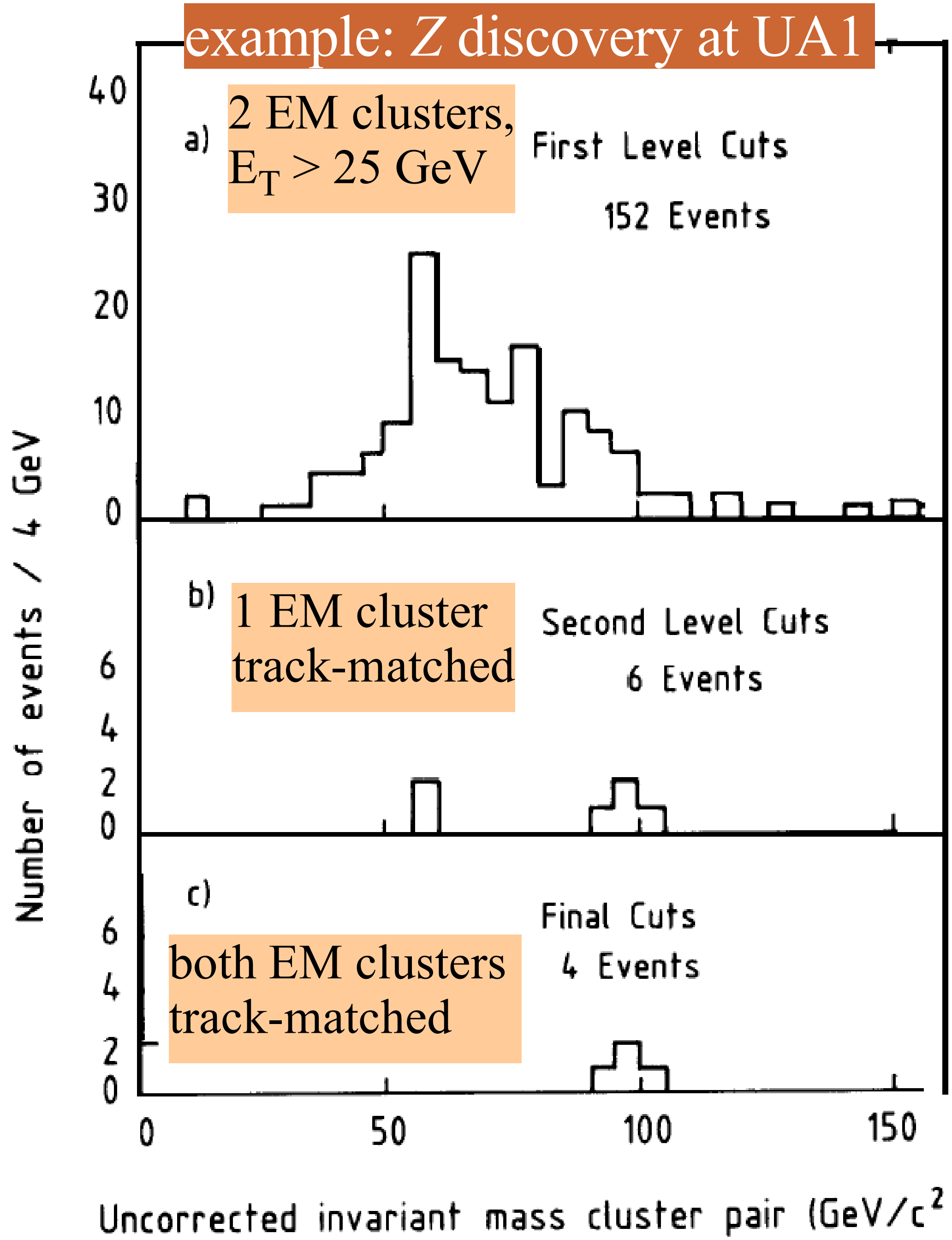
3. Statistical analysis
   – Measurement with uncertainty
   – Confidence limit

Basic Event Analysis Procedures

1) Cut-based event counting
2) Peak in a characteristic distribution

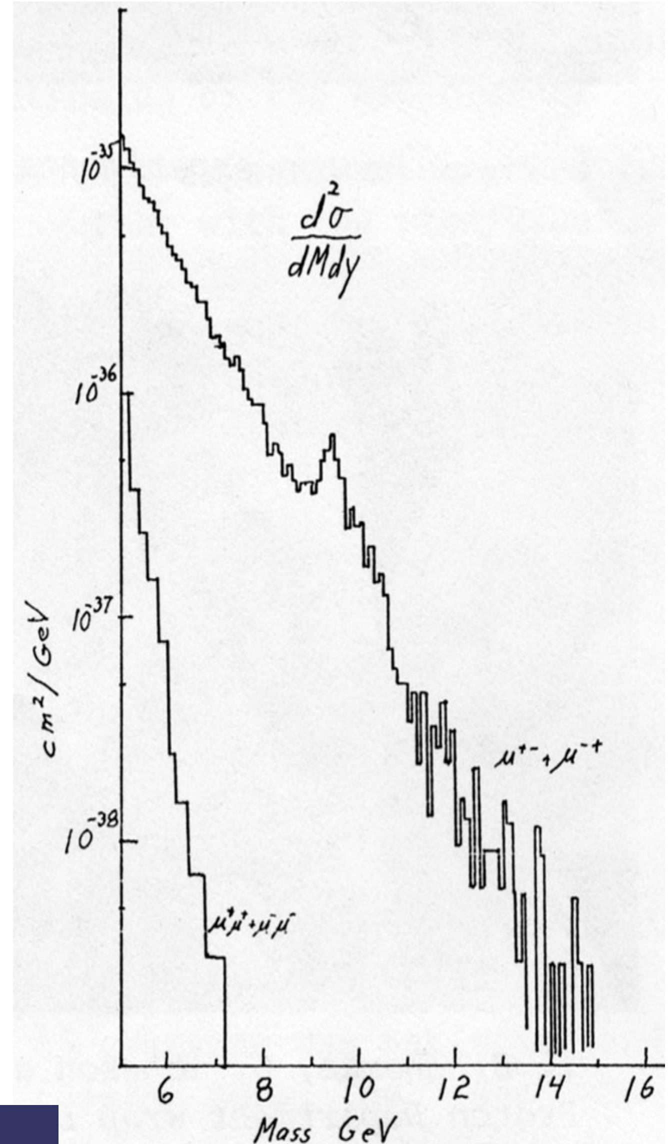
Event counting

- Apply cuts to variables describing the event
  - Object identification
  - Kinematic cuts on objects
  - Event kinematics
- Goal: cut until the signal is visible
  - No background left
  - Or large S/√B
- Sensitive to any signal with this final state
- Requires understanding of background



Peak in a characteristic distribution

- Find a variable that has a smooth distribution for background
  - Typically invariant mass
- Measure this distribution over a large range of possible values
- Look for possible resonance peaks
  - Example: b-quark discovery at Fermilab
- Sensitive to any resonance with this final state

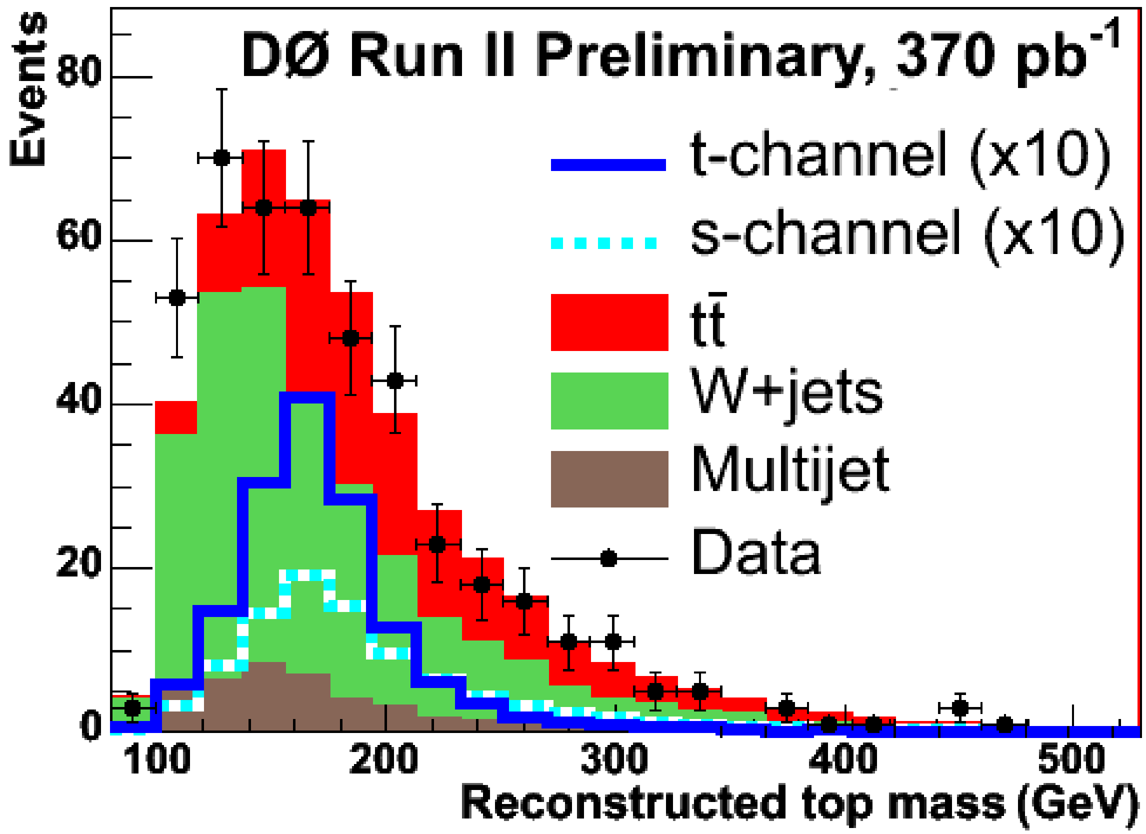


"Bump Hunting"

When Event Counting and Bump Hunting don't work

Example: DØ Single Top in 370 pb^{-1}

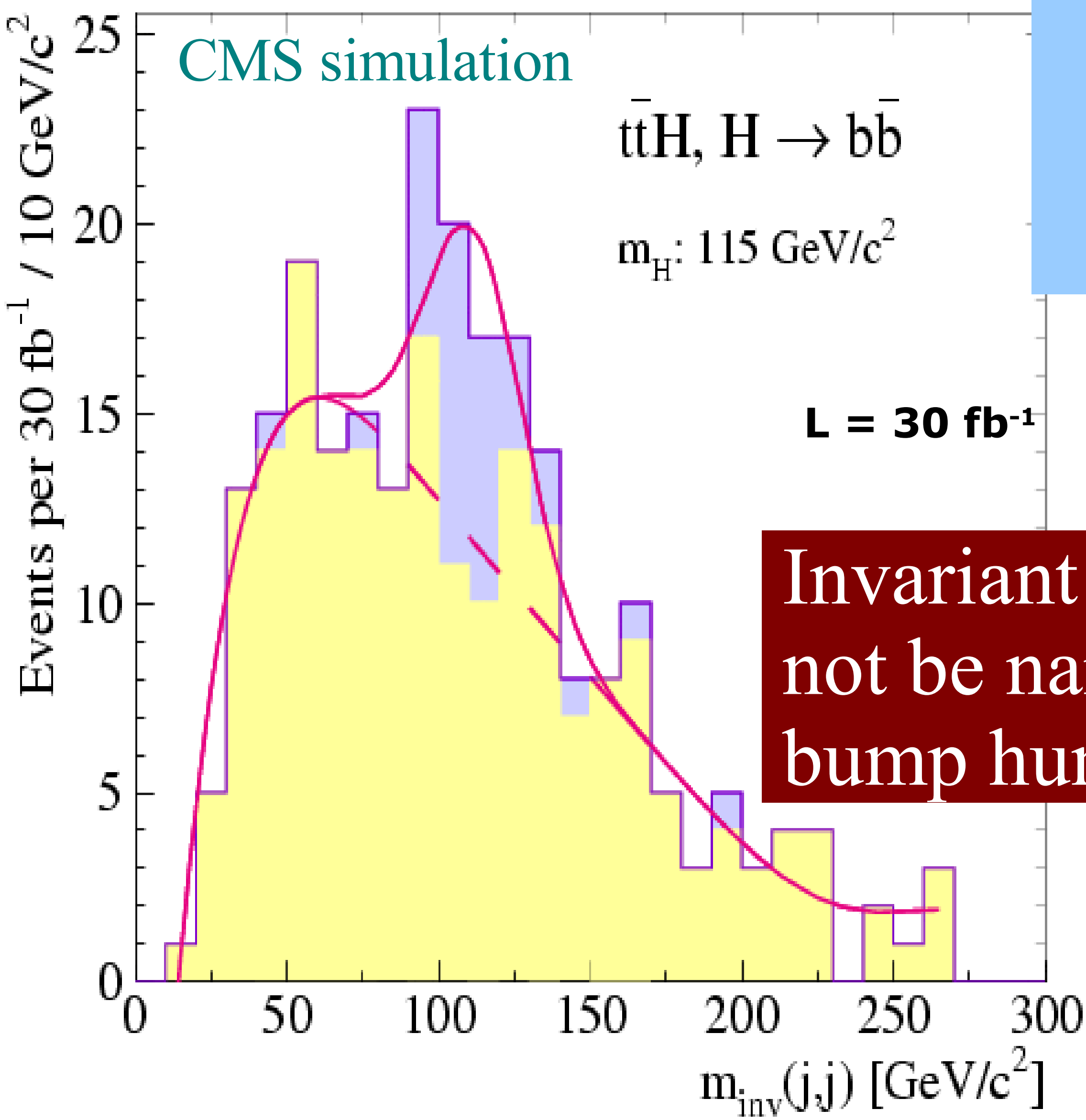
	s-channel	t-channel
Signal yield	9.5	15.0
Bkgnd yield	452	
Data	443	
Signal/bkgnd	1:50	1:30



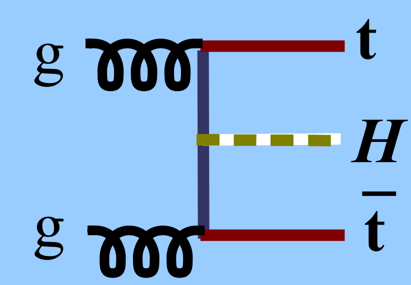
Signal/Background too small for event counting

Invariant mass too broad for bump hunting

Example: Light Higgs at the LHC



Top-Higgs production



Invariant mass peak might not be narrow enough for bump hunting

How to improve upon

Event Counting
and
Bump Hunting

Optimized Event Analysis

Optimized = {
- Optimize signal-background separation
- Exploit full event information
  - Event kinematics, angular correlations, ...
- Take all correlations into account

}

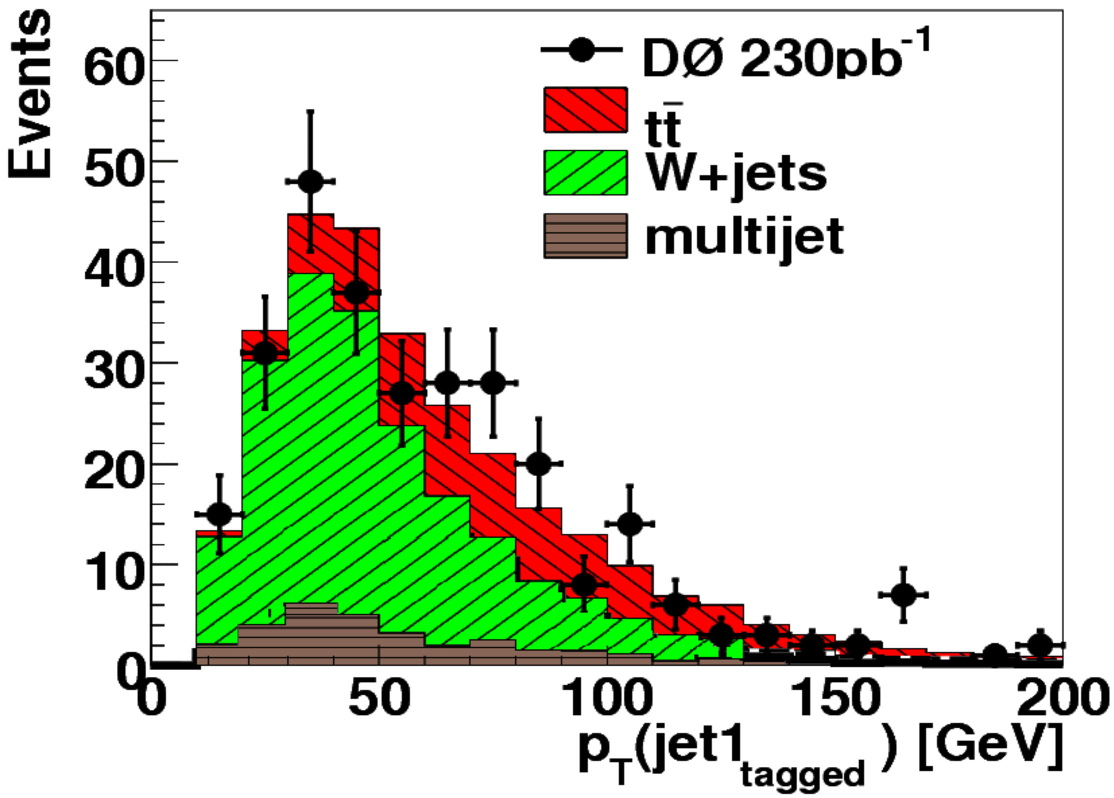
- Requires detailed expectation for signal and background
  - Only applicable to searches for a specific signal or measurements of a specific process
- Limited by background and signal modeling
  - MC statistics, MC model, background composition, shape, ...

Wrong signal model: search is not sensitive 😐

Wrong background model: find something that isn't there ☹

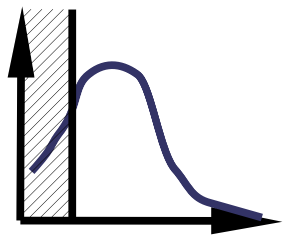
Optimizing the Event Analysis

- Find discriminating variables
  - Using physics intuition, analyzing Feynman diagrams
  - Brute force trial and error
  - Define smallest set that covers all of phase space
- Check that background model matches data for these variables
  - In background dominated samples
  - In cross-check samples free of signal
  - Also check correlations

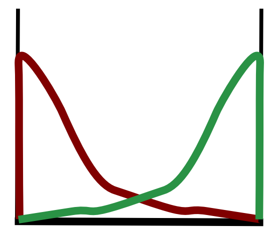


Event Analysis Techniques

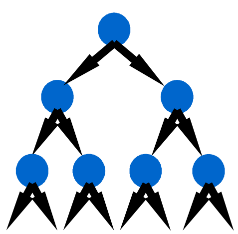
Cut-Based



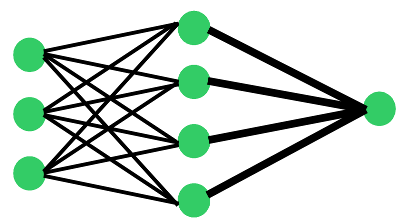
Likelihoods



Decision Trees



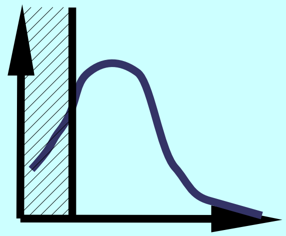
Neural Networks



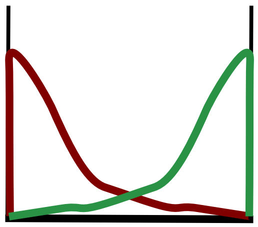
Many others: Kernel methods, support vector machines, Matrix element, ...

Event Analysis Techniques

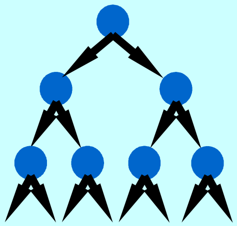
Cut-Based



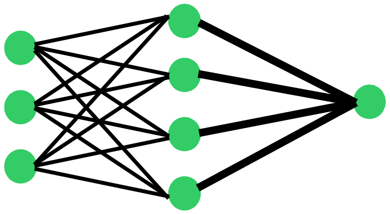
Likelihoods



Decision Trees

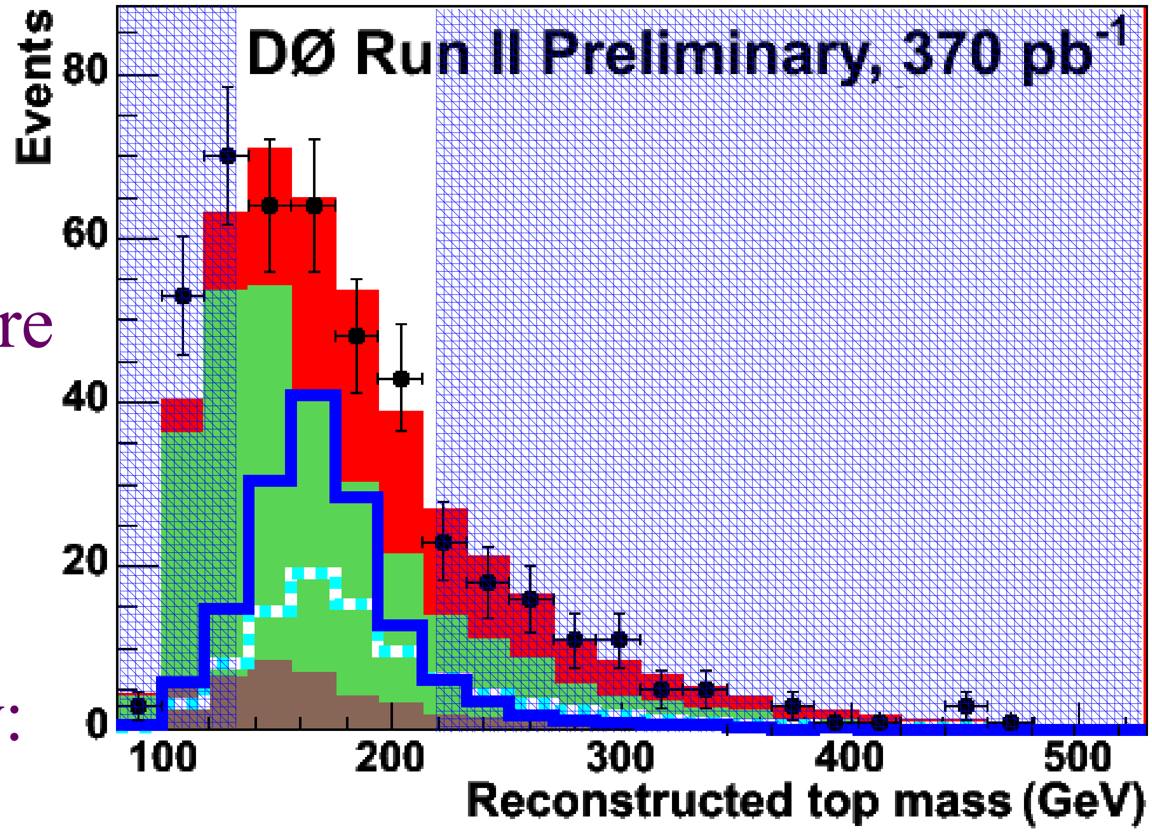


Neural Networks

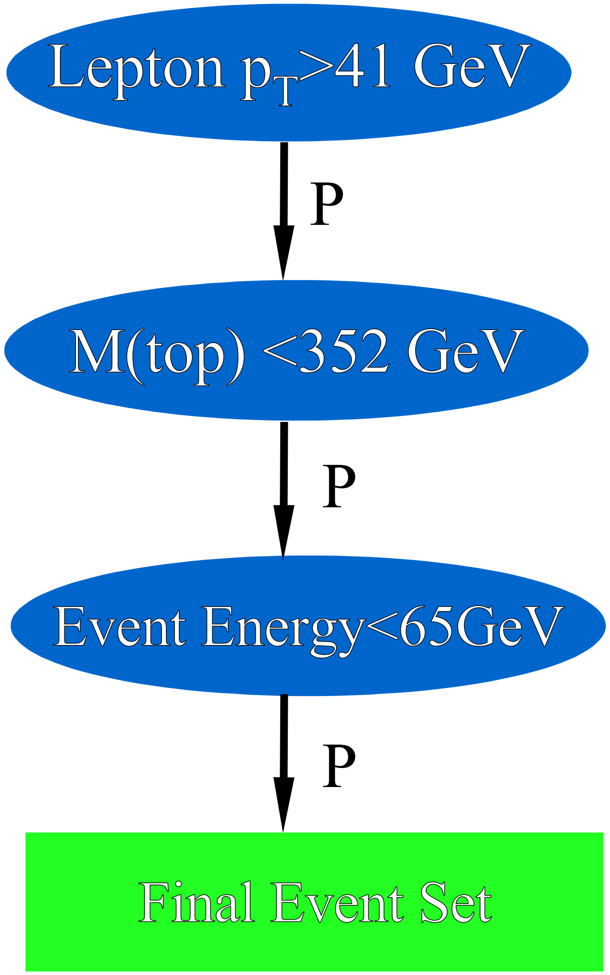


Cut-Based Analysis

- Cut on several discriminating variables
  - Systematically explore possible cuts
- Optimize each cut based on
  - Expected uncertainty: maximize $S/\sqrt{B}$
  - Expected confidence limit



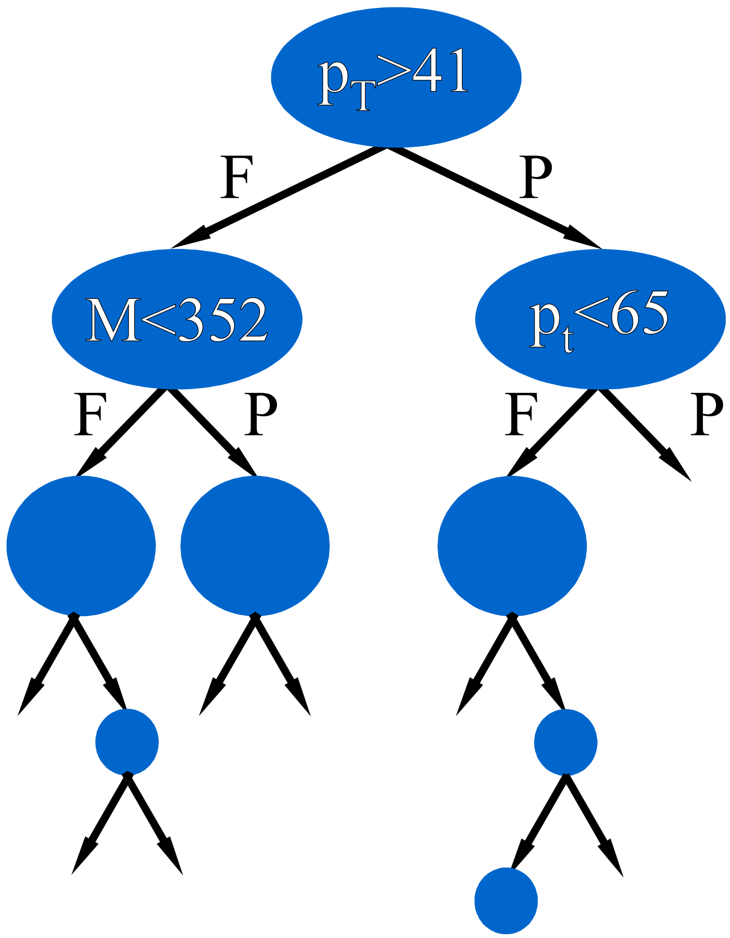
Cut-Based Analysis

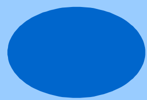


In the final event set

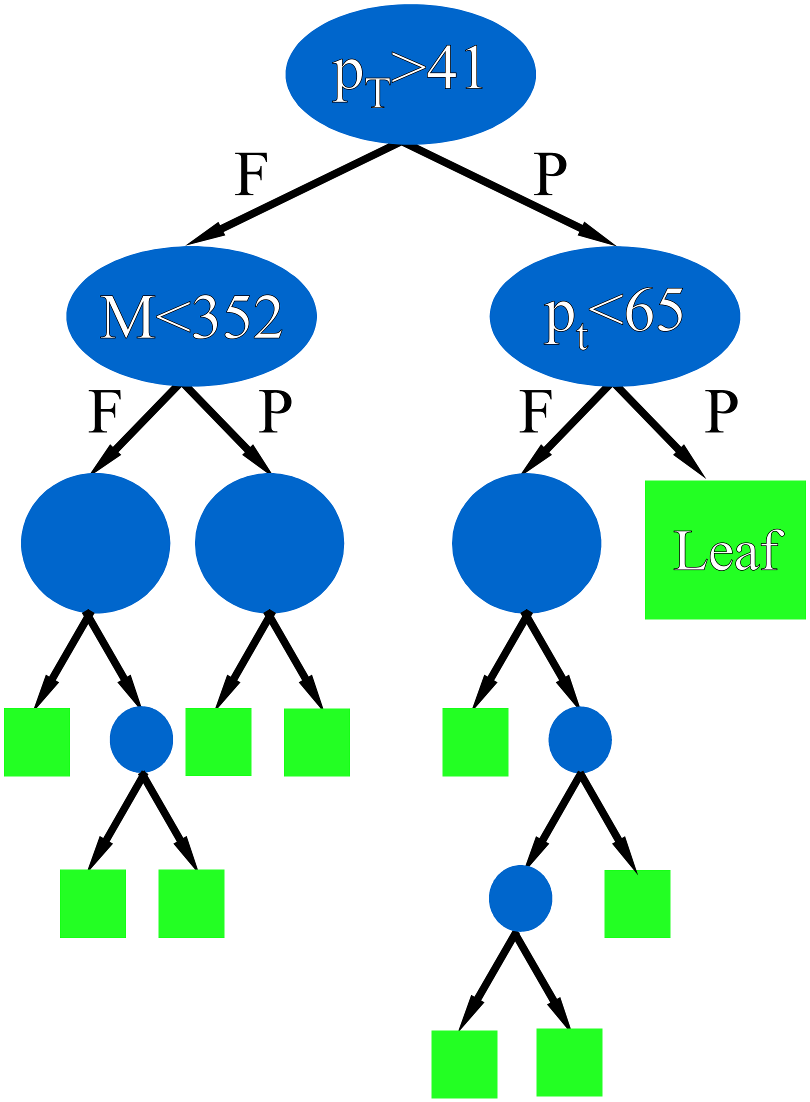
- Estimate background yield
- Compare to data
  $N_{obs} = N_{data} - N_B$
- Calculate signal acceptance
  $\sigma = N_{obs} / (A*L)$

Including Events that fail a Cut



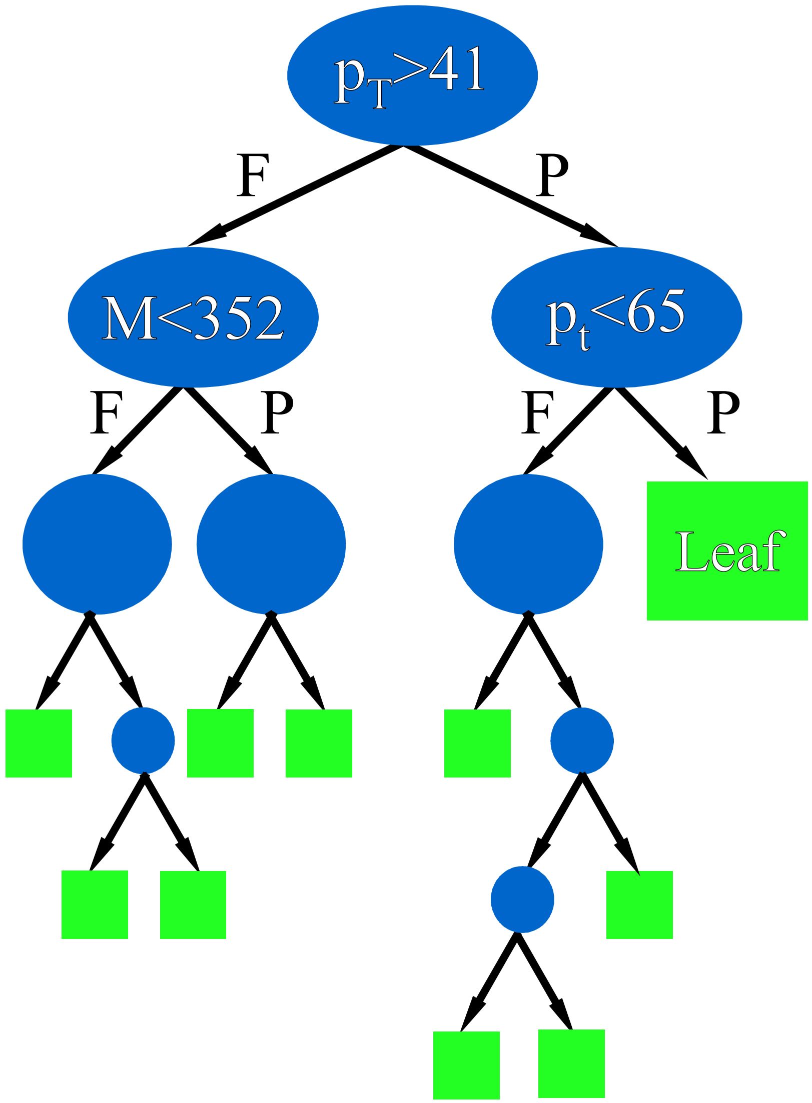
– Create a tree of cuts
– Divide sample into "pass" and "fail" sets
– Each node  corresponds to a cut (branch)

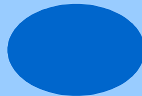
Trees and Leafs



- Create a tree of cuts
- Divide sample into "pass" and "fail" sets
- Each node corresponds to a cut (branch)
- A leaf corresponds to an end-point
- For each leaf, calculate purity (from MC):
  purity = $N_S/(N_S+N_B)$

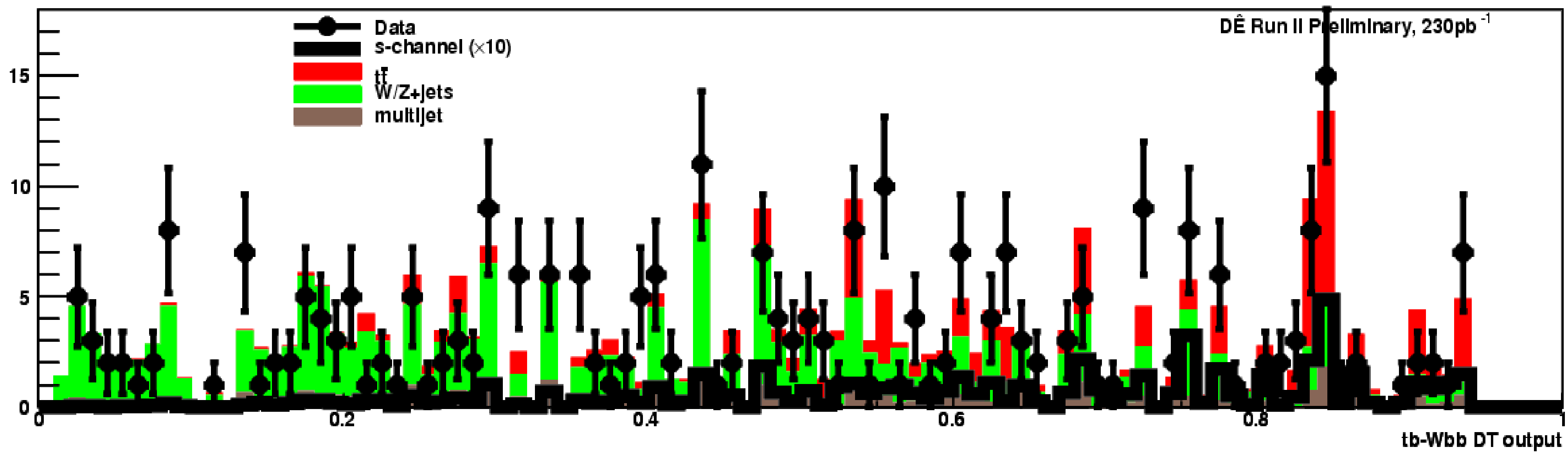
Decision Tree



- Create a tree of cuts
- Divide sample into "pass" and "fail" sets
- Each node  corresponds to a cut (branch)
- A leaf  corresponds to an end-point
- For each leaf, calculate purity (from MC):
  purity = $N_S/(N_S+N_B)$
- Train the tree by optimizing the Gini improvement:
  - Gini = $2\, N_S N_B /(N_S + N_B)$
  - Each leaf will be either background- or signal-enhanced

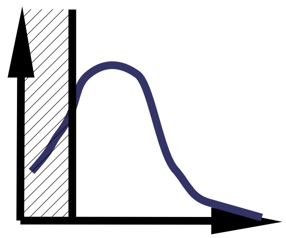
Decision Tree Output

- Train on signal and background models (MC)
  - Stop and create leaf when N_{MC}<100
- Compute purity value for each leaf
- Send data events through tree
  - Assign purity value corresponding to the leaf to the event
- Result is a probability distribution

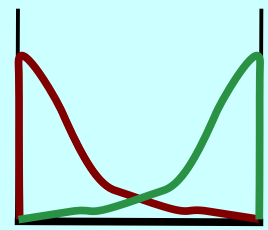


Event Analysis Techniques

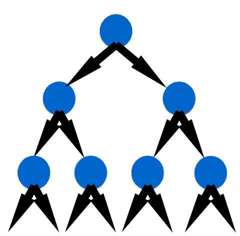
Cut-Based



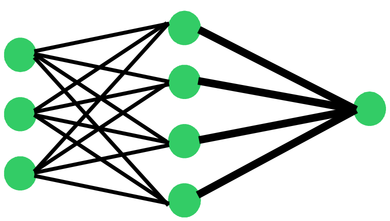
Likelihoods



Decision Trees



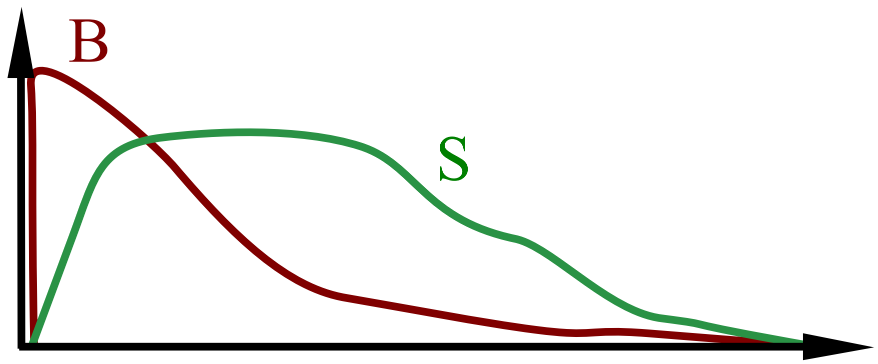
Neural Networks



Likelihood Analysis

- Convert any variable into a probability distribution function:

$$p = \frac{N_S}{N_S + N_B}$$



  - Determine pdf from signal and background MC

- Likelihood: product of pdf values for each discriminating variable

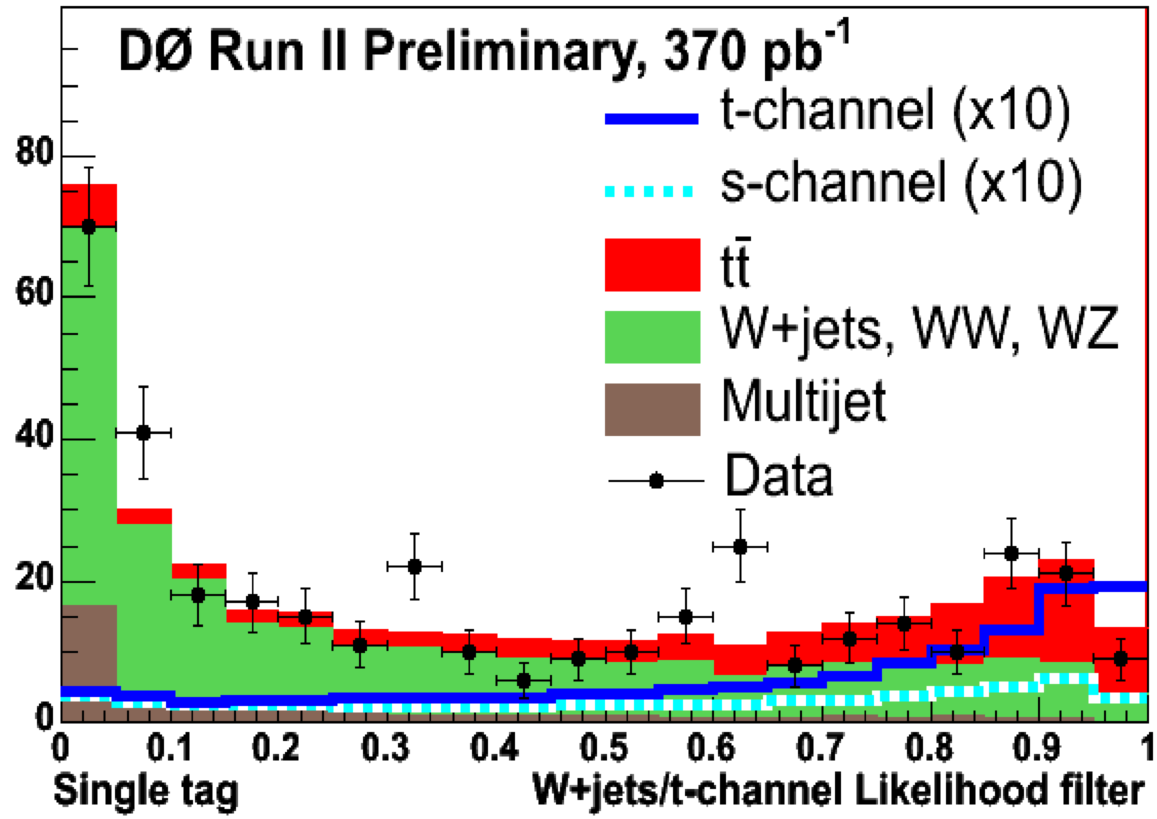
$$L = p_1 \times p_2 \times p_3 \times \ldots$$

  - Valid if variables are uncorrelated

Also called “simple Bayes”

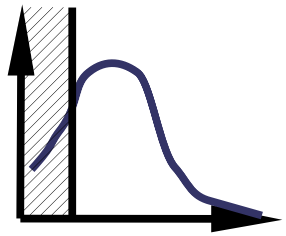
Likelihood Output

- No training required
- For each data event, evaluate likelihood
  - From the discriminating variables
- Result is a probability distribution

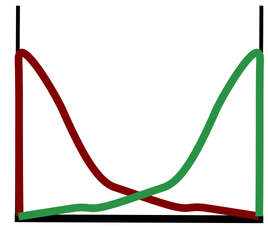


Event Analysis Techniques

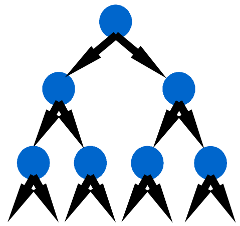
Cut-Based



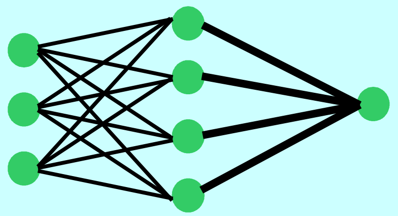
Likelihoods



Decision Trees

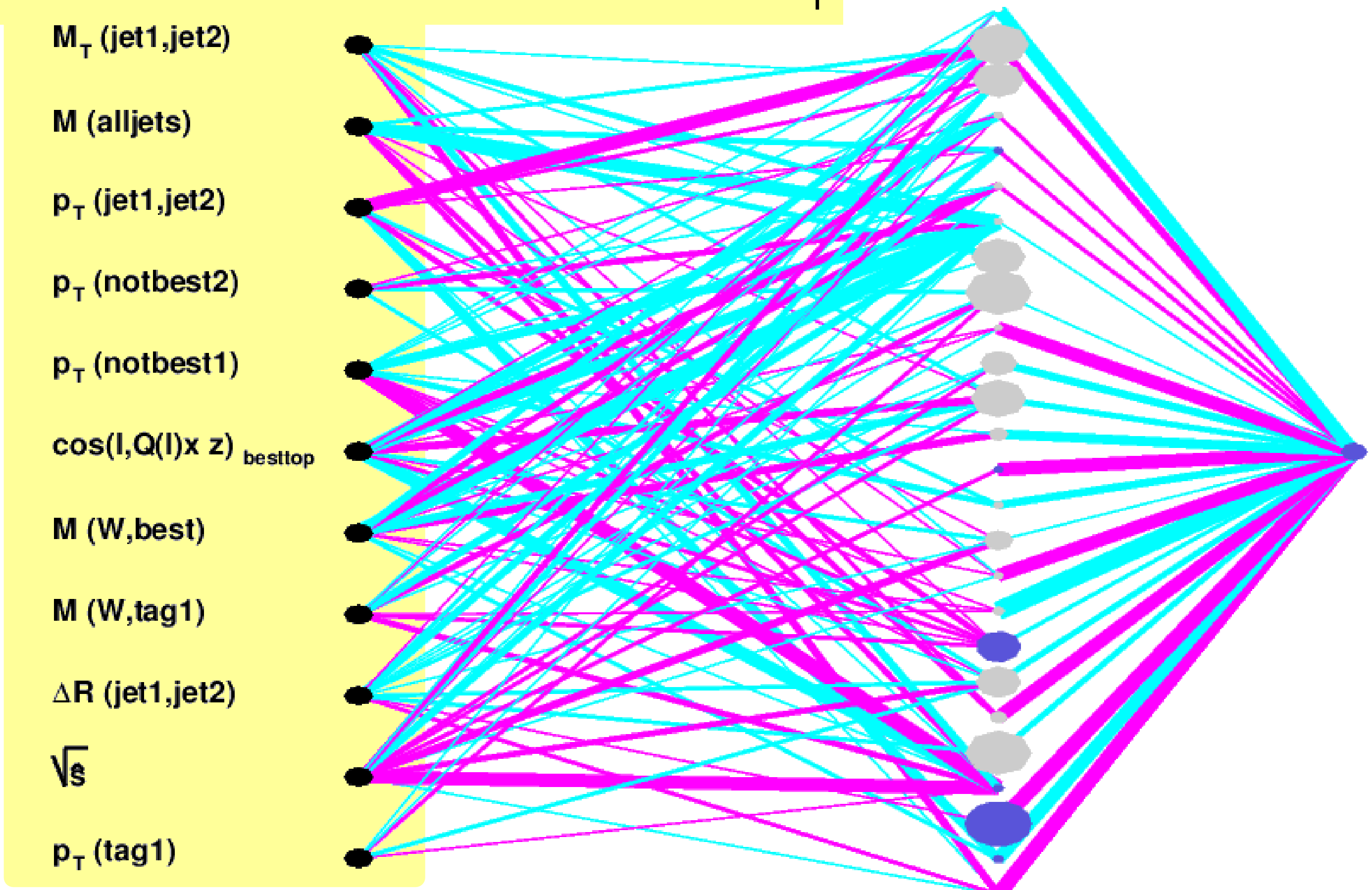


Neural Networks



Neural Networks

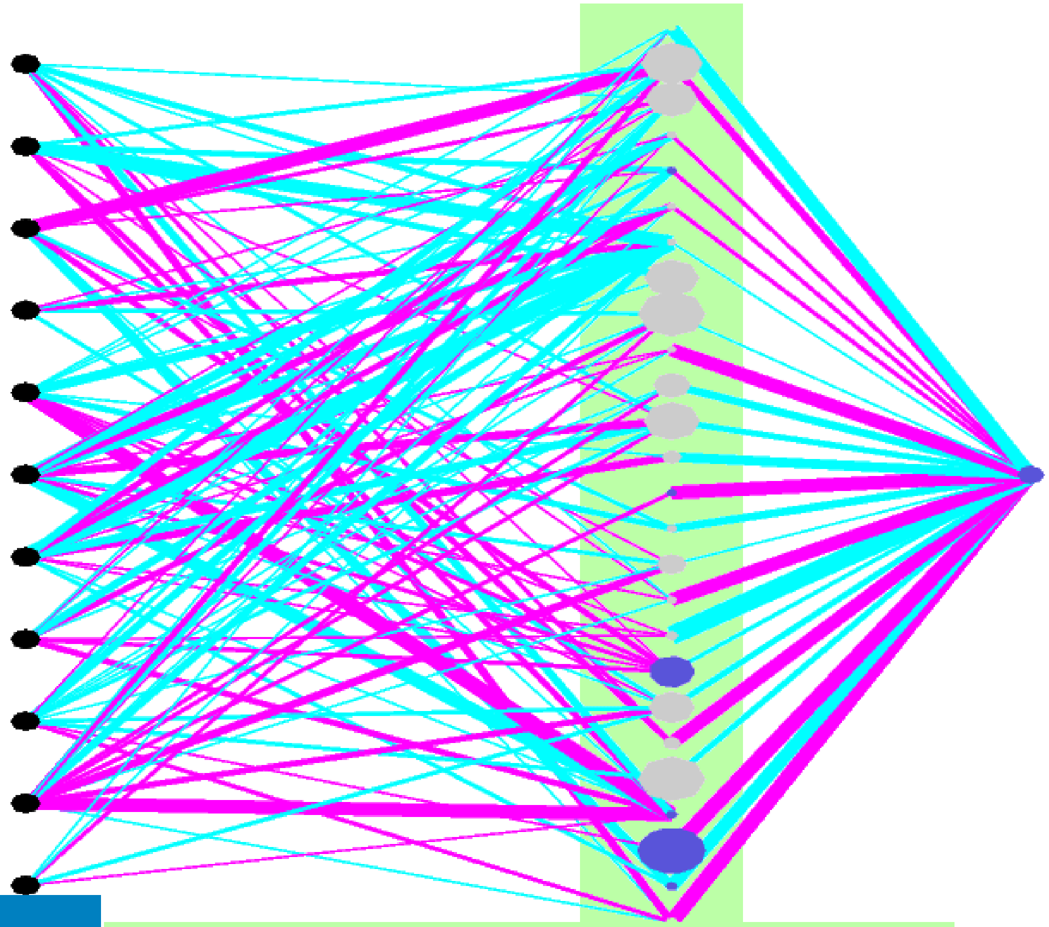
Input Nodes: One for each variable x_i

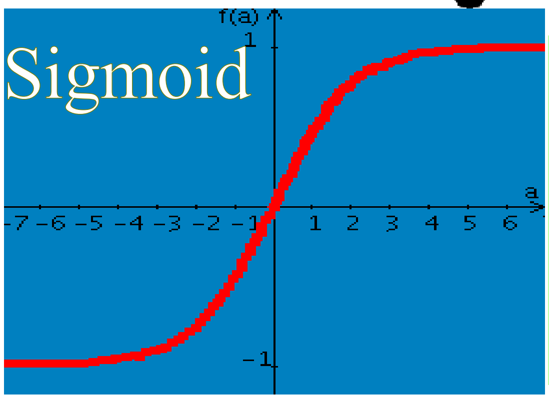


Goal: approximate signal probability

$$f(\vec{x}) \approx P(S|\vec{x})$$

Neural Networks

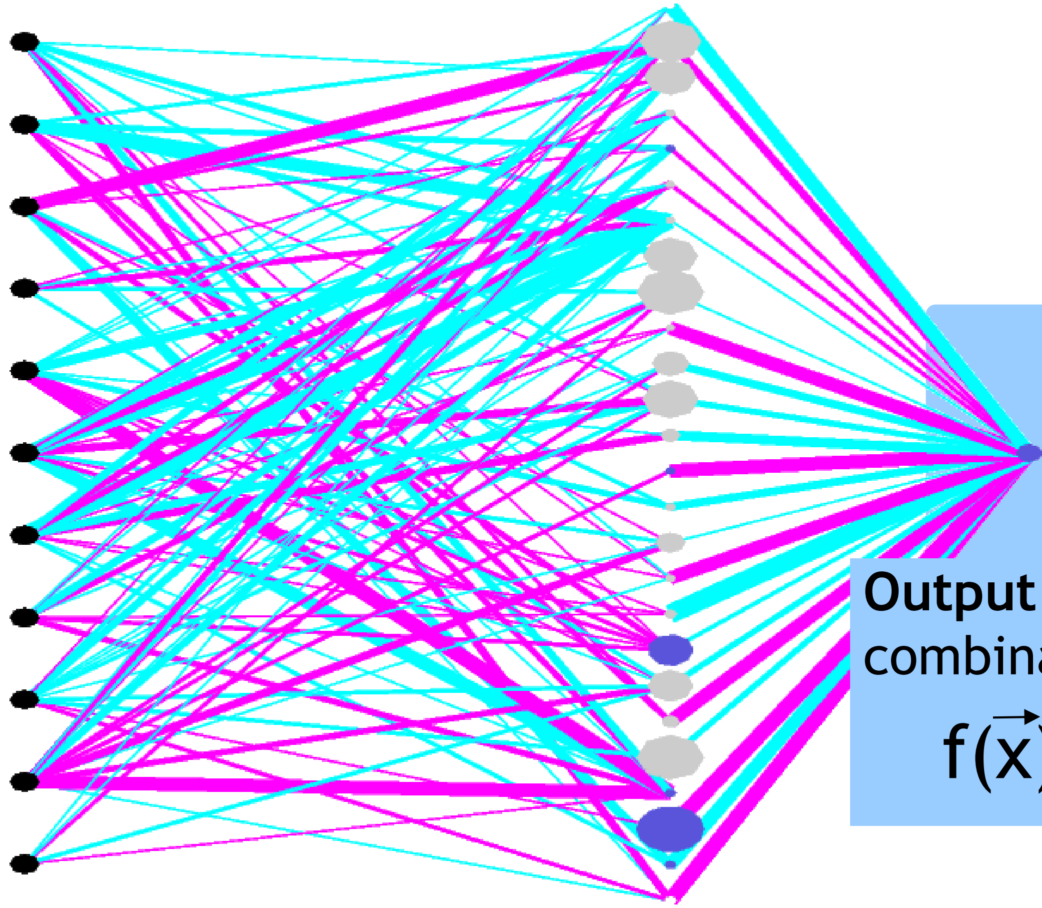




Hidden Nodes: Each is a sigmoid dependent on the input variables

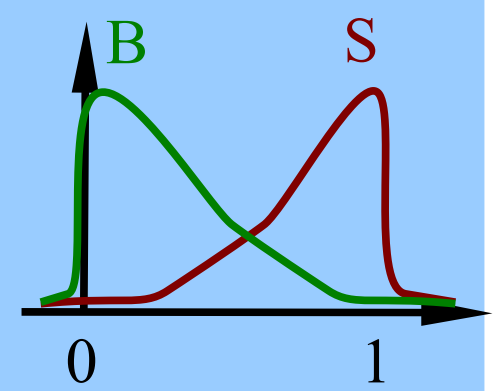
$$n_k(\vec{x},\vec{w}_k) = \frac{1}{1 + e^{-\sum w_{ik} x_i}}$$

Neural Networks



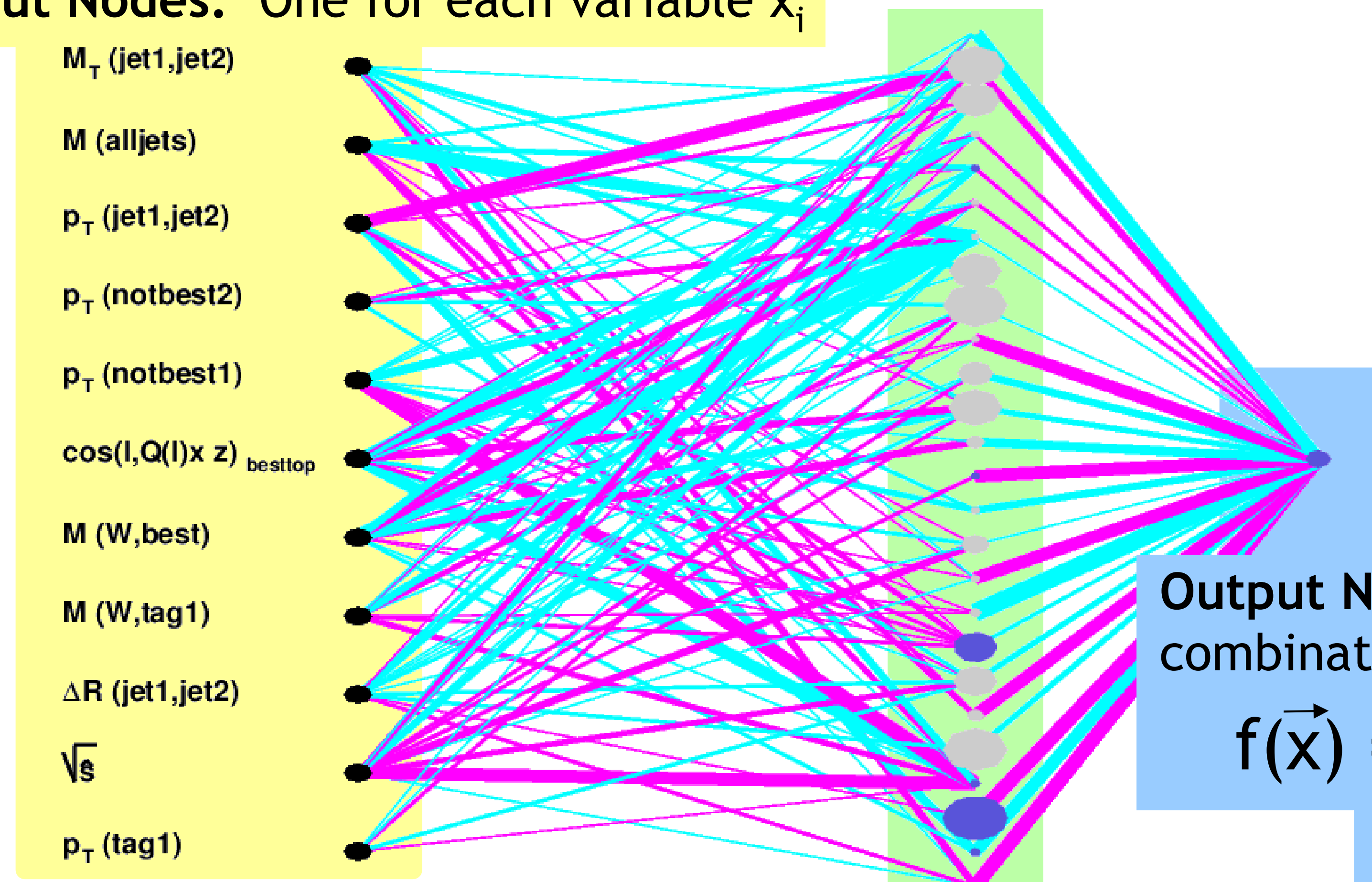
Output Node: linear combination of hidden nodes

$$f(\vec{x}) = \Sigma\, w'_k\, n_k(\vec{x},\vec{w}_k)$$



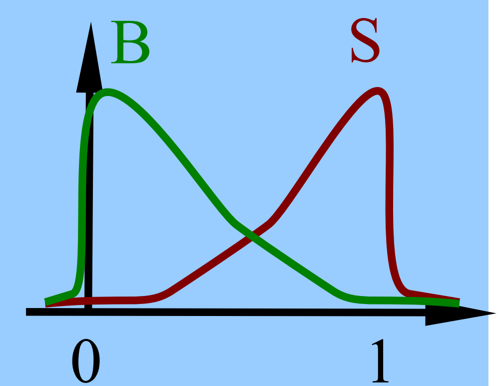
Neural Networks

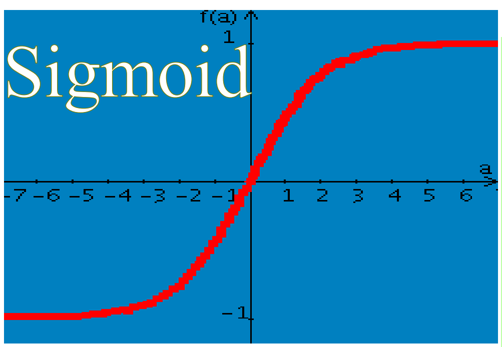
Input Nodes: One for each variable x_i



Output Node: linear combination of hidden nodes

$$f(\vec{x}) = \Sigma\, w'_k\, n_k(\vec{x},\vec{w}_k)$$



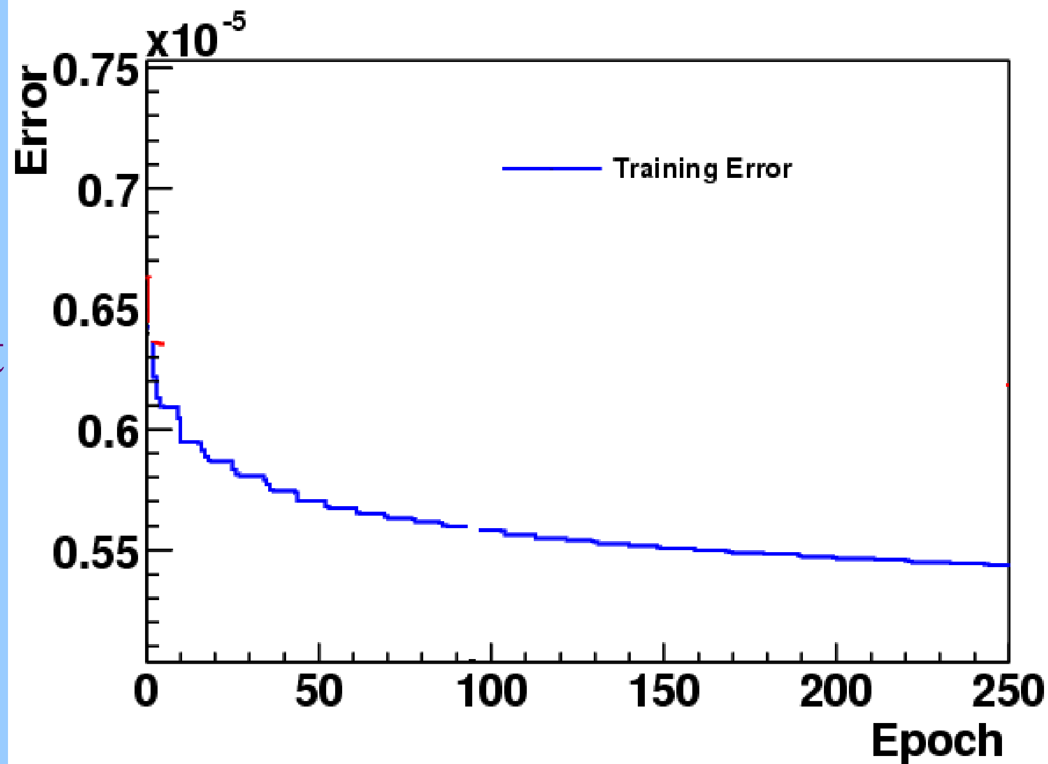


Hidden Nodes: Each is a sigmoid dependent on the input variables

$$n_k(\vec{x},\vec{w}_k) = \frac{1}{1 + e^{-\Sigma\, w_{ik}\, x_i}}$$

Neural Network Training

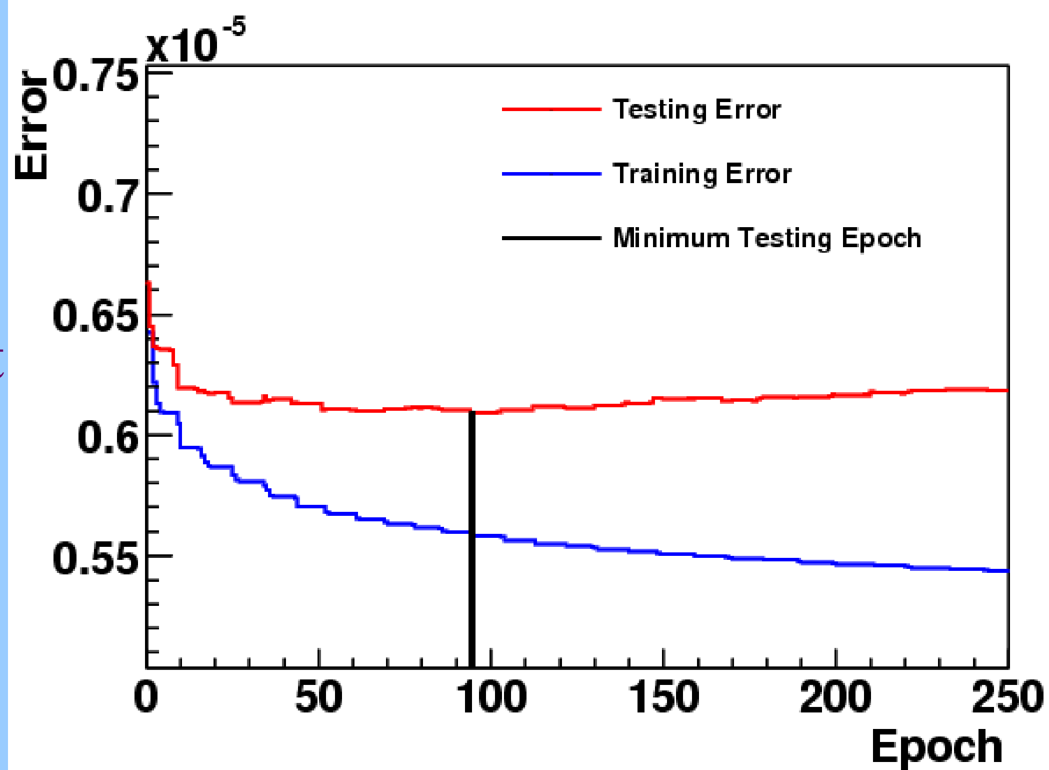
- Initialize NN weights
- Read in signal and background model events
  - Training sample
- Compute NN error
  - $\sum (f_{observed} - f_{expected})$
- Adjust all NN weights as result
- Compute NN error again
- Repeat until ...



DØ single top search

Neural Network Training

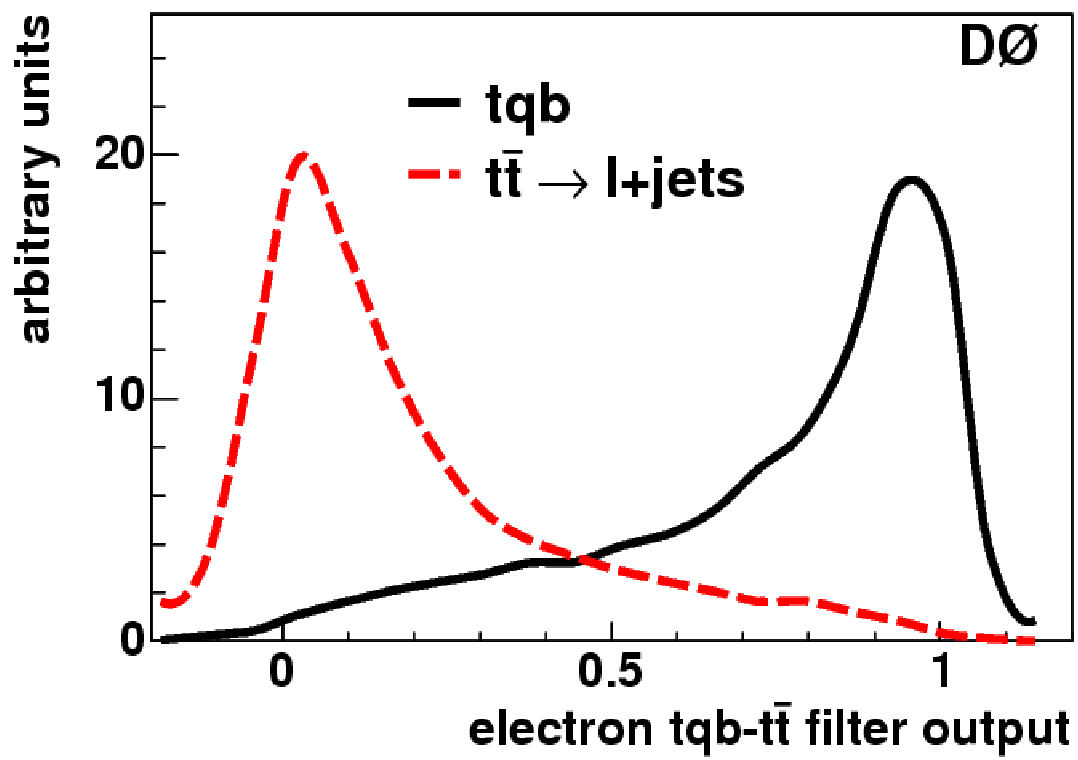
- Initialize NN weights
- Read in signal and background model events
  - Training sample
- Compute NN error
  - $\sum (f_{\text{observed}} - f_{\text{expected}})$
- Adjust all NN weights as result
- Compute NN error again
- Apply NN to independent set of signal and background
  - Testing sample
- Stop training when error from testing sample starts increasing



DØ single top search

Neural Network Result

- Train on signal and background models (MC)
  - Stop when signal-background separation stops improving
    - Independent MC training sample
- For each data event, compute NN output
- Result is almost a probability distribution
  - But not necessarily constrained to [0,1]



DØ single top search

Boosting

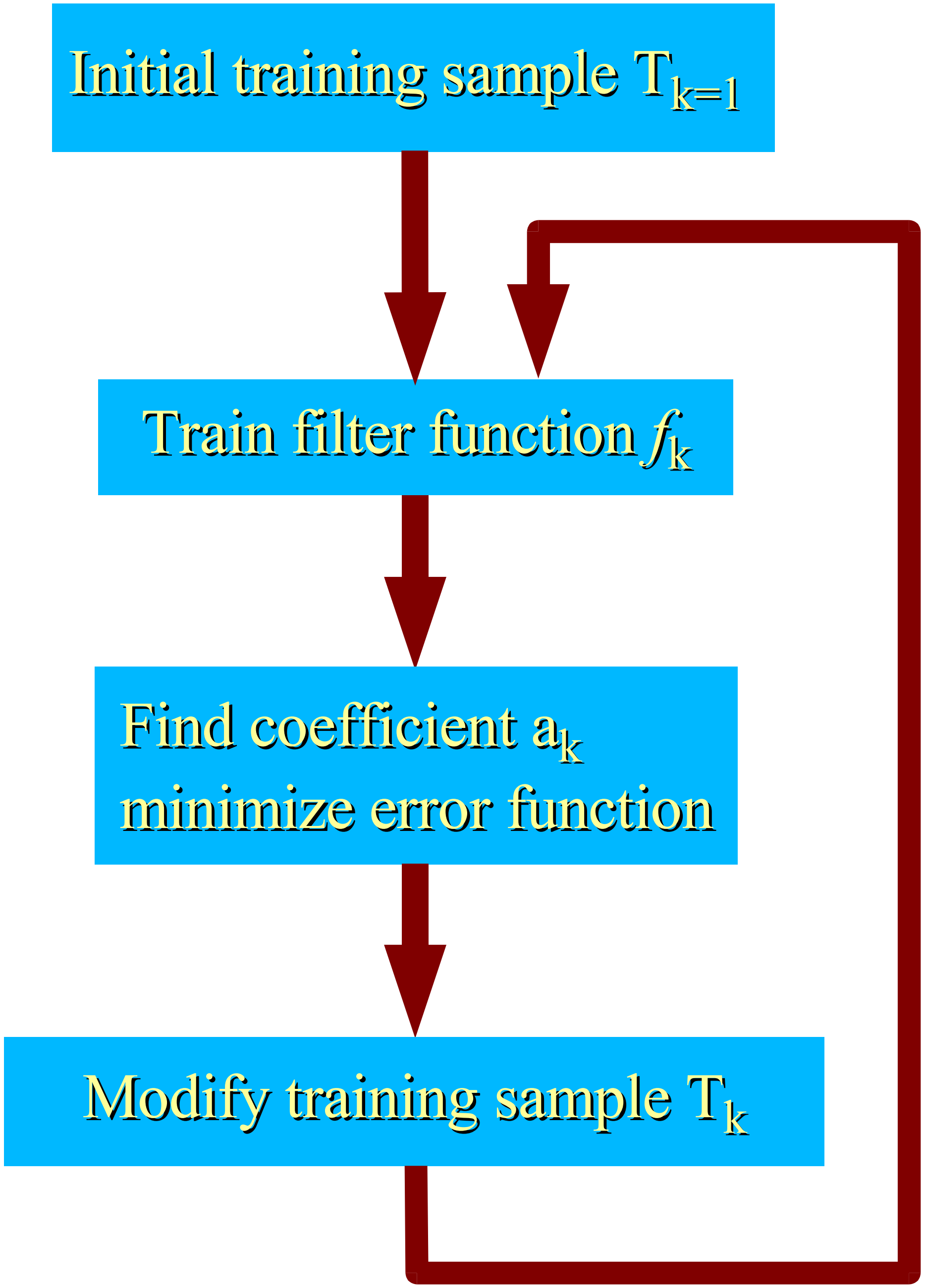
Boosting

- A general method to improve the performance of any weak qualifier
  - Decision trees, neural networks, ...
- Linear combination of many filter functions

$$F(x) = \sum_k a_k f_k$$

  - a_k: coefficient, typically result of minimization of error function

Boosting Procedure

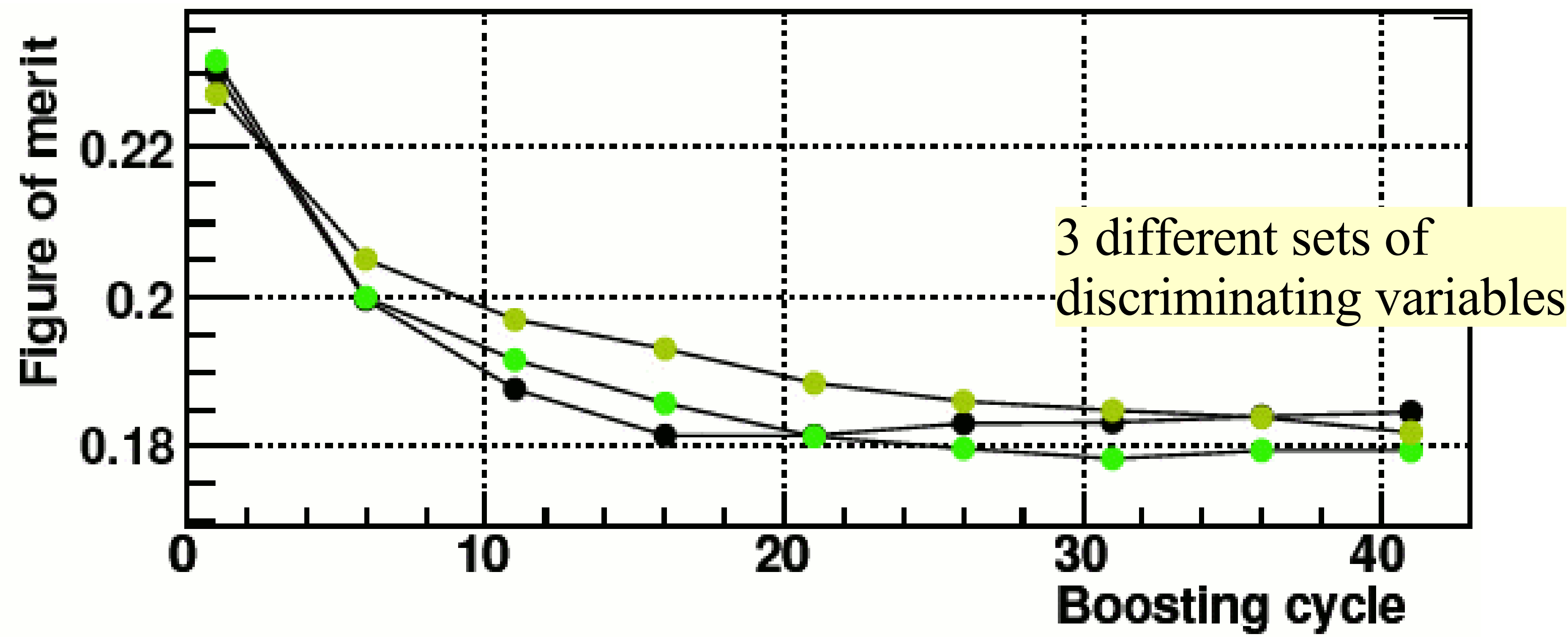


$$F(x) = \sum_k a_k f_k$$

Adaptive Boosting

- In each iteration, update coefficient a_k
  - From minimizing error function
  - coefficients decrease at each iteration
- Update weight for each event in training sample T_k
  - Figure out which events have been misclassified
    - Signal events should have purity ≥ 0.5
    - Background should have purity <0.5
  - Increase event weight for those events that have been misclassified

Boosting Performance



DØ single top search
with decision trees

Comparing Multivariate Methods

How optimal can an optimal event analysis be?

Bayesian Limit

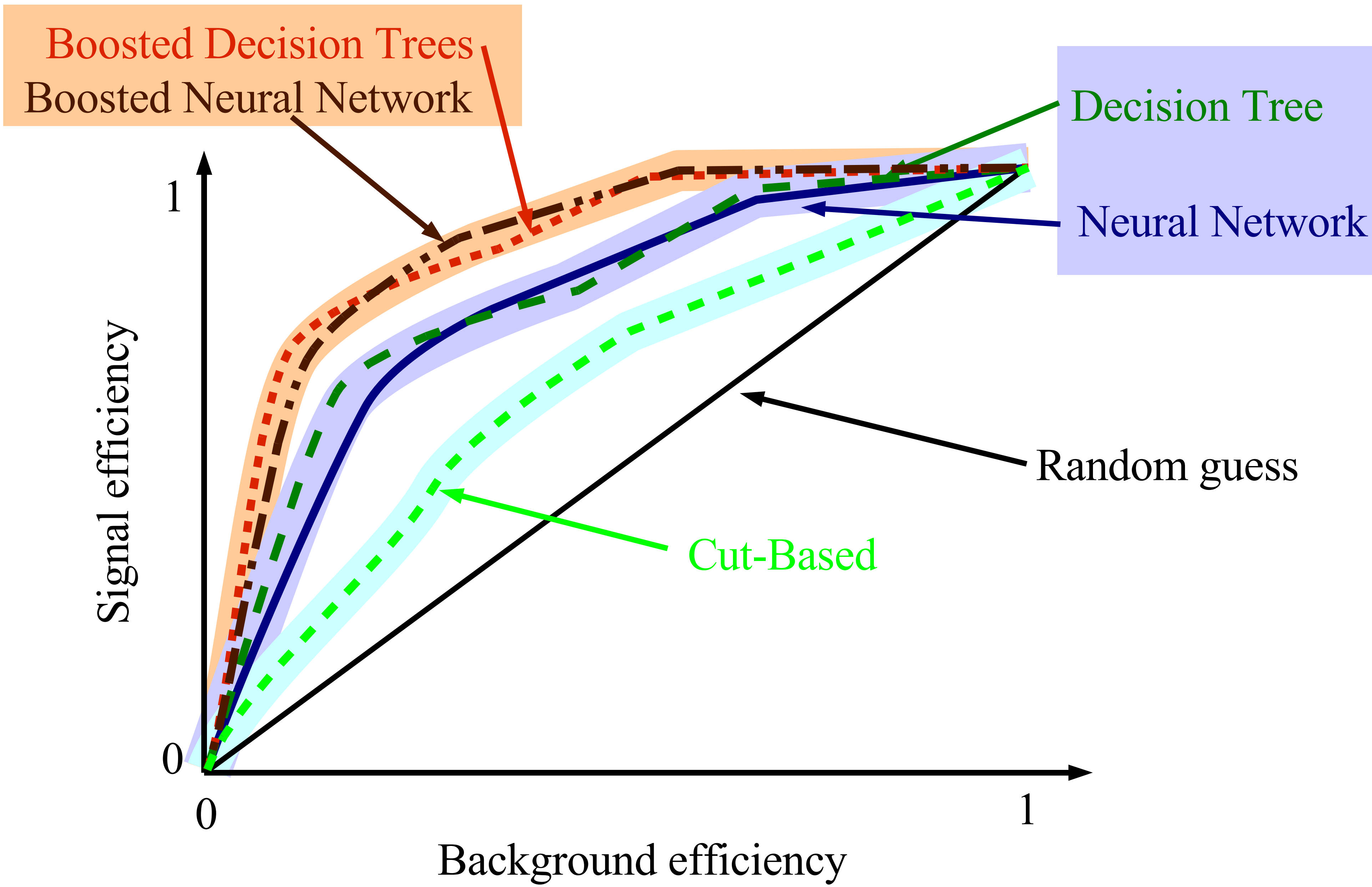
- For each analysis, there exists a fully optimized signal-background separation
  - Bayesian limit, also called target function

$$L(x) = \frac{P(x|S)}{P(x|B)}$$

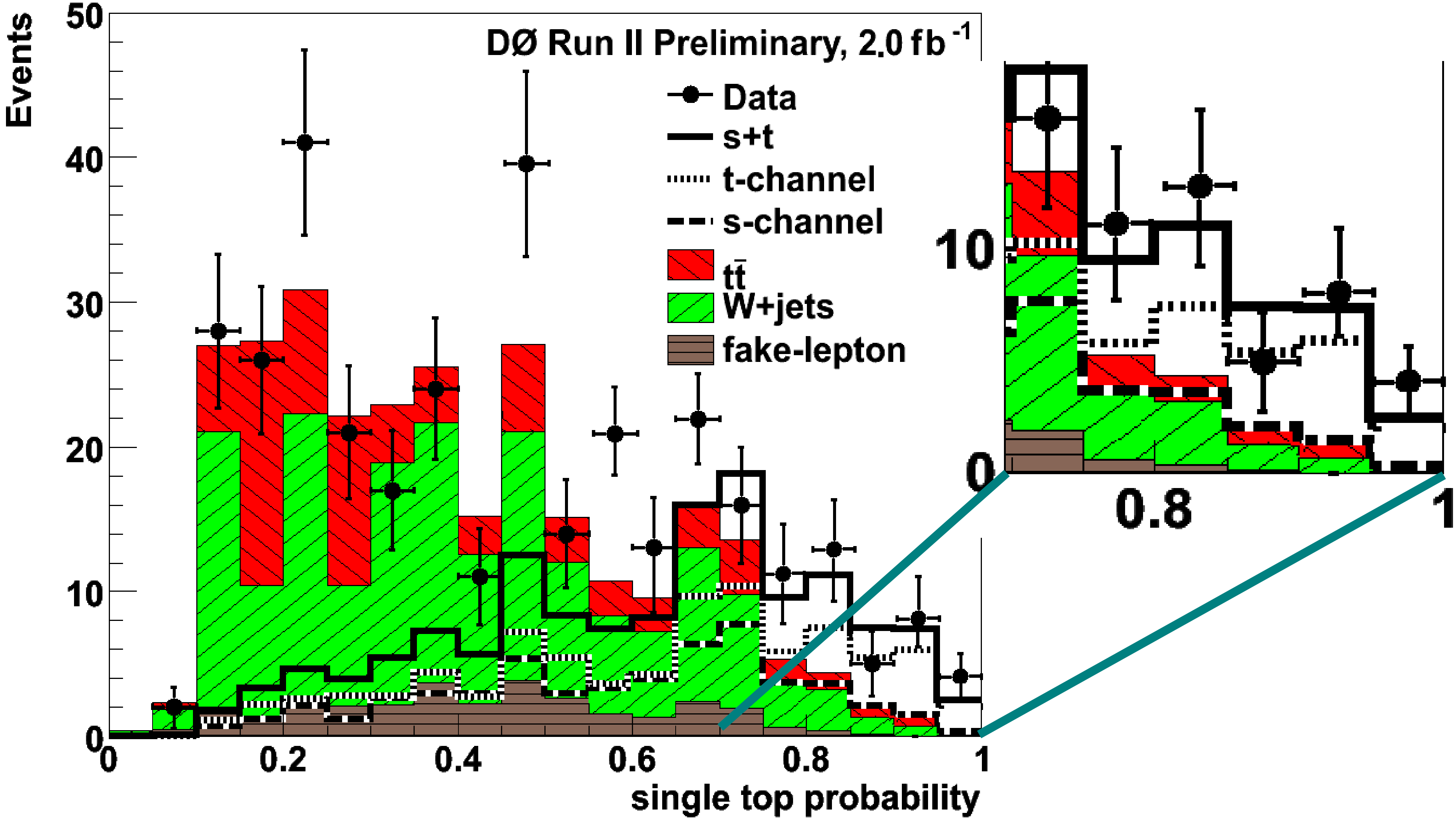
- For a single discrimintating variable, this is equivalent to the pdf
- For many discriminating variables, this isn't possible anymore
  - Typically not enough MC statistic to compute a multi-dimensional pdf

When do we reach the Bayesian limit?

Comparison


Boosted Decision Trees
Boosted Neural Network
Decision Tree
Neural Network
1
0
Signal efficiency
Random guess
Cut-Based
0
1
Background efficiency

Discovery using a Multivariate Analysis



Conclusions

- Multivariate event analysis techniques are a common tool in HEP
  - So far mostly neural networks, now also decision trees
    - Glast, MiniBoone ID, Atlas ID, Dzero
- Boosting significantly improves weak qualifiers
- Accurate background modeling is very important

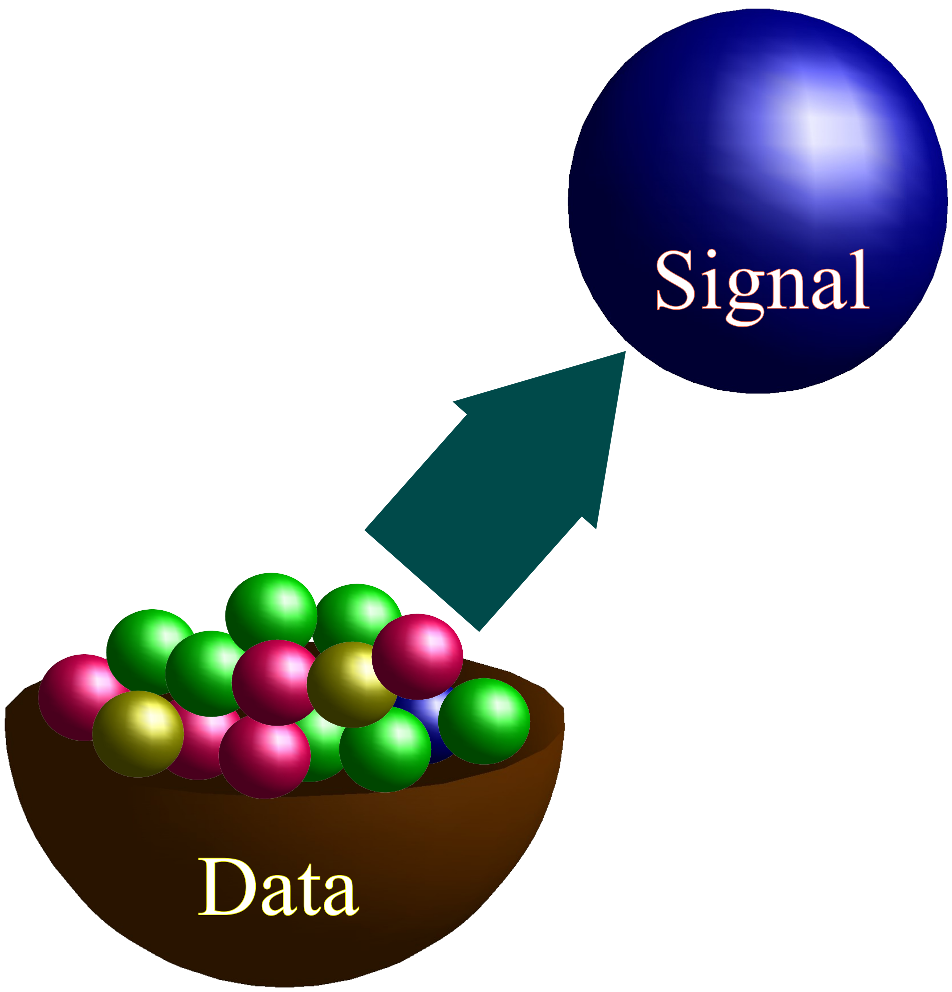
Advanced Event Analysis enables Discoveries

Resources

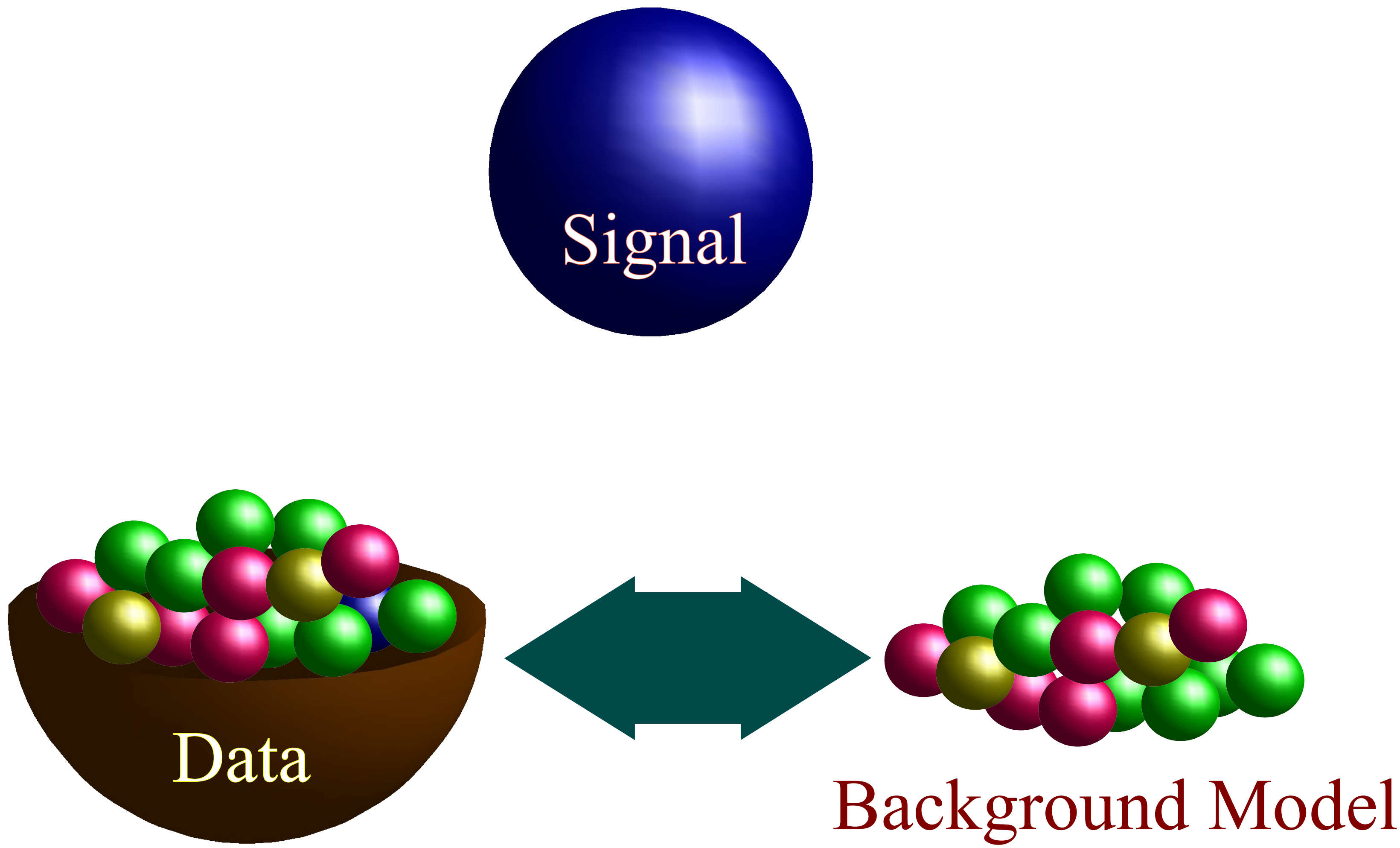
- PhyStat 2005 conference
  http://www.physics.ox.ac.uk/phystat05/
- Jim Linnemann's collection of statistics links:
  http://www.pa.msu.edu/people/linnemann/stat_resources.html
- Neural Networks in Hardware
  http://neuralnets.web.cern.ch/NeuralNets/nnwInHep.html
- Neural Network package JetNet
  http://www.thep.lu.se/public_html/jetnet_30_manual/jetnet_30_manual.html
- Neural Network package MLPFit
  http://schwind.home.cern.ch/schwind/MLPfit.html
- Boosted Decision Trees in MiniBoone
  http://arxiv.org/abs/physics/0508045
- Decision Tree Introduction
  http://www.statsoft.com/textbook/stcart.html
- GLAST Decision Trees
  http://scipp.ucsc.edu/~atwood/Talks%20Given/CPAforGLAST.ppt

Backup Slides

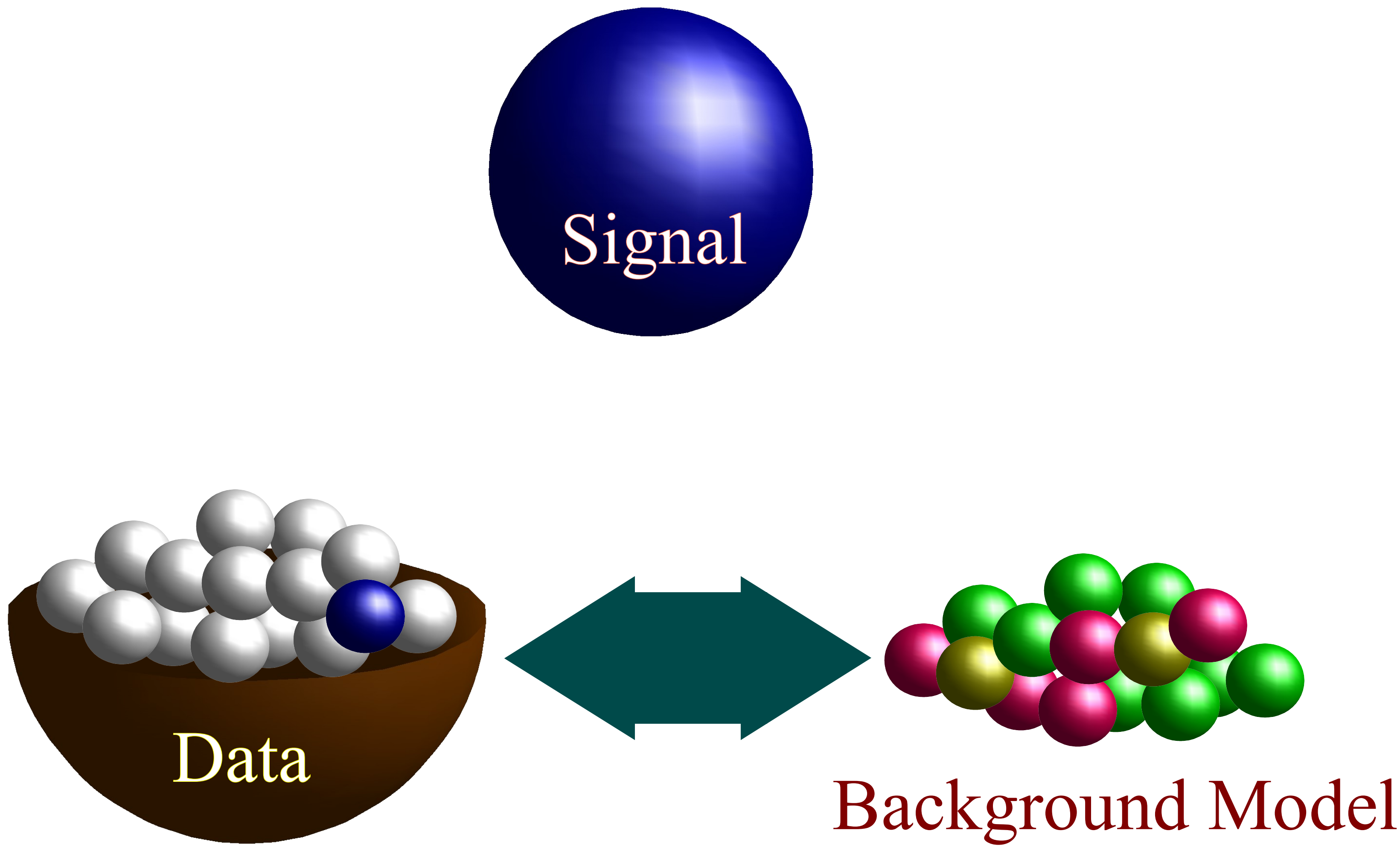
Analysis Outline



Analysis Procedure

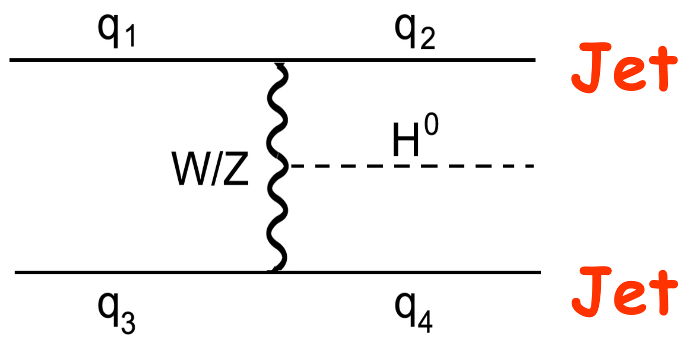


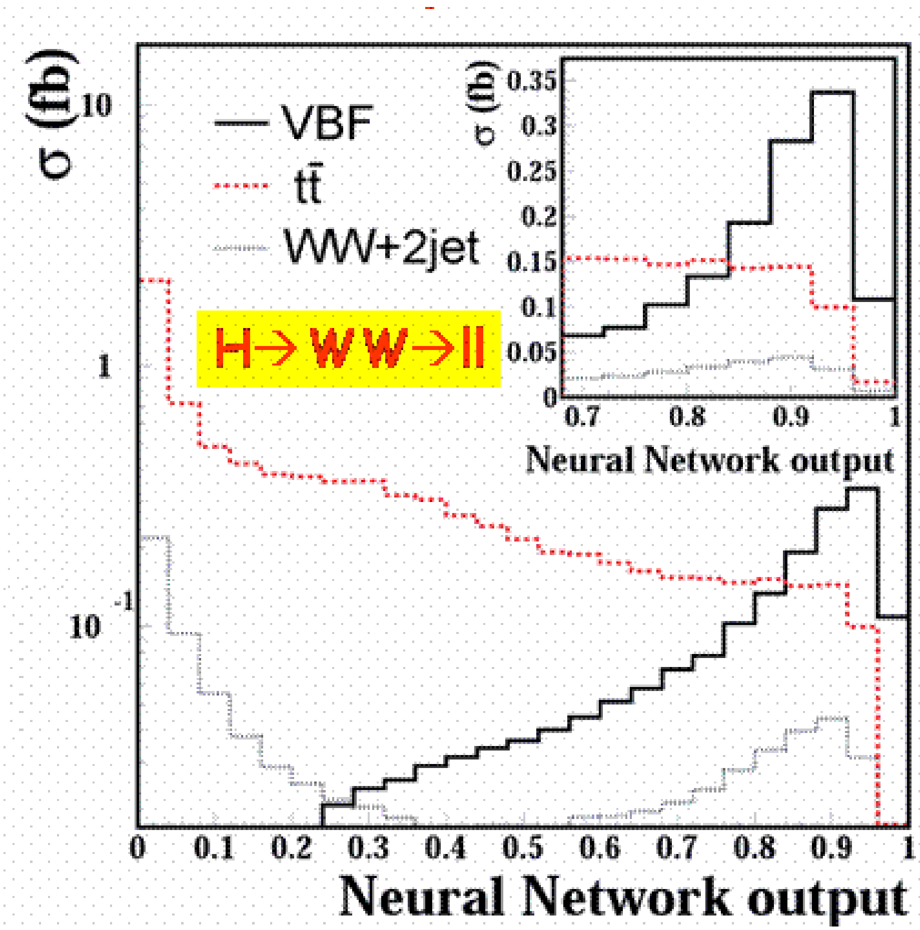
Analysis Procedure



Example: Atlas VBF Higgs Search

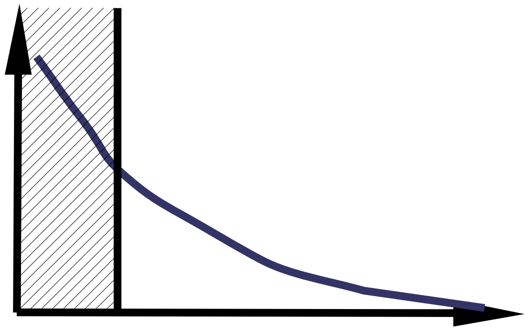
- Higgs boson production through vector boson fusion



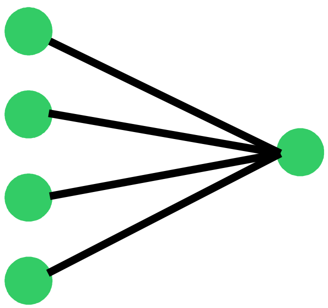


Atlas Preliminary, Bruce Mellado, Wi

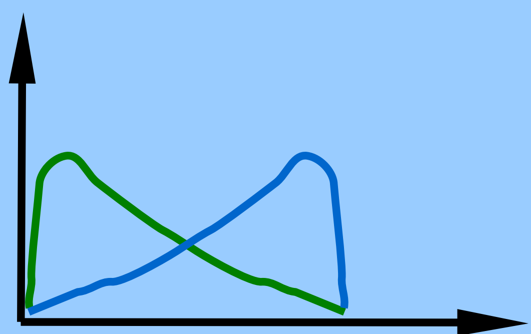
Analysis outline



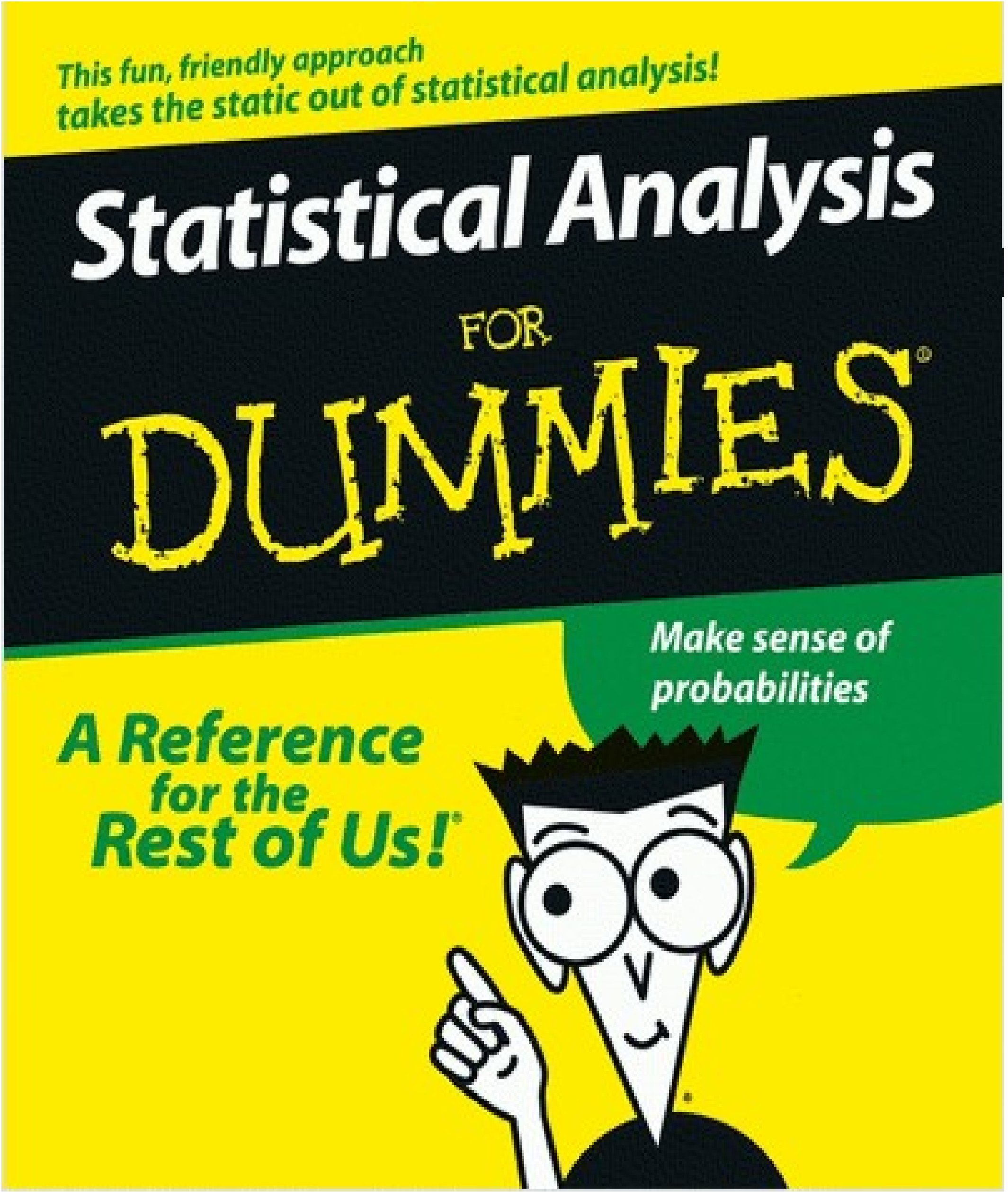
1. Event selection
   - Object identification
   - Background modeling



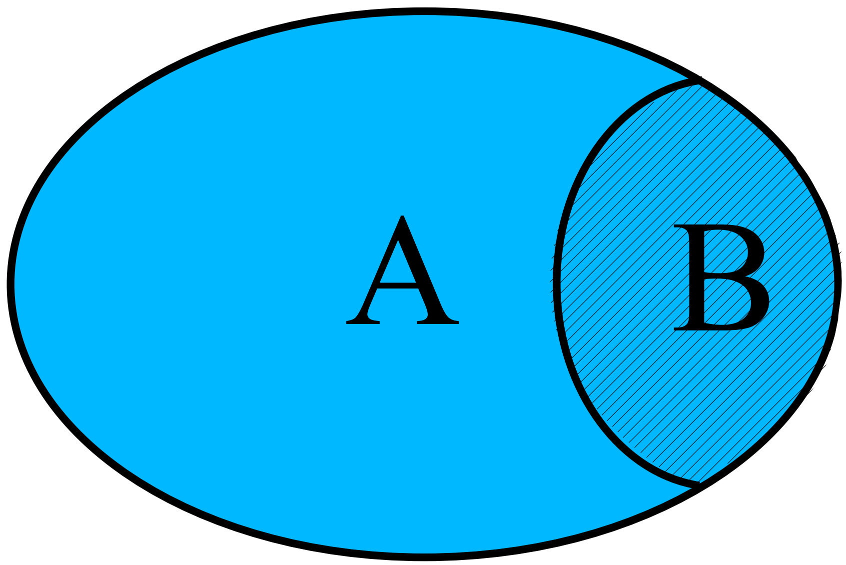
2. Event analysis
   - Discriminating variables
   - Cut/combine in multivariate analysis



3. Statistical analysis
   - Measurement with uncertainty
   - Confidence limit

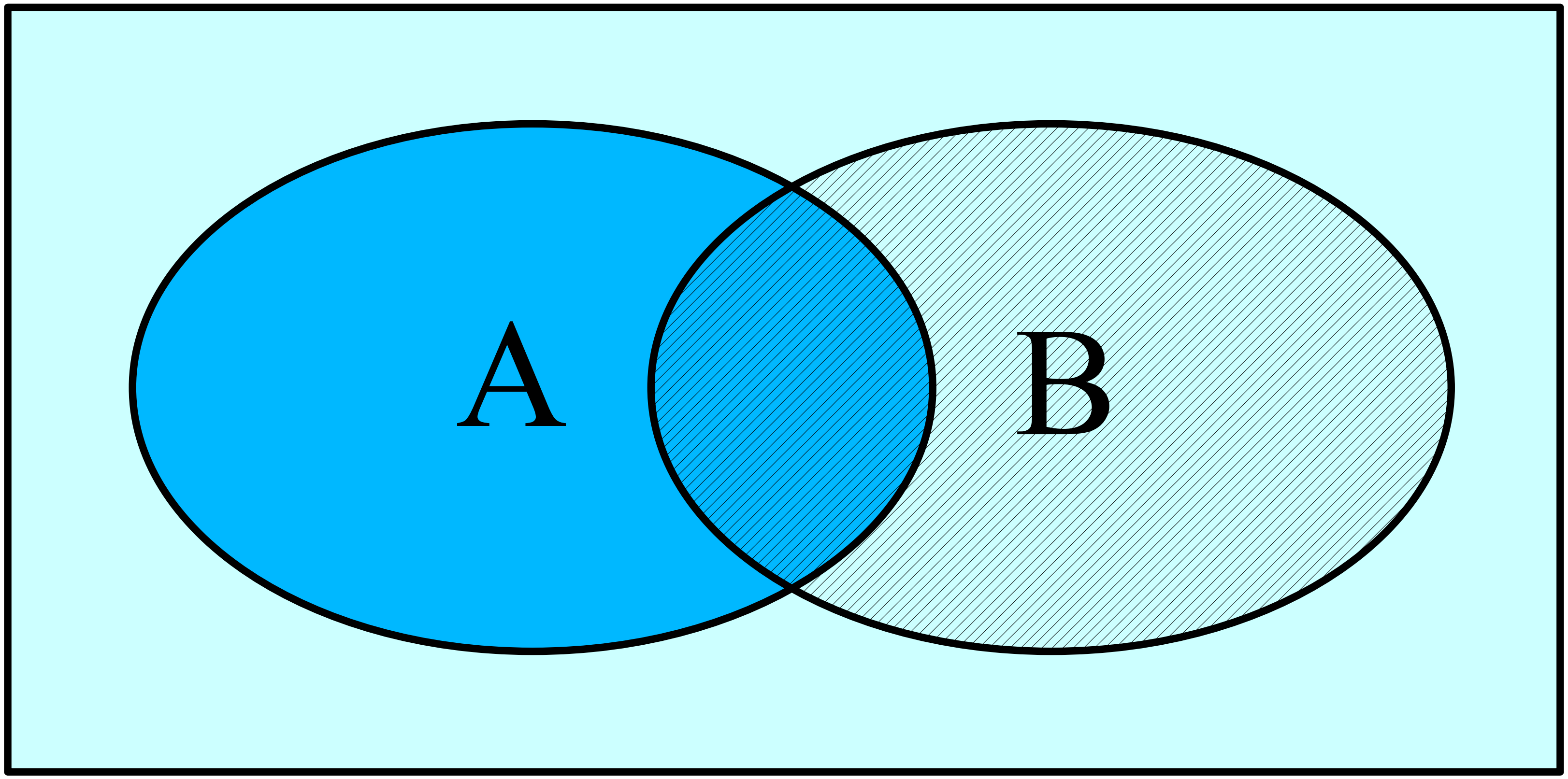

This fun, friendly approach
takes the static out of statistical analysis!
Statistical Analysis
FOR
DUMMIES
Make sense of
probabilities
A Reference
for the
Rest of Us!

Bayes Theorem



P(B|A): conditional probability for B, given that A is true.

Bayes Theorem



$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

Bayesian Statistical Analysis

$$P(\text{signal}|\text{data}) = \frac{P(\text{data}|\text{signal}) \times P(\text{signal})}{P(\text{data})}$$

Bayesian Statistical Analysis

$$P(\text{signal}|\text{data}) = \frac{P(\text{data}|\text{signal}) \times P(\text{signal})}{P(\text{data})}$$

"Posterior probability"

Bayesian Statistical Analysis

$$P(\text{signal}|\text{data}) = \frac{P(\text{data}|\text{signal}) \times P(\text{signal})}{P(\text{data})}$$

"Posterior probability"

Likelihood

Bayesian Statistical Analysis

$$P(\text{signal}|\text{data}) = \frac{P(\text{data}|\text{signal}) \times P(\text{signal})}{P(\text{data})}$$

“Posterior probability”

Likelihood

Normalization factor

“prior probability”

Bayesian Statistical Analysis

$$P(\text{signal}|\text{data}) = \frac{P(\text{data}|\text{signal}) \times P(\text{signal})}{P(\text{data})}$$

"Posterior probability"

Likelihood

Normalization factor

"prior probability"

- Procedure
  1) Determine likelihood from signal and background models
  2) Make assumption for prior and find normalization factor
  3) Compute posterior from actual data
- Measurements based on posterior
  - Cross section (peak) and uncertainty (width)
  - Confidence limit from integrating 95% of posterior area

Statistical Data Analysis

- Event counting
  - Likelihood is a Poisson function
    - Mean $\mu \rightarrow$ Signal + Background
    - Width $\sigma = \sqrt{\mu}$
  - Measurement uncertainty ~ $S/\sqrt{(S+B)}$
  - Search: how far is data away from background?
    - Sensitivity ~ $1/\sqrt{B}$
    - "5 sigma discovery" $\rightarrow S/\sqrt{B} > 5$
- Combining channels
  - Independent datasets
  - Multiply likelihoods $L = L_1 \times L_2$
- Binned likelihood
  - Each bin of a given distribution is a separate channel

$$L(x,\mu) = \frac{\mu^x e^{-x}}{x!}$$