第九章 神经元网络方法及其应用举例

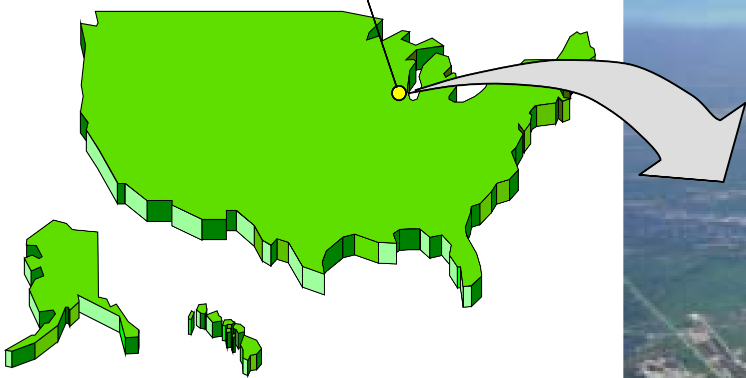
Questions: what can-do or can-not-do of a von Neumann machine?

Good at	**Not so good at**
Fast arithmetic	Interacting with noisy data or data from the environment
Doing precisely what the programmer programs them to do	Massive parallelism
	Fault tolerance
	Adapting to circumstances

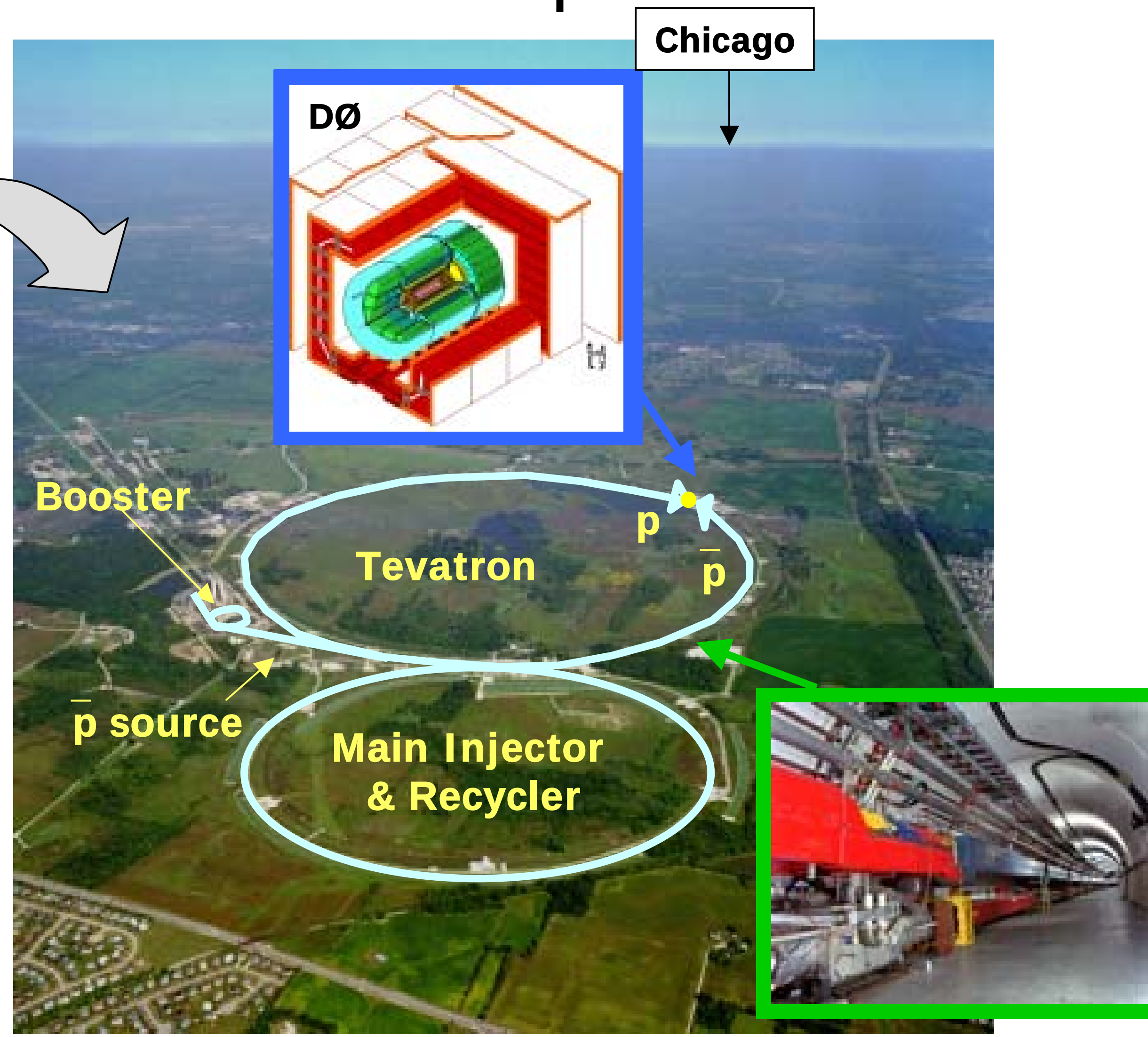
Where can ANN systems or “Brain” help?
- where we can't formulate an algorithmic solution.
- where we *can* get lots of examples of the behaviour we require, 大样本训练
- where we need to pick out the structure from existing data.
  e.g. Pattern recognition (recognizing handwritten characters?)

无数学模型描述的复杂系统识别，YES or NOT？

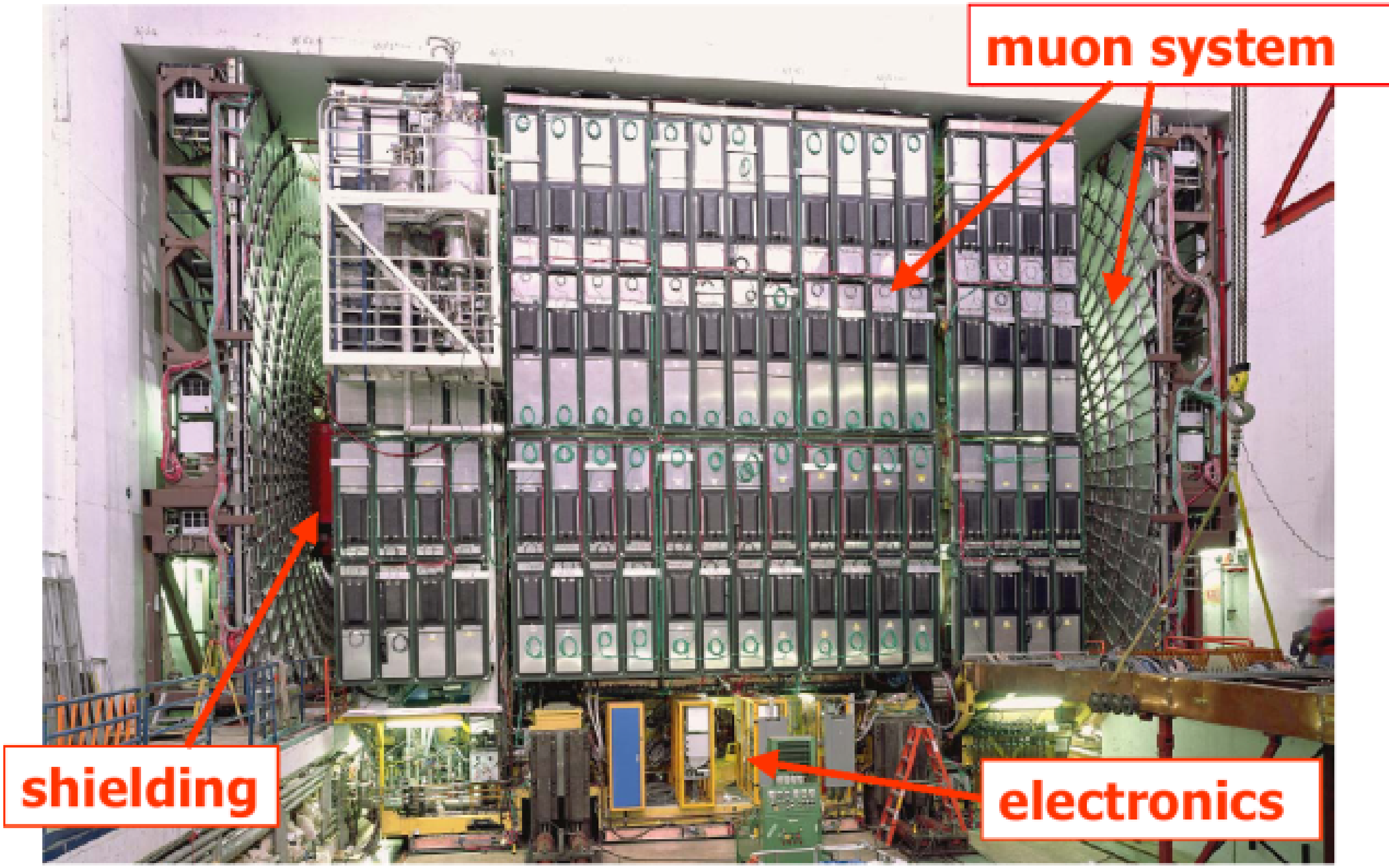
<u>Tevatron</u>：Fermilab Proton-Antiproton Collider



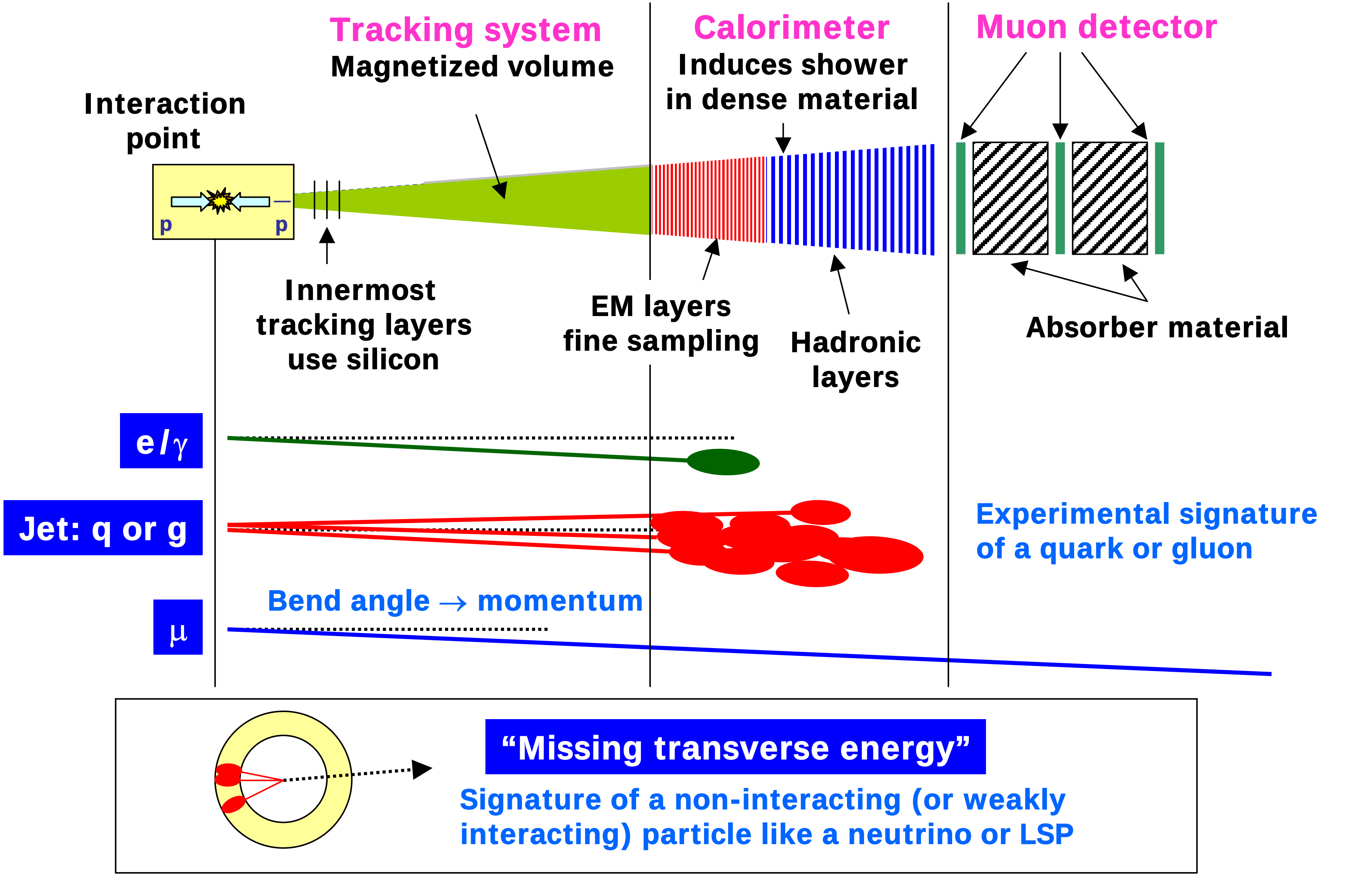
1)Linear Accelerator
2)Booster
3)Main Injector
4)Antiproton Source
5)Tevatron @1.96TeV
6)CDF+DØ



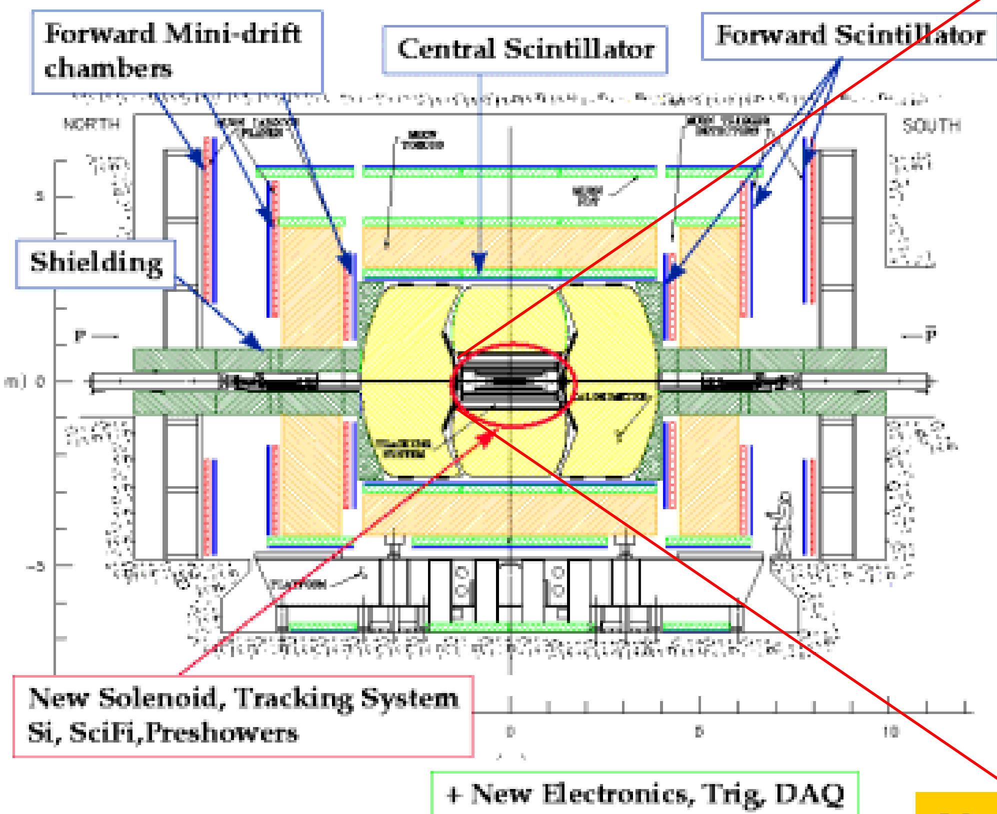
DØ Detector

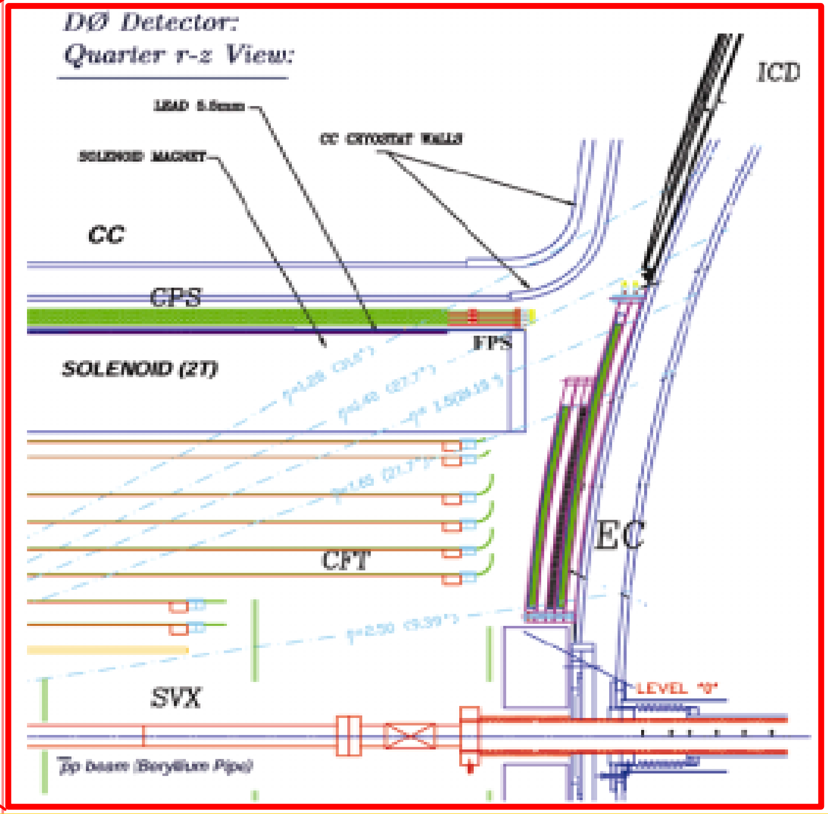


A Schematic Hadron Collider Detector



The DØ Run II Detector





New for Run II:

- Magnetic Tracker: **SMT, CFT, 2T Solenoid**
- Preshower
- Forward Muon
- Trigger & DAQ

Retained from Run I:

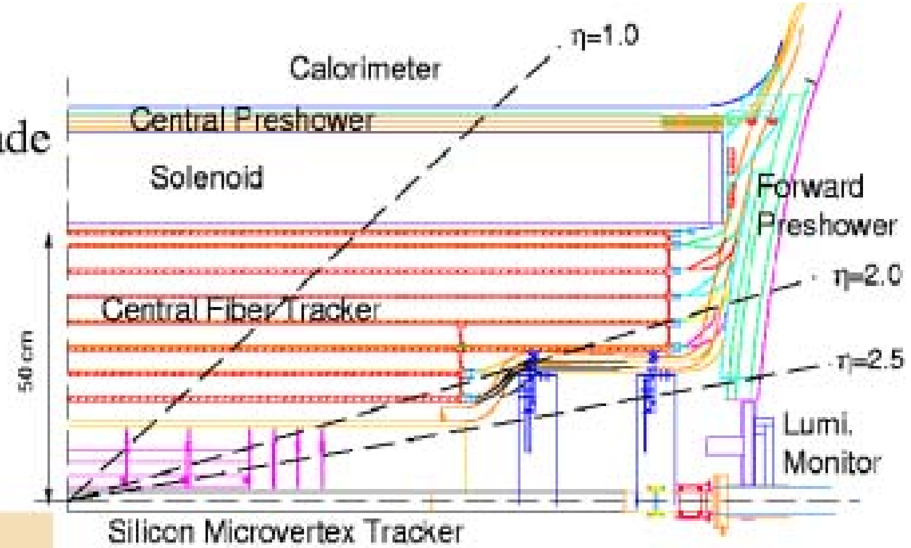
- U/LAr Calorimeter
- Central Muon Detectors
- Muon Toroid

Vertex & Central Tracking

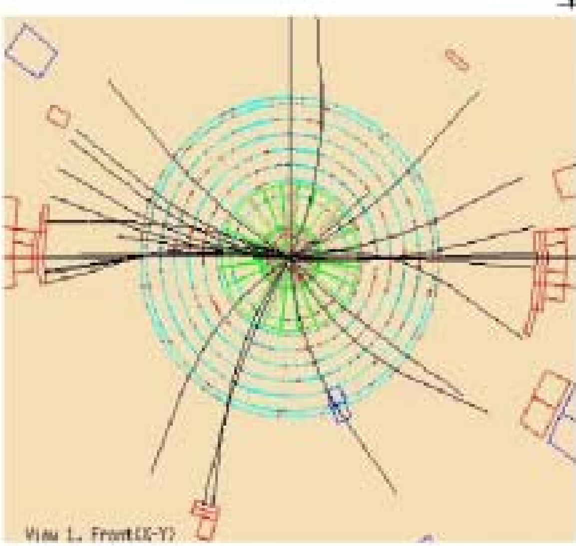
1. Silicon Microvertex Tracker : ~ 10μm(design)
2. Central Fiber Tracker: scintillator fiber

Major part of Run II upgrade

- 2 T solenoid magnet
- 8 layer CFT
- Silicon tracker



$\eta = -\ln(tg(\theta/2))$

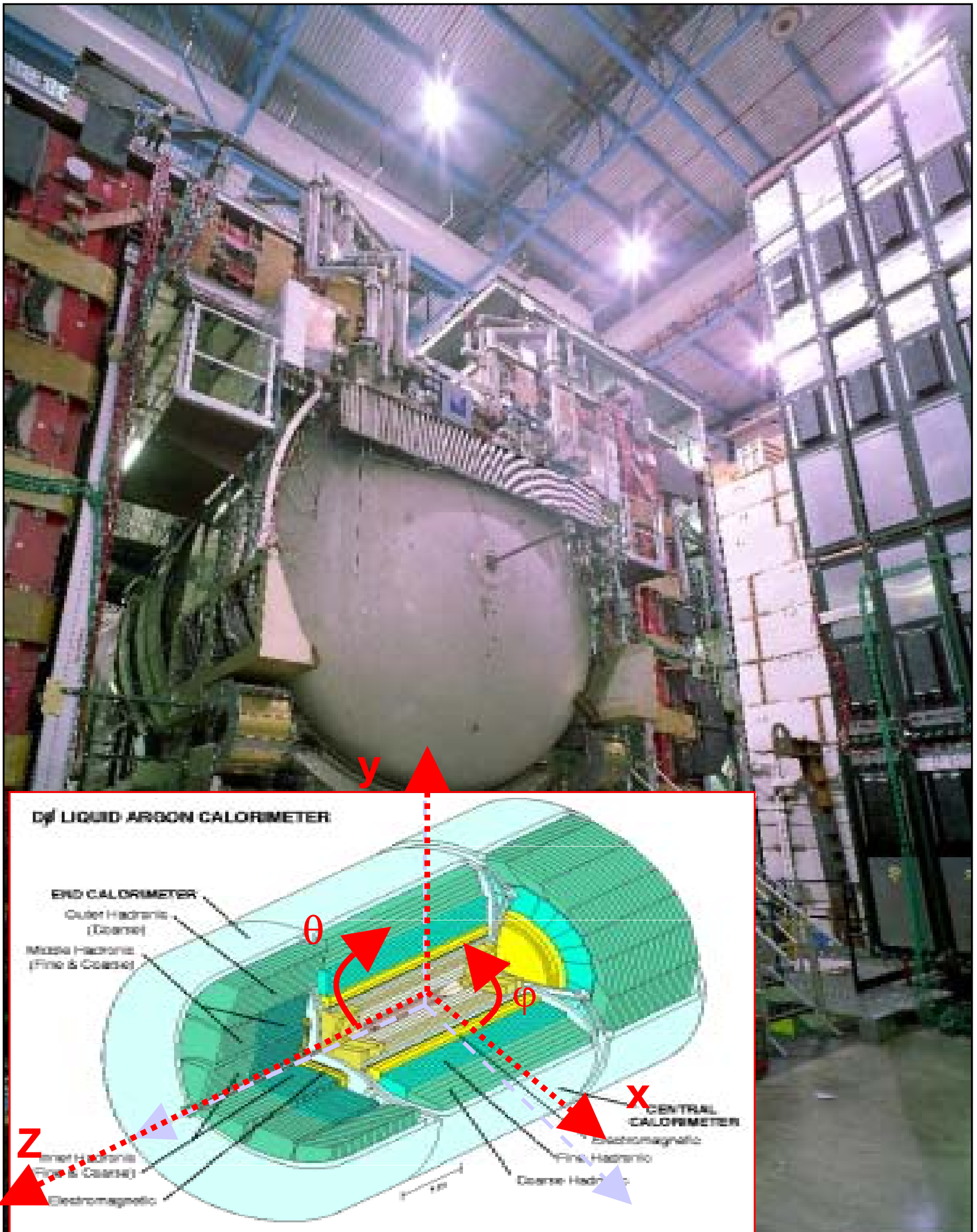


Momentum resolution (design):

$$\frac{\sigma_{p_T}}{p_T} = 0.015 \oplus 0.0014 p_T$$

<10% at 40 GeV

The Calorimeter



- Liquid Argon sampling
  - ✓ Stable, uniform response
  - ✓ LAr purity
- Uranium absorber (Cu/Fe for coarse hadronic)
  - ✓ dense absorber hence can be compact
  - ✓ Compensated EM and hadronic response
  - ✓ Linear response
- Hermetic with full coverage
  - ✓ $|\eta| < 4.2$ ($\theta \approx 2^o$)

Resolution:

$\sigma/E \sim 15\%/\sqrt{E(GeV)}$ "fine" EM

$50\%/\sqrt{E(GeV)}$ "coarse" jet

$\sigma_{MET} \sim a + b*S_T + c*S_T^2$ (run1)

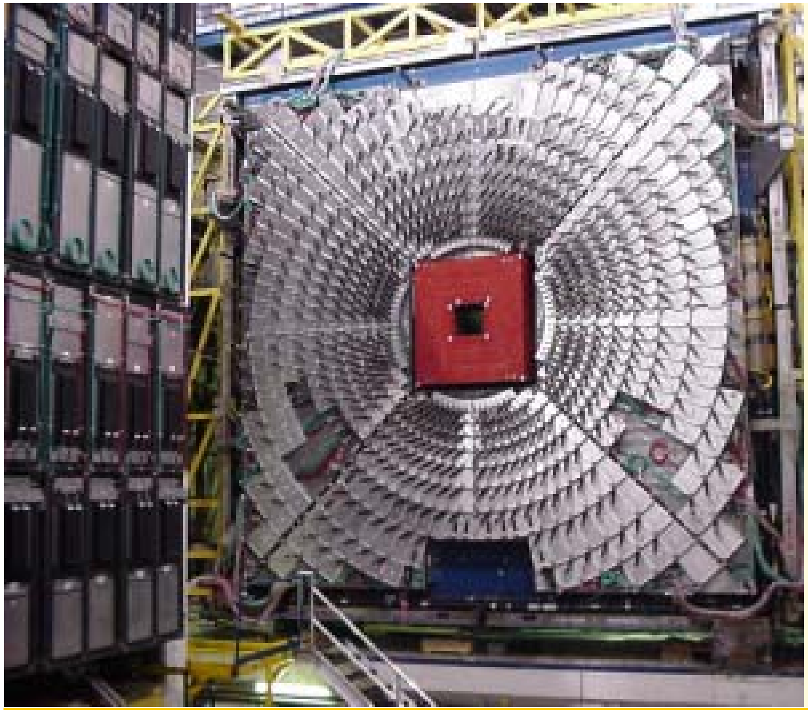
S_T scalar sum of ET

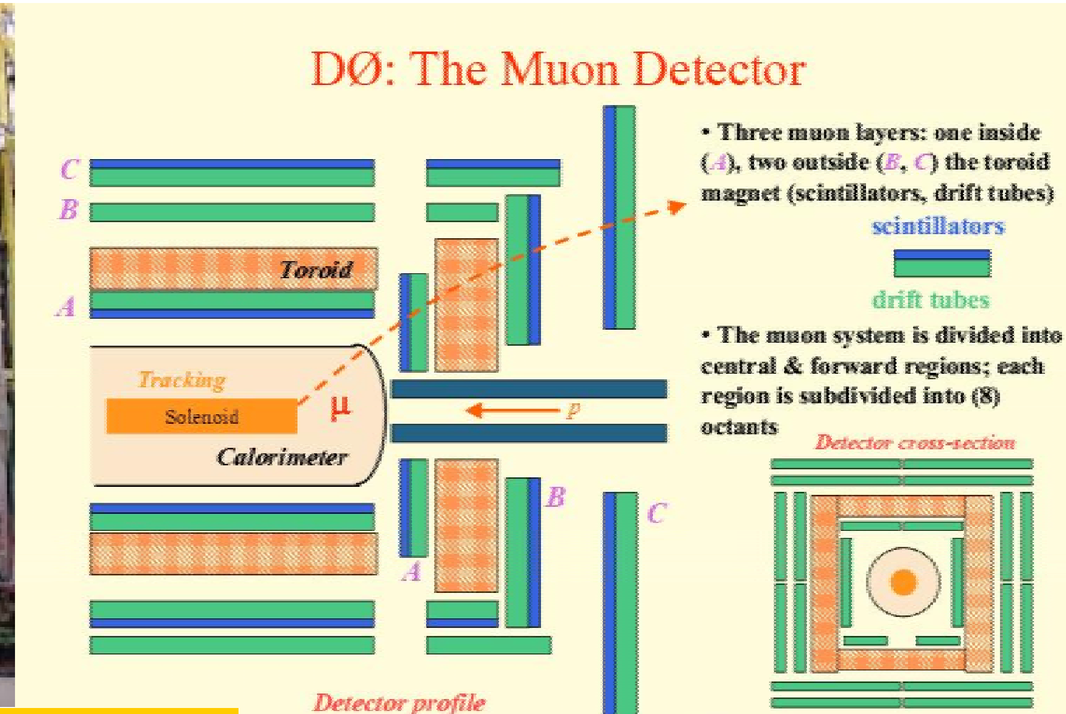
a ~1.89GeV,

b ~6.7E-3,

c ~9.9E-6/GeV

Muon Detector



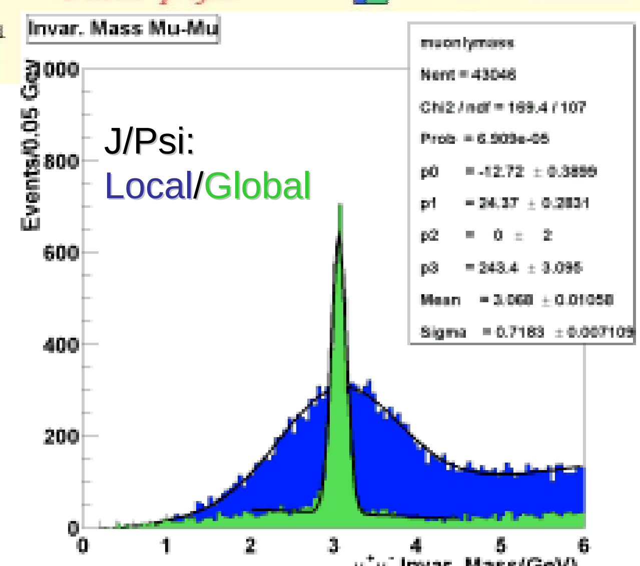


- Two regions & Three layers of Scintillator and Drift Tubes
  - Central and Forward
  - A Layer inside Toroid magnet
  - B & C Layer outside Toroid magnet
- Muon rapidity coverage to ±2
- Shielding reduces backgrounds by 50-100x
- Coarse Local pT resolution

Momentum resolution (design) :

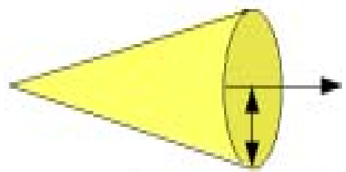
$$\frac{\sigma_{1/p_T}}{1/p_T} = 0.18 \oplus 0.005p$$

~40% at 40 GeV



What we can "see"?

1. "Jet" (CAL) :

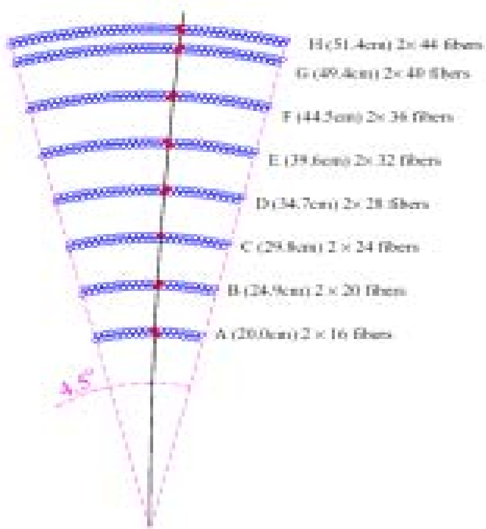


$$\Delta R = \sqrt{(\Delta\phi^2 + \Delta\eta^2)} < 0.5$$

cell -> "Tower" -> "Jet"

hadronic(0.5) vs. γ/e(0.2)

2. "Track" (CFT+SMT) :



3. "Vertex" (SMT) : PVrt/Svrt

4. "muon" (Muon)

5. "MET" (CAL) :

$\Sigma E_{cell} * \sin\theta$ w/o muon

Good at
M.C.问题：e,μ,τ,夸克信号特性？

Generator level:
Z→ee,$\mu\mu$,$\tau\tau$,qq(di-jet)

↓

GEANT level:
- 不稳定粒子衰变($e^{-it/\tau}$)
- 粒子与物质相互作用
  电离,韧致辐射,强作用→能量损失

↓

Digitize level:
能量沉积→探测器unit/cell
电子学ADC读出

↓

Reconstruction level:
unit/cell → grouped → cluster
→ 信号shape/pattern,能量+方向

Not-so-Good at
实际问题：Pattern Recognition

粒子鉴别ID

↑

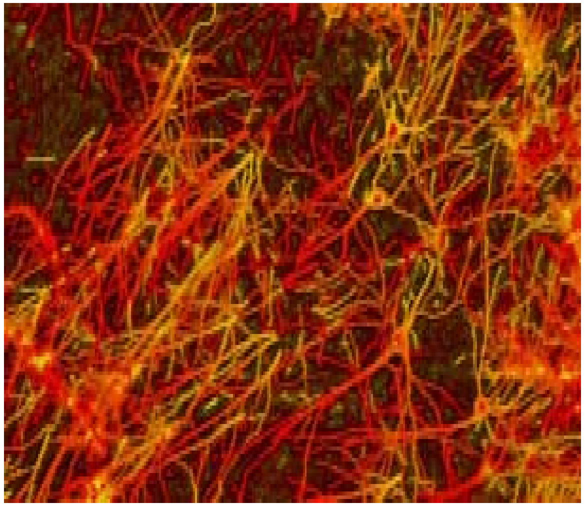
?
No formula, no theory

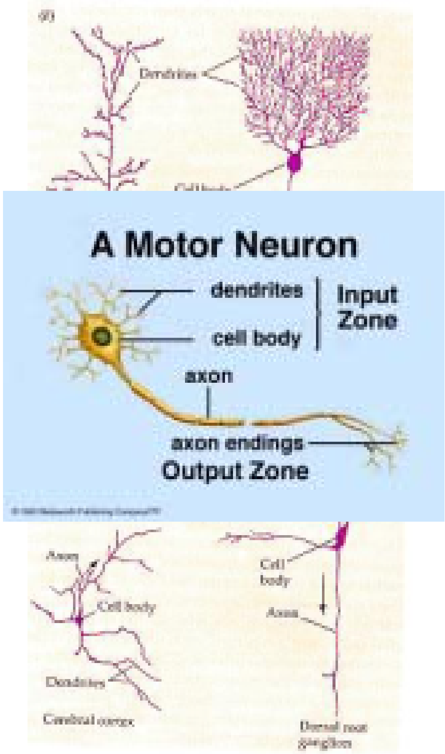
ANN:
经验与训练

↑

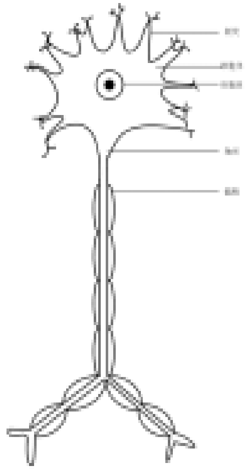
探测器中unit/cell ADC readouts
~O(10^6), Sigal/Background entangled

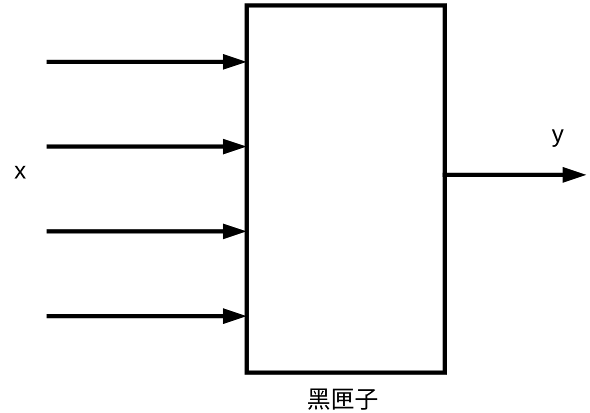
人脑：学习与经验主义





Neural networks are a form of multiprocessor computer system, with
- simple processing elements
- a high degree of interconnection
- simple scalar messages
- adaptive interaction between elements





1) 多输入x_i，描述被识别对象

2) 单输出y，被识别对象ID

3) “黑匣子”：

- 输入整合：

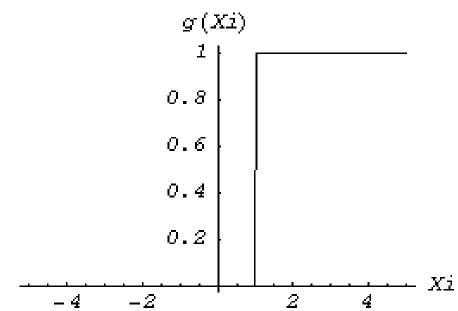
$$a_j = \sum_k w_{jk} x_k + t_j$$

w_{jk}：各输入道权重

t_j：输出道阈值

- 输出active function, **$y_j = g(a_j)$,** eg

$$g(x) = \begin{cases} 1, & \text{当} x \geq \theta \\ 0, & \text{当} x < \theta \end{cases}$$



4) 极值原理

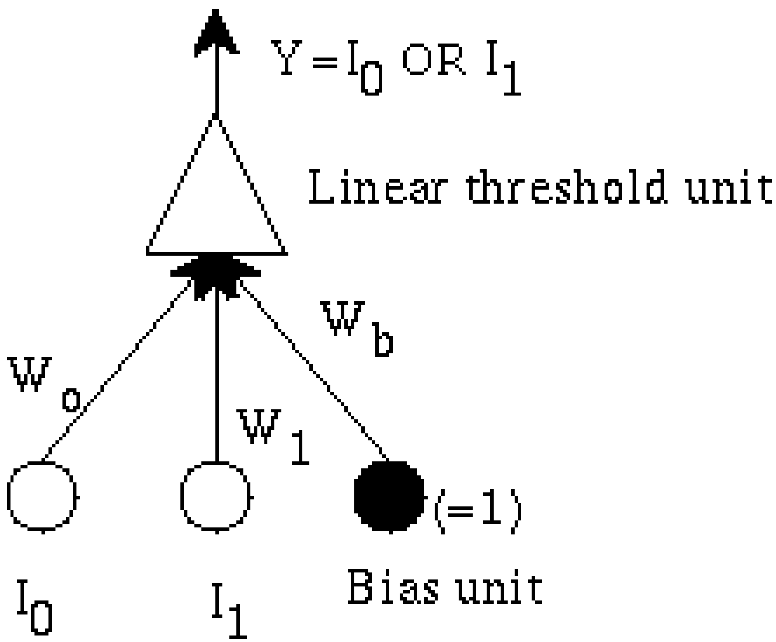
$$E = \frac{1}{2} \sum_P \sum_i \left(y_i(P) - A_i(P) \right)^2$$

$y_i(P)$：功能函数输出

$A_i(P)$：实验测量真值

学习：

- 正向计算y_i, E
- 反向调节w_{ij},t_j, ie Back-Propagated (BP)
- 直至E最小



- 输入I_0,I_1=0,1

- 权重W_0,W_1，阈W_b

  $a=W_0 * I_0 + W_1 * I_1 + W_b$

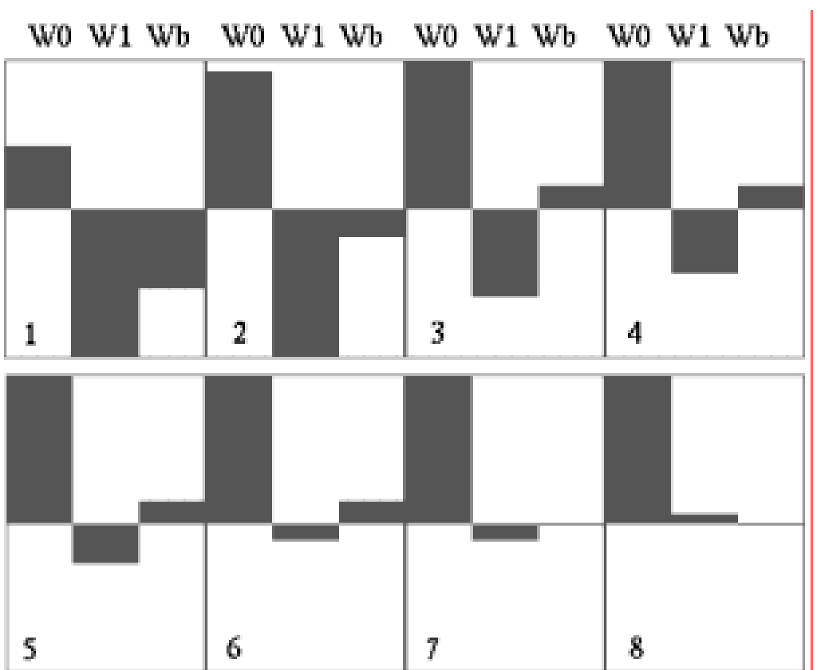
- 激发函数：step function

  $y=1 \quad a>0$

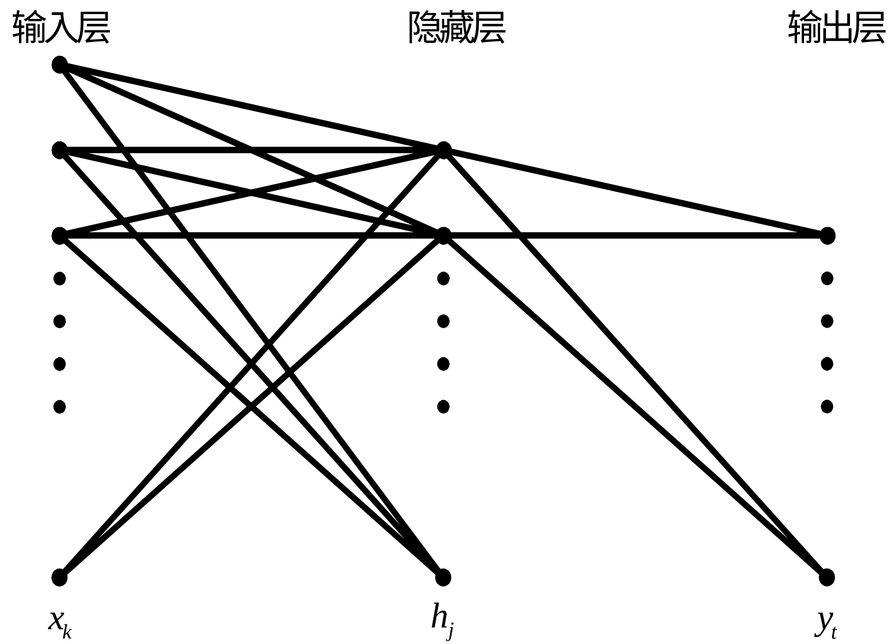
  $\quad 0 \quad a<=0$

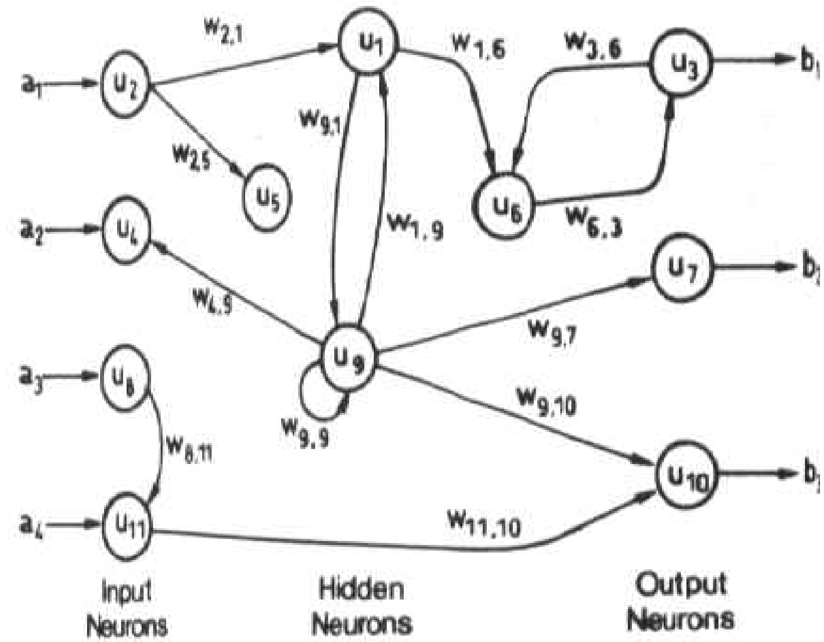
- 真值(训练目标)

  $A = I_0 \text{ OR } I_1$



Not function well
Need more complex network





输入→隐藏：

$$a_j = \sum_k w_{jk} x_k + t_j \qquad h_j = g(a_j)$$

隐藏→输出：

$$\tilde{a}_i = \sum_k \tilde{w}_{ik} h_k + \tilde{t}_i \qquad y_i = g(\tilde{a}_i)$$

相同的基本构成与
active function

反向传播网络训练Back-Propagated (BP)

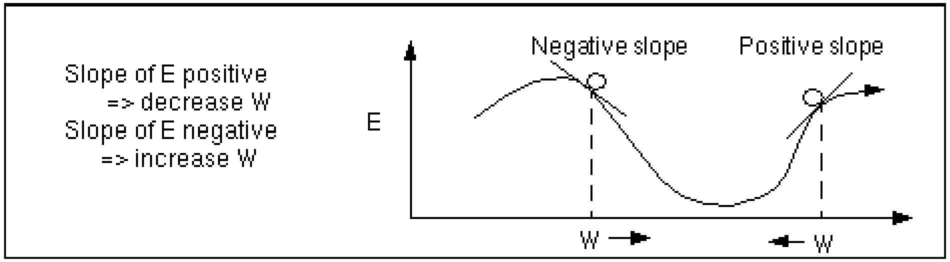
0. 极值原理：误差函数E最小→gradient descent

1.输出层→隐藏层：

$$\Delta \tilde{w}_{ji} = -\eta \frac{\partial E}{\partial \tilde{w}_{ji}} \qquad \frac{\partial E}{\partial \tilde{w}_{ji}} = \sum_p (y_i - A_i) g'(\tilde{a}_i) h_j = \sum_p \delta_i g'(\tilde{a}_i) h_j$$

学习强度

$$\tilde{w}_{ji} += \Delta \tilde{w}_{ji} \to 0$$

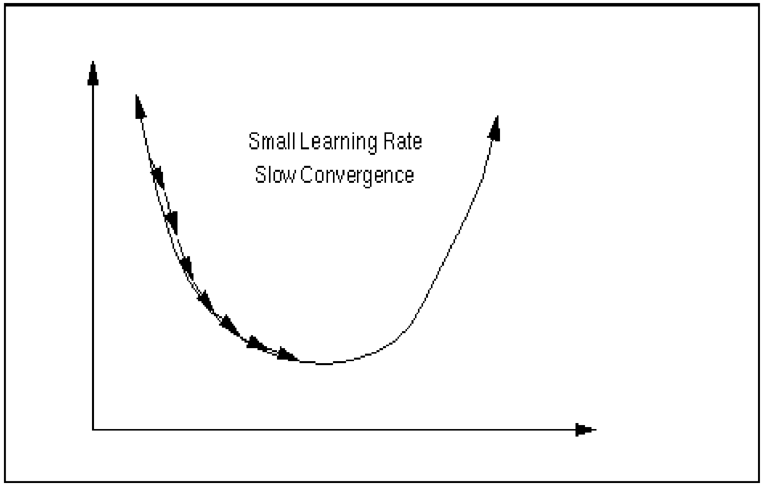


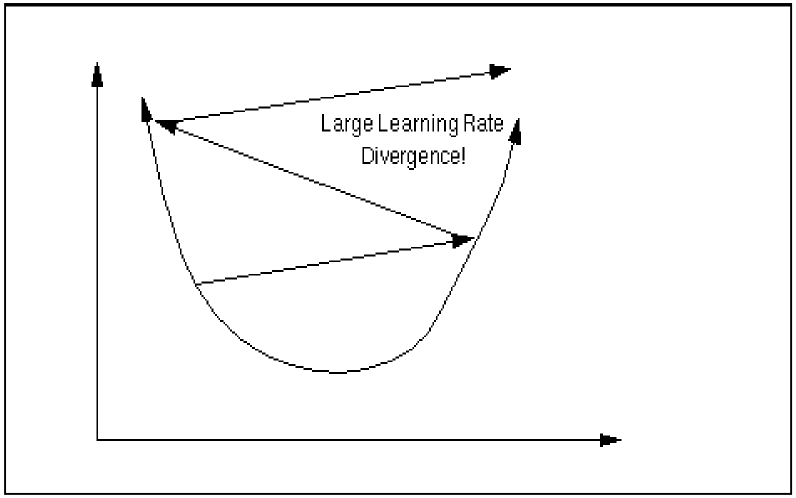
2.隐藏层→输入层：

$$\frac{\partial E}{\partial w_{kj}} = \sum_P \sum_i \delta_i g'(\tilde{a}_i) \tilde{w}_{ji} g'(a_j) x_k = \sum_P \delta'_j g'(a_j) x_k$$

$$w_{kj} += -\eta \frac{\partial E}{\partial w_{kj}} \to 0$$

学习强度η (Learning Rate)





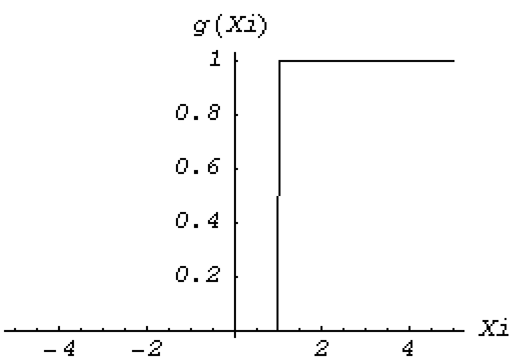
If too small, long time to converge

If too large, cause the algorithm to diverge, an overflow error in the computer's floating-point arithmetic

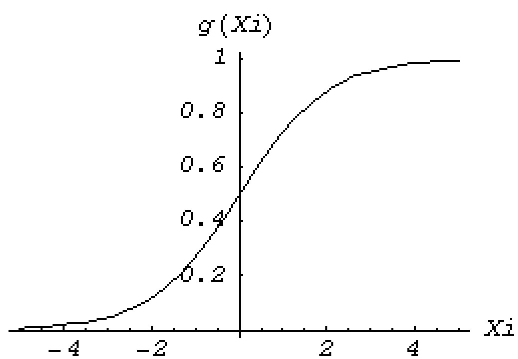
[0.001,0.5]，先大后小

summary

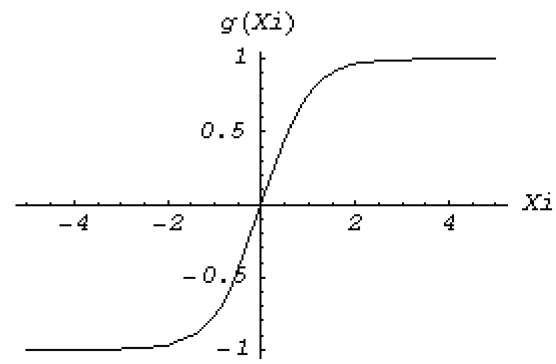
- 真值A_i==true || false
- 输入→隐含层→输出
- 输入与权重$a_i = w_{ij} * x_j$
- 激发函数与输出$y_i = g(a_i)$, 强制性是非选择与放大
- 真值比较 → 反向传播 → 误差梯度递减调节 → 固定w_{ij}



$$g(x) = \begin{cases} 1, & 当 x \geq \theta \\ 0, & 当 x < \theta \end{cases}$$

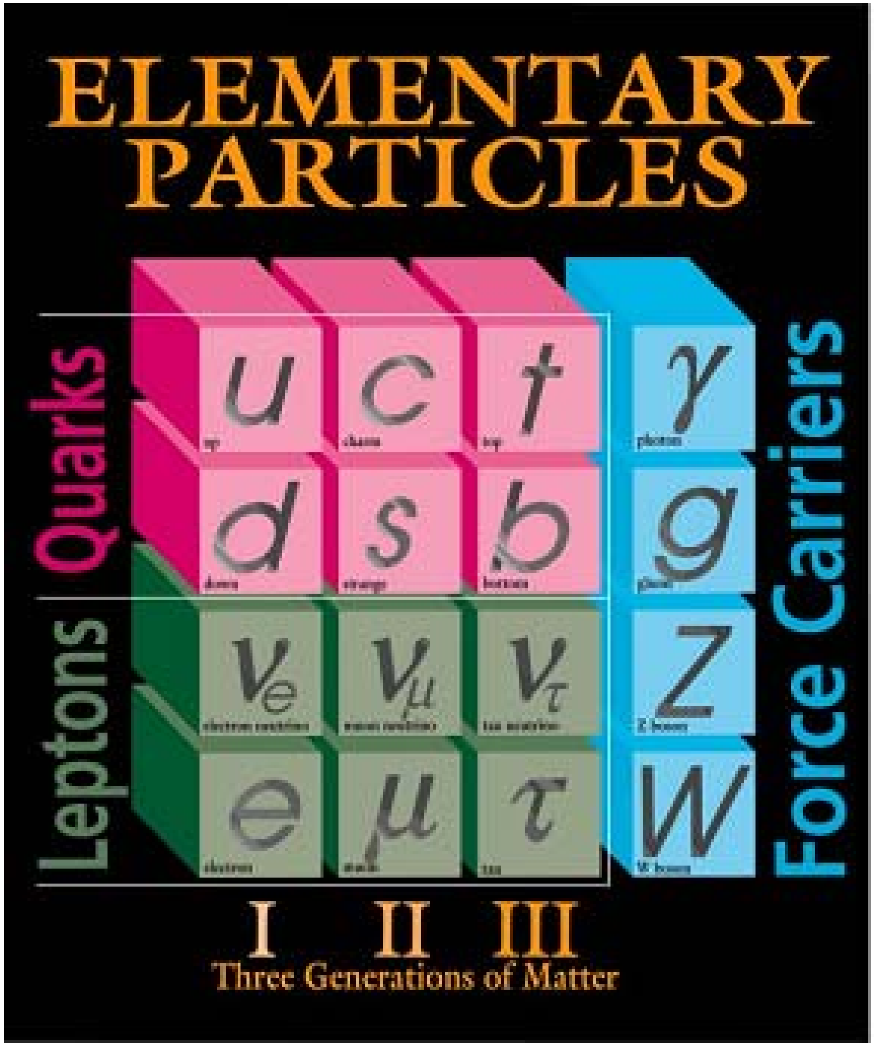


$$g(x) = \frac{1}{1 + e^{-x/T}}$$



$$g(x) = \tanh(x/T)$$

高能强子对撞中的τ鉴别



- 3 generations of quarks
- 3 generations of leptons

} "matter"

- 3 types of interactions
  - Electromagnetic γ
  - Weak W/Z
  - Strong g

} "force carrier"

- Last missing piece: m=f/a?

Higgs? Vacuum?

真空对称性自发破缺

基本粒子与探测

强子谱(1)：

- gluon胶子 ：m==0
  - g → qqbar/gg
- up, down夸克：m~1MeV, π介子, η,ρ,ϕ,ω …etc,etc
  - $\pi^{\pm}$, $c\tau$ = 8m
  - π^0 →2γ, Br~100%
- strange夸克：m~100MeV, K介子
  - $K^{\pm}$, $c\tau$ = 4m
  - K^0-K^0bar mixing → K_L-K_S CP-violation
- charm夸克：m~1GeV
  - D介子, 不稳定
  - ccbar→J/ψ(1S)→$\mu\mu$, Br~6%

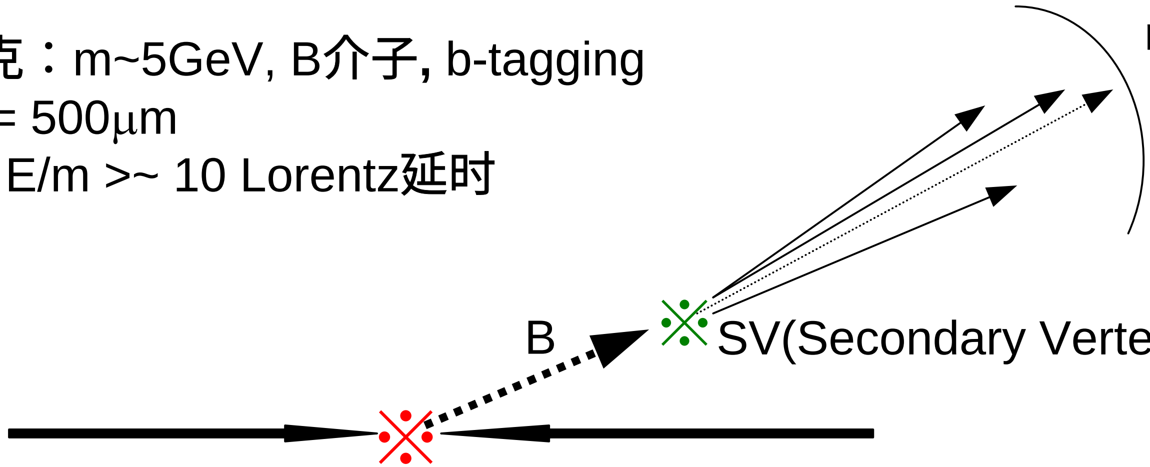
} 原初“jet”喷注
Cone 0.5

强子谱(2)：

- bottom夸克：m~5GeV, B介子, b-tagging

  $c\tau$ = 500μm

  γ = E/m >~ 10 Lorentz延时



2次顶点判选 & 电荷中心法±1/3 etc ~ 50% efficiency

- top夸克：m~180GeV, Γ ~ 3GeV, 寿命极短，不强子化

  t → b + W (semi-lepton decay lν, l=e,μ)　　~100% (10%)

轻子谱(1)：

- 中微子ν_e, ν_μ,ν_τ： m~0, weak interaction → missing transverse energy

- 电子e：m~0.511MeV, stable
  strong bremsstrahlung in high Z material → compact in Calorimeter, ie
  lose most energy in Cone 0.2
  cut on isolation,
  em-fraction,
  shower shape hmatrix

- μ：m~113MeV, cτ=658m
  Track + MIP in Calorimeter + penetrate out-most Muon chamber

轻子谱(2)：

- τ：m~1.776GeV, cτ=90μm

tau -> $l^- \nu_l \nu$ (l=e,μ) ~17.5%
$h^- \nu$ + neutral ~50%
} -> π-"1-prong"

$h^- h^+ h^- \nu$ + neutral ~15% -> ρ-"3-prong"

1-prong problem:

- 1 track, no 2nd Vertex reconstruction
- 50% hadron jet, a little more compact and isolated than initial quark/gluon jet

…but how to describe, Cone 0.4, 0.5 or 0.7?

Challenge to id hadronic tau decay on hadron collider!

Di-tau study: light Higgs decay

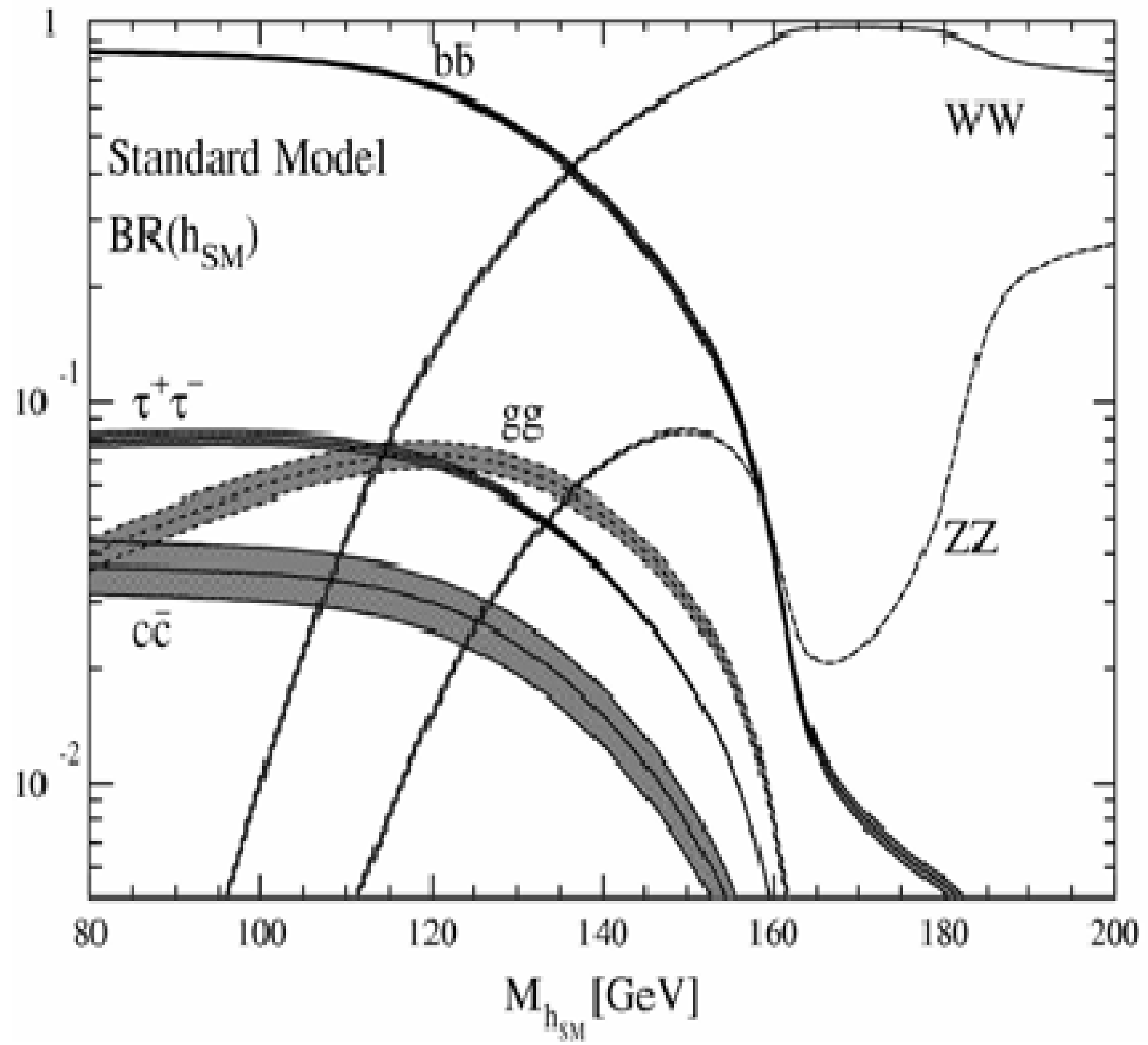
BR of SM Higgs

- Yukawa couplings

$$\Gamma \sim m^2/M_W^2$$
$$= m^2/80^2\ [\text{GeV}^2]$$

- bb channel:
  90% Br, O(1pb)
  v.s.
  background O(1mb)
  S/B ~ 1:1E9

- $\tau\tau$ channel:
  10% Br, O(0.1pb)
  v.s.
  background O(100pb)
  S/B ~ 1:1E3



高能强子对撞中ANN-τ鉴别

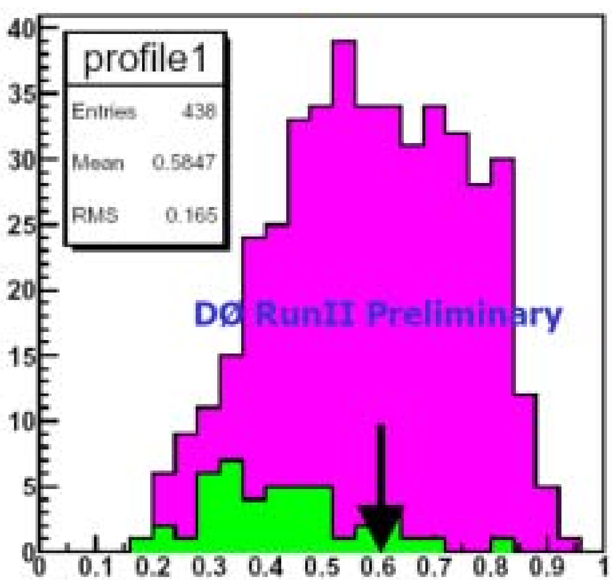
输入变量：

- *profile* = ET Tower(1+2)/ETtot (i.e. broader than electron)
- *trkiso* = PT of tracks, excl the τ 1-prong, in Cone0.7
- *Et/pt, ringiso, e1e2, dalpha, EM12fr* etc

真值比较：

- MC：signal $\tau \rightarrow \pi^{\pm}\nu, \pi^{\pm}\pi^{0}\nu$ vs. q/g/b jet background
- data：signal Z $\rightarrow \tau\tau \rightarrow e/\mu + h$, cuts as E_T(l)>12, M(lh)<60, df>2.5, unlike-sign(l*h)
  background all selections except like-signal instead

Tuning the w-est variable：



$\sigma(Z\tau\tau, \pi\text{-type}) = 235 \pm 137$ pb

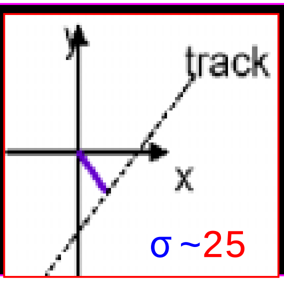
$\sigma(Z\tau\tau, \rho\text{-type}) = 222 \pm 71$ pb

$\sigma(Z\mu\mu) = 261.8 \pm 5.0 \pm 8.9 \pm 26.2$ pb

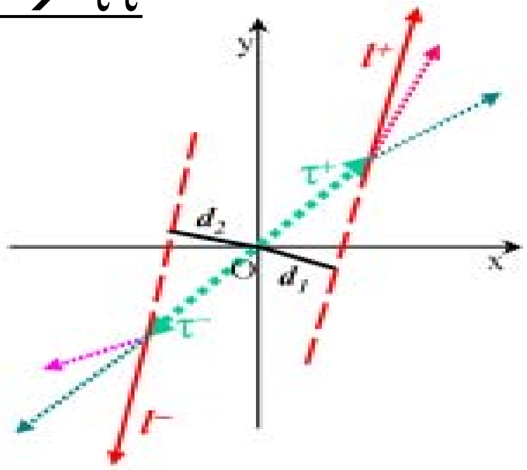
- No way from MC $\rightarrow$ fast MC ID, but a *substitute*
- ANN gives the best result of $Z\tau\tau$ π-type as calibration with efficiency ~ 80%

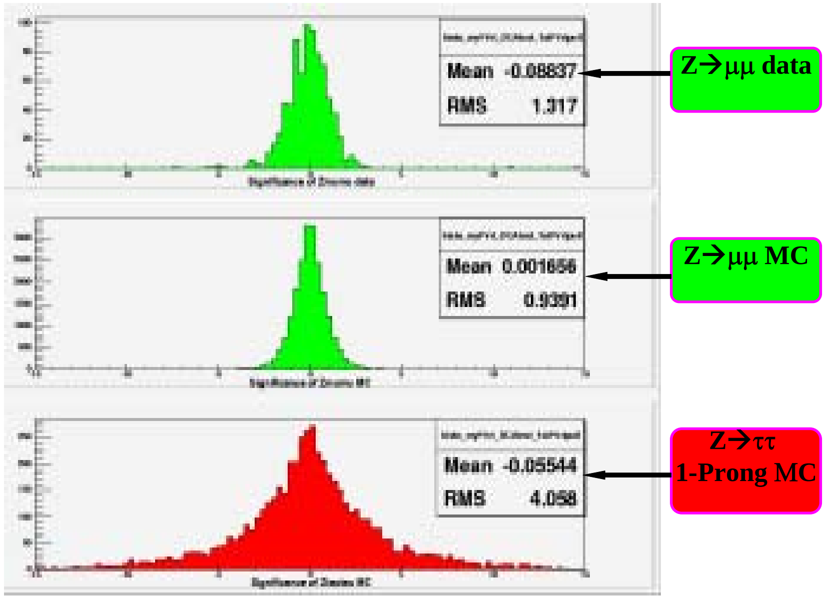
A way to use lifetime for tau id?

CT[tau] ~ 80 μm DistanceClosestApproch~40



The goal Higgs → ττ





- double lifetime information by sumDCA of di-tau， might help
- sumDCA>>resolution, should discriminate from no-lifetime di-track system.
- ANN is the only choice to model, "exaggerate" and tune the cut