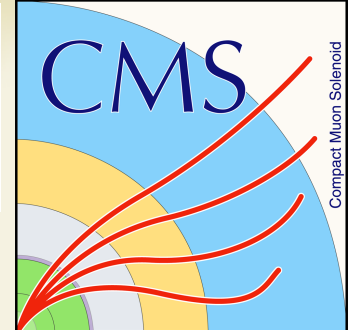




Методология применения нейронных сетей в анализе данных коллайдерных экспериментов

- ~ История вопроса и формулировка задачи
- ~ Обобщенные методы применения NN
- ~ Текущие сложности и вопросы

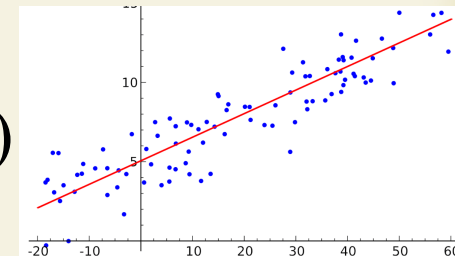
Л.В. Дудко, П.В. Волков, Г.А. Воротников,
А. Заборенко, Э. Абасов, М.Белоброва, Е. Сивакова

НИИЯФ, Физ. ф-т МГУ

История применения искусственных Нейронных сетей (NN)

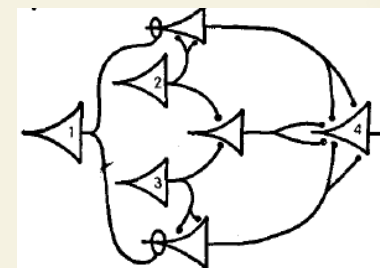
2

~ Линейная регрессия, Лежандр (1805), Гаусс (1809)



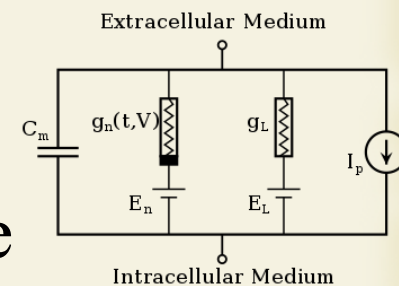
~ 13-я проблема Гильберта (1900г.): возможно ли представление функции нескольких переменных в виде суперпозиции функций меньшего количества переменных.

~ Ramón y Cajal, Нобелевская премия
“structure of the nervous system” (1906)



~ McCulloch, Pitts (1943), модель перцептрона

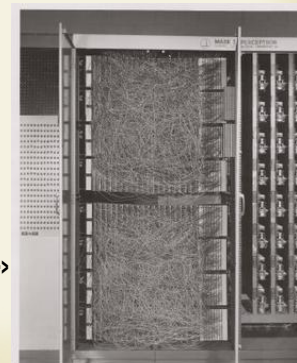
~ Hodgkin & Huxley (1952), (Нобелевская премия 1963), модель распространения сигнала в нейроне



~ Von Neumann “Probabilistic logics and the synthesis of reliable organisms from unreliable components” (1956) (Spiking NN)

~ F. Rosenblatt «Mark I Perceptron at the Cornell Aeronautical Laboratory»

[Cornell University Library, [1] F. Rosenblatt, «The perceptron: A probabilistic model for information storage and organization in the brain.» Psychological Review, Vol 65(6), Nov 1958, 386-408; Rosenblatt, Frank (1957), «The Perceptron--a perceiving and recognizing automaton.» Report 85-460-1, Cornell Aeronautical Laboratory.]



История NN, математические основы

~ 13-я проблема Гильберта была решена В. И. Арнольдом совместно с А. Н. Колмогоровым “Любая непрерывная функция любого количества переменных представляется в виде суперпозиции непрерывных функций одной и двух переменных.”

$$F(x_1, x_2, \dots, x_n) = \sum_{j=1}^{2n+1} g_j \left(\sum_{i=1}^n h_{ij}(x_i) \right)$$

В.И.Арнольд, «О функциях трех переменных»
ДАН СССР. - 1957. - Т.114. - N4. - С.679-681;

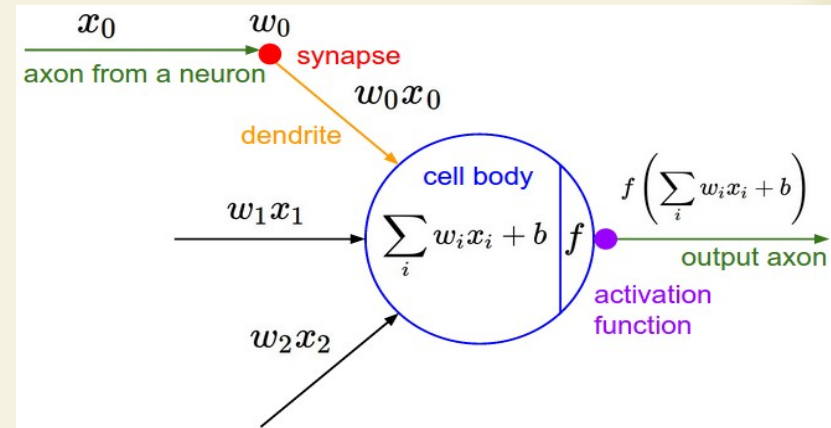
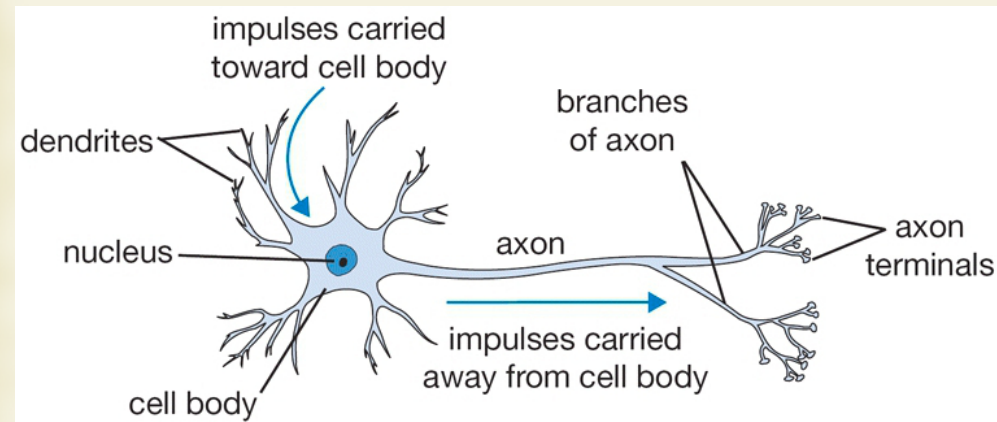
В.И.Арнольд, «О представлении непрерывных функций трех переменных суперпозициями непрерывных функций двух переменных», Матем. сб., 48(90):1 (1959), 3–74, <http://mi.mathnet.ru/msb4884>;

А.Н.Колмогоров, «О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных» ДАН СССР. - 1956. - Т.108. - N2. - С.179-182;

А.Н.Колмогоров, «О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одной переменной и сложения», ДАН СССР. — 1957. — Т. 114, вып. 5. — С. 953—956

Концепция перцептрона и NN

4



Пусть $F(x_1, x_2, \dots, x_n)$ – любая непрерывная функция, определенная на ограниченном множестве, и $\varepsilon > 0$ – любое сколь угодно малое число, означающее точность аппроксимации.

Теорема. Существуют такое число k , набор чисел w_{ij} , θ_i и набор чисел ν_i , что функция

$$f(x_1, x_2, \dots, x_n) = \sigma\left(\sum_{i=1}^k \nu_i \sigma\left(\sum_{j=1}^n w_{ij} x_j + \theta_i\right) + \theta\right) \quad (1.2)$$

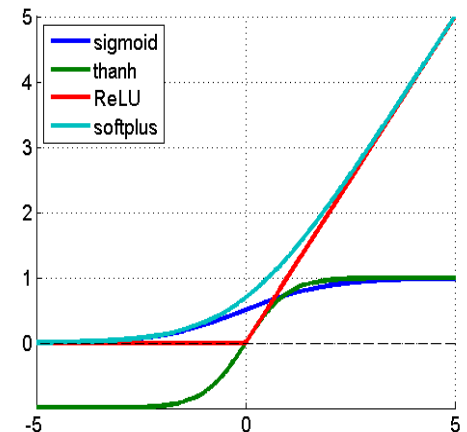
приближает данную функцию $F(x_1, x_2, \dots, x_n)$ с погрешностью, не более ε на всей области определения, где функция σ – некоторая нелинейная функция, например сигмойд:

$$\sigma(x) = \frac{1}{1 + e^{-2x}}. \quad (1.3)$$

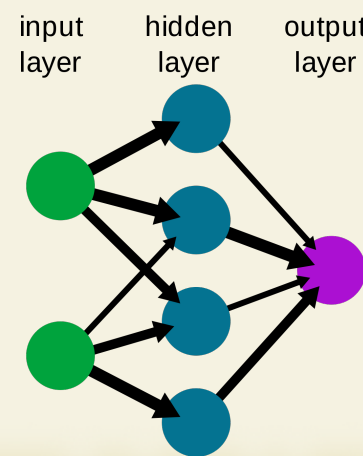
Наиболее популярным критерием сходимости тренировки выбирается функция ошибки:

$$\chi^2 = \frac{1}{2N} \sum_{i=1}^N (f_i - t_i)^2, \quad (1.5)$$

здесь f_i реальный выход сети (1.2) для i -ого события, t_i желаемый выход, а N число тренировочных событий.

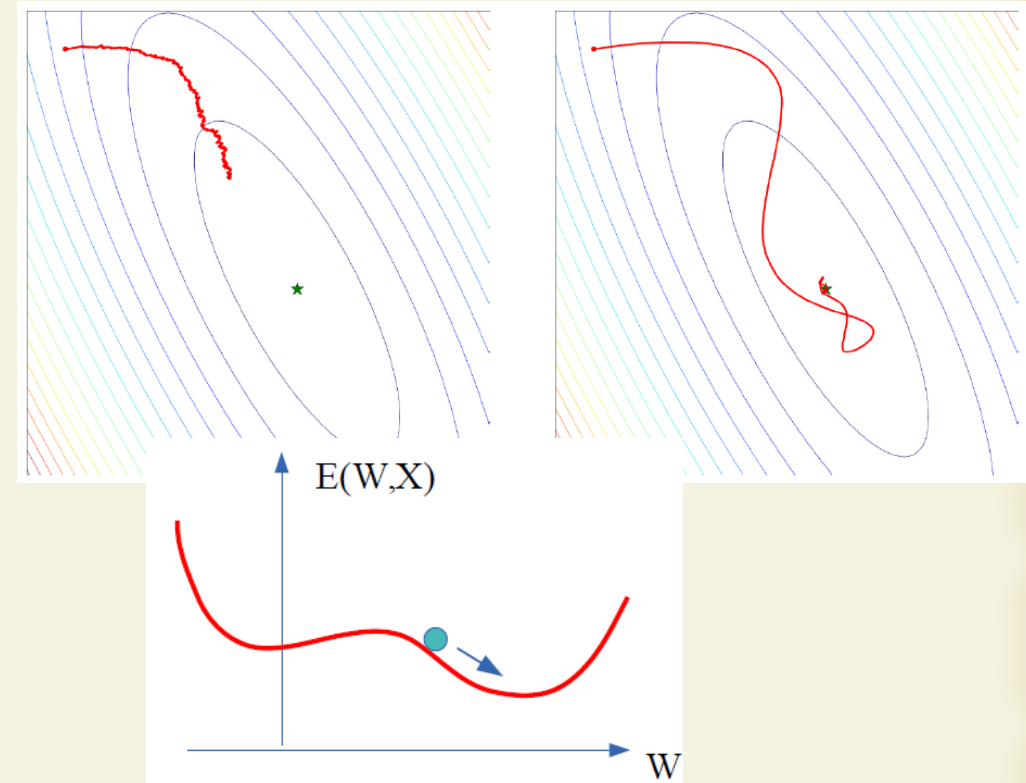
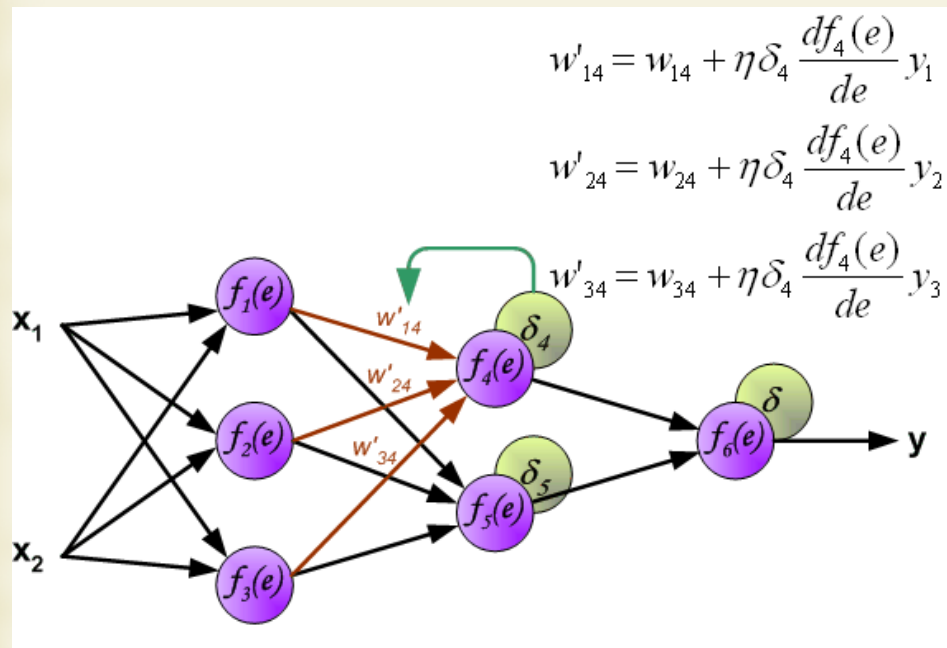


A simple neural network



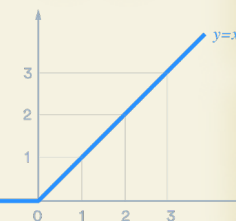
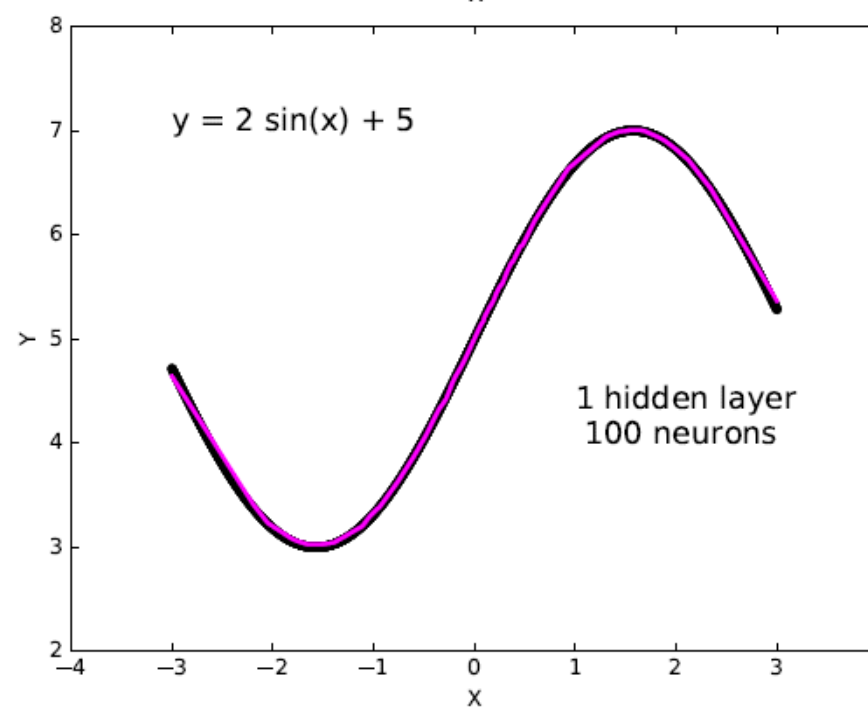
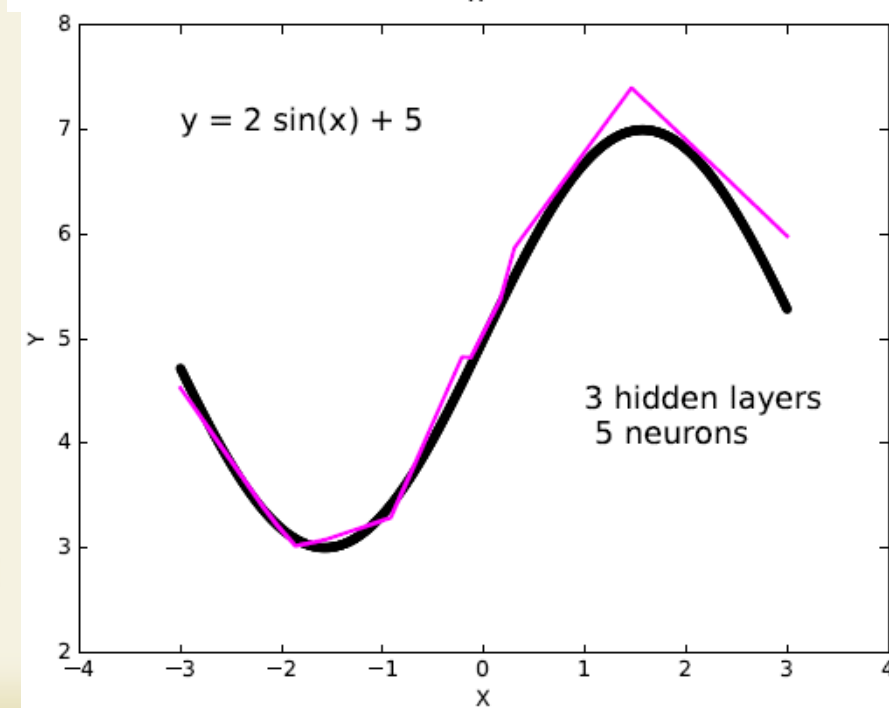
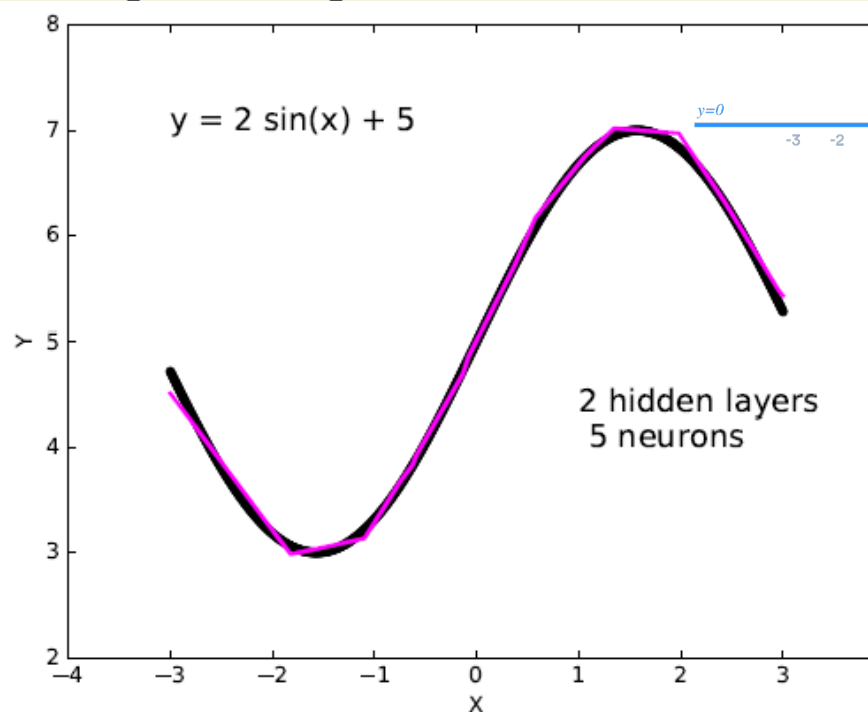
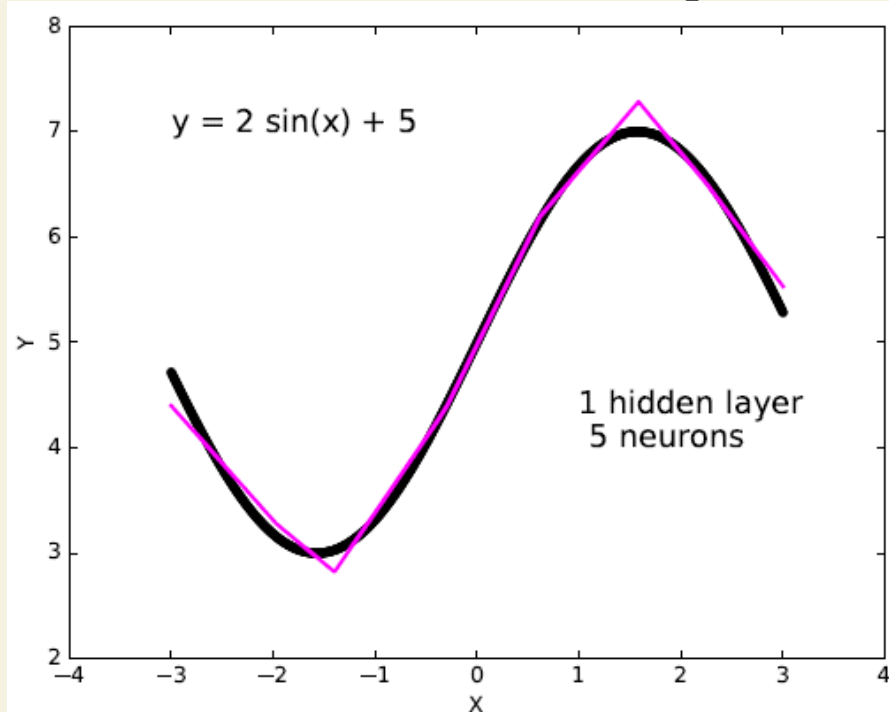
Обратное распространение ошибки (back-propagation) и тренировка NN

5



- 1) Hornik, Stinchcombe, White. Multilayer Feedforward Networks are Universal Approximators. Neural Networks, 1989, v. 2, 5. ;
- 2) Cybenko. Approximation by Superpositions of a Sigmoidal Function. Mathematical Control Signals Systems, 1989, 2.;
- 3) Funahashi. On the Approximate Realization of Continuous Mappings by Neural Networks. Neural Networks, 1989, v. 2, 3.

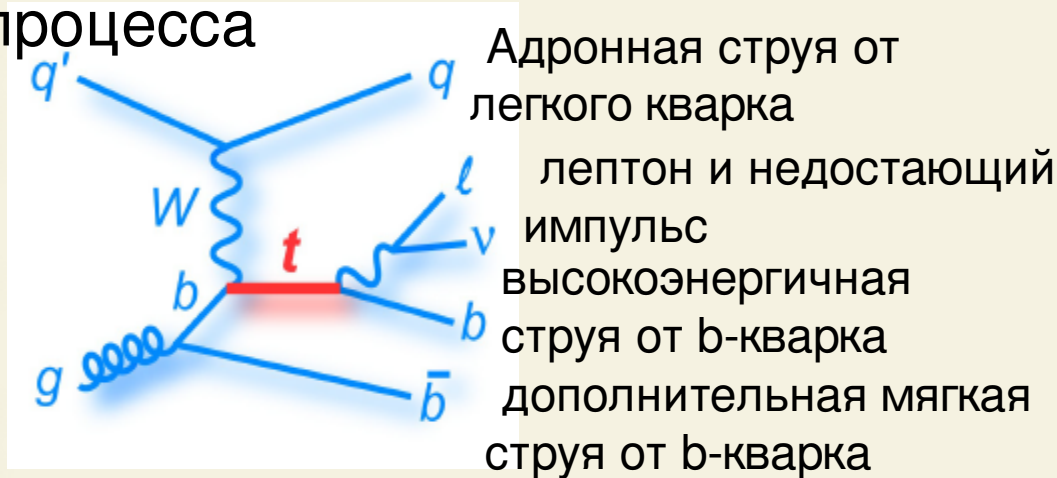
Простой пример



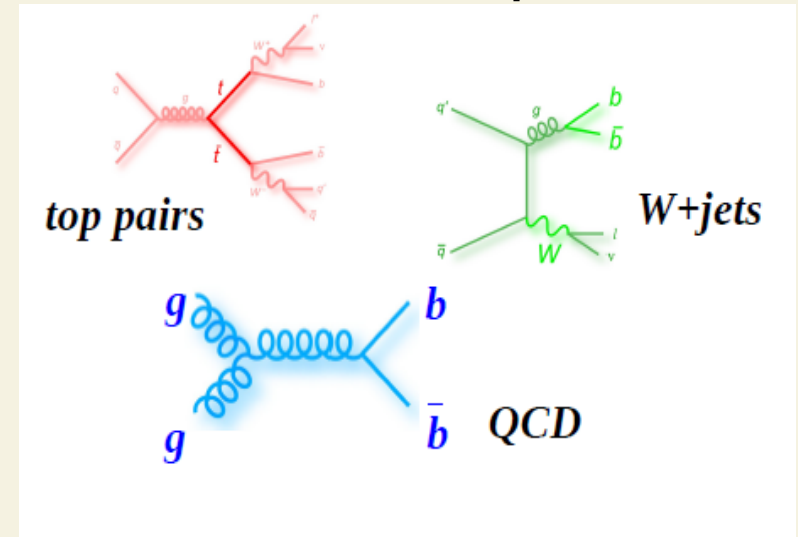
Формулировка стандартной задачи анализа данных коллайдерного эксперимента

7

Конечная сигнатура сигнального процесса



Фоновые процессы



Полные и дифференциальные сечения $d\sigma \sim M^2(p_i, p_f, s, t, u)$ являются функциями скалярных произведений 4-х импульсов частиц и мандельштамовских переменных

Пример квадрата матричного элемента для процесса $u, d \rightarrow t, b$:

$$|M|^2 = V_{tb}^2 V_{ud}^2 (g_W)^4 \frac{(p_u p_b)(p_d p_t)}{(\hat{s} - m_W^2)^2 + \Gamma_W^2 m_W^2},$$

$$|M|^2 = V_{tb}^2 V_{ud}^2 (g_W)^4 \frac{\hat{t}(\hat{t} - M_t^2)}{(\hat{s} - m_W^2)^2 + \Gamma_W^2 m_W^2}.$$

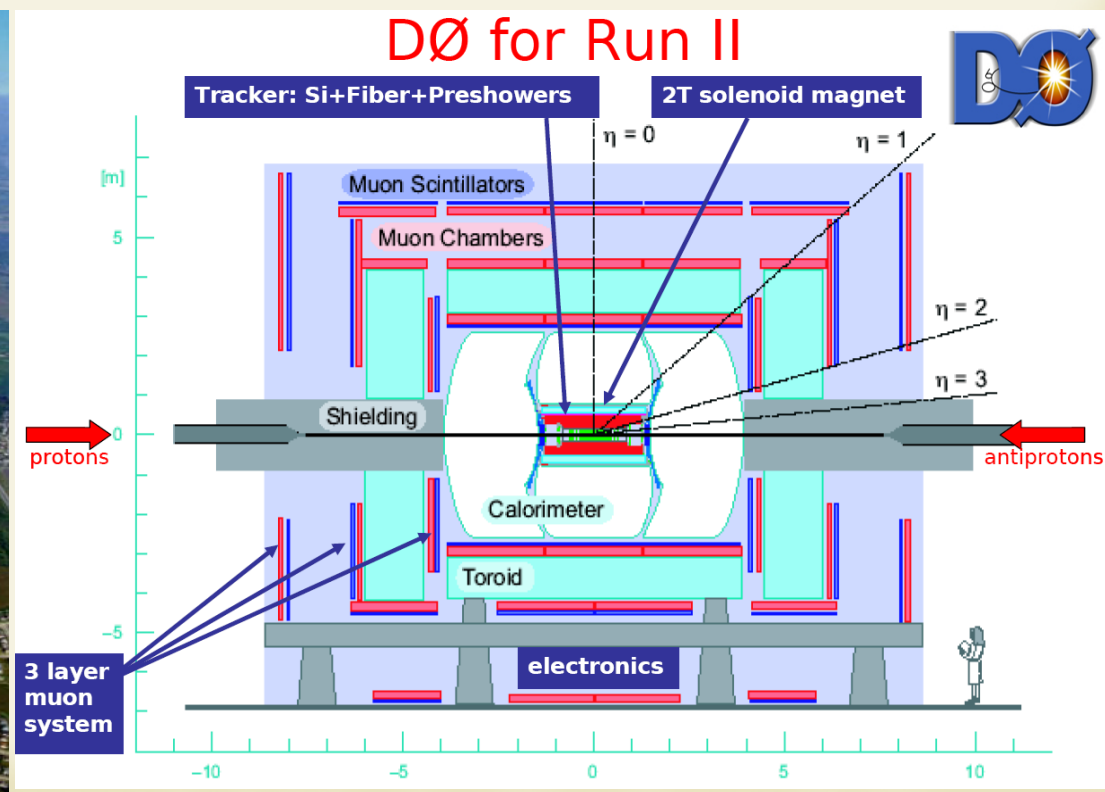
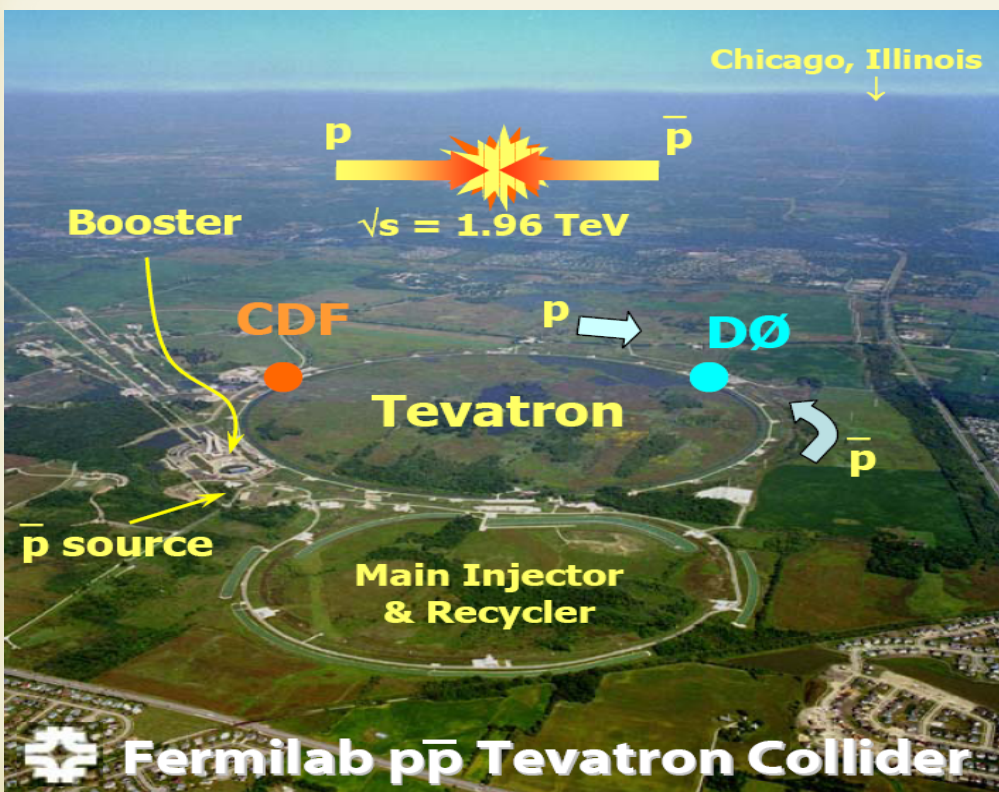
Задача классификации событий в коллайдерном эксперименте и вероятностное выделение событий сигнального процесса.

Tevatron, D0

Протон-антипротонный коллайдер
 Run I: 1992-1995, 1.8 ТэВ, 100 пб⁻¹
 Открытие топ кварка 1995г.
 Phys.Rev.Lett. 74 (1995) 2632-2637 -
 3224 цит.

Run II: 2001-2011, 1.96 ТэВ, 10 фб⁻¹
 Наблюдение одиночного топ кварка
 Phys.Rev.Lett. 98 (2007) 181802 – 270 цит.
 Открытие одиночного топ кварка
 Phys.Rev.Lett. 103 (2009) 092001 – 538 цит.

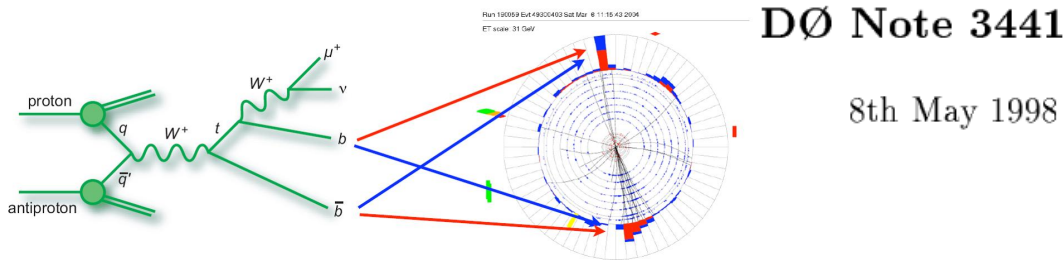
538 публикаций D0 со словами топ кварк
 в заголовке.



Вклад НИИЯФ в топ физику в D0 коллаборации

9

1993 г. - создание группы НИИЯФ для анализа данных D0 эксперимента
1995 г. - создание совместной с UCR группы D0 для исследования одиночного топ кварка.
Phys.Rev.D 63 (2000) 031101 – 86 цит., первая публикация D0 по одиночному топ кварку
Phys.Lett.B 517 (2001) 282-294 – 104 цит.
применение нейронных сетей, одиночный топ.
Всего 114 публикаций D0 по одиночному топ кварку.



Status of Single Top Analysis for $e + jets/\mu$

A. Belyaev, E. Boos, L. Dudko, P. Ermolov, N. Sotnikova
Institute of Nuclear Physics, Moscow State University, Russia

A. Heinson

University of California, Riverside, USA

Abstract

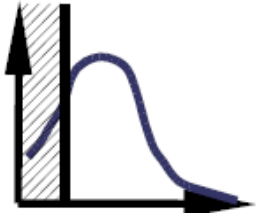
"We present the analysis and the preliminary results from a search for single top quark production ($t+X$ and $\bar{t}+X$) in the electron+jets/muon channel ¹. The data set used for this analysis corresponds to approximately 108 pb^{-1} from Runs 1a, 1b, and 1c. We present the event

В 2002г. исследования одиночного топ кварка в D0 отмечены Шуваловской премией МГУ.

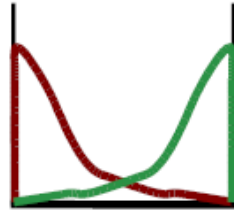
В 2007 г. первое наблюдение одиночного топ кварка отмечено Ломоносовской премией МГУ



Cut-Based



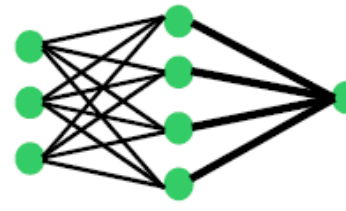
Likelihoods



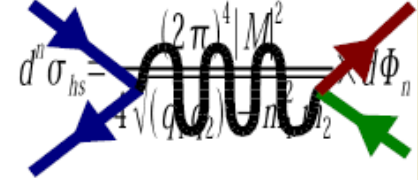
Decision Trees



Neural Networks



Matrix Elements

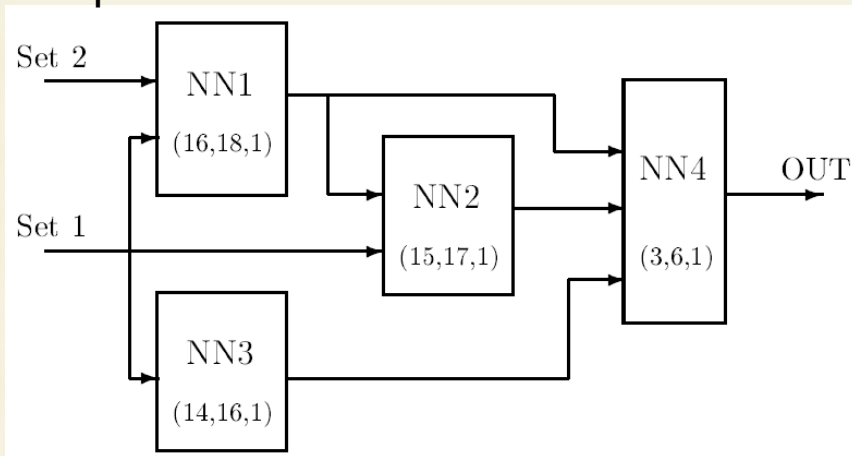


- Поиск редких процессов с высоким уровнем фона вынуждает применять наиболее эффективные методы анализа**
 - Начиная с 1995 года группа НИИЯФ МГУ развивает методы применения нейронных сетей для исследования процессов на коллайдерах.
 - D0-Note 3612 (1999) E.Boos, L.Dudko; AINENP'99 (первые публикации)
- Разработан метод «оптимальных наблюдаемых» и общий рецепт формирования пространства переменных для применения нейронных сетей предложены рецепты применения в задачах поиска топ кварка и бозона Хиггса**
 - E.Boos, L.Dudko, T.Ohl Eur.Phys.J. C11 (1999) 473-484 — 118 цитирований
 - E.Boos, L.Dudko Nucl.Instrum.Meth. A502 (2003) 486-488 — 25 цит.
 - E.Boos, V.Bunichev, L.Dudko, A.Markina, M.Perfilov Phys.Atom.Nucl. 71 (2008) 388-393
 - Int.J.Mod.Phys.A 35 (2020) 21, 2050119
 - Методы использованы в большинстве публикаций D0 по одиночному топ кварку, в некоторых публикациях CMS, и в Physics TDR будущего коллайдера FCC

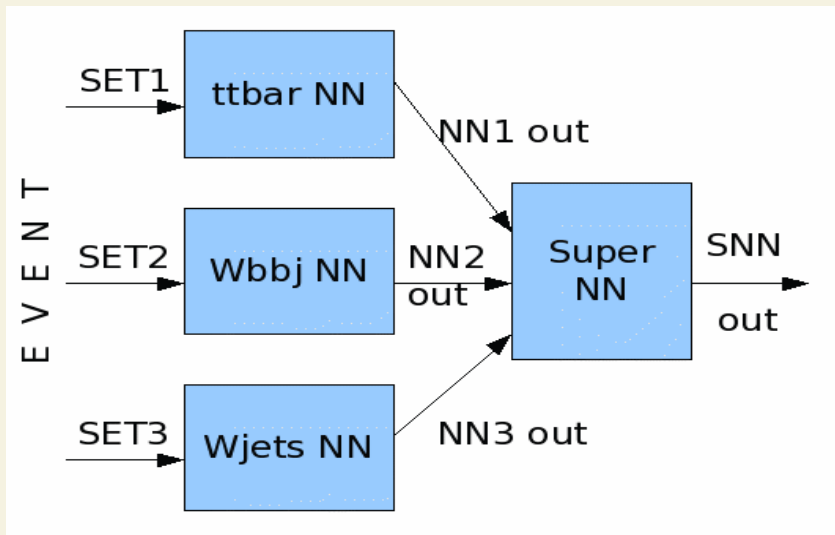
Оптимизация NN в D0 для поиска одиночного топ кварка

11

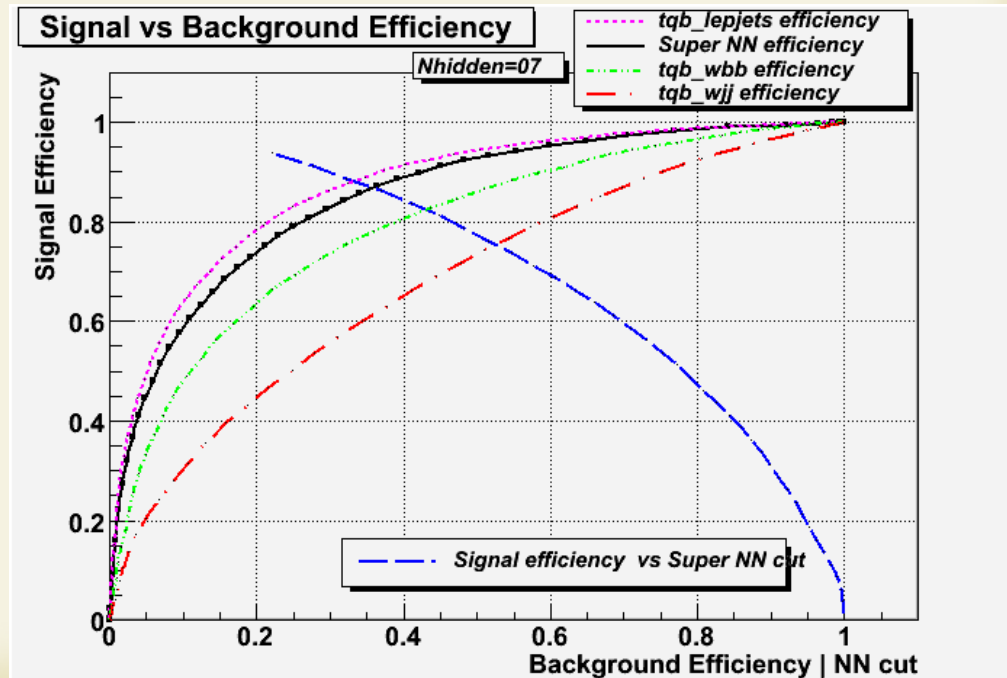
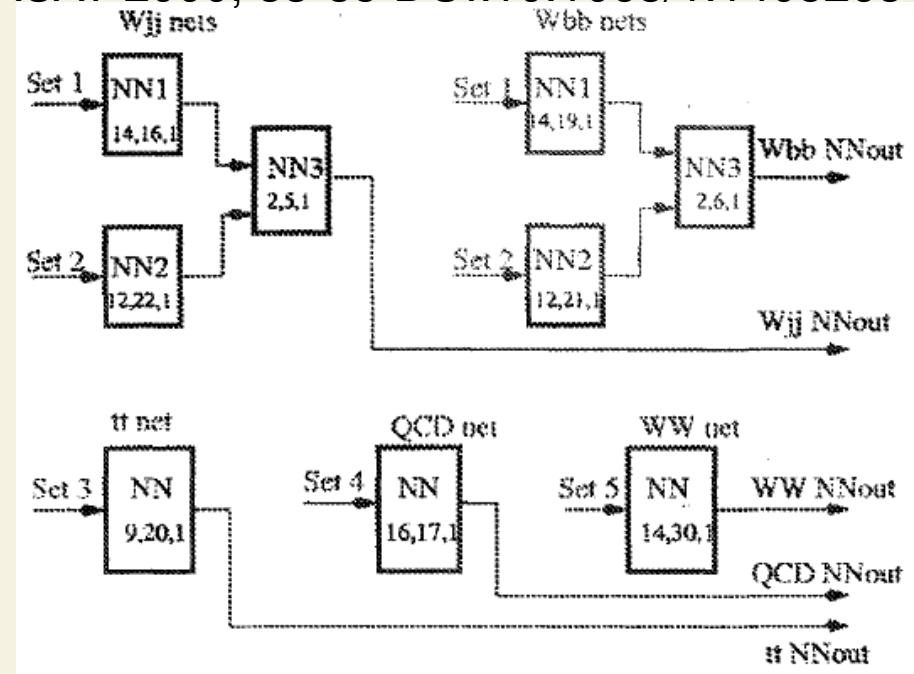
hep-ex/9907041



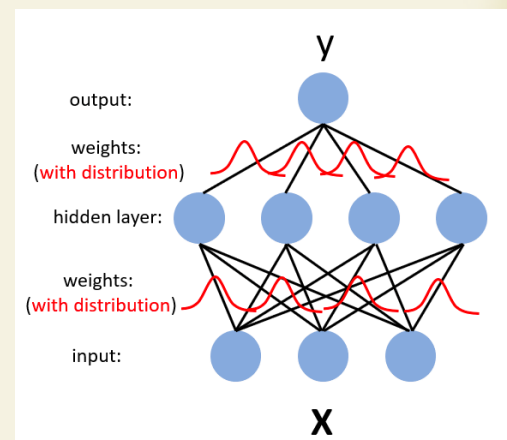
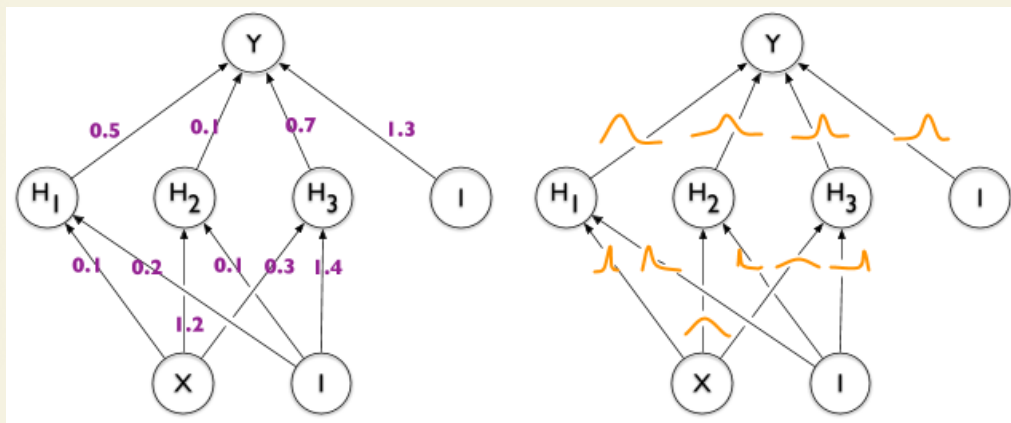
Phys.Lett.B 517 (2001) 282-294



ACAT 2000, 83-85 DOI:10.1063/1.1405268



Байесовские нейронные сети (BNN)

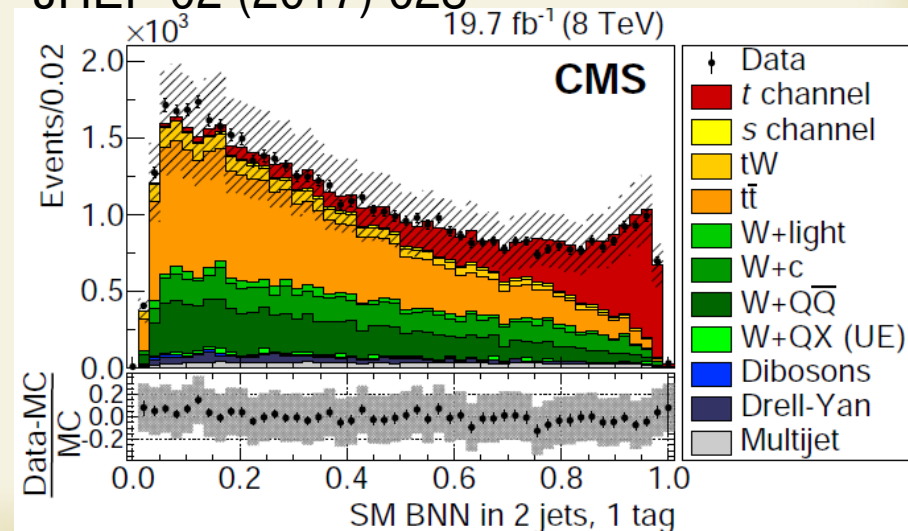


P. C. Bhat and H. B. Prosper, "Bayesian Neural Networks" PHYSTAT 2005;
R. M. Neal, Bayesian Learning of Neural Networks (1996); FBM package;

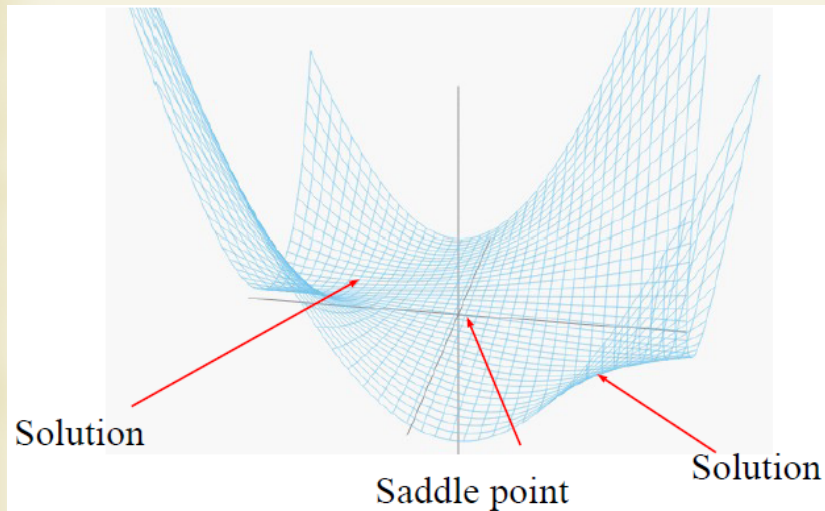
Все анализы D0 после 2005 использовали BNN а не NN
e.g. D0, Observation of Single Top Quark Production
Phys.Rev.Lett. 103 (2009) 092001

Возможна реализация в
глубоких сетях
(tensorflow_probability,
variational dropout, ...), но
с заранее заданной формой
распределения, возможна
реализация через ансамбли
сетей.

Реализация BNN в анализе CMS (LHC)
JHEP 02 (2017) 028

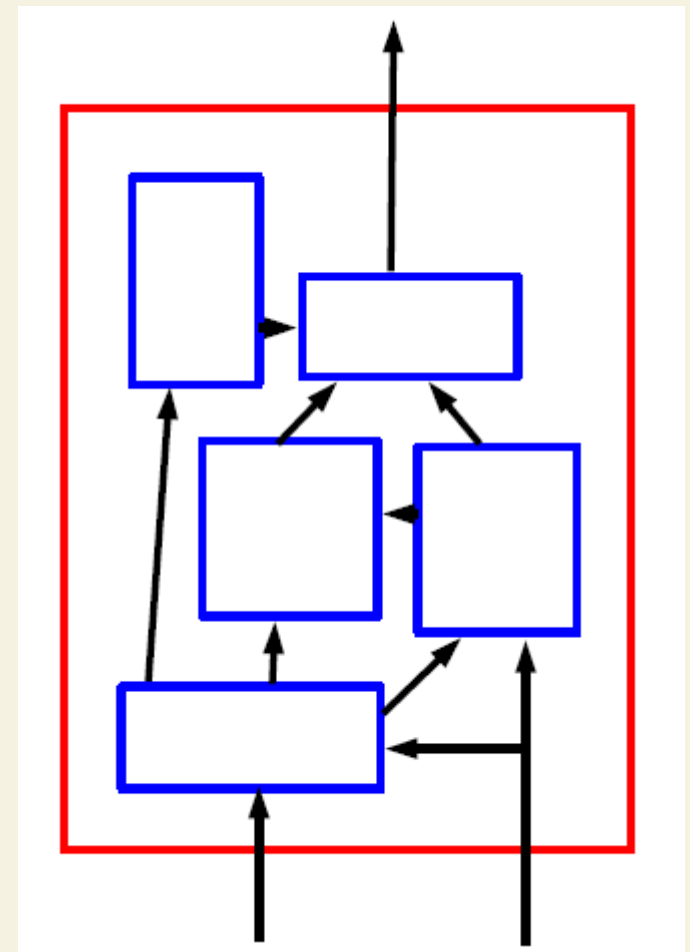


Глубокие нейронные сети (Deep Learning Neural Networks, DNN)



Hinton, G. E., Osindero, S., & Teh, Y. W. (2006).
A fast learning algorithm for deep belief nets.
Neural computation, 18(7), 1527-1554.

Основное преимущество глубоких нейронных сетей
(много слоев, узлов) — возможность
анализировать сырую информацию.

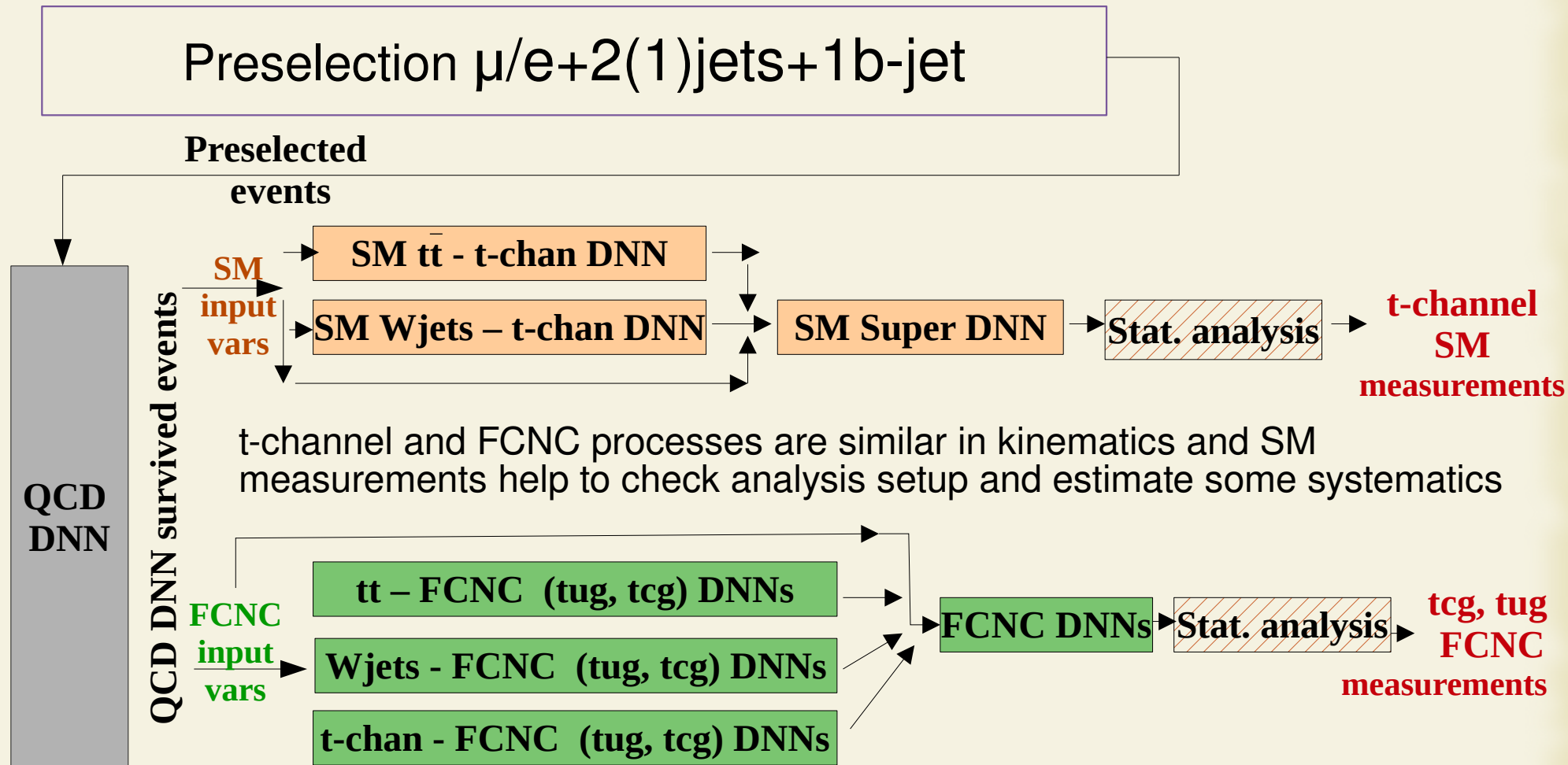


Первое использование в ФВЭ: Nature Commun. 5 (2014) 4308

Technique	Discovery significance		
	Low-level	High-level	Complete
NN	2.5σ	3.1σ	3.7σ
DN	4.9σ	3.6σ	5.0σ

$$gg \rightarrow H^0 \rightarrow W^\mp H^\pm \rightarrow W^\mp W^\pm h^0$$

Текущие реализации DNN в анализе в CMS (LHC)



Run II: combination of 2016 (pre,post VFP), 2017, 2018 datasets [CMS-AN-16-371].

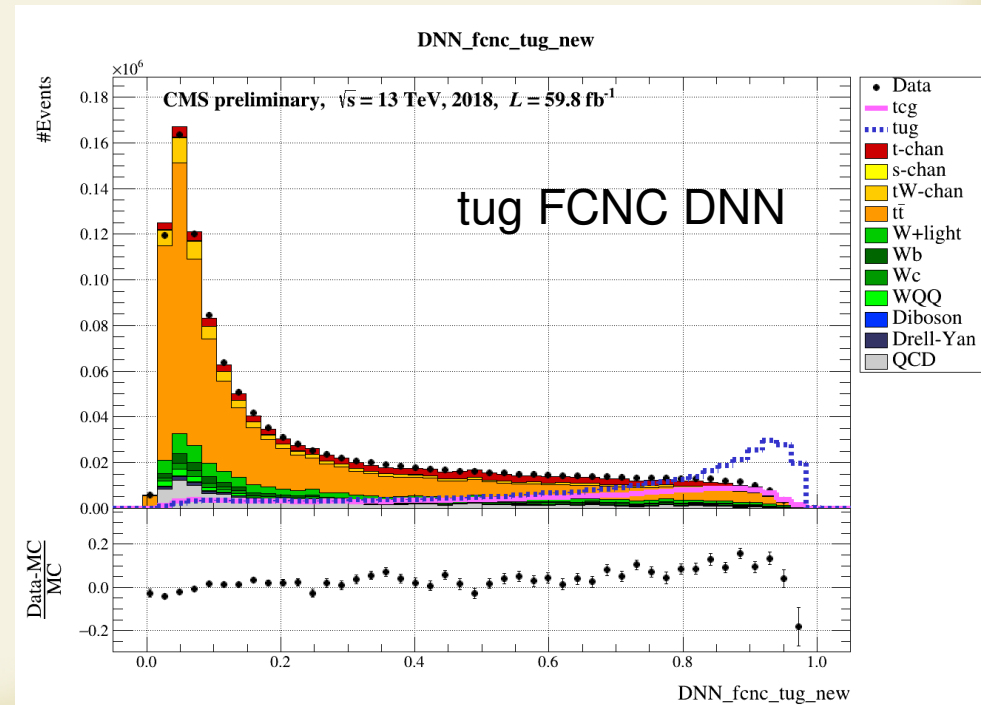
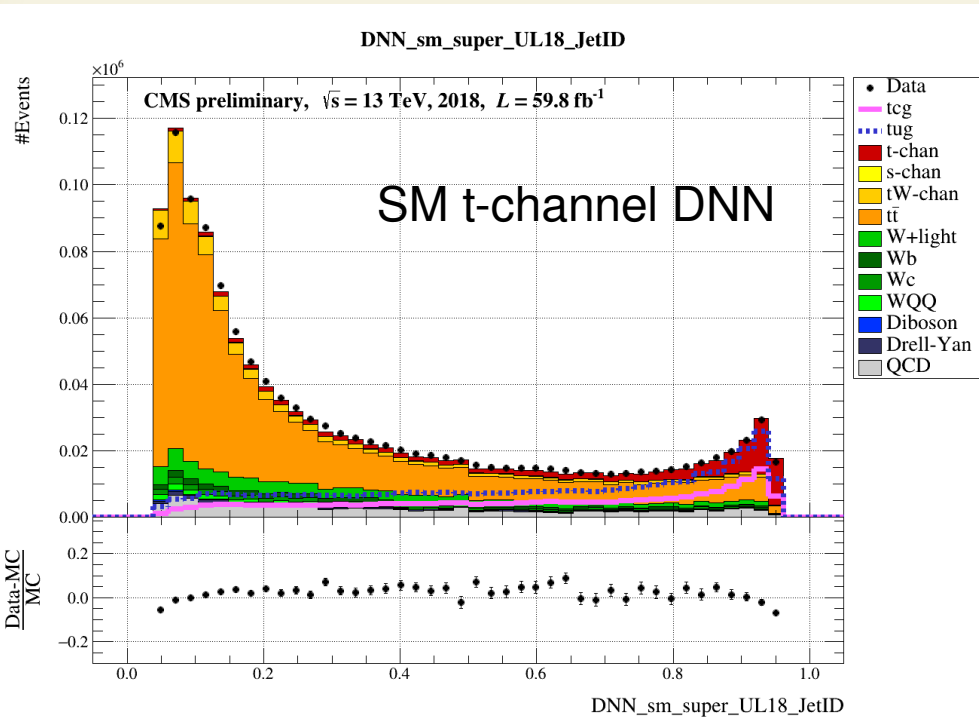
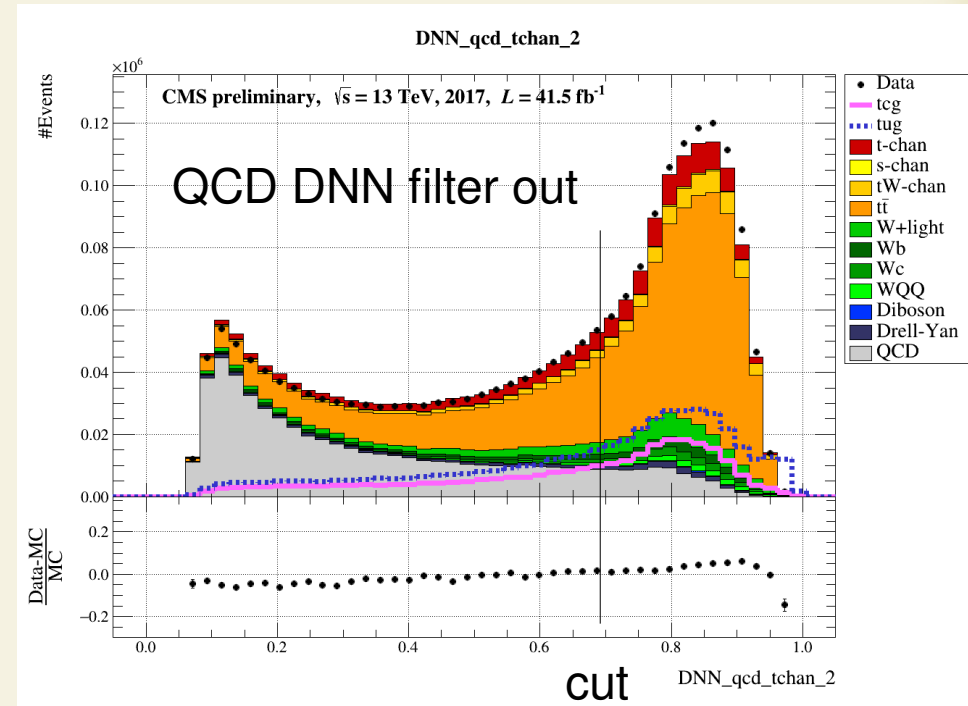
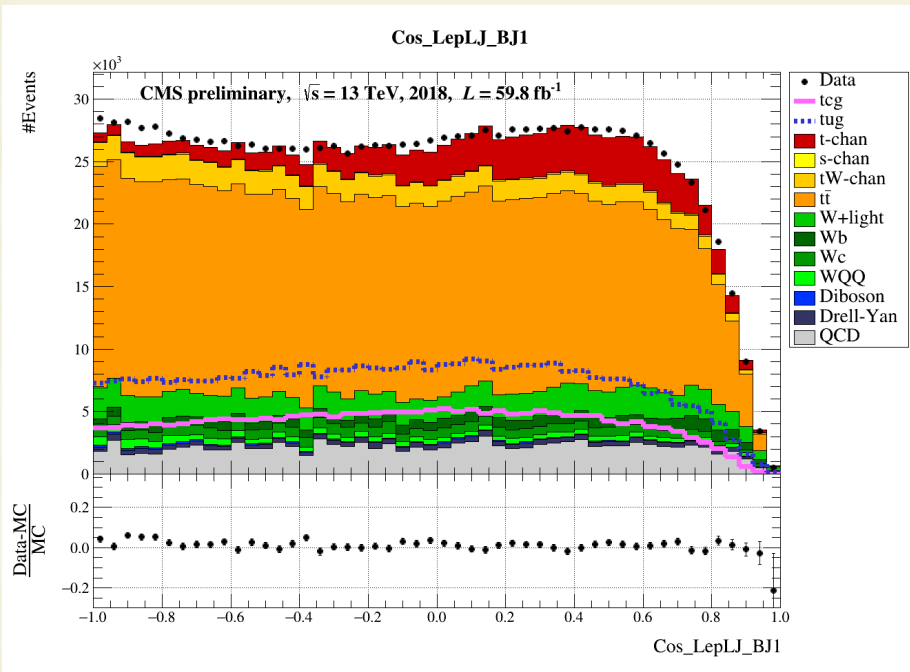
Published results:

- 1) 7&8 TeV JHEP02(2017)028 (FCNC tqg & aWtb)
- 2) 14 TeV HL-LHC YR, PAS-FTR-18-004; extrapolation to HE-LHC (FCNC tqg)
- 3) FCC 100 TeV, CDR vol.1 (FCNC tqg)

Possible SM measurements: t-channel total cross section, fiducial cross section, differential cross sections, V_{tb} , $R(t/\bar{t})$ measurements

Детали настройки и оптимизации DNN в докладе А. Заборенко

Примеры распределений реализации DNN в CMS



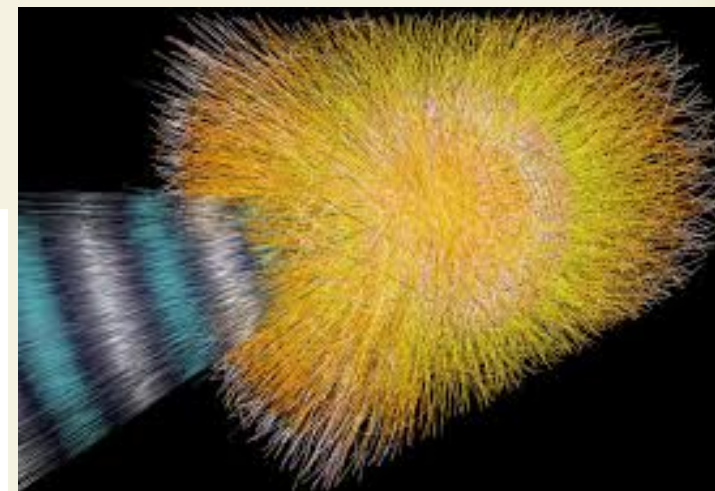
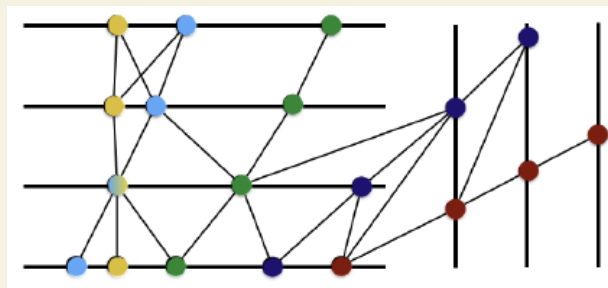
Общие рекомендации использования NN

- 1) Увеличивая размер сети (количество входов, узлов, слоев) мы увеличиваем количество свободных параметров (весов) и усложняем тренировку.
- 2) Входная информация для сети должна отражать полный набор отличительных признаков, но исключать избыточную информацию усложняющую тренировку.
- 3) Минимизируя степень нелинейности в формулировке задачи для DNN можно упростить тренировку $F(x)=x^2 \rightarrow NN(x^2)$
- 3) Наиболее эффективная полносвязная сеть — полный набор однородных признаков на входе (нормализация, логарифмическая трансформация) и минимально-достаточное количество узлов в архитектуре.
- 4) Необходим контроль влияния систематических неопределенностей существующих в задаче. В идеале, пропагирование неопределенностей должно быть включено в функцию ошибки для минимизации.

Основные направления использования DNN в коллайдерной физике

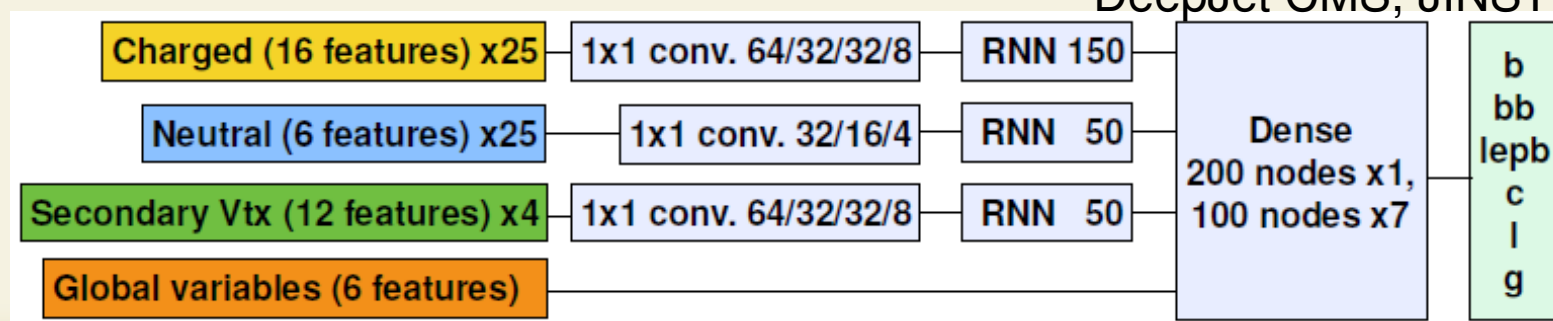
~ Анализ данных. Задача классификации событий и выделение сигнальных. Задача регрессии, измерение параметра - выход NN это функция интерпретируемая в качестве значения параметра.

~ Восстановление треков в детекторе (GraphNN, ...)



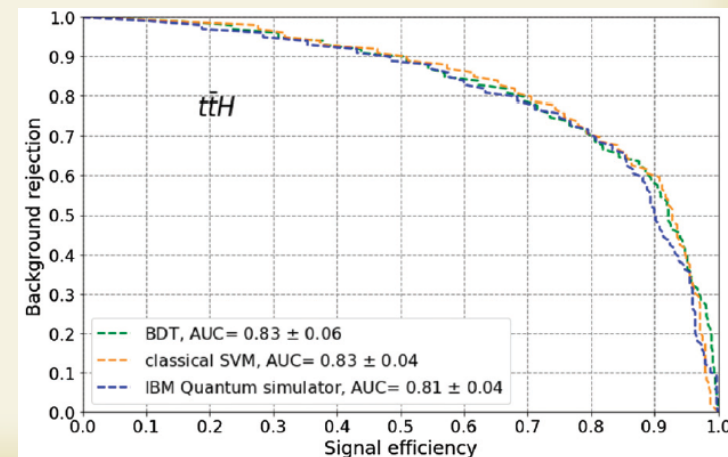
~ Идентификация объектов (тагирование аромата адронных струй, boosted top, ...)

DeepJet CMS, JINST 15 (2020) 12, P12012



Типы реализаций искусственных NN (ANN)

- ~ Вычисляемые ANN, память и процессор разделены, различные типы DNN (GPU, TPU)
- ~ Нейроморфные реализации (Spiking NN), вычисления и память объединены,
<https://indico.cern.ch/event/1070931>
- ~ Мемристорные сети, на основе мемристоров чьи Вольт-Амперные характеристики зависят от предыдущих значений тока; интеграция мемристорных сетей и живых клеточных структур.
- ~ Осцилляторные сети, нейронная сеть моделируется через возбуждение полей осцилляторов.
- ~ Реализация на квантовых компьютерах;
 IBM 10qubits
 DOI: 10.1088/1361-6471/ac1391



The Brain: an Amazingly Efficient "Computer"

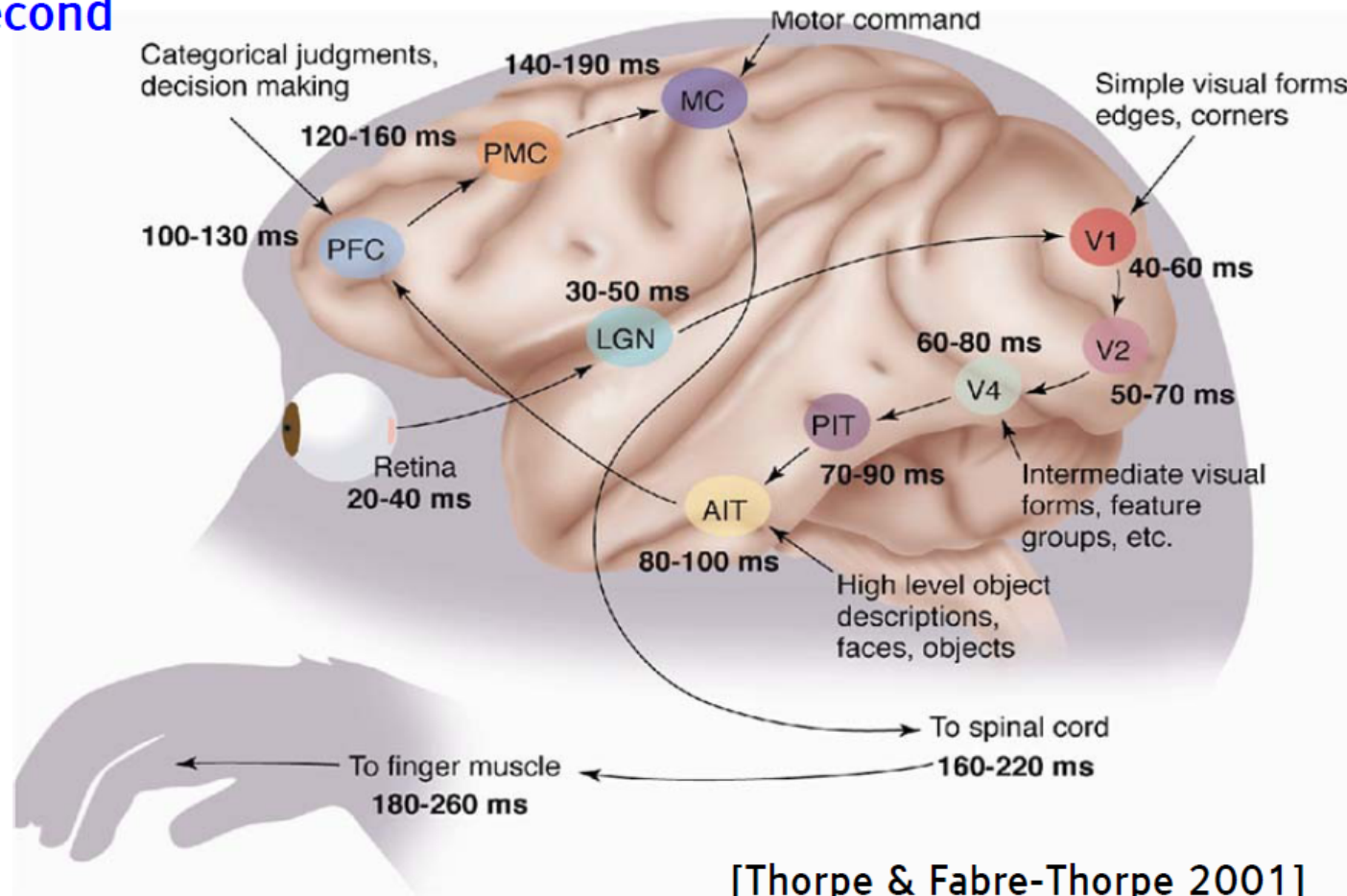
Y LeCun

- 10^{11} neurons, approximately
- 10^4 synapses per neuron
- 10 "spikes" go through each synapse per second on average
- 10^{16} "operations" per second

■ 25 Watts
▶ Very efficient

■ 1.4 kg, 1.7 liters

■ 2500 cm²
▶ Unfolded cortex

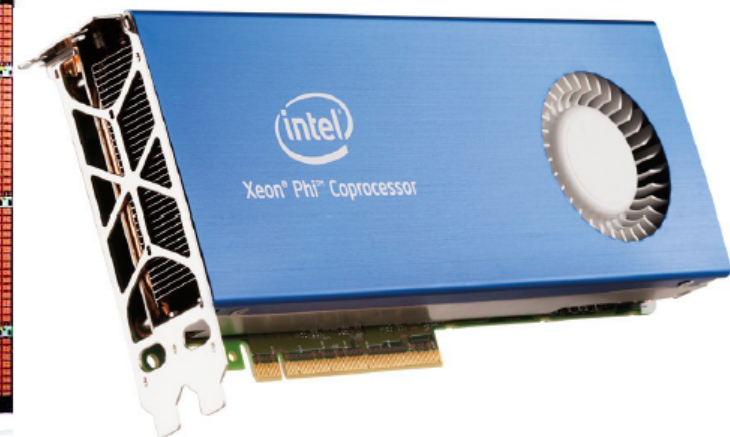
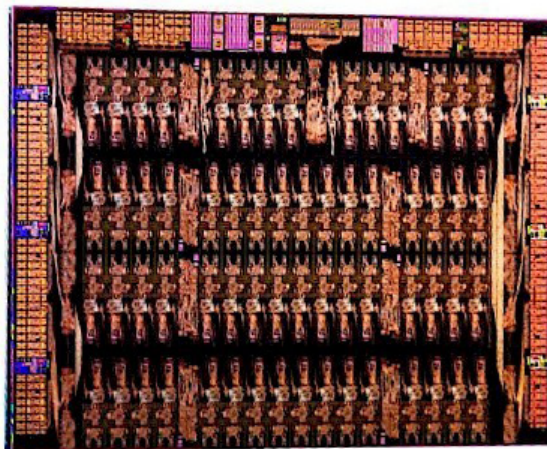


Fast Processors Today

Y LeCun

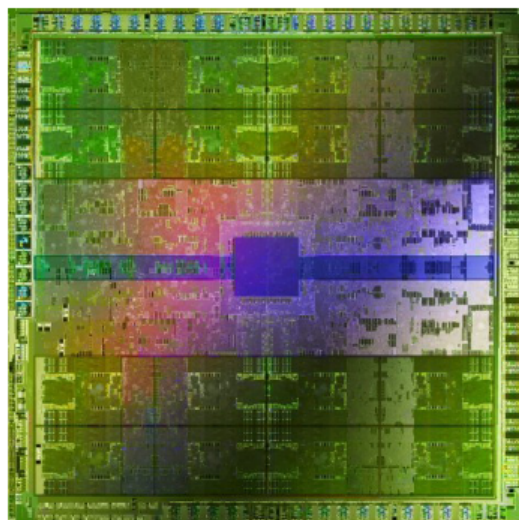
Intel Xeon Phi CPU

- ▶ 2×10^{12} operations/second
- ▶ 240 Watts
- ▶ 60 (large) cores
- ▶ \$3000



NVIDIA Titan-Z GPU

- ▶ 8×10^{12} operations/second
- ▶ 500 Watts
- ▶ 5760 (small) cores
- ▶ \$3000



Are we only a factor of 10,000 away from the power of the human brain?

- ▶ Probably more like 1 million: synapses are complicated
- ▶ A factor of 1 million is 30 years of Moore's Law
- ▶ 2045?