History of Machine Learning in High Energy Physics

Pushpa Bhat
Fermilab

ARISA

Intelligence

René Descartes

"Cogito ergo"
" I think, therefore I am"

Leibniz, Hobbes and Descartes explored if rational thought could be made systematic

From times immemorial,
humans have built tools
and fancied "intelligent" machines

"The Thinker"

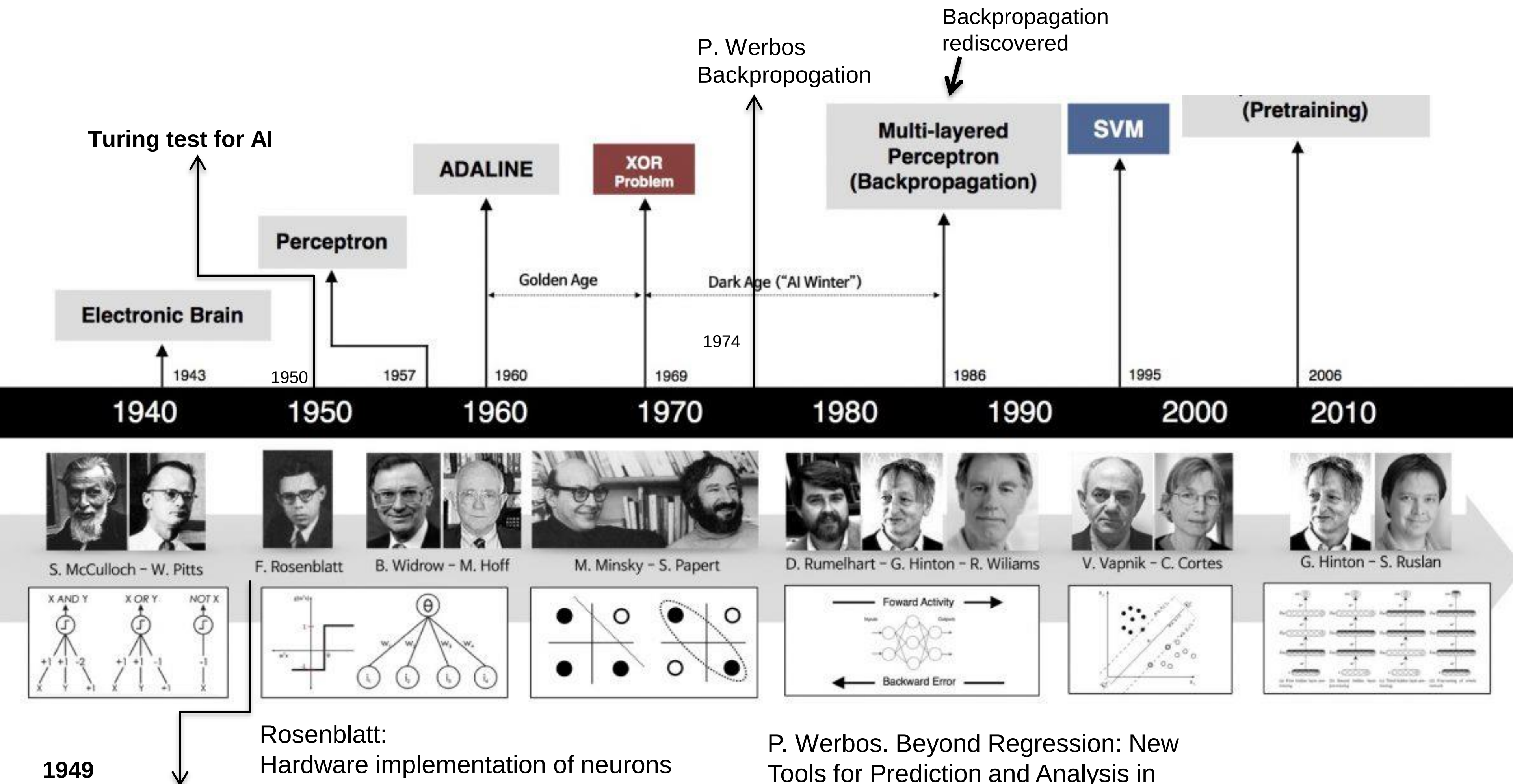
Rodin's "Le Penseur"

Alan Turing

"Can machines think?"

Turing Test (1950):
If a machine could carry on a conversation that is indistinguishable from that with a human being, then it is reasonable to say it is "thinking"

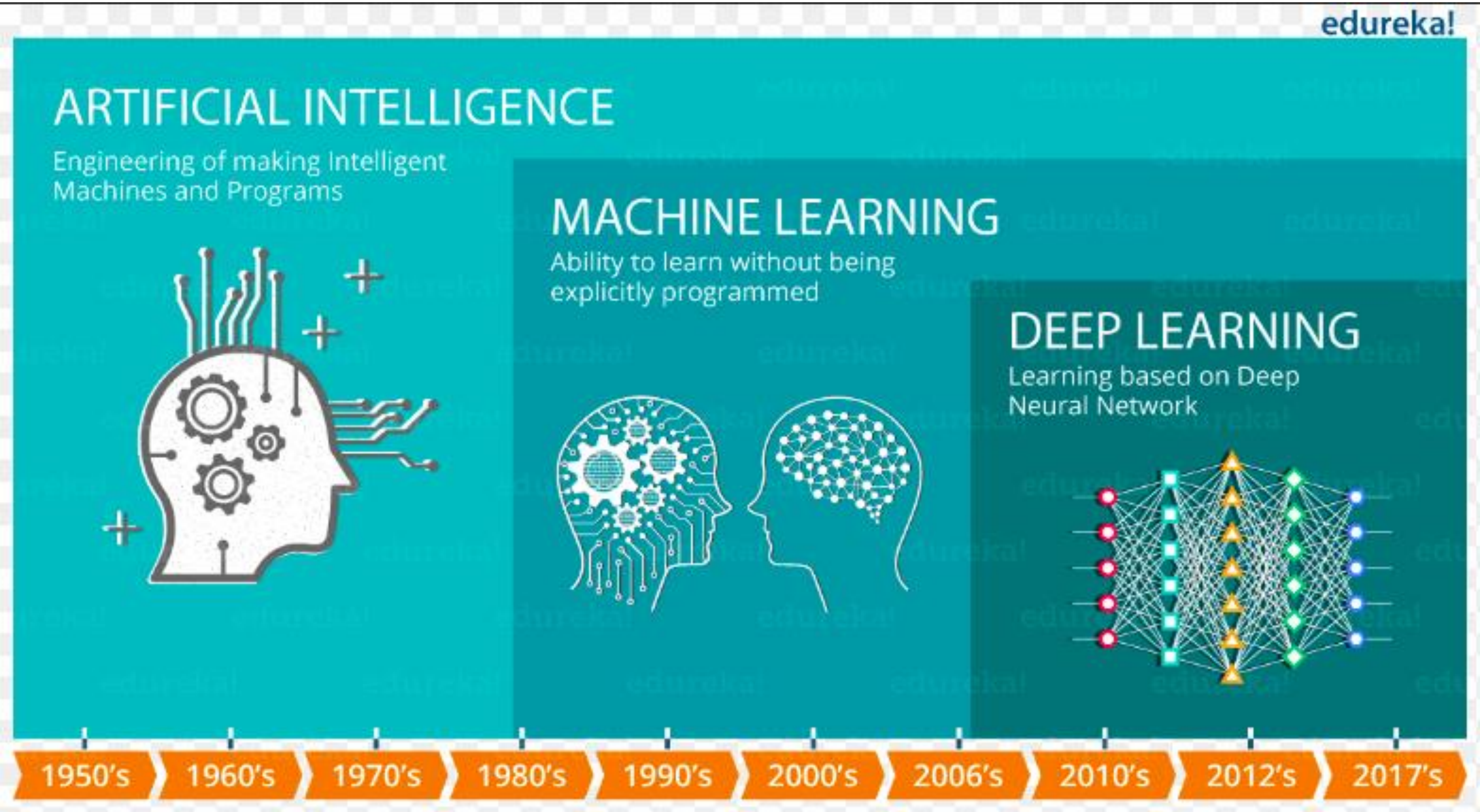
Machine Learning Milestones



1949
Hebb's learning

Rosenblatt:
Hardware implementation of neurons
First visual pattern recognition device

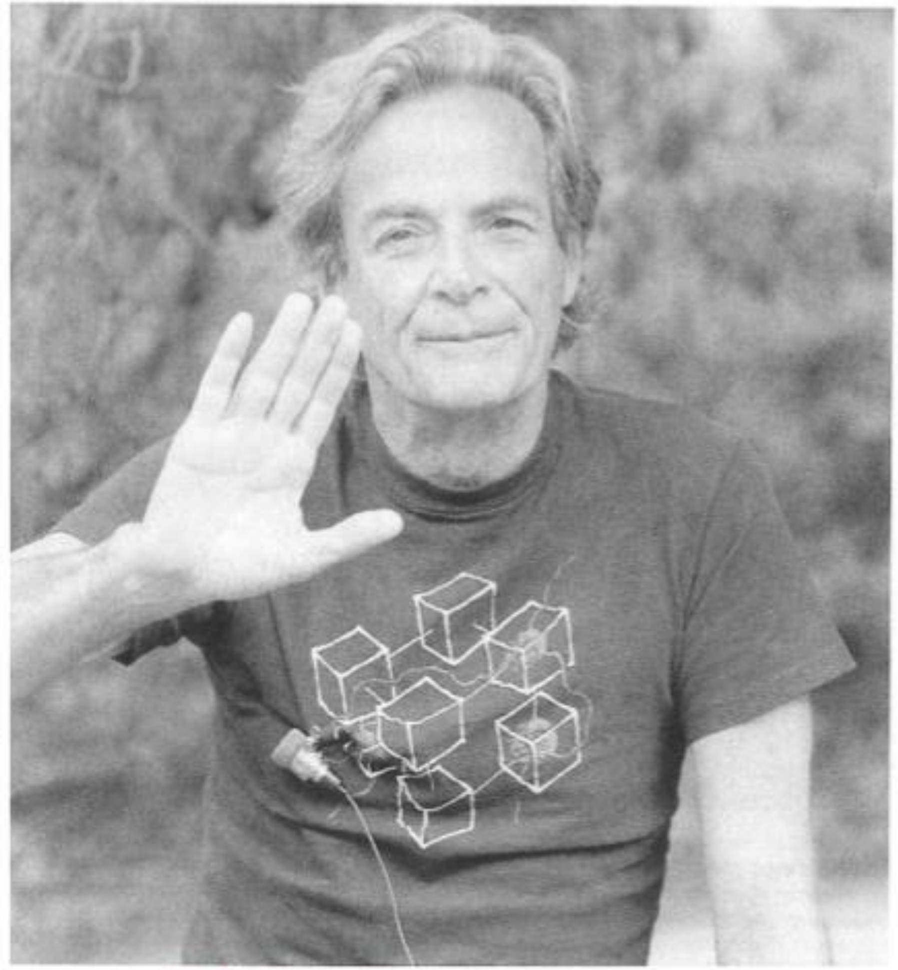
P. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974.


edureka!
ARTIFICIAL INTELLIGENCE
Engineering of making Intelligent Machines and Programs
MACHINE LEARNING
Ability to learn without being explicitly programmed
DEEP LEARNING
Learning based on Deep Neural Network
1950's
1960's
1970's
1980's
1990's
2000's
2006's
2010's
2012's
2017's

Richard Feynman at the
Thinking Machines, Inc., (1983)

The schematic representation of the **connection machine** that he helped design on his T-shirt

Feynman worked out in detail the program for computing Hopfield's neural network on the connection machine



Richard Feynman

Feynman's reaction on hearing about the Thinking Machine
"That is the dopiest idea that I have ever heard"
And went ahead and signed up to work.

ML in HEP: Early History

- Backpropagation in MLPs was a breakthrough that brought renewed interest in neural networks and their applications in research
- 1987-89
  - Applications in particle tracking
  - NN hardware for triggering
  - B. Denby, C. Lindsey (Fermilab)
    - Track finding, track triggers
  - C. Peterson, L. Lönnblad, et al (Lund group)
    - MLP implementation in JETNET package
    - NN for track finding, classification

Lindsey and Denby

- 1990-
  - P. B. (with H. Prosper) initiated NN for top search @D0 after q/g discrimination study at Snowmass
- 1990-92 LEP experiments started using NNs in data analysis

Machine Learning in HEP, The Early Days

- 1988 - B. Denby, Comp. Phys. Comm.49:429 (1988)
  - C. Peterson NIM A279, 537 (1989); LU-TP-88-8(1988)
- 1990 L. Lönnblad, C. Peterson, T. Rögnvaldsson, *Phys. Rev. Lett.* 65:1321 (1990); JETNET
- 1990 P. Bhat, L. Lönnblad, K. Meier, K. Sugano, *Snowmass Meeting*
- 1990 B. Humpert, Comp. Phys. Comm. 56:299
- 1992 C. Peterson, CHEP 92, B. Denby, FERMILAB-CONF-92-269-
- 1994 P.C. Bhat (for the DØ Collaboration) on top search, DPF'94
- 1994 C. Peterson et al., Computer Physics Communications, 81:185
- 1997 Moneti (CLEO Collaboration) Nuclear Physics B (Proc. Suppl.) 59:17 (1997).
- Many examples of applications at the Tevatron, LEP, DESY, SLAC, CESR, KEK, ...

International Workshop Series on Software Engineering, Artificial Intelligence and Expert Systems in High Energy and Nuclear Physics

Initiated by Denis Perret-Gallix

AIHENP Workshops

- 1990 Lyon, France, March 19-24 (2 NN papers)
- 1992 La Londe Les Maures, France, Jan. 13-18
- 1995 Pisa, Italy, April 3-8
- 1996 Lausanne, Sep. 2-6
- 1999 Heraklion, Crete, April 12-16
- 2000 @ Fermilab, AIHENP → ACAT

  Advanced Computing and Analysis Techniques in Physics Research

  (Organized by P Bhat, M Kasemann)

AIHENP → ACAT

VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research

ACAT 2000
(Formerly AIHENP)

October 16-20, 2000

Fermi National Accelerator Laboratory

Artificial Intelligence, Innovative Software Algorithms and Tools, Symbolic Problem solving and Large Scale Computing
in High Energy Physics, Astrophysics, Accelerator Physics and Nuclear Physics

http://conferences.fnal.gov/acat2000/
Checkout videos of plenary talks

Rene Brun inspired at ACAT2000 to launch TMVA in ROOT

ACAT2017

International Workshop

AI Reloaded

August 21-25, 2017

University of Washington, Seattle, USA

Empowering the Revolution:
Brining Machine Learning to High Performance Computing

https://indico.cern.ch/event/708041/

ML/MVA in the early 1990s
Skepticism in the Community

Needed a lot of patience, perseverance for years, and some very good quotations!

Examples of quotes on my title slides

A reasonable man adapts himself to the world.
An unreasonable man persists to adapt the world to himself.
So all progress depends on the unreasonable one.
- Bernard Shaw

We are riding the wave of the Future!

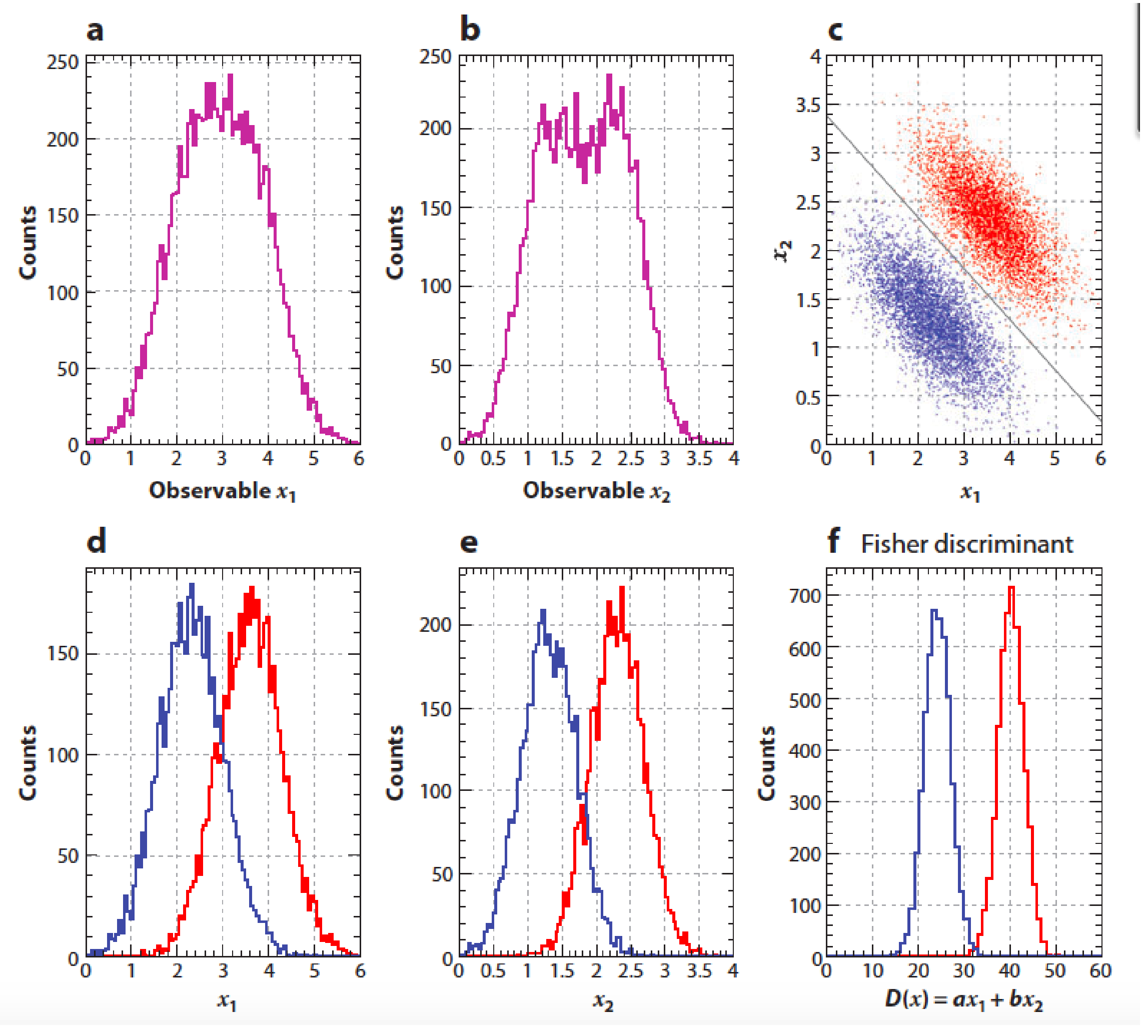
Truth shall set you free!

Keep it simple
As simple as possible
Not any simpler
- Einstein

Tevatron Run I

- DZero did not have the SVX detector in Run I and so MVA/ML methods became critical.

Why Multivariate Methods?

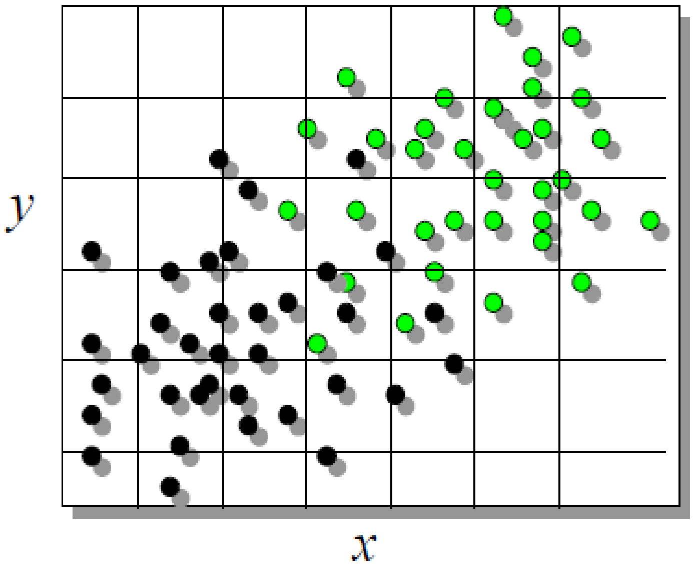


- ❖ **Because they are optimal!**

Optimizing "Rectangular" Cuts

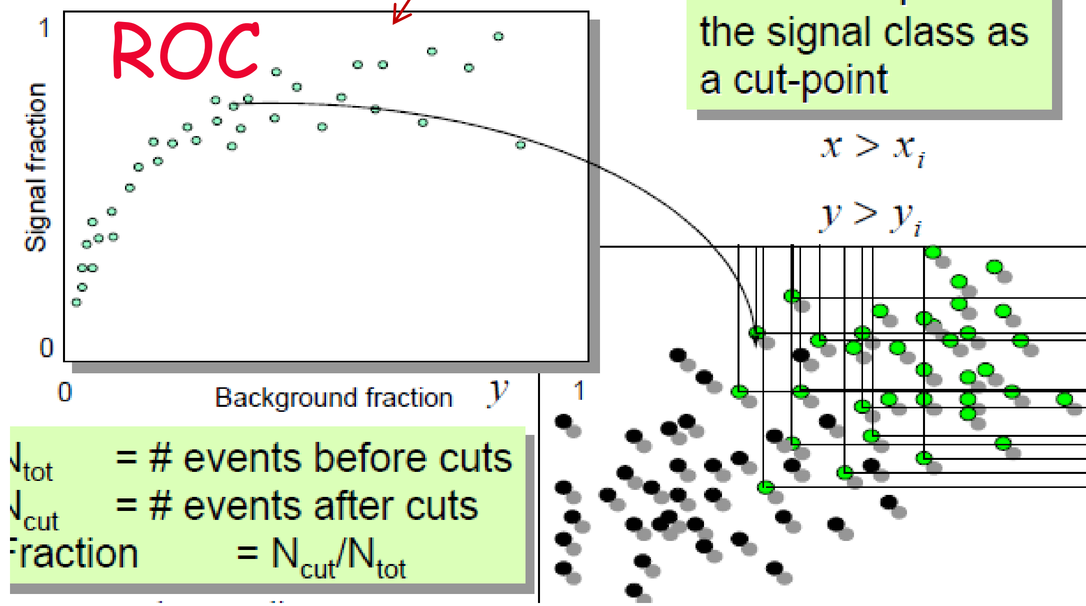
Bhat, Prosper, Stewart, 1993-94

Regular Grid search



Signal eff. Vs bkgd. eff

Take each point of the signal class as a cut-point



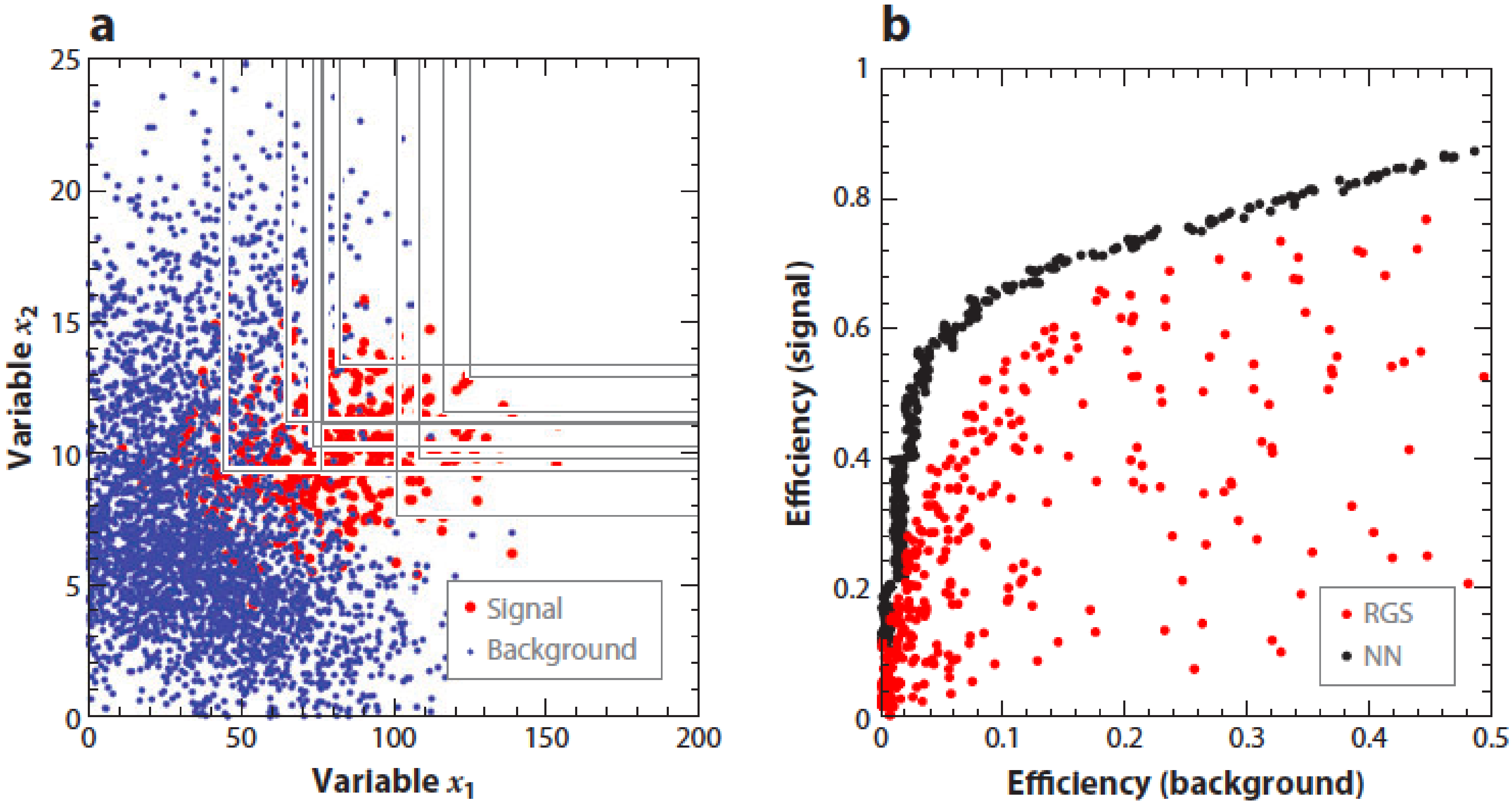
RGS can serve as a benchmark for comparisons of efficacy of variables, variable combinations, and classifiers

Random Grid search (RGS)

Find "best" conventional cuts

PB, et al., Computer Phys. Commun. 228 (2018)245

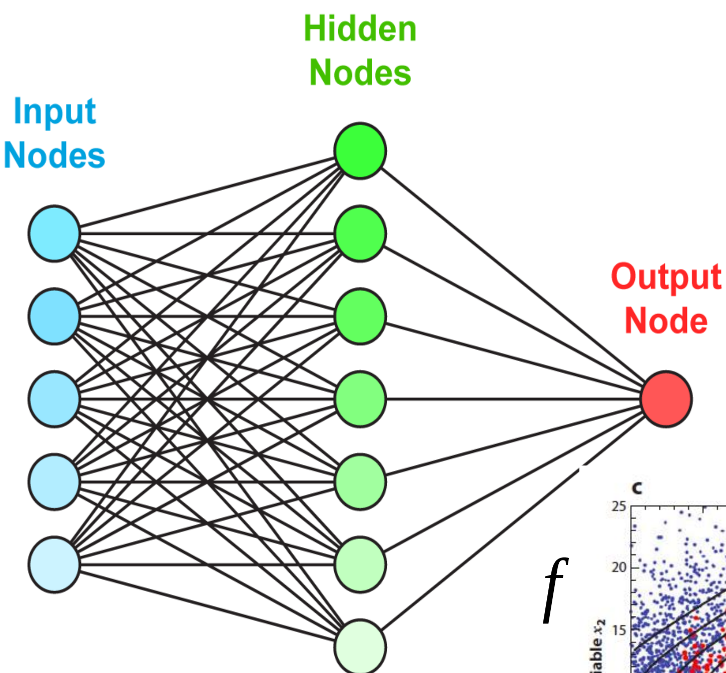
Random Grid Search



P. B., Annu. Rev. Nucl. Part. Sci. 2011. 61:281–309
(Multivariate Analysis Methods in Particle Physics)

Neural Networks
The Bayesian Connection

- The output of a neural network can approximate the Bayesian posterior probability p(s|x):

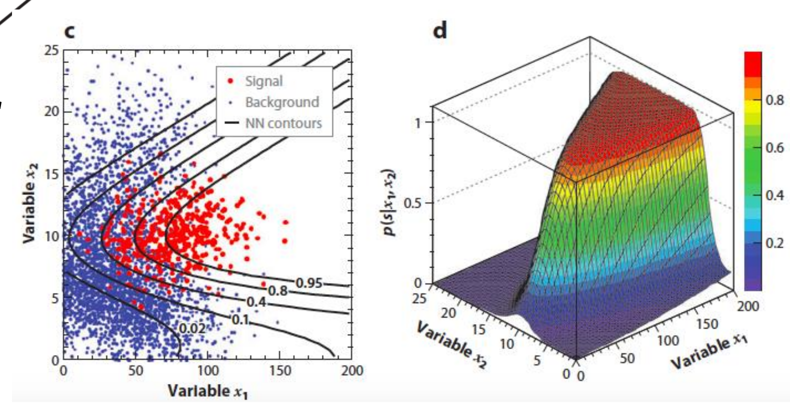


$$f(\mathbf{x},\mathbf{w}) = g(\textstyle\sum_j w_j h_j + \theta) = p(s\,|\,\mathbf{x})$$

$$h_j = g(\textstyle\sum_j w_{ij} x_i + \theta)\,; \quad g(a) = \frac{1}{1+e^{-a}}$$

$$f(x,\hat{w}) \text{ » } p(s\,|\,x) = \frac{r}{1+r}$$

f



$$r = \frac{p(x\,|\,s)p(s)}{p(x\,|\,b)p(b)}$$

Flexible, non-linear model

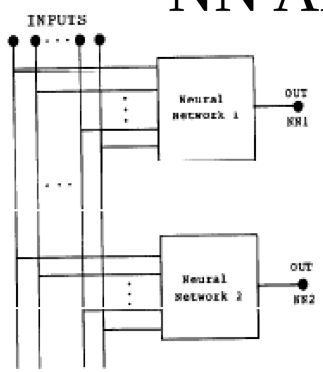
Search for the Top Quark at DØ using Multivariate Methods

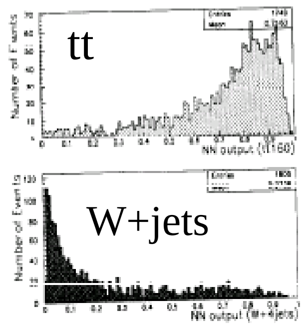
FERMILAB-Conf-94/261-E

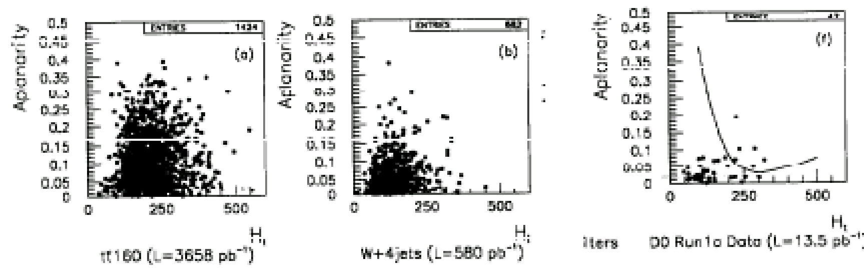
P.B. for DØ   DPF94

Analysis with 2 and 5 variables

NN Analysis tt → e+jets channel







tt160   W+jets   Data

- Features:
  - Used two independent networks to suppress separate backgrounds
- If DØ had pursued the NN analysis in other channels, the evidence and/or discovery could have been had sooner?

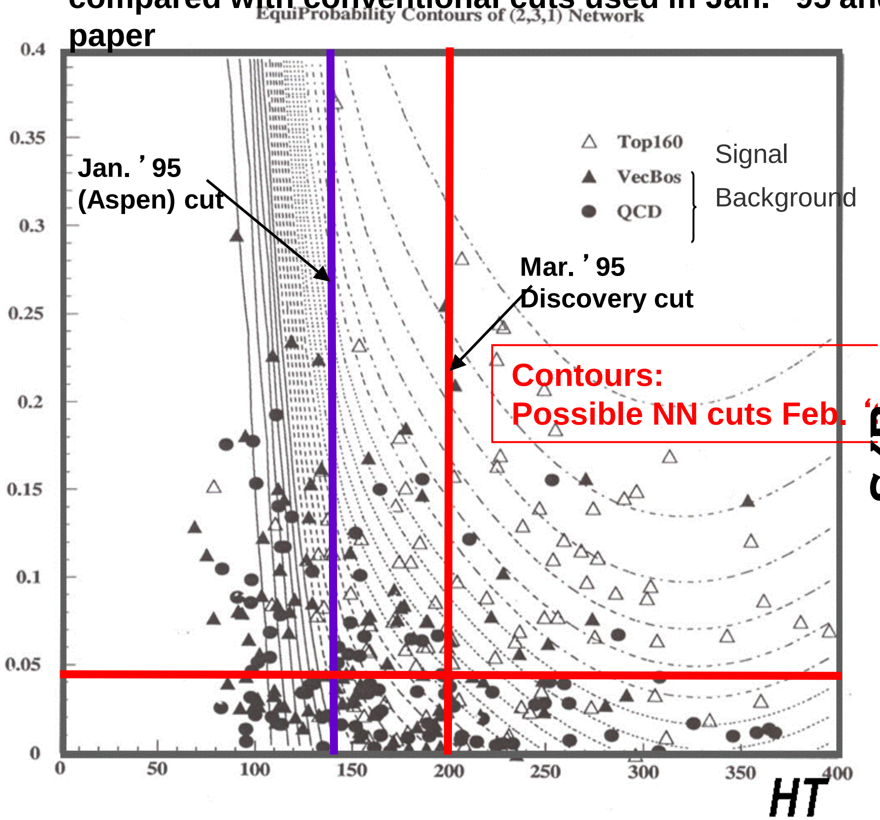
hep-ex/9507007

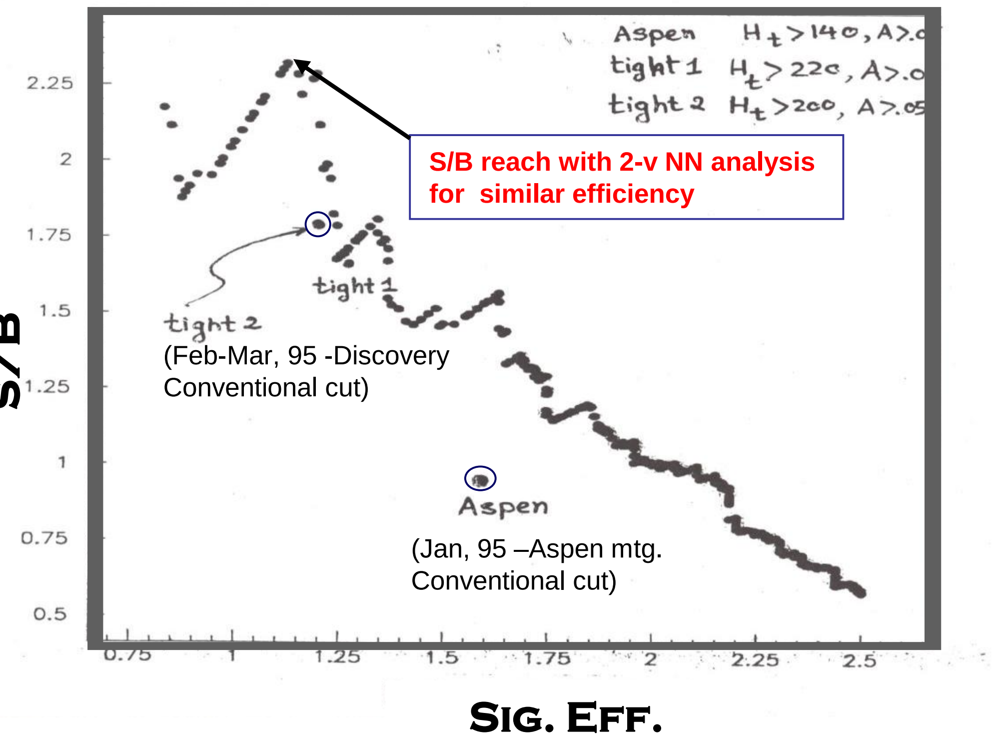
Top Quark Discovery: Cut Optimization Feb. 1995

P. Bhat, H.Prosper, E. Amidi
D0 Top Marathon, Feb. ‘95

Aplanarity & HT variables
Letpon+jets channels

Neural Network Equi-probability Contour cuts from 2-variable analysis compared with conventional cuts used in Jan. ’95 and in Observation paper





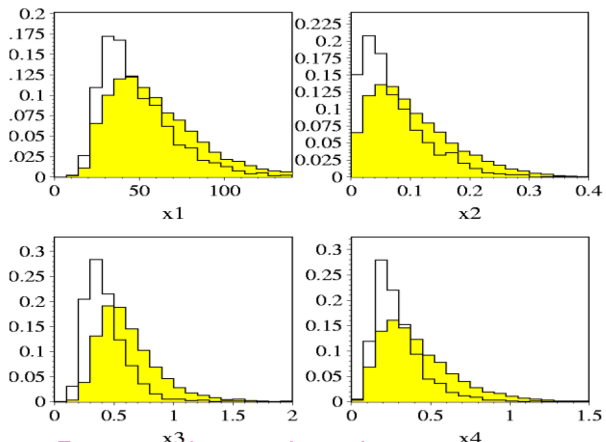
1996

Direct Measurement of the Top Quark Mass

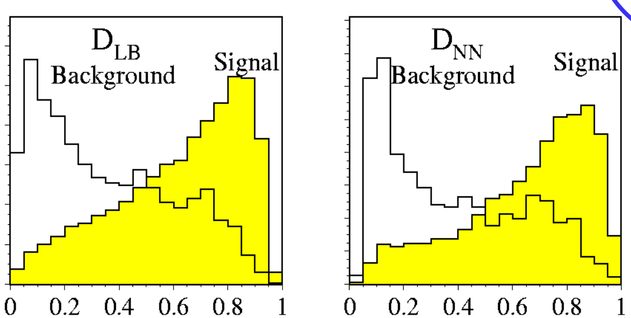
S. Abachi *et al.* (D0 Collaboration)
Phys. Rev. Lett. **79**, 1197 – Published 18 August 1997

First significant physics result using multivariate methods

Discriminant variables

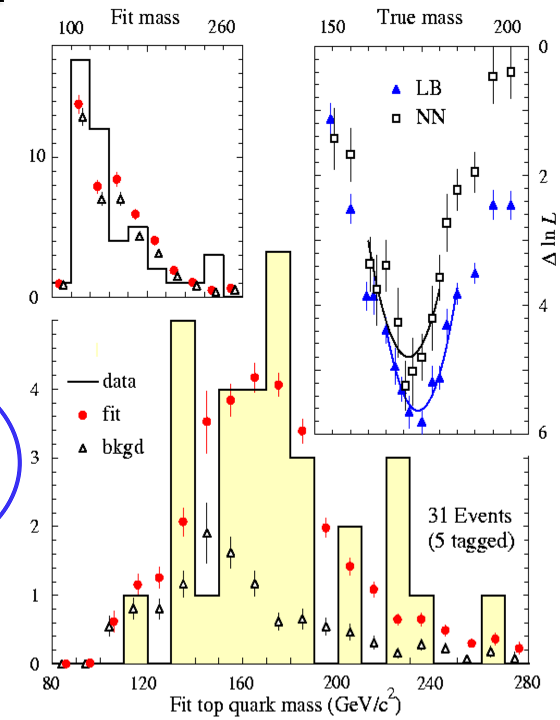


The Discriminants



$$D = \frac{P_s}{P_s + P_b}$$

DØ Lepton+jets



m_t = 173.3 ± 5.6(stat.) ± 6.2 (syst.) GeV/c^2

Fit performed in 2-D: ($D_{LB/NN}$, m_{fit})

LB: Low-bias maximum likelihood
NN: Neural Networks

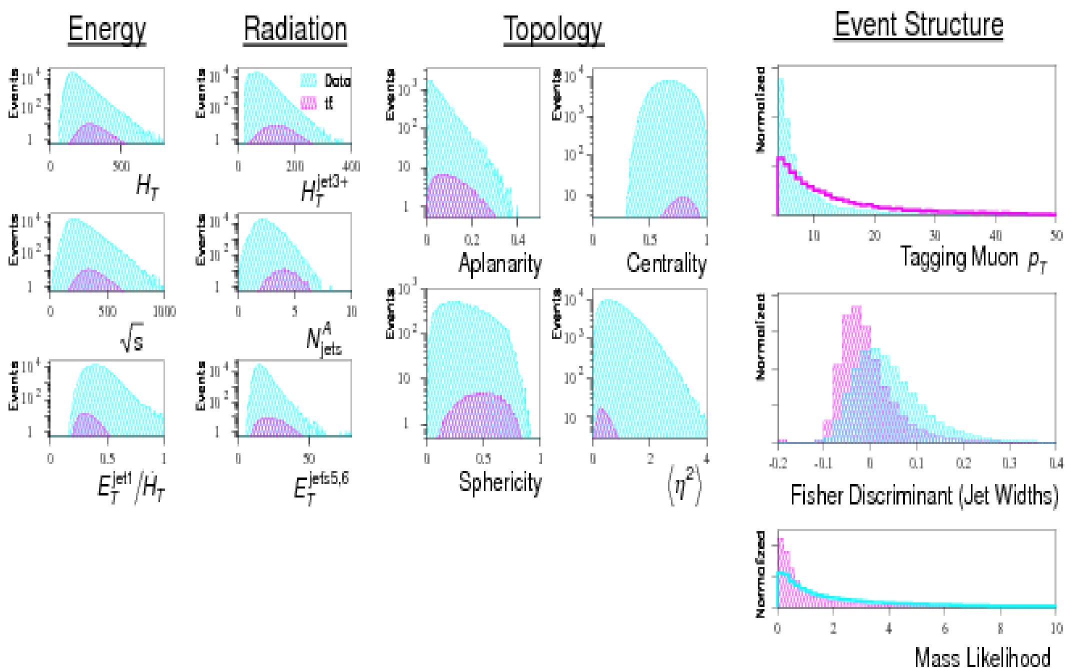
Statistical error for the same data sample reduced from 11.7 GeV to 5.6 GeV!

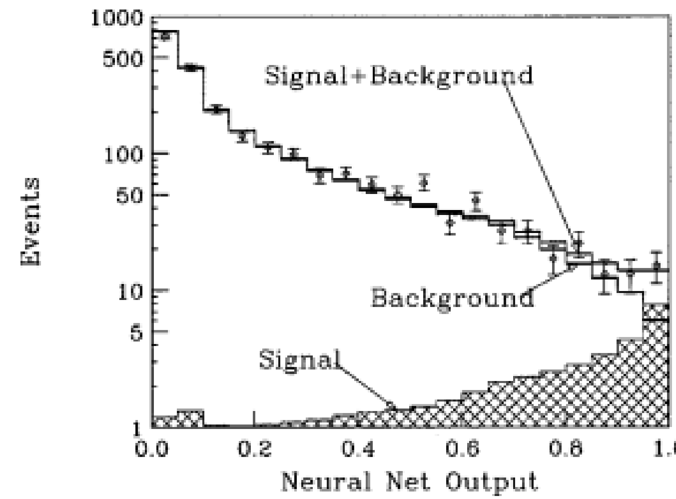
Top → All jets

B. Abbott *et al.* (DØ Collaboration)
Phys. Rev. D **60**, 012001 – Published 26 May 1999

- 18 variables!

110 pb^{-1}





igure 2 Neural network variables for the DØ $t\bar{t} \to$ alljets analysis. The first 10 variables are used in one network, and the output from that network is used together with the last three variables in a second network.

Multijet data used as background for training

Matrix Element Method

- Maximal use of information in each event by calculating event-by-event signal and background probabilities based on the respective matrix element

$$P_{evt}(x; m_{top}, JES) = f_{top} P_{sig}(x; m_{top}, JES) + (1 - f_{top}) P_{bkg}(x; JES)$$

x: reconstructed kinematic variables of final state objects

JES: jet energy Scale from Mw constraint

- Signal and background probabilities from differential cross sections

$$P(x; m_{top}) = \frac{1}{\sigma} \int d^n\sigma(y; m_{top})\, dq_1\, dq_2\, f(q_1)\, f(q_2)\, W(x, y)$$

- Write combined likelihood for all events

$$-\ln L(x_1, \ldots, x_n; m_{top}, JES) = -\sum_{i=1}^{n} \ln P_{evt}(x_i; m_{top}, JES)$$

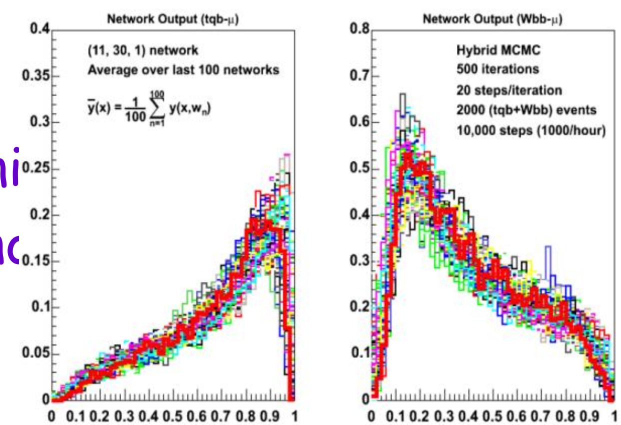
- Maximize likelihood w.r.t. m_{top}, JES

Bayesian Neural Networks

- Instead of attempting to find a single "best" network, i.e., a single "best" set of network parameters (weights), with Bayesian training we get a posterior density for the network weights, $p(w|T)$, $T\equiv$ Training data
- The idea here is to assign a probability density to each point w in the parameter space of the neural network. Then one takes a weighted average over all points, i.e., over all possible networks.

$$\tilde{y}(\mathbf{x}) = \int f(\mathbf{x},\mathbf{w})\, p(\mathbf{w}\,|\,T)\, d\mathbf{w}$$

- Advantages:
  - Less likely to be affected by "over traini
  - No need to limit the number of hidden no
  - Good results with small training sample



P.C. Bhat, H.B. Prosper Phystat 2005, Oxford

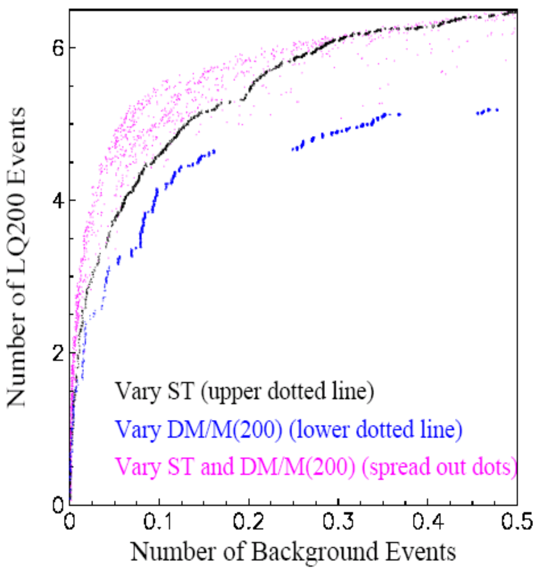
Leptoquark Searches

- H1 and ZEUS at HERA reported ('97) an excess of events at high Q^2 which could be interpreted as due to first generation scalar leptoquarks (LQ1) with a mass around 200 GeV with 100% branching to eq.
- If LQ1 exists, they can be pair produced at the Tevatron. So, CDF and DØ went to work to look into data for evidence for their presence.
- Signature: 2 electrons + 2 jets.
- Backgrounds: Drell-Yan, ttbar → dilepton channel, Multijet with fake leptons.

Leptoquark Searches

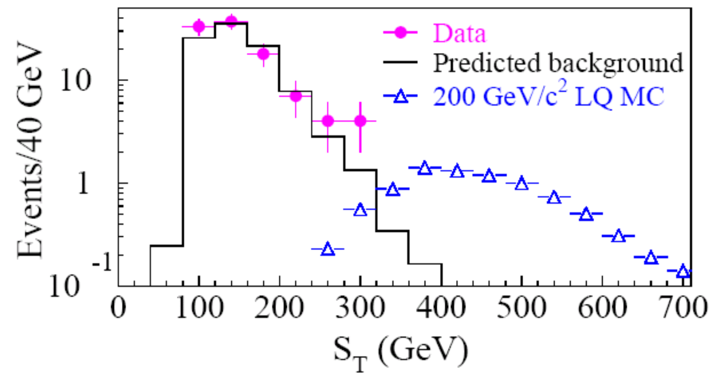
Phys. Rev. Lett. 79 (1997)4321

1st Generation Scalar $LQ\ \overline{LQ} \longrightarrow eejj$

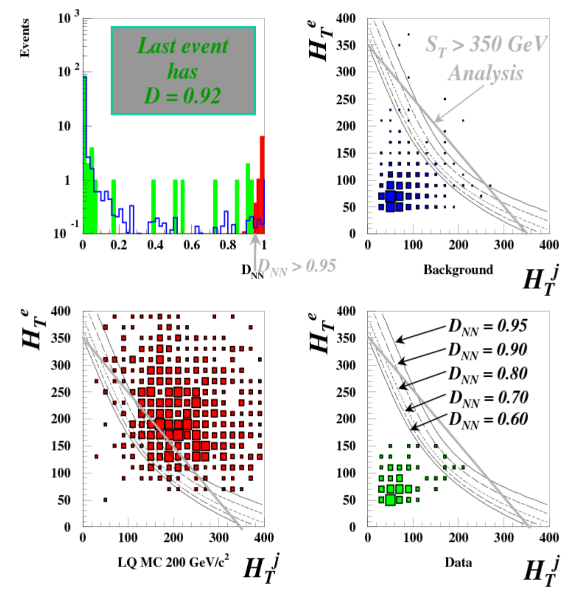


>=ETE1 20.
>=ETE2 20.
>=ETJ1 15.
>=ETJ2 15.

$\int Ldt = 123\ pb^{-1}$



S_T> 350 GeV ➔ ~0.4 bkgnd.
0 events onserved.



$\beta = 1$:
$M_{LQ} > 225\ GeV/c^2$ at 95% CL

Random Grid Search
Maximize signal
for a given background

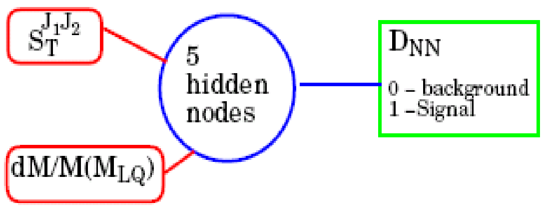
$$DM/M(M_{LQ}) = sqrt((Mej_1 - M_{LQ})^2 + (Mej_2 - M_{LQ})^2)/M_{LQ}$$

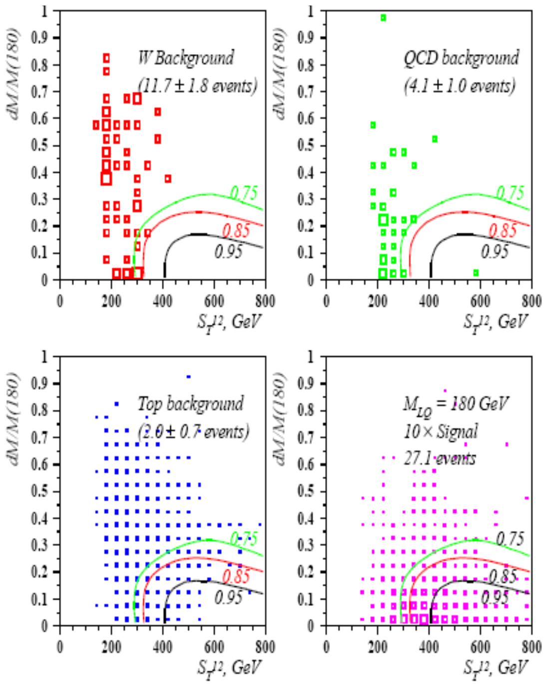
$$S_T = (E_T(e_1) + E_T(e_2) + \Sigma E_T(j_i))$$

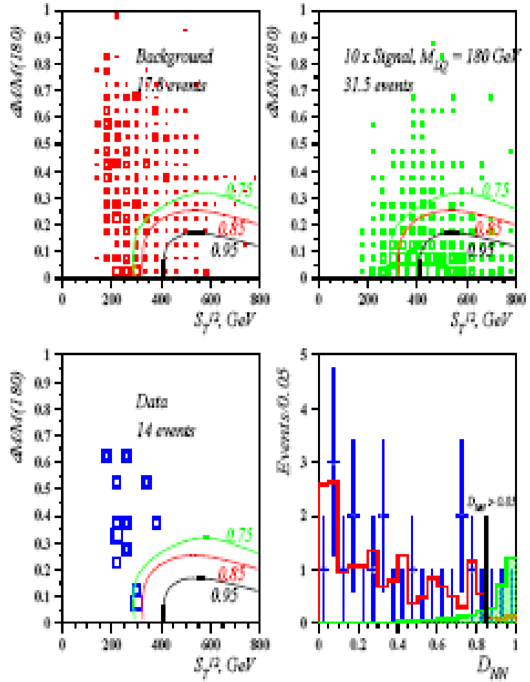
DØ Leptoquark Searches

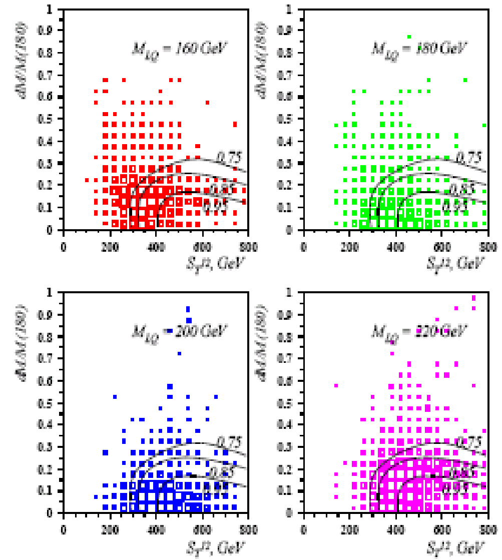
Phys. Rev. Lett. 80 (1998) 2051

Neural Network Analysis: 1st Generation Scalar $LQ\ \overline{LQ} \longrightarrow e\nu jj$









B=.29±.25 evt, 0 events observed.

$\beta = 1/2$:
$M_{LQ} > 204$ GeV/c² at 95% CL

Background, data, and signal after $M_T(e\nu)$ cut

Excludes interpretation of HERA excess as
Due to first generation leptoquarks

More in Run I

- Other analyses that used Multivariate methods in Run I
  - Top mass measurement in dilepton analysis uses KDE.
  - Top cross section analyses with topological variables
  - Other Leptoquark searches
  - Technicolor searches
  - Analyses with Taus
  - Single top search at CDF

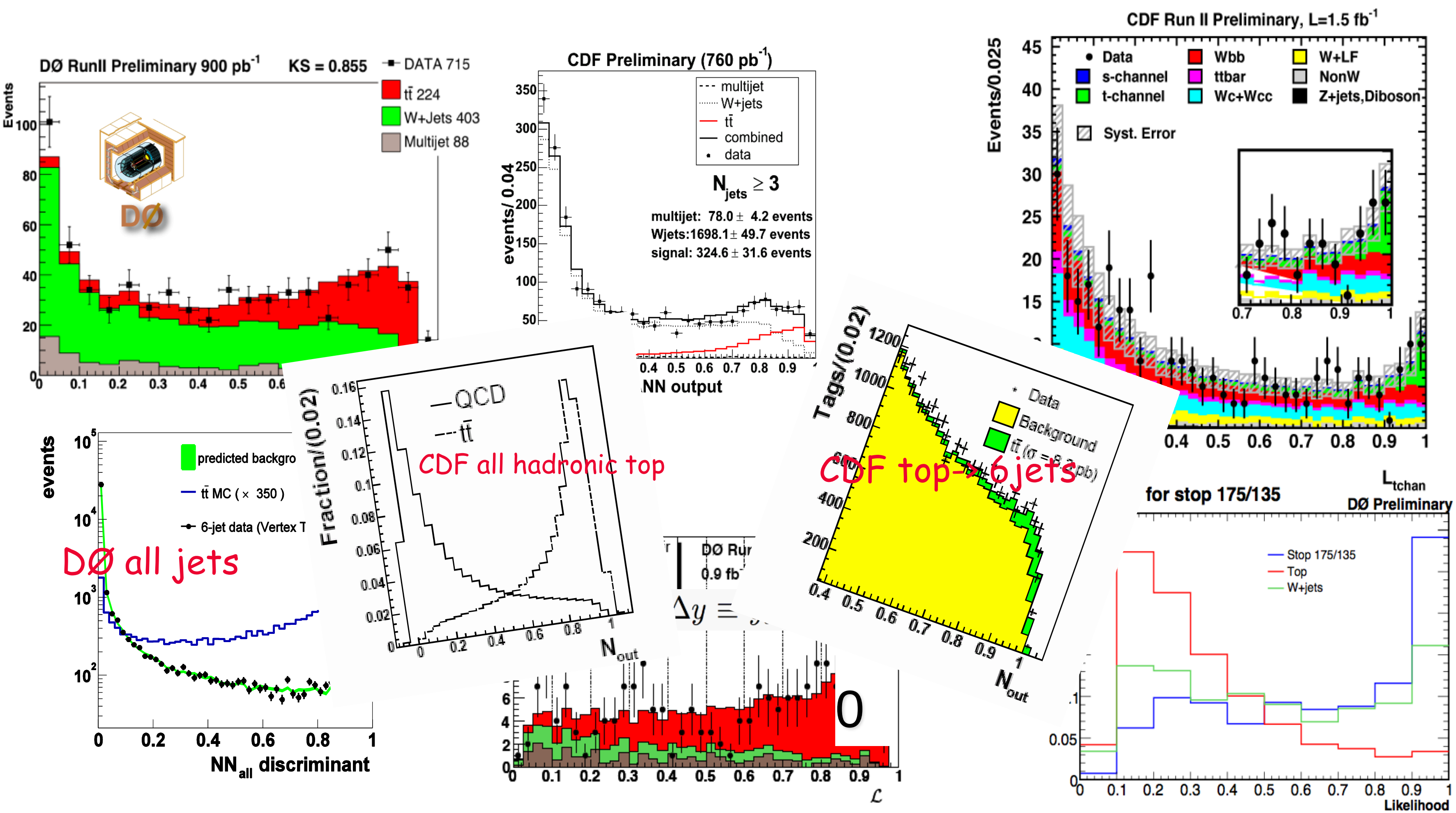
Run II

Fermilab RUN II Advanced Analysis Group created; meetings, workshops, tutorials 2000-2004

Multivariate Methods in Run II

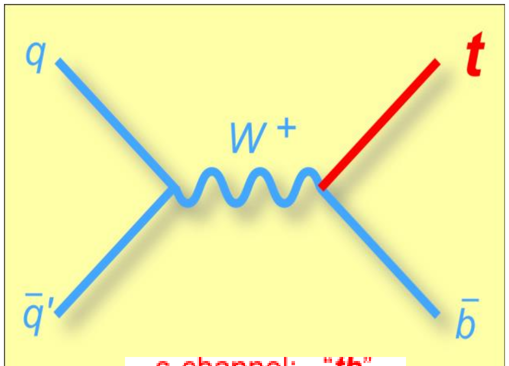
- The use of Multivariate Methods is ubiquitous in Run II !!!
- NN used extensively and routinely in
  - b-tagging
  - Tau ID
  - Jet energy corrections
  - Top Cross section and mass measurements
  - Top decay properties – W helicity, forward-backward asymmetry,..
  - Higgs Searches
  - New particle searches
  - + other Multivariate methods such as DT, ME, likelihoods
  - BNN, BDT (and ME) became popular methods

Ubiquitous!



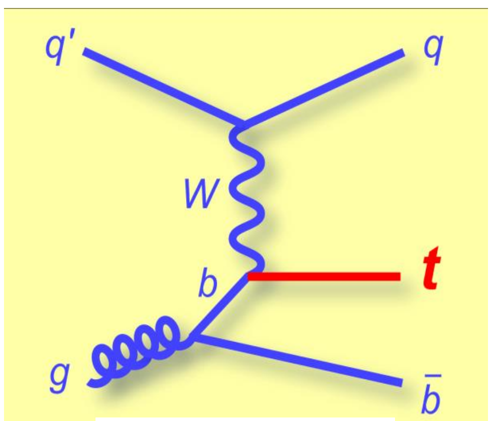
DØ RunII Preliminary 900 pb⁻¹
KS = 0.855
DATA 715
tt̄ 224
W+Jets 403
Multijet 88
CDF Preliminary (760 pb⁻¹)
multijet
W+jets
tt̄
combined
data
N_jets ≥ 3
multijet: 78.0 ± 4.2 events
Wjets:1698.1 ± 49.7 events
signal: 324.6 ± 31.6 events
NN output
CDF Run II Preliminary, L=1.5 fb⁻¹
Data
s-channel
t-channel
Wbb
ttbar
Wc+Wcc
W+LF
NonW
Z+jets,Diboson
Syst. Error
Events/0.025
CDF all hadronic top
QCD
Fraction/(0.02)
N_out
CDF top->6jets
Data
Background
Tags/(0.02)
DØ all jets
predicted backgro
tt̄ MC (× 350)
6-jet data (Vertex T
events
NN_all discriminant
for stop 175/135
L_tchan
DØ Preliminary
Stop 175/135
Top
W+jets
Likelihood

Observation of Single Top Production



s-channel: "*tb*"

$\sigma = 1.12 \pm 0.05$ pb



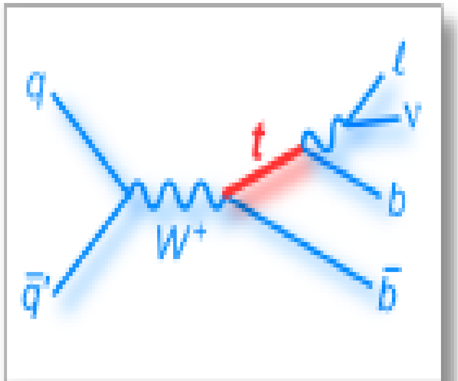
t-channel: "*tqb*"

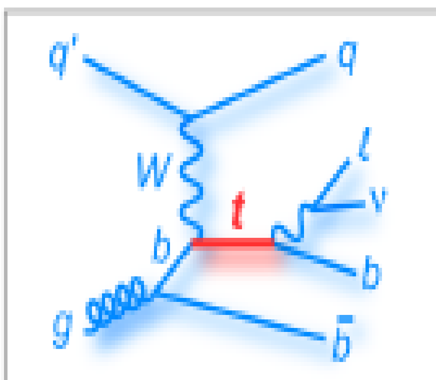
$\sigma = 2.34 \pm 0.13$ pb

- $\sigma_t \sim 3.4$ pb, $\sigma_{tt} \sim 6.8$ pb
- Yet, single top production observed 14 years after top observed in pair production!
- Top discovery (1995) in tt used ~50 pb^{-1} and single top observation required x50-60 more data!

  (DØ ~ 2.3 fb^{-1}, CDF ~ 3.2 fb^{-1})
- Fewer features than ttbar, since only one top per event ← harder to separate signal from backgrounds
- Multivariate methods indispensable!!

Analysis Strategy





- Final state channels involving leptonic decays of the W boson and at least one b-tagged jet are considered by both experiments.
- Use NN to enhance b-tag efficiency and purity.
- Use many multivariate methods. Combine discriminants into a single, final/super discriminant.

DØ's Search History

25x more data

Many improvements in analysis methods

9

Searches, upper limits

PRD **63**, 031101	(2000)	0.09 fb^{-1}	Cuts	TOPCITE = 50+
PLB **517**, 282	(2001)	0.09 fb^{-1}	Neural networks (28 variables)	TOPCITE = 50+
PLB **622**, 265	(2005)	0.23 fb^{-1}	NNs (25 variables) Bayesian likelihoods	TOPCITE = 50+
PRD **75**, 092007	(2007)	0.23 fb^{-1}	Long write-up	

>3σ Evidence

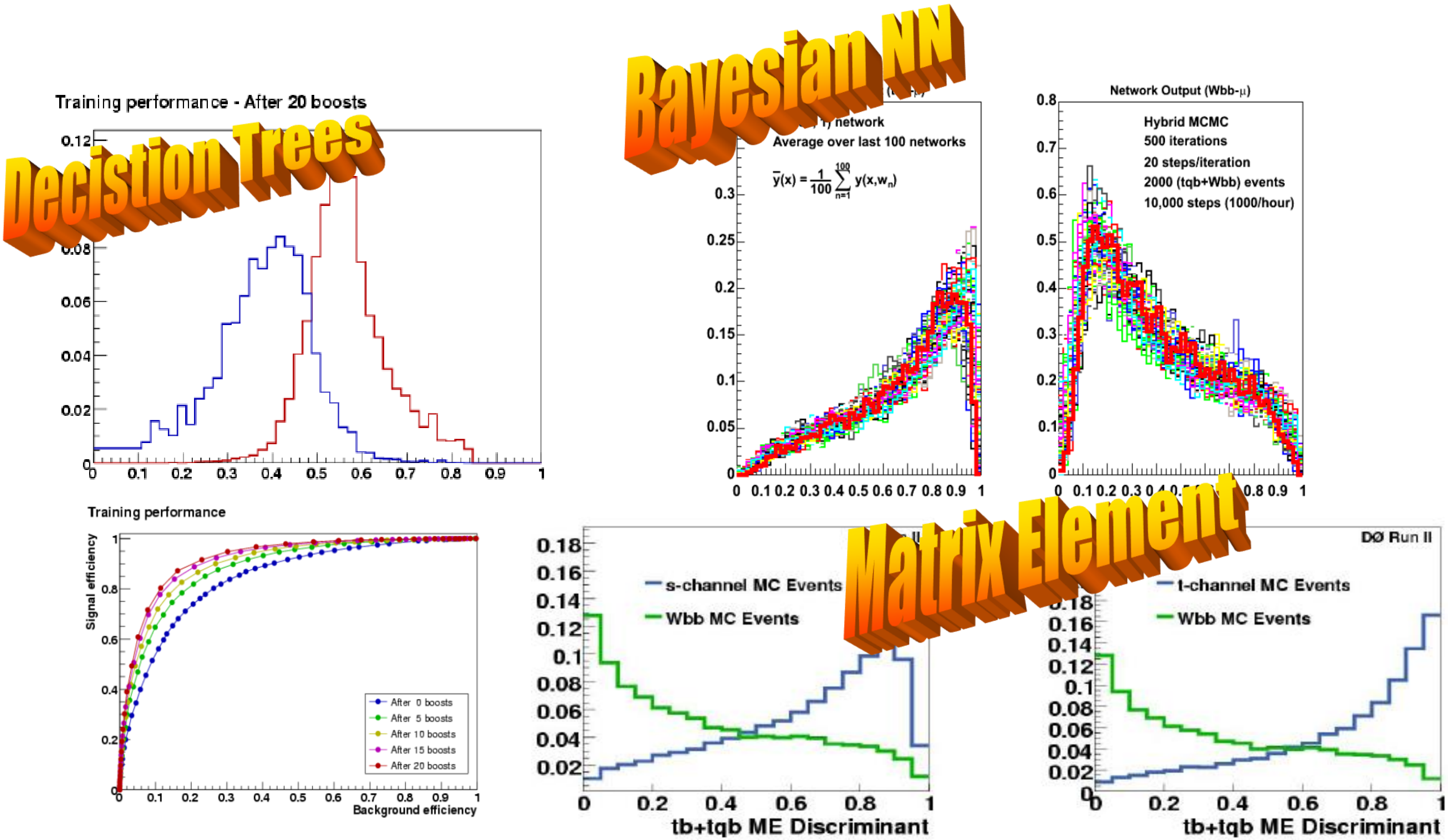
PRL **98**, 181802	(2007)	0.9 fb^{-1}	Boosted decision trees (49 variables) Bayesian neural networks (25 variables) Matrix elements Bayesian likelihoods	TOPCITE = 100+
PRD **78**, 012005	(2008)	0.9 fb^{-1}	Long write-up	TOPCITE = 50+

5σ Observation

PRL **103**, 092001	(2009)	2.3 fb^{-1}	Boosted decision trees (64 variables) Bayesian NNs (18–28 variables) Matrix elements Bayesian-NN combination Bayesian likelihoods	TOPCITE = 50+

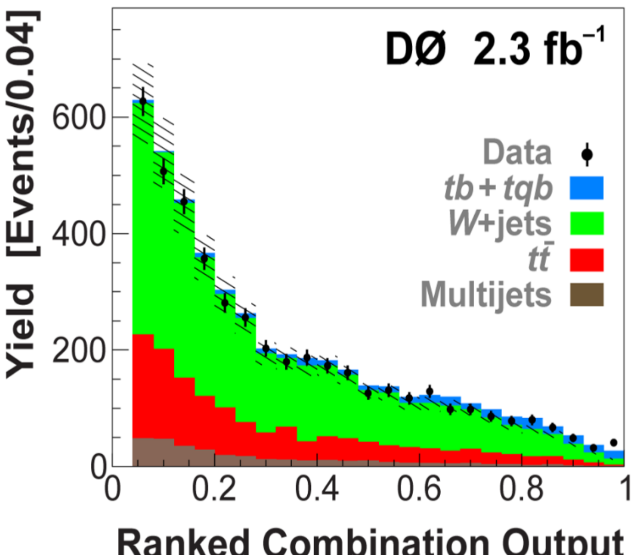
Ann Heinson (UC Riverside)

Discrimination Performance



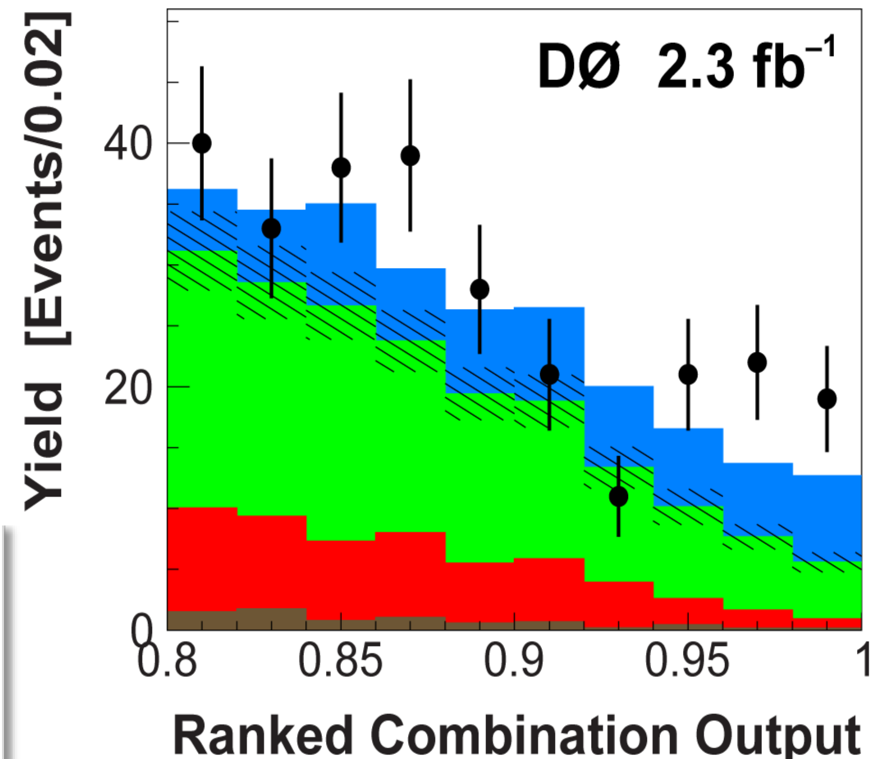
Combination Output

Final Discriminant

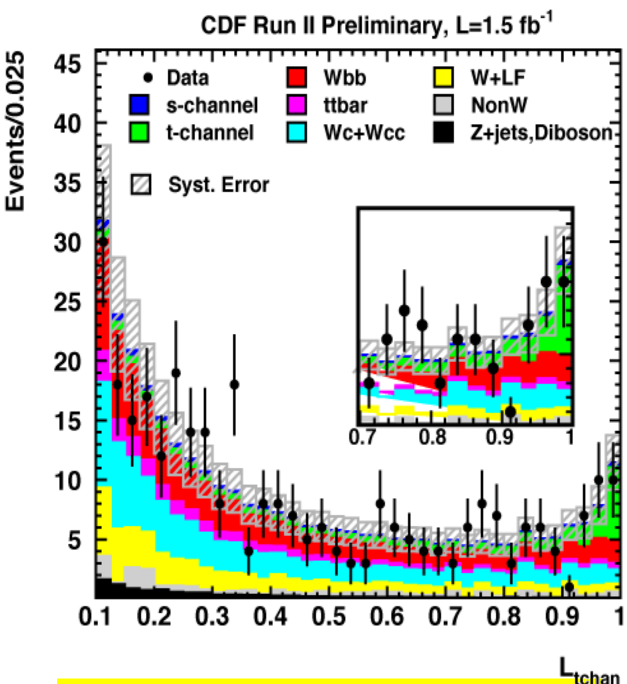


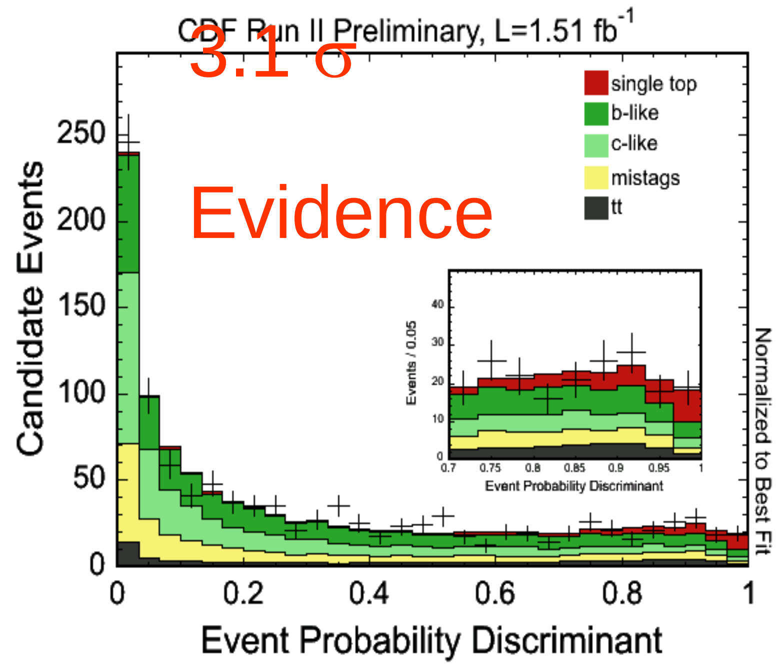
DØ 2.3 fb^{-1} Single Top Results			
Analysis Method	Single Top Cross Section	Significance Expected	Significance Measured
Boosted Decision Trees	$3.74\ ^{+0.95}_{-0.79}$ pb	4.3 σ	4.6 σ
Bayesian Neural Networks	$4.70\ ^{+1.18}_{-0.93}$ pb	4.1 σ	5.4 σ
Matrix Elements	$4.30\ ^{+0.99}_{-1.20}$ pb	4.1 σ	4.9 σ
Combination	3.94 ± 0.88 pb	4.5 σ	5.0 σ

Signal Region



Evidence for Single top at CDF





σ_{s+t}= 2.7 ± 1.2 pb

σ_s= 1.1, σ_t =1.3 pb

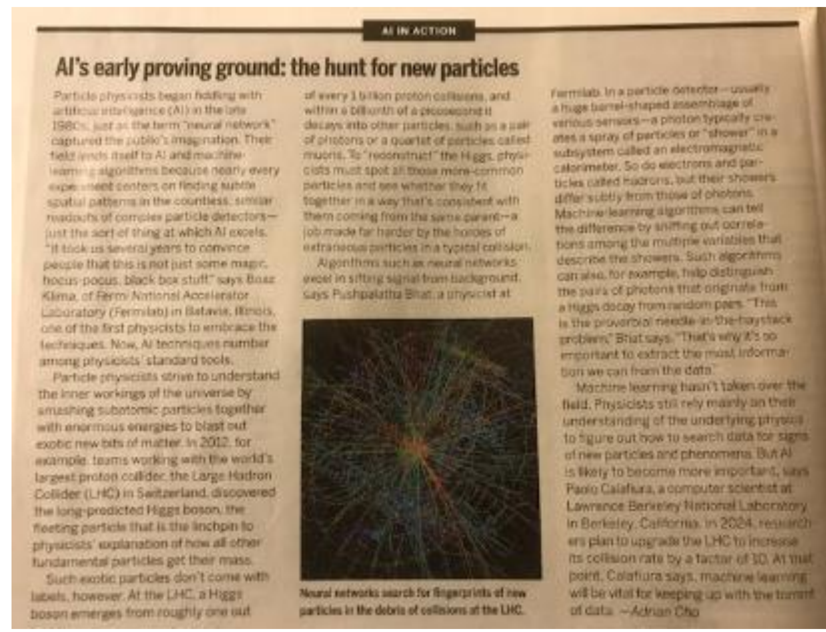
Expected sensitivity: 2.9σ

Observed significance: 2.7σ

σ_{s+t}= 3.0 ± 1.2 pb

σ_s= 1.1, σ_t =1.9 pb

Expected sensitivity: 3.0σ



AI IN ACTION

AI's early proving ground: the hunt for new particles

Particle physicists began fiddling with artificial intelligence (AI) in the late 1980s, just as the term "neural network" captured the public's imagination. Their field lends itself to AI and machine-learning algorithms because nearly every experiment centers on finding subtle spatial patterns in the countless, similar readouts of complex particle detectors—just the sort of thing at which AI excels. "It took us several years to convince people that this is not just some magic, hocus-pocus, black box stuff," says Boaz Klima, of Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, one of the first physicists to embrace the techniques. Now, AI techniques number among physicists' standard tools.

Particle physicists strive to understand the inner workings of the universe by smashing subatomic particles together with enormous energies to blast out exotic new bits of matter. In 2012, for example, teams working with the world's largest proton collider, the Large Hadron Collider (LHC) in Switzerland, discovered the long-predicted Higgs boson, the fleeting particle that is the linchpin to physicists' explanation of how all other fundamental particles get their mass.

Such exotic particles don't come with labels, however. At the LHC, a Higgs boson emerges from roughly one out of every 1 billion proton collisions, and within a billionth of a picosecond it decays into other particles, such as a pair of photons or a quartet of particles called muons. To "reconstruct" the Higgs, physicists must spot all those more-common particles and see whether they fit together in a way that's consistent with them coming from the same parent—a job made far harder by the hordes of extraneous particles in a typical collision.

Algorithms such as neural networks excel in sifting signal from background, says Pushpalatha Bhat, a physicist at Fermilab. In a particle detector—usually a huge barrel-shaped assemblage of various sensors—a photon typically creates a spray of particles or "shower" in a subsystem called an electromagnetic calorimeter. So do electrons and particles called hadrons, but their showers differ subtly from those of photons. Machine-learning algorithms can tell the difference by sniffing out correlations among the multiple variables that describe the showers. Such algorithms can also, for example, help distinguish the pairs of photons that originate from a Higgs decay from random pairs. "This is the proverbial needle-in-the-haystack problem," Bhat says. "That's why it's so important to extract the most information we can from the data."

Machine learning hasn't taken over the field. Physicists still rely mainly on their understanding of the underlying physics to figure out how to search data for signs of new particles and phenomena. But AI is likely to become more important, says Paolo Calafiura, a computer scientist at Lawrence Berkeley National Laboratory in Berkeley, California. In 2024, researchers plan to upgrade the LHC to increase its collision rate by a factor of 10. At that point, Calafiura says, machine learning will be vital for keeping up with the torrent of data. —Adrian Cho

Neural networks search for fingerprints of new particles in the debris of collisions at the LHC.

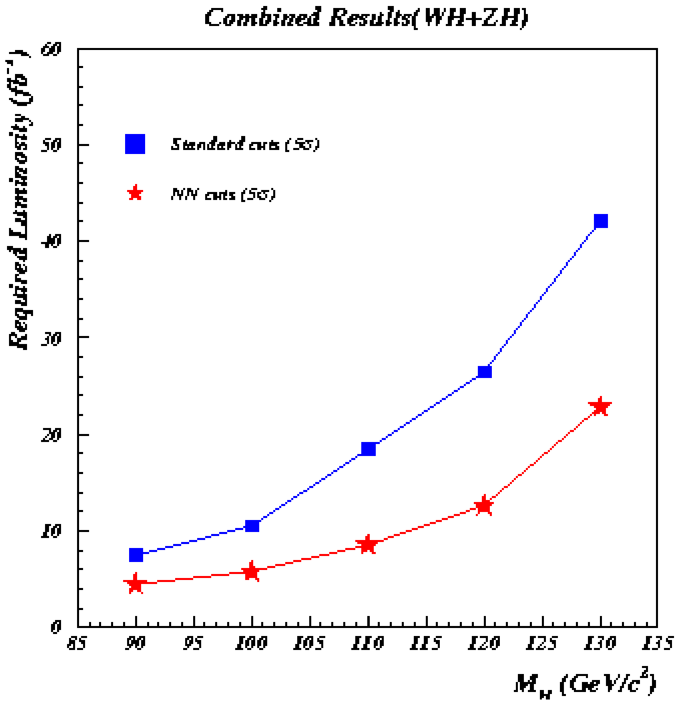
Higgs @ Run II

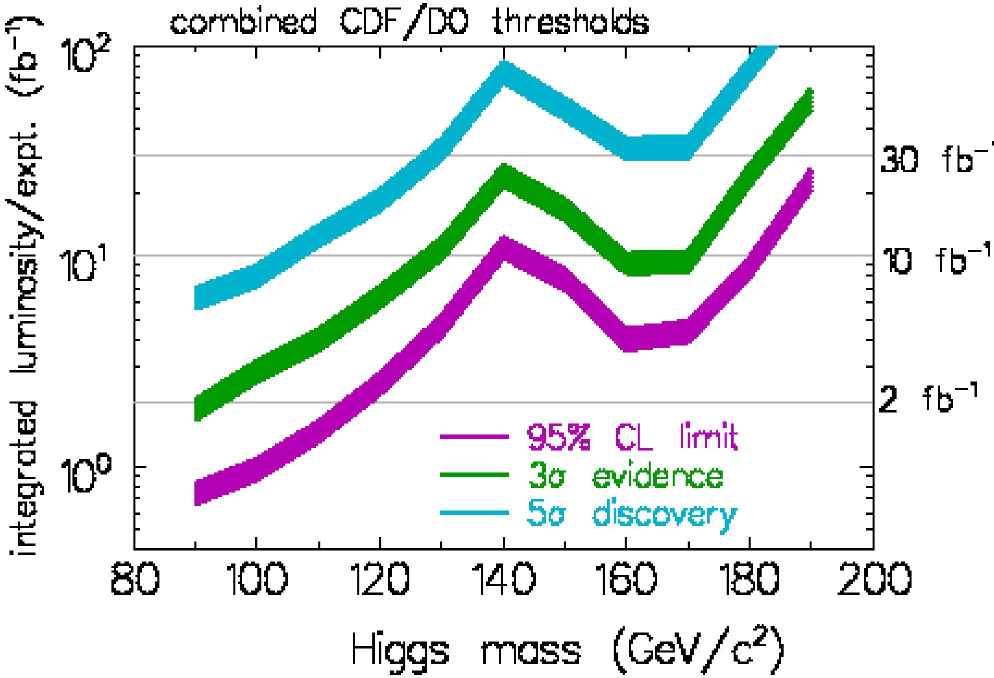
The SM Higgs Boson

WH/ZH Channels

Discovery Reach at the Tevatron

- The challenges are daunting! Using NN provides same reach with a factor of 2 less luminosity w.r.t. conventional analysis
- Improved bb mass resolution & b-tag efficiency crucial





- * P.C.Bhat, R.Gilmartin, H.Prosper, Phys.Rev.D.62 (2000) 074022
- Run II Higgs study hep-ph/0010338 (Oct-2000)

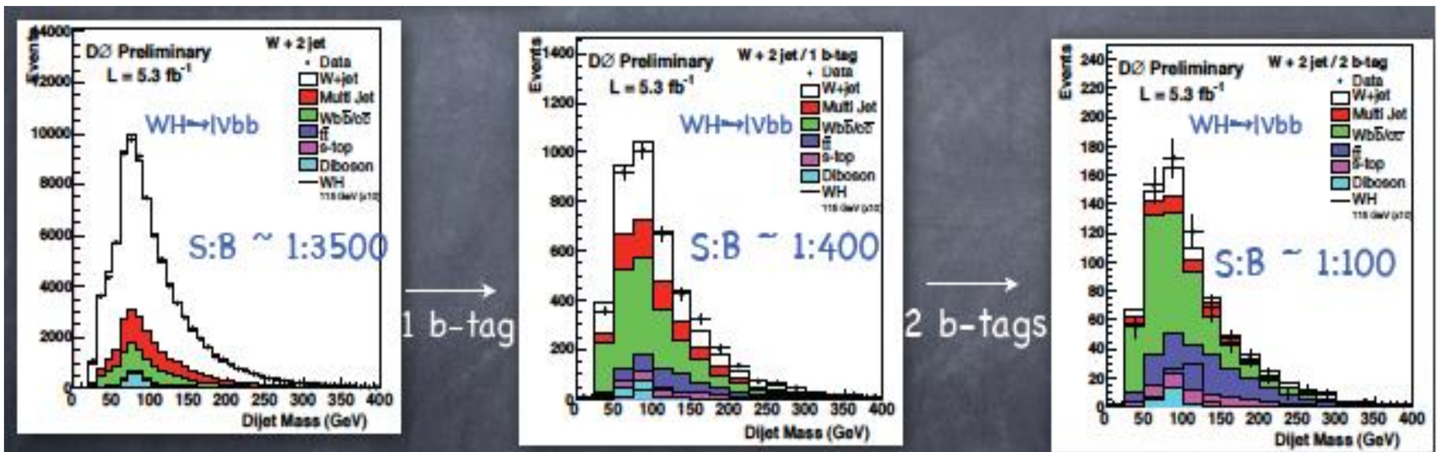
Tevatron Higgs storyline

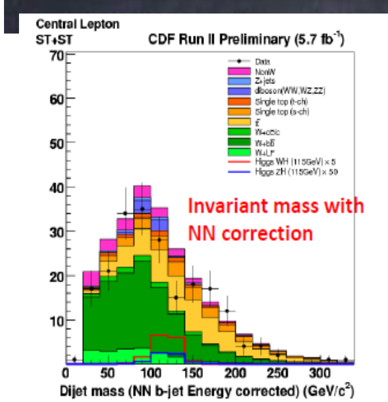
- How to build an advanced Higgs analysis program
  - Start with basic analysis for particular channel
  - Bootstrap special techniques to gain sensitivity
    - Improve acceptance
      - Loosen lepton ID & b-tag requirements
      - Add backup triggers
      - Relax kinematic selection
    - But... backgrounds increase & become more difficult to model
      - Incorporate specialized background rejection techniques
      - Don't cut, separate out events into categories with alike S/sqrt(B)
        - High S/sqrt(B) gives best signal sensitivity
        - Low S/sqrt(B) gives best background constraints
      - Use multivariate techniques to distinguish signal events from bkgd
      - Background modeling checks ! Data must stay well modeled !
- Repeat for each Higgs topology per grad student
- Combine modes taking into account uncertainties correlated between backgrounds

Ben Kilminster, ICHEP 2010 10

b-tagging, dijet mass

- Good b-tagging efficiency crucial!





Dijet mass resolution important to discriminate from background

The two most sensitive variables

Multivariate techniques

- Multivariate analysis techniques
  - Used in all TeV Higgs analyses
- Functions transform multiple inputs into single discriminant tuned for identifying a single process
- Algorithms have similar performance :
  - NN = Neural Net
  - ME = Matrix Element
  - BDT = Boosted Decision Trees
    - RF = "random forest" of decision trees

- Improve analyses by ~20% with respect to leading two variables
  - Correlations useful
    - ie, if M_{jj} is consistent with Higgs, so better be sum E_T and missing transverse energy
  - Caveat : our primary sensitivity gains in recent years don't come from multivariate techniques
    - Mainly from improved signal acceptance
      - Looser lepton ID
      - Better b-tagging, etc.

Improvements in b-tagging and dijet Mass resolution also achieved with MVA

Ben Kilminster, ICHEP 2010 15

Deep Learning NN

- Use raw data inputs instead of derived "intelligent" variables (or use both)
  - Pre-processing or feature extraction in the DNN
- Pre-train initial hidden layers with unsupervised learning
- Multi-scale Feature Learning
  - Each high-level layer learns increasingly higher-level features in the data
- Final learning better than shallow networks, particularly when inputs are unprocessed raw variables!
- **However, need a lot of processing power (implement in GPUs, time (and training examples)**

Summary

- Machine Learning (MVA) methods have been critical in major discoveries and rich harvest of physics from the Tevatron and at the LHC.
- The new revolution in Deep Learning in the new century is advancing our physics goals and ambitions
  - Boosted Objects, Jet ID, Jet calibration
  - Pile-up mitigation
  - End-to-end Event Reconstruction
  - Triggering and data acquisition
  - …
- Innovation must and will continue! Infusion of new ideas should be welcome! Trust but verify?
- The promise of ML being realized and the future is bright!

EXTRAS

Tools used at the Tevatron

- Jetnet
- MLPFit
- SNNS (Stuttgart NN Simulator)
- TMVA
- Stat Pattern Recognition
- FBM (Radford Neal's Bayesian NN)
- NeuroBayes
- NeuroEvolution

Tevatron MVA Strategy

- Use MV methods for b-tagging, energy corrections, signal/background discrimination, anywhere it can help!
- Bootstrap many multivariate techniques
- Validate background modeling; extensive cross-checks
  - Split into 2-3 categories
  - High $s/\sqrt{B}$ ← best signal sensitivity
  - Low $s/\sqrt{B}$ ← gives good background constraints
  - Cross-check modeling of many variables in these regions
  - Control regions and control samples

Cross Checks

- Huge number of variable distributions
- Discriminant distributions
  - Very signal-like, very background-like, intermediate
  - Variable distributions for different discriminant regions
- Variables in and out of the analysis
- Varying the parameters of the methods to study the changes in the final results

Which Method is best?

- The "no free lunch" theorem tells you that there is no one method that is superior to all others for all problems.
- In general, one can expect Bayesian neural networks (BNN), Boosted decision trees (BDT) and random forests (RF) should provide excellent performance over a wide range of problems.
- BDT is popular because of robustness, noise resistance (psychological comfort)

Strategy for use in CMS

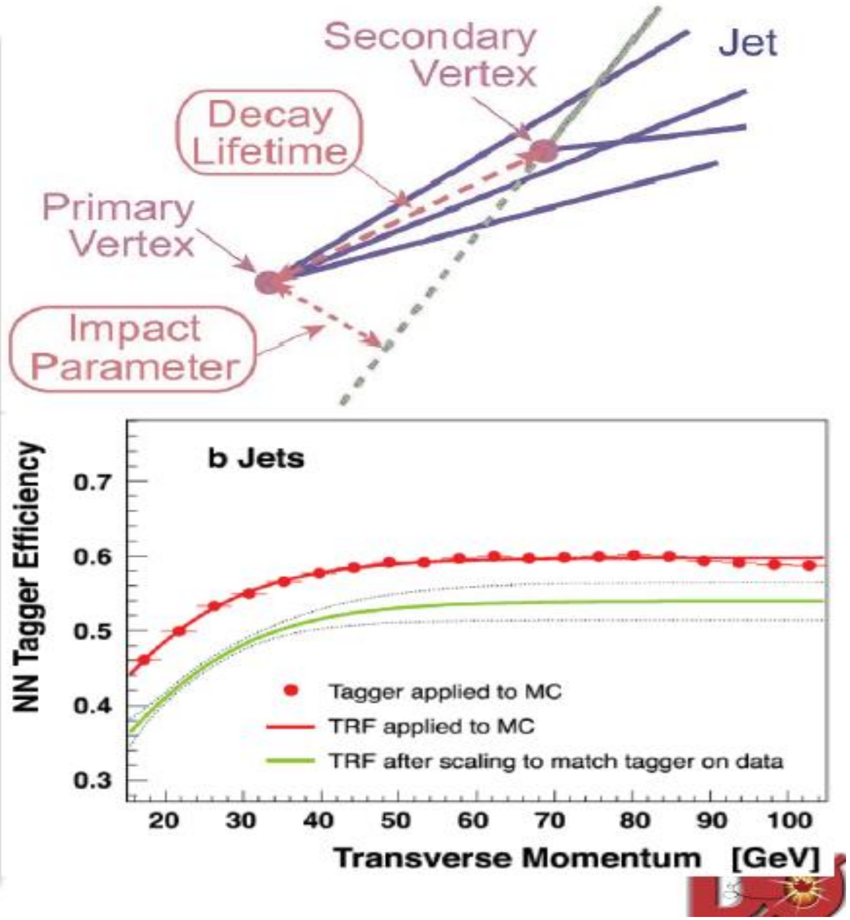
- Even a small number of variables treated in a multivariate manner can provide significant gains
- Combining simple classifiers based on a few variables can significantly boost the final performance
- One can make use of two or more methods for cross-checks.
- One can use data as the background model in processes where signal-to-background ratio is very small. An advantage of this approach is that the data models both the physics and instrumental backgrounds

Prospects for use in CMS, cont'd

- Should be used in applications to improve basic detector measurements such as jet energy scale, b-tagging
- These methods can be used safely where it is possible to cross-check the modeling using well-known physics processes such as Z-decays, QCD bb, etc.
- MVA will be crucial to discover low mass Higgs Boson at CMS during this run.

D0 b-tagger

- NN trained on 7 input variables from existing taggers.
  - secondary vertices
  - impact parameter
- Much improved performance:
  - fake rate reduced by 1/3 for same *b* efficiency relative to previous tagger
  - smaller systematic uncertainties
- Tag Rate Functions (TRFs) in η, p_T, z-PV applied to MC
- Operating point:
  - *b*-jet efficiency $\sim 50\%$
  - *c*-jet efficiency $\sim 10\%$
  - light jet efficiency $\sim 0.5\%$

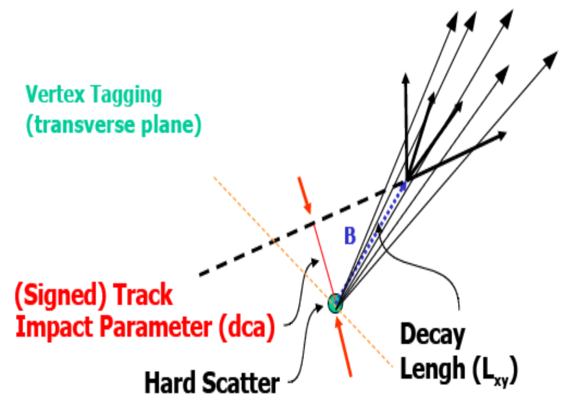


b-tagging

- Critical for low mass H→ bb
  - Improves S/B by > 10
- Use lifetime information
  - Correct for MC / data differences
    - Measured at given operating points

- **DØ: Neural Net tagger**
- Secondary vertex & dca based inputs,
- derived from basic taggers
- High efficiency, purity
- **Loose = 70% eff, 4.5% mistag**
- **Tight = 50% eff, 0.3% mistag**

- Analyse separately ("tight") single & ("loose") double tags

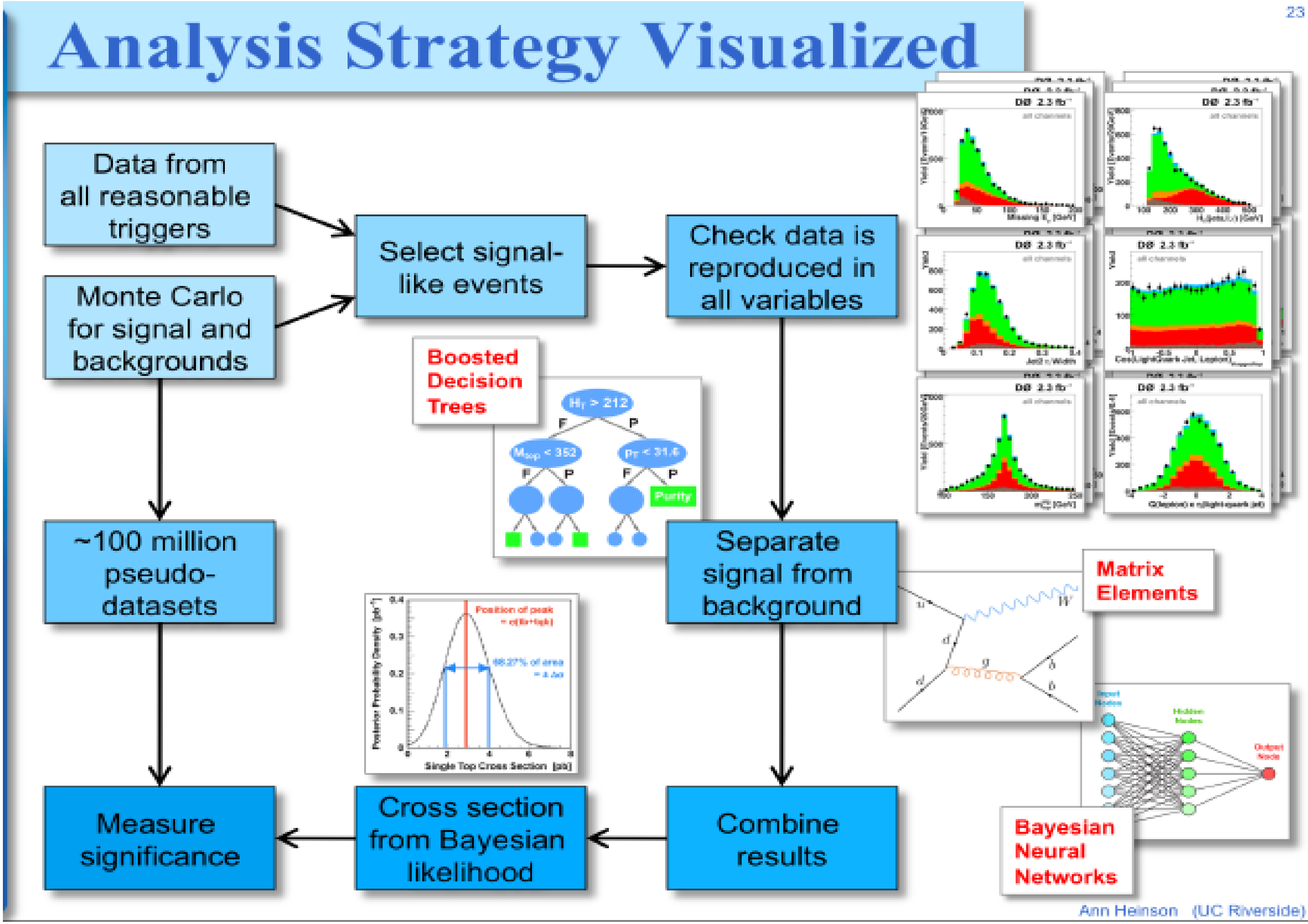


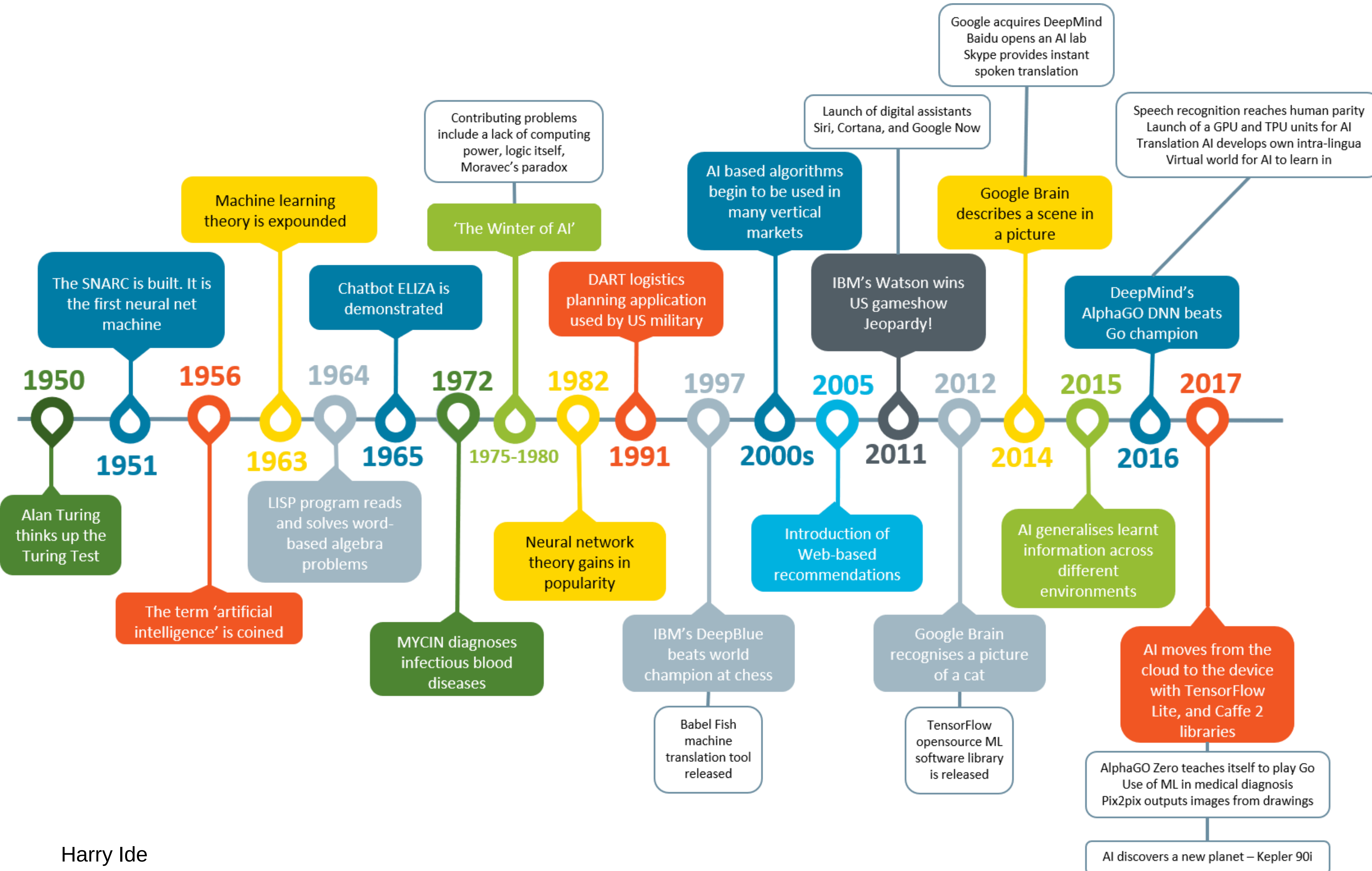
CDF: Secondary vertex reconstruction

- Neural Net - improves purity
- Inputs: track multiplicity, p_T, vertex decay length, mass, fit
- **Loose = 50% eff, 1.5 % mistag**
- **Tight = 40% eff, 0.5 % mistag**

Optimal Data Analysis

- The goal: Best possible physics!
  - Maximize sensitivity to searches
  - Extract signals, known and new, with high efficiency
  - Improve precision of measurements
- How do we achieve that?
  - Make maximal use of information in the data!!!
  - If data is multivariate, use multivariate analysis methods!


23
Analysis Strategy Visualized
Data from all reasonable triggers
Monte Carlo for signal and backgrounds
Select signal-like events
Check data is reproduced in all variables
Boosted Decision Trees
$H_T > 212$
F
P
$M_{top} < 352$
$p_T < 31.6$
Purity
~100 million pseudo-datasets
Separate signal from background
Matrix Elements
Bayesian Neural Networks
Measure significance
Cross section from Bayesian likelihood
Combine results
Position of peak
Single Top Cross Section [pb]
DØ 2.3 fb⁻¹
Ann Heinson (UC Riverside)



Harry Ide
On Innovation Observatory

Event Yields

Before b-tagging

Event Yields in 2.3 fb^{-1} of DØ Data	
e,μ, 2,3,4-jets, pretag	
tb + *tqb*	444
W+jets	98,444
Z+jets, dibosons	8,631
$t\bar{t}$ pairs	1,895
Multijets	5,798
Total background	114,777
Data	114,777

S:B ~ 1:260

After b-tagging

Event Yields in 2.3 fb^{-1} of DØ Data	
e,μ, 2,3,4-jets, 1,2-tags combined	
tb + *tqb*	223 ± 30
W+jets	2,647 ± 241
Z+jets, dibosons	340 ± 61
$t\bar{t}$ pairs	1,142 ± 168
Multijets	300 ± 52
Total prediction	4,652 ± 352
Data	4,519

S:B ~ 1:20

Useful Variables

Best Variables to Separate Single Top from W+Jets DØ 2.3 fb^{-1} Analysis	
Object kinematics	$\not{E}_T$
	p_T(jet2)
	p_T^{rel}(jet1,tag-μ)
	E(light1)
Event kinematics	M(jet1,jet2)
	M_T(W)
	H_T(lepton,$\not{E}_T$,jet1,jet2)
	H_T(jet1,jet2)
	H_T(lepton,$\not{E}_T$)
Jet reconstruction	Width$_\phi$(jet2)
	Width$_\eta$(jet2)
Top quark reconstruction	M_{top}(W,tag1)
	ΔM_{top}^{min}
	M_{top}(W,tag1,S2)
Angular correlations	cos(light1,lepton)$_{btaggedtop}$
	Δϕ(lepton,$\not{E}_T$)
	Q(lepton) x η(light1)

Best Variables to Separate Single Top from Top Pairs DØ 2.3 fb^{-1} Analysis	
Object kinematics	pT(notbest2)
	pT(jet4)
	pT(light2)
Event kinematics	M(alljets–tag1)
	Centrality(alljets)
	M(alljets–best1)
	H_T(alljets–tag1)
	H_T(lepton,$\not{E}_T$,alljets)
	M(alljets)
Jet reconstruction	Width$_\eta$(jet4)
	Width$_\phi$(jet4)
	Width$_\phi$(jet2)
Angular correlations	cos(lepton$_{btaggedtop}$, btaggedtop$_{CMframe}$)
	Q(lepton) x η(light1)
	ΔR(jet1,jet2)