

PAUL SCHERRER INSTITUT

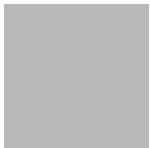


Andreas Adelmann:: LSM :: Paul Scherrer Institut

# Machine Learning & Particle Accelerators

Acknowledgement:

A.L. Edelen (UCO), J. Snuverink, Ch. Baumgarten & D. Reggiani (PSI)



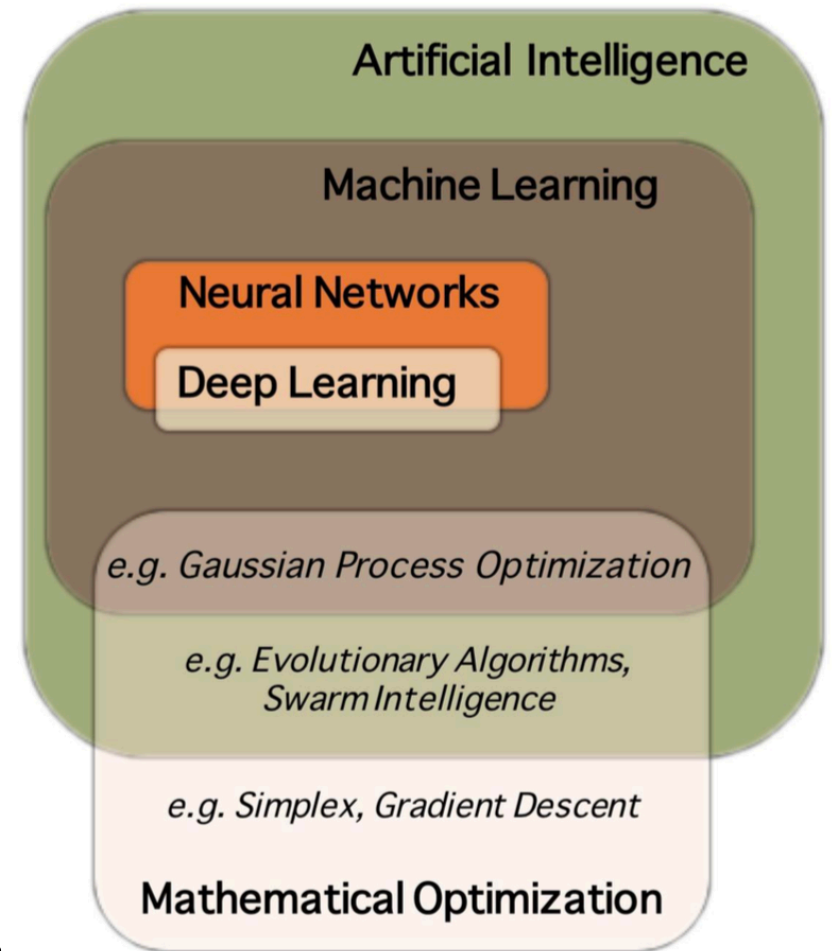


# Content

- Short History
- A bit of definition & theory
- Examples
- Ideas for ML and HIPA (blueprint for SLS & SwissFEL)

# Field Taxonomy (not a rigid definition)

- Artificial Intelligence (AI)
  - enabling machines to exhibit aspects of human intelligence
- Machine Learning (ML)
  - enables machines to fulfill tasks without explicitly programmed
- Neural Networks (NNs or ANNs)
  - one approach within ML
- Deep Learning (DL)
  - synonymous with deep (many-layered) NNs



A.L. Edelen

## Short History

- AI: Machine Learning (ML) and Neural Networks (NN) are old techniques [1]
  - **1950** — Alan Turing creates the “Turing Test” to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human.
  - **1957** — Frank Rosenblatt designed the first neural network for computers (the perceptron)
  - **1967** — The “nearest neighbor” algorithm was written (for the TSP)
  - **1979** — Students at Stanford University invent the “Stanford Cart” (navigate around obstacles)
  - **1985** — Terry Sejnowski invents NetTalk, which learns to pronounce words at the level of a baby.
  - **1990s** — Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions — or “learn” — from the results.
  - **1997** — IBM’s Deep Blue beats the world champion at chess.
  - **2006** — Geoffrey Hinton coins the term “deep learning” to explain new algorithms that let computers “see” and distinguish objects and text in images and videos.
  - **2010** — The Microsoft MSFT - Kinect can track 20 human features at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.
  - **2015** — AlphaGo first program to beat a professional Go player (ML & TreeSearch)
  - **now** ..... ?

# Why do we have such a meeting **now**?

**Increased computational capability** enables more complicated NN architectures and faster training + larger data sets

GPUs



Accessibility of HPC clusters

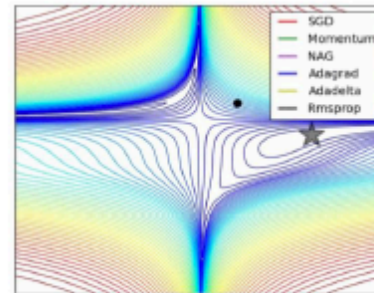
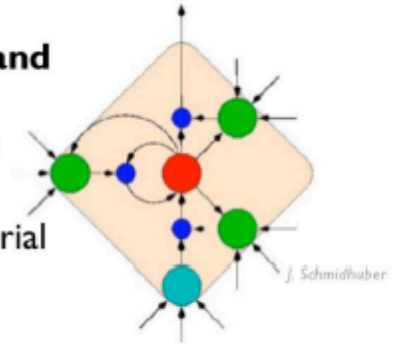


Can **easily share** large data sets, code, and computing setups (e.g. via cloud computing services)

Up-and-coming advancements: neuromorphic hardware



**New network architectures and training paradigms**, such as long short term memory (LSTM) networks, neural Turing machines, and generative adversarial networks (GANs)



A. Radford

**Better theoretical understanding of NNs and improved optimization methods**

**Applications** have driven a lot of advancement (both algorithmic and practical/heuristic)

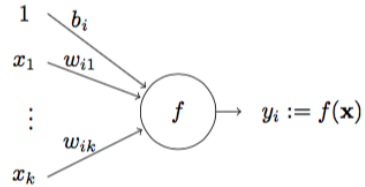


Google

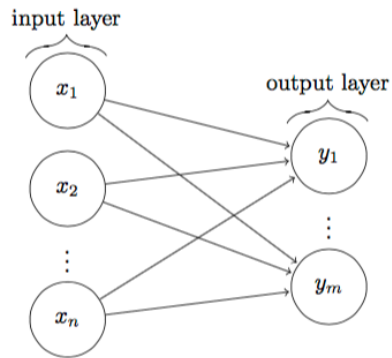
A.L. Edelen

# A bit of definition & theory

## 1 Neuron or Node



## 2 The Perceptron



The perceptron consists of  $n$  input units and  $m$  output units. Every input unit is connected to every output. For  $1 \leq i \leq m$  the  $i^{\text{th}}$  output unit computes the output

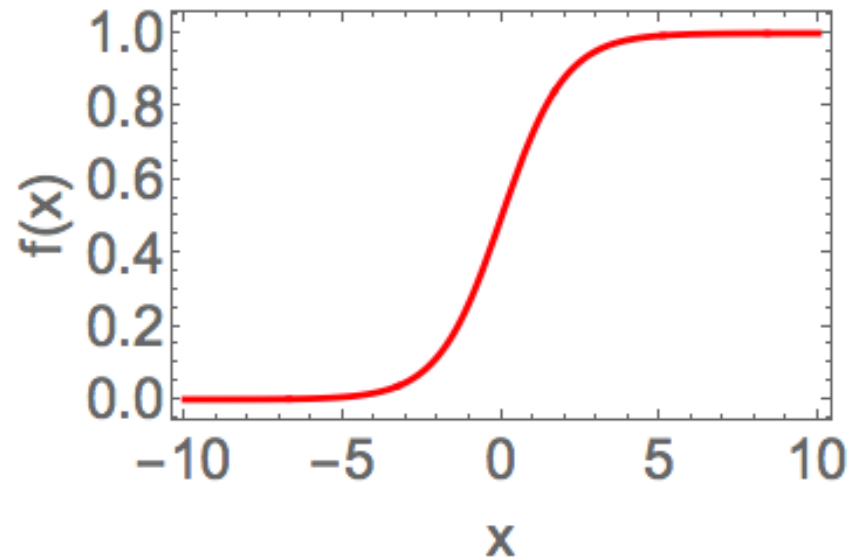
$$y_i = f(z_i) \text{ with } z_i = b_i + \sum_{k=1}^n w_{ik}x_k, \quad i = 1 \dots m, \quad (1)$$

where  $x_j$  is the input of the  $j^{\text{th}}$  input unit. In this case the propagation rule is the weighted sum over all inputs with weights  $w_{ik}$ .

## 3 Activation Function

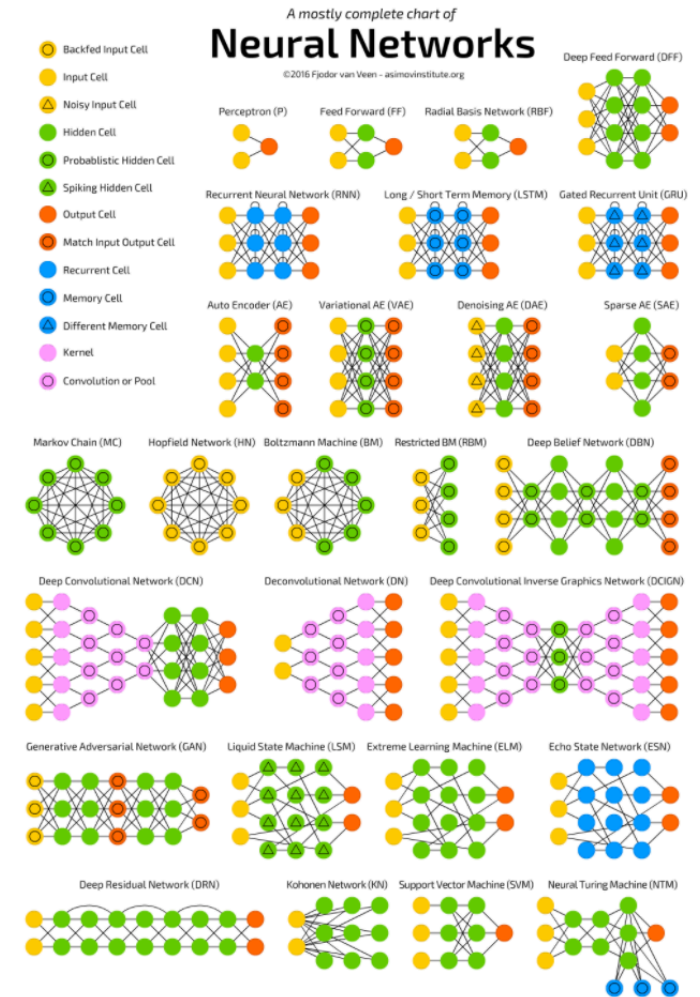
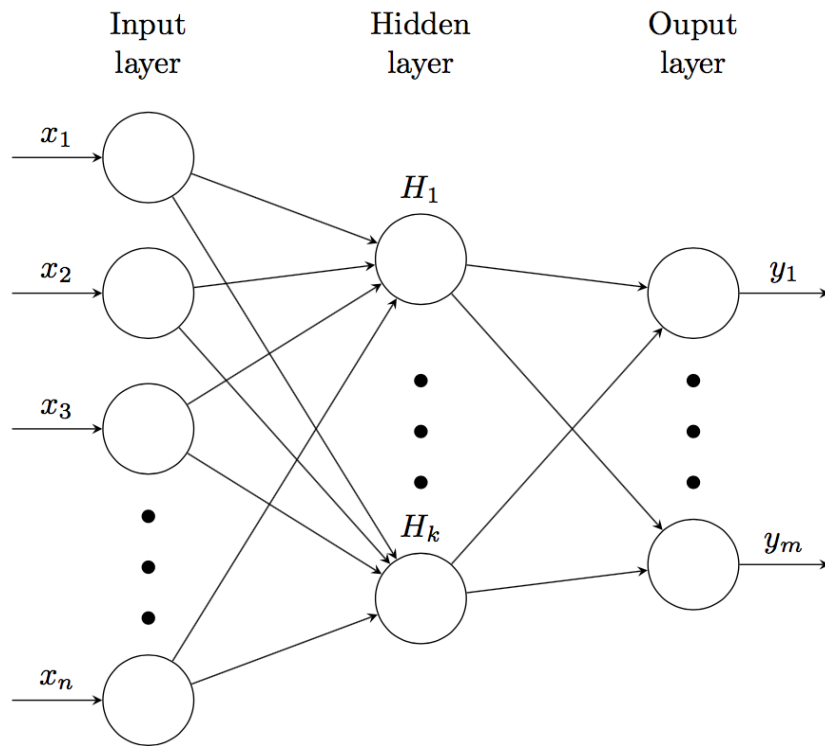
A sigmoid function is a commonly used s-shaped function. The logistic sigmoid is given by

$$f(x) = \frac{1}{1 + \exp(-x)}. \quad (2)$$



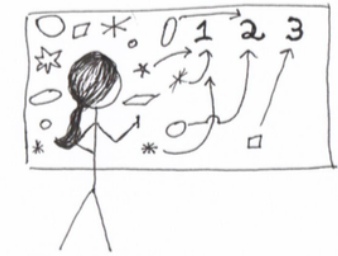


# A bit of definition & theory ...



## A bit of definition & theory ...

And now learning (training) ....



- **Unsupervised learning**
  - the training set provides only input values. The neural network has to find similarities by itself.
- **Reinforcement learning**
  - after processing an input value of the training set the neural network gets feedback whether the result is considered good or bad.
- **Supervised learning**
  - the training set provides both input values and desired target values (labeled training set).





# The Learning Process

For supervised learning tasks, the objective of training NNs is to minimise the errors between the desired output signal and actual output signal. This error can be typically defined as:

$$E(\mathbf{W}, \mathbf{b}) = \sum (\|\mathbf{y}(\mathbf{W}, \mathbf{b}, \mathbf{x}_t) - \mathbf{y}_t\|). \quad (6)$$

- $\mathbf{x}_t$  is a vector of input signals
- $\mathbf{y}_t$  is the desired output
- $\mathbf{W}$  is the weight matrix that is a combination of the input weight matrix, hidden weight matrix, and output weight matrix
- $\mathbf{b}$  is the bias vector
- $\mathbf{y}(\mathbf{W}, \mathbf{b}, \mathbf{x}_t)$  is the actual output signal for each neuron

The remaining task is to minimise  $E$ .

The most commonly used supervised learning algorithms for NNs include

- gradient descent and
- backpropagation (a special case of gradient descent).

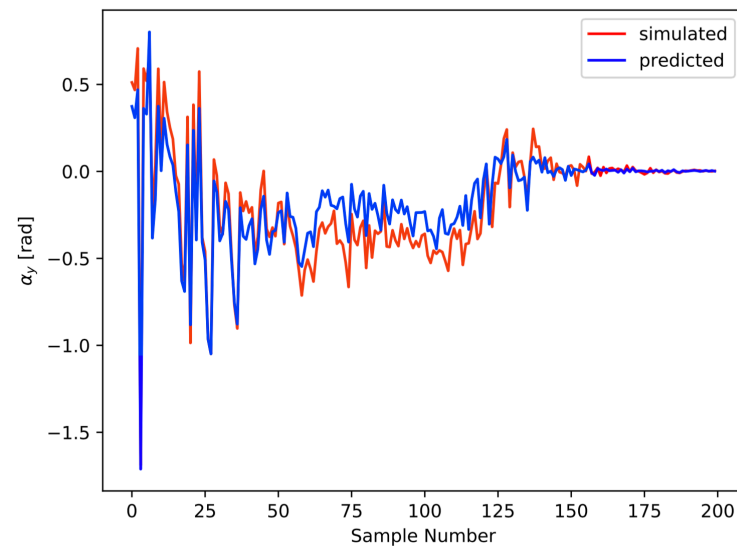
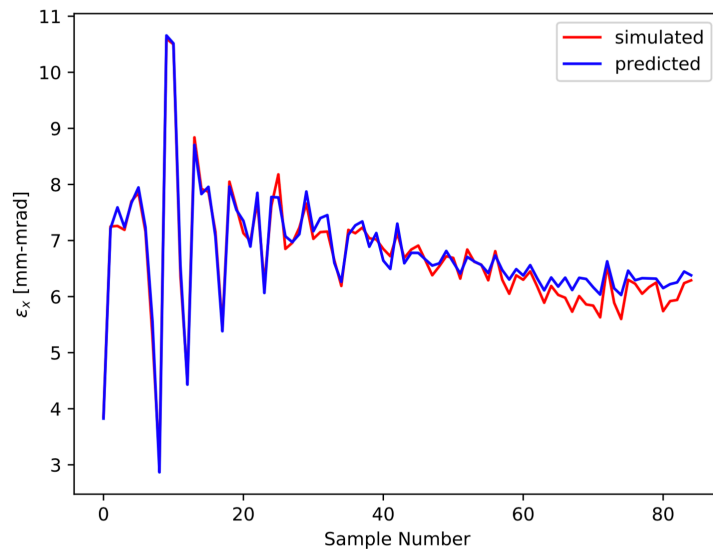
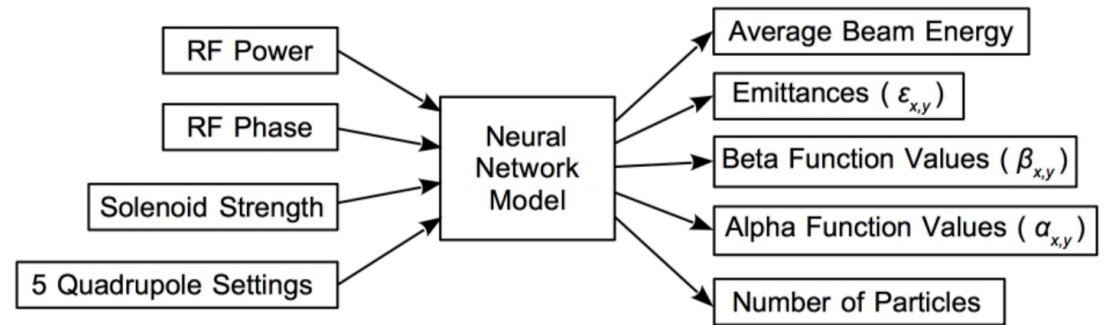
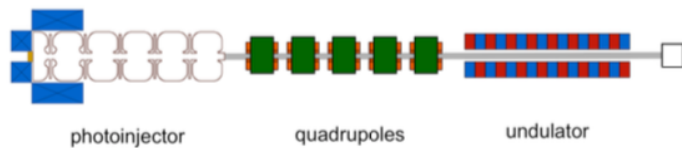
## Examples

- INITIAL EXPERIMENTAL RESULTS OF A MACHINE LEARNING-BASED TEMPERATURE CONTROL SYSTEM FOR AN RF GUN ([arXiv:1511.01883](#), [A.L.Edelen et.al](#))
- NEURAL NETWORK MODEL OF THE PXIE RFQ COOLING SYSTEM AND RESONANT FREQUENCY RESPONSE ([arXiv:1612.07237](#), [A.L.Edelen et.al](#))
- FIRST STEPS TOWARD INCORPORATING IMAGE BASED DIAGNOSTICS INTO PARTICLE ACCELERATOR CONTROL SYSTEMS USING CONVOLUTIONAL NEURAL NETWORKS ([arXiv:1612.05662](#), [A.L.Edelen et.al](#))
- USING A NEURAL NETWORK CONTROL POLICY FOR RAPID SWITCHING BETWEEN BEAM PARAMETERS IN AN FEL ([LA-UR-17-28069](#), [A.L.Edelen et.al](#))
  - The neural network architecture consists of four hidden layers containing 50, 50, 30, and 30 nodes, respectively. Each node in the hidden layers uses a hyperbolic tangent activation function and a dropout probability of 10%
- Construction of Hamiltonians by machine learning of energy and entanglement spectra ([arXiv:1705.05372](#), [H. Fujita et al.](#))

# USING A NEURAL NETWORK CONTROL POLICY FOR RAPID SWITCHING BETWEEN BEAM PARAMETERS IN AN FEL

A. L. Edelen et al. (LA-UR-17-28069)

Light with a wave-length that is tunable between 200  $\mu\text{m}$  and 800  $\mu\text{m}$ . It consists of a 5.5-cell, 1.3-GHz photocathode RF gun, a fixed-gap THz undulator  $K=1$



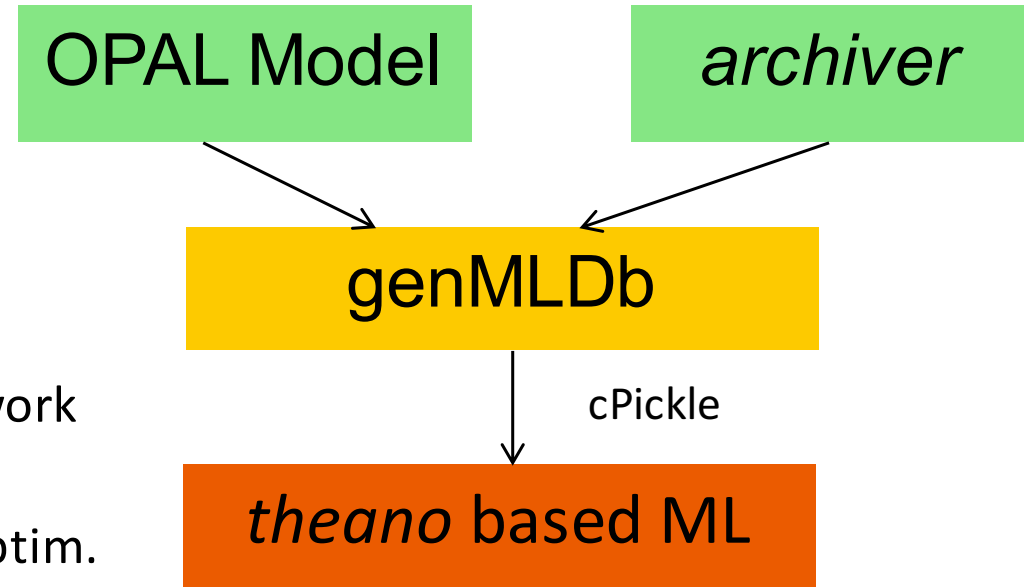


# Ideas for ML and HIPA (supervised learning)

1. Improving on mitigation/modeling of interlocks and (total) losses
  - even more hours to the experiments
  - less activation
  - protection of delicate devices such as electrostatic septa
2. Indirect predictions: model for spot size prediction on the SINQ target
  - use ionization monitors
  - other upstream data

# What we have done so far

- **Exchange data base** for allowing:
    - seamless scaling
    - easy interface to ML frameworks
  - Use A.L.Edelen *theano* based ML framework
    - **AWA linac**
      - learning data from OPAL GA-optim.
- 
- $$(RMS1,EMIT1,RMS2,EMIT2) = F(ISOL,RFPHASE,FWHM,K1,K3,K4,K5)$$
- HIPA
    - collection of learning data from the *archiver* (J. Snuverink)
      - from high resolution data taking just before 2017 shutdown





## **Machine learning and scientific computing on latest generation GPUs**

**Peter Messmer (NVIDIA)**

The latest generation NVIDIA GPUs offer a wealth of features for scientific computing and machine learning applications. In this presentation, I will introduce some of these features, including our TensorCores and NvLink and show how these technologies can efficiently be used for scientific computing.