Optimization in Neural Network analysis of Higgs Search at the Tevatron

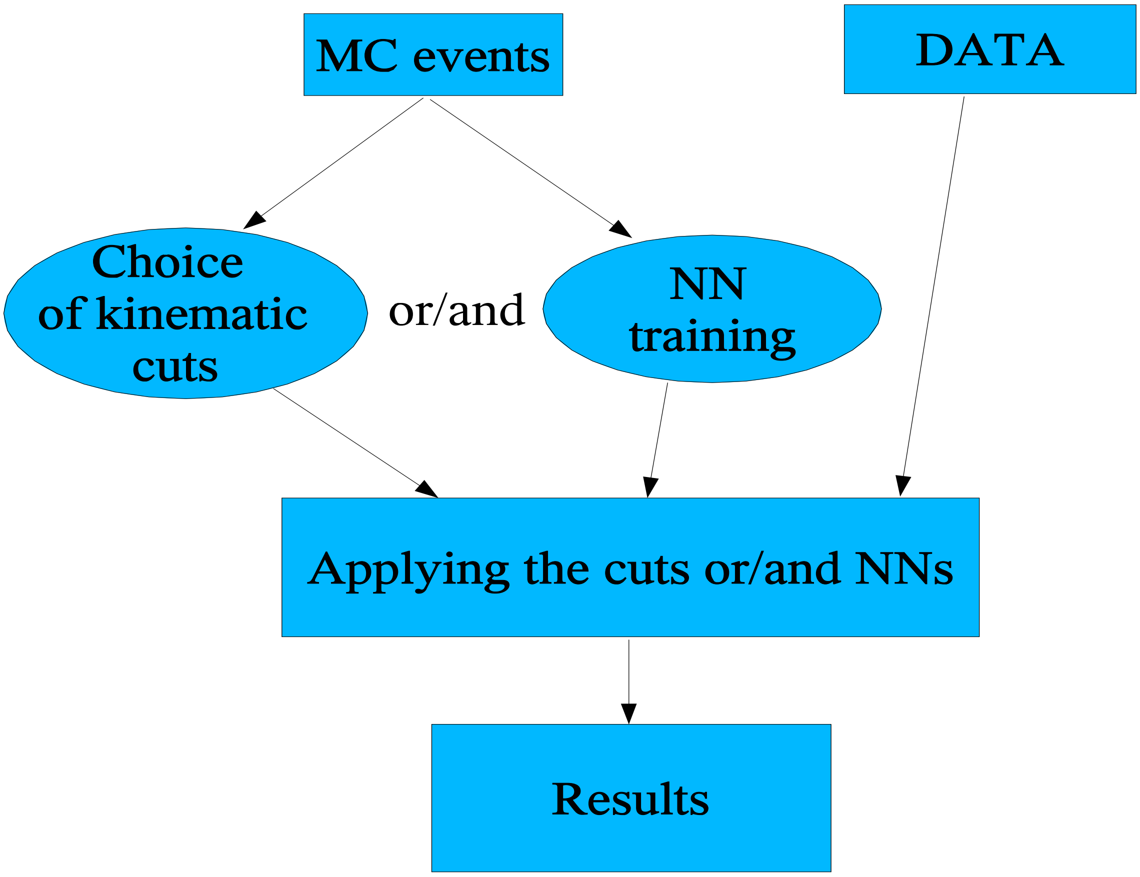
E. Boos, L. Dudko

Scobeltsyn Institute of Nuclear Physics,

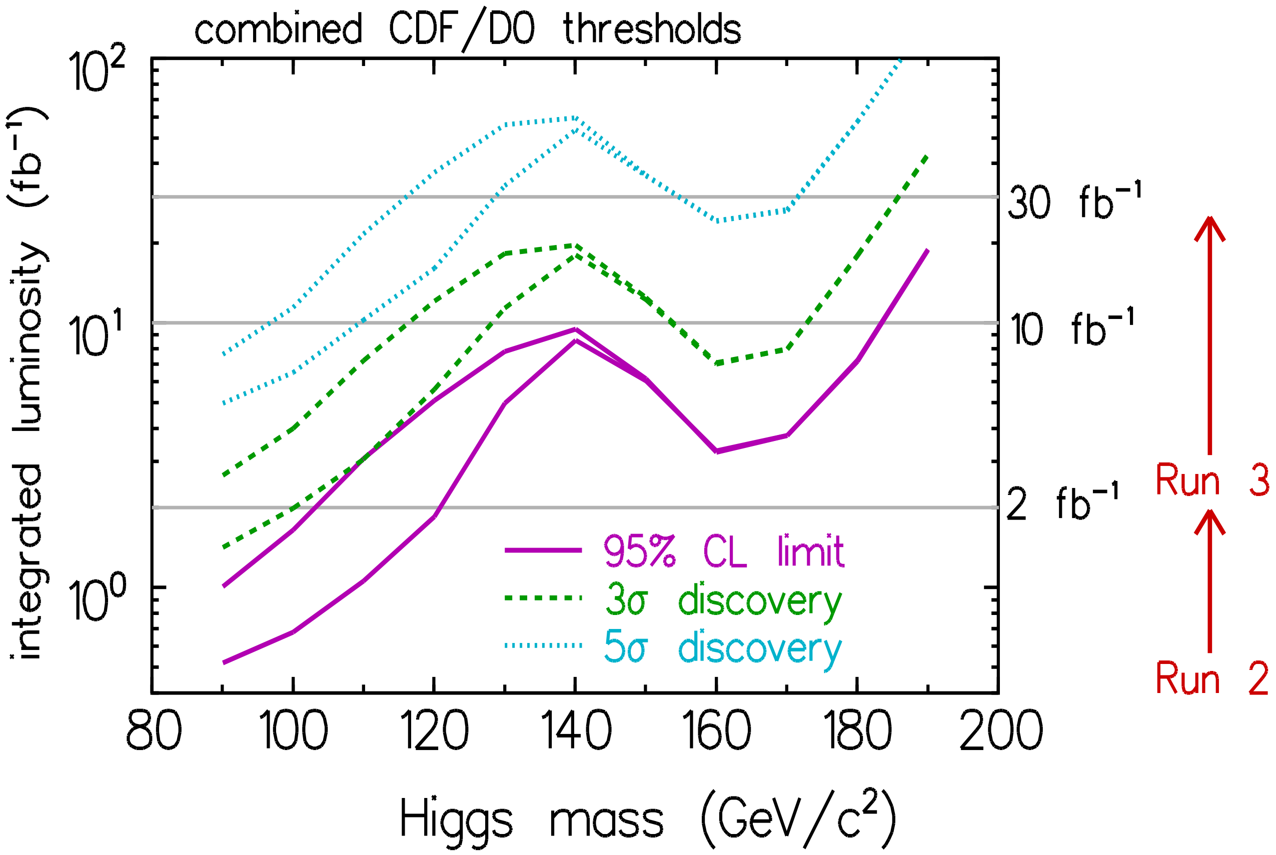
Lomonosov Moscow State University

- Introduction
- Classes of sensitive variables
- Comparison and expectations

Standard Scheme of Collider Physics Analysis



Tevatron Expected efficiency of the SM Higgs search

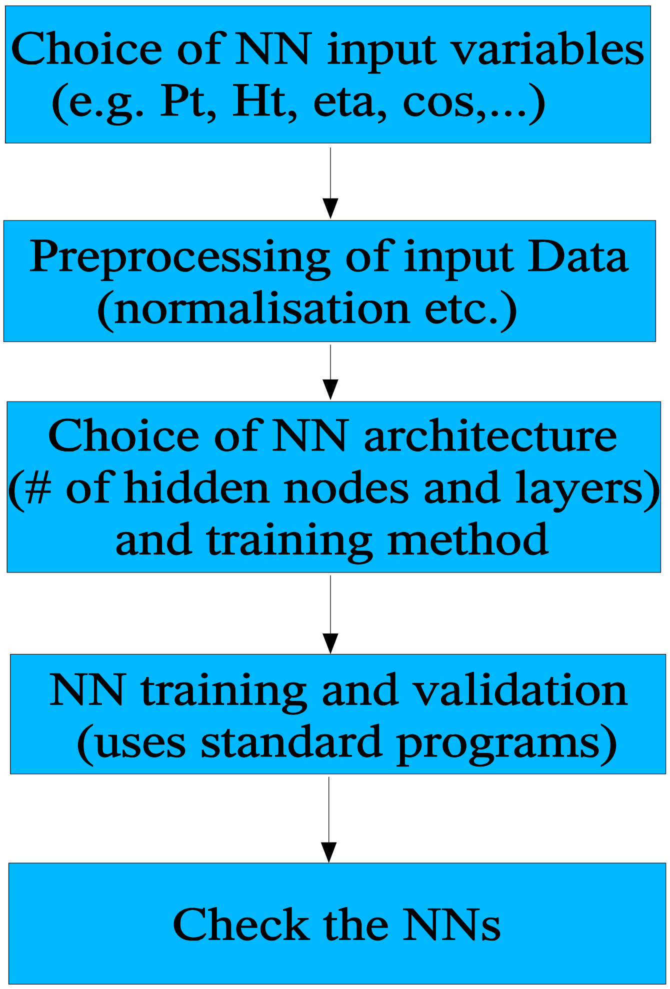


P.Bhat, R. Gilmartin, H.B. Prosper

Phys.Rev. **D62** (2000) 074022

hep-ph/0006162

Standard Scheme of NN Training



Three classes of sensitive variables

- "Singular" variables
  (denominator of Feynman diagrams)

possible s-channel singularity:

$$M^2_{f1,f2} = (p_{f1} + p_{f2})^2 \tag{1}$$

possible t-channel singularity:

$$\hat{t}_{i,f} = (p_f - p_i)^2 = -\sqrt{\hat{s}}e^Y p_T^f e^{-|y_f|} \tag{2}$$

Most of the rates for the signal and for the backgrounds come from the integration over the phase space region close to these singularities. If some of the singular variables are different or the positions of the singularities are different the corresponding distributions will differ most strongly.

Three classes of sensitive variables

- “Angular” variables (numerator of Feynman diagrams, spin effects)

$cos(b1, b1b2)|_{b1b2}$ cosine of the angle between highest P_T b-quark and vector sum of the two highest P_T b-quarks in the rest frame of these two b-quarks.

Scalar (Higgs) and vector (gluon, Z-boson) particle decays lead to significantly different distributions on this variable, this is also very much different for the case when b-quarks come from the decay of top and anti-top quarks;

$cos(z, b1)|_{lab}$ and $cos(W, e)|_W$ reflect the difference in t-channel Z-boson and s-channel Higgs-boson production topologies where $_{lab}$ means the laboratory rest frame and z means the z-axis;

- “Threshold” variables (denominator of Feynman diagrams)

$\hat{s}$ and H_T related to the fact that various signal and background processes may have very different energy thresholds.

Method was approved in the NN analysis of the single top production in DØ (Run I)

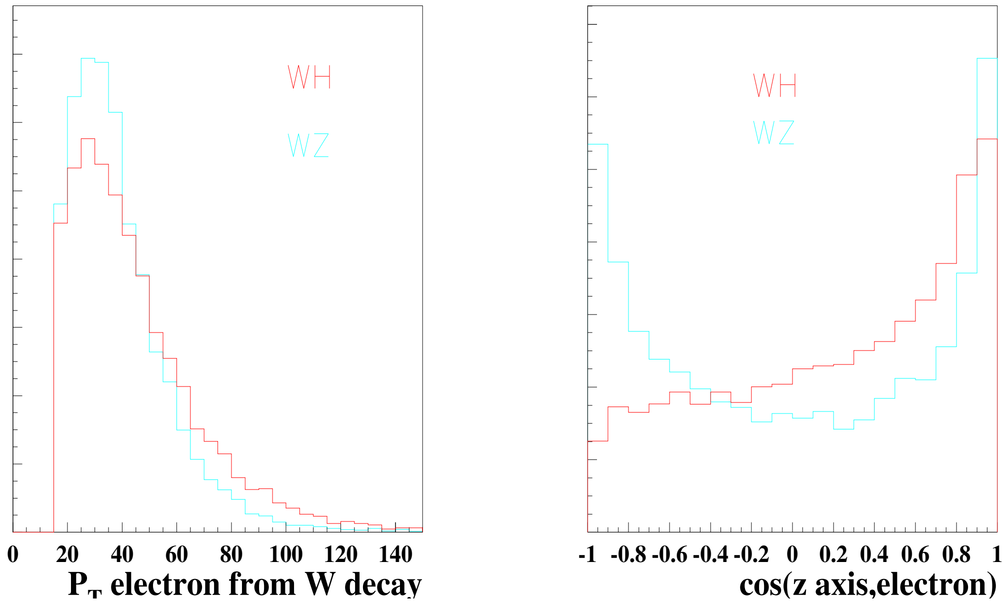
(Phys. Lett. **B517**, 282 (2001),
DØ Note 3856, DØ Note 3612, AIHENP 99, ACAT 2000)

Application for the search of 115 GeV Higgs

- MC model – CompHEP
  (WH, $Wb\bar{b}b$, WZ, $t\bar{t}$, tb, tbj)
  All events are available in MCDB
  `http://www-d0.fnal.gov/~dudko/mcdb`
- Detector smearing – SHW
- NN training package – MLPfit
- Train 4 NNs for the WH and each type of Backgrounds.
- Train with old and new set of input variables
- Compare the efficiency of the NNs

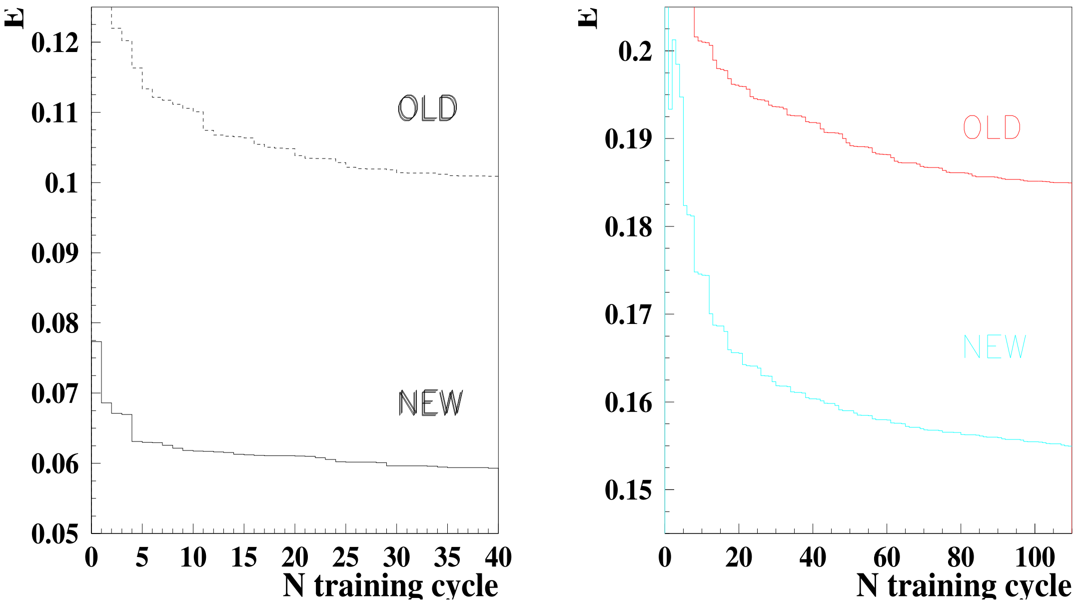
Example of distributions:

Old ineffective (left plot) and new (right plot) variables



NN Error function for the $WH - t\bar{t}$ (left plot) and $WH - WZ$ networks (right plot).

$$E = \frac{1}{N}\sum_{i=1}^{N}(d_i - o_i)^2$$



Final Choice of the Optimal Set of NN Input Variables

- $Wb\bar{b} - WH$
  NN: $M_{b\bar{b}},\ P_T^{b1},\ P_T^{b2},\ P_T^{bb},\ \hat{s},\ H_T^{jets},$
  $cos(b1,b2)|_{lab},\ cos(b1,b1b2)|_{b1b2}$

- $WZ - WH$
  NN: $M_{b\bar{b}},\ P_T^{b1},\ P_T^{b2},\ H_T^{jets},$
  $cos(b1,b2)|_{lab},\ Q\times cos(z,b1)|_{lab},$
  $cos(W,e)|_W$

- $t\bar{t} - WH$
  NN: $M_{b\bar{b}},\ M_{Wb},\ \hat{s},\ M_{Wjets-b},\ H_T^{jets},$
  $Q\times cos(\psi_{axis},e)|_{top},\ cos(b1,b1b2)|_{b1b2}$

- $tbj,\ tb - WH$
  NN: $M_{b\bar{b}},\ M_{Wb},\ P_T^{b2},\ \hat{s},$
  $M_{Wjets-b},\ P_T^{top},\ H_T^{jets},\ cos(z,e)|_{lab},$
  $Q\times cos(z,b1)|_{top},\ cos(e,j)|_{top}$

Conclusion

- Based on the analysis of Feynman diagrams we can construct the optimal set of kinematical variables
- The method is efficient and universal
- Direct implementation to the Higgs search, significantly improve the efficiency of the NNs
- To estimate the improvement in Tevatron potential need carefull analysis of real detector efficiecies
- The constructed set of input variables are ready for implementation in real physics analysis.
- detailes are in hep-ph/0302088