Tutorial: Unsupervised Deep Learning

Harold Szu,
Research Ordinary Prof. BME, Catholic Univ. Am. Wash DC;
Fellows of IEEE, AIMBE, OSA, INNS, SPIE;
A Founder of INNS 1988,
Foreign Academician of Russian Academy of NL Sciences:#135, Jan 15, 1999 St. Petersburg;
Senior Scientist, Army Night Vision ESD, Ft. Belvoir, VA

Reference Websites:

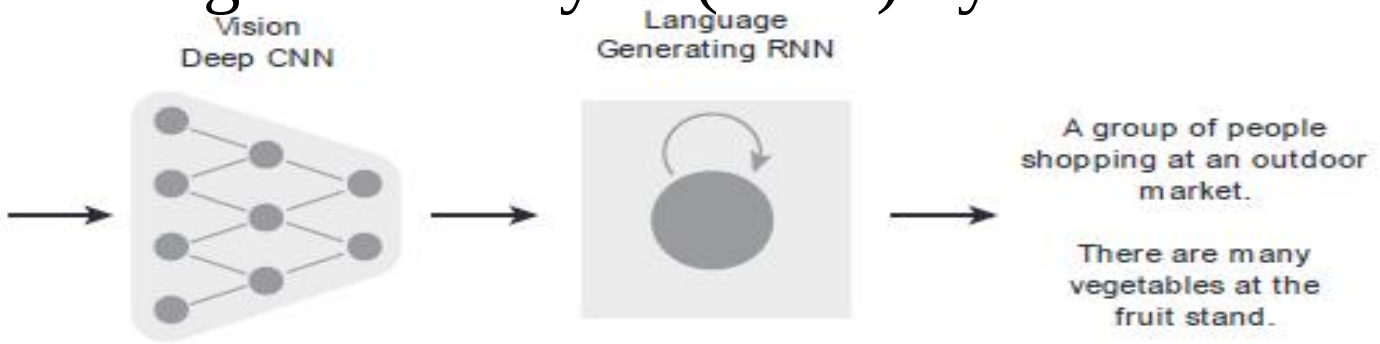
1. https://www.Researchgate.net/profile/Harold_Szu2
2. http://www.mathgeneology.org/harold_szu
3. http://www.ica-wavelet.org

Gulf Mexico Spring School (GMSS)
Sponsored by ONR, INNS, & IEEE/CIS
Tallahassee FL. April 15-18 ,2017;
Free web registration http://www.ica-wavelet.org
First come to sign up first serve up to 80 (qualified with extended summary to defray registration fee, & final paper will be due May 1st 2017 published by Elsevier 2018 May 15).

All begin with Mini-super MPD Computers !!! e.g. Graphic Processor Unit (GPU) for MPD ANN AI, adopted by Internet Giants Google, Face Book, YouTube for Big Data Analysis(BDA)

- ***What is AI? Alan Turing*** *introduced* ***AI*** *as that "one cannot tell the difference whether the other end of computer terminal is human or machine,"* *circa WWII.*
- ***Now the other end of computer can beat human in a chess game, face recognition , video perception.***
- **Google** co-founder ***Sergey Brin*** sponsored ***AI Alpha Go*** (围棋**)** was surprised by the intuition, the beauty & the communication skills displayed by Alpha Go **which** had beat Korean Go Grand Master ***Lee Sedol (***李世石**)** ***in*** 4 to 1 score on March 9-15, 2016, as millions watched on Internet**).**
- **YouTube** has applied AI **Deep Learning** to annotate videos automatically & discovered the favorite video on YouTube turns out to be about favorite pet Cats. (why? led to DARPA XAI)
- **Facebook** applied AI **Deep Learning wishes to achieve** aging & e-IQ independent facial expressions.
- BDA computing by GPU is supervised by the output classification, Supervised Deep learning.
- DARPA I2O Dir. David Gunning observed "effectiveness of AI will be limited by the machine's inability to explain its decisions and actions to users"; Explainable AI (XAI) RFP: Aug. 2016; May 2017+4 Yrs
- ***Emergent Technology with Gap:*** A Strategy for Explainable AI (XAI) follows how human understand their decision space and modeled it in Biological Neural Network (BNN) to ***separate 2 major blocks (1) Feature Extraction (FE) Deep Learning {V1 color (R,G,B) in 3 layers, V2 edge, V3 counter, V4 texture, & V5 contrast ;V6 emotion (GM Cells), V7 TBD} . (2) Machine Classification (MC) and scores of Output by feature histogram (not pixel histogram) that explain behind MC with FE.***

Big Data Analysis (BDA) by AI ANN on GPU Appendix



Yann LeCun (Facebook AI, NYU), Yoshua Bengio (Montreal) , Geoffrey Hinton (Google, Toronto) . “Deep Learning”, Nature 521, pp. 436-444,May 28, 2015

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood □or.

A stop sign is on a road with a mountain in the background

A little girl sitting on a bed with a teddy bear.

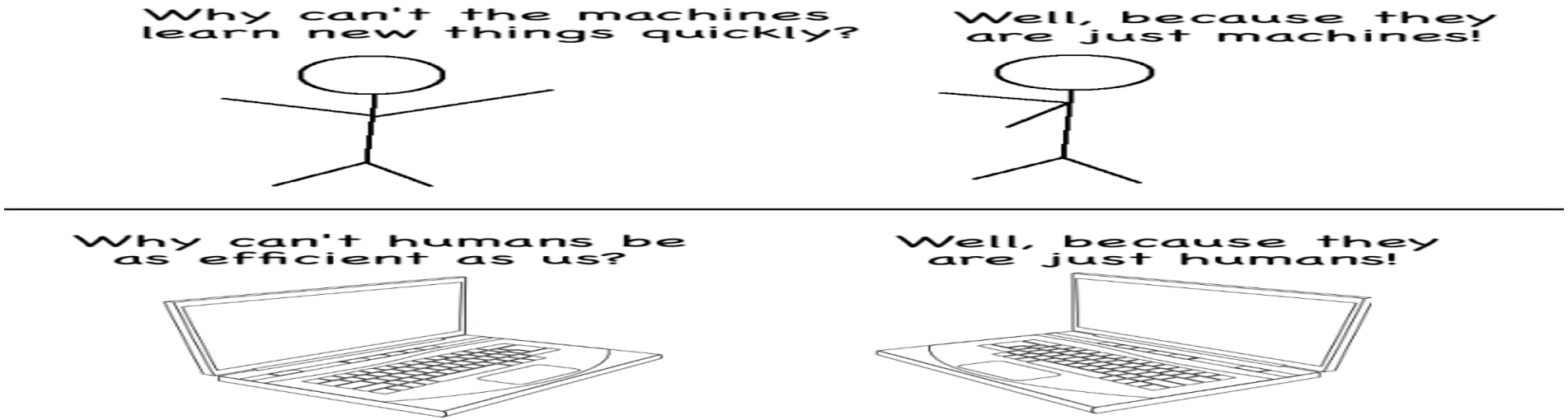
A group of people sitting on a boat in the water.

A gira□e standing in a forest with trees in the background.

Review:

1. How Graphic Processor Units (GPU) Mini-Super MPD Computers add wings to Artificial Intelligence (AI) .

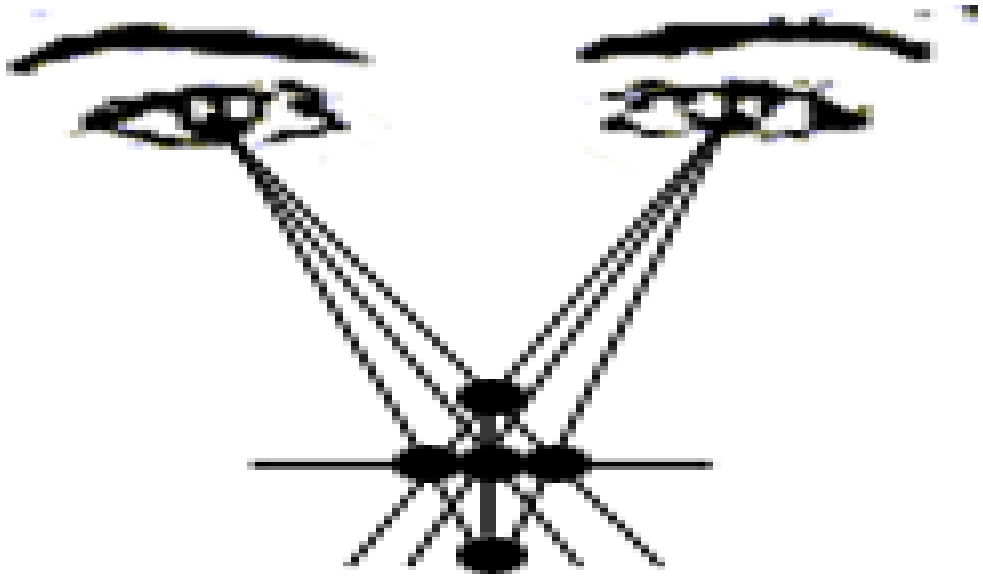
$M Nvidia made of 64 racks and each rack has 8 GPU in a small room size for air cooling. Current machine can support 100 layers & each layer with thousands by thousands millions nodes computed Artificial Neural Nets (ANN) in a pseudo-real time. Whether AI decision is explainable and trustworthy or not is DARPA I2O Dir. David Gunning $10M question as explainable AI, or XAI, for the next 4 years. In general, we concur that ANN for Man-Machine interface requires an **Explainable Computational Intelligence (ECI)** to be important for peaceful co-existence between Man and Machines (SciFi: Terminator);



2. Emerging AI Technology Gaps: Machine Autopilots work fine and dainty on the roads; though consumer report said that autopilots have been terrorized by some human drivers on the road. All the other AI applications may require Explainable Computational Intelligence (ECI), esp. for DoD Aided Target Recognition (AiTR).

3. How Natural Intelligence (NI) can fix it? We shall provide a success Biological Neural Nets (BNN) paradigm to endowed ECI with **two separate functional blocks (1) Feature Extraction (FE)** using the **Power of Sensor Pairs (PSP)** relaxing the noise energy at constant temperature HVS: V1~V7 Deep Learning at Cortex 17 (back of the head) and (2) the organized storage **Hippocampus Associative Memory (HAM) maps** features to categories **reversibly** as ECI.

Quiz: Why do we have two eyes, two ears, nostrils, all sensory pairs? Note that why receptors ears eyes nostrils input prefer pairs, while emitters are single mouth to talk and eat?



Answers:

(1)Hardware fault tolerance;

(2)Stereovision—no range info David Marr Binocular Paradox

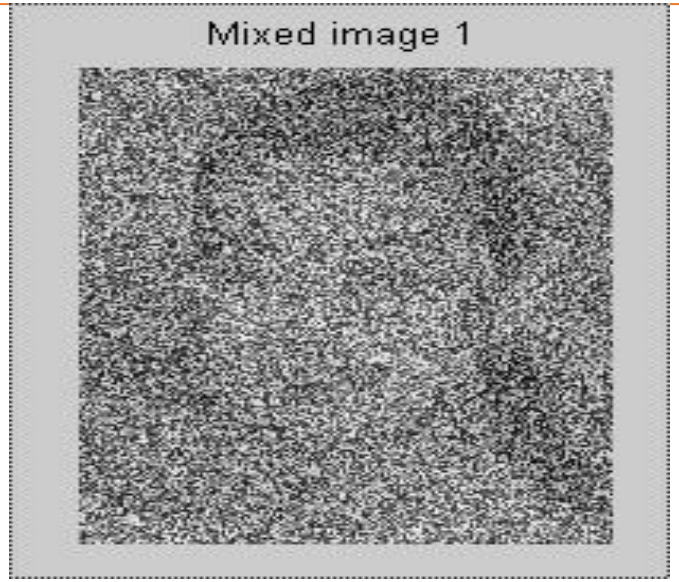
(3)Unsupervised learning by Coincidence Account

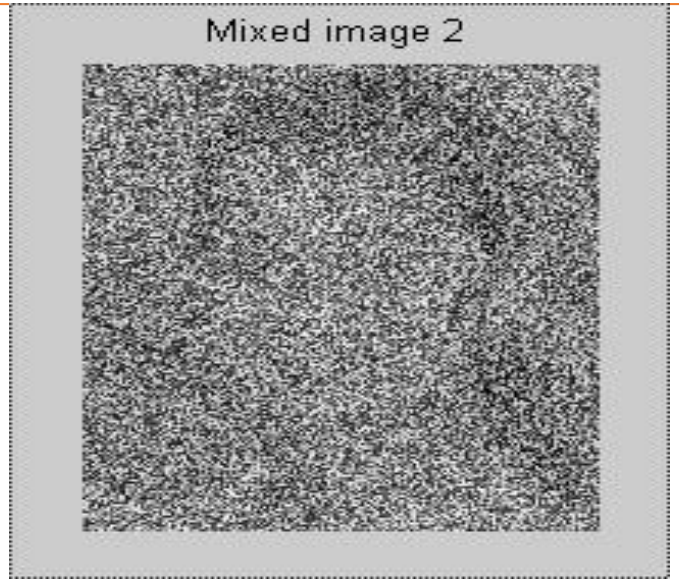
(4) All above (correct answer)

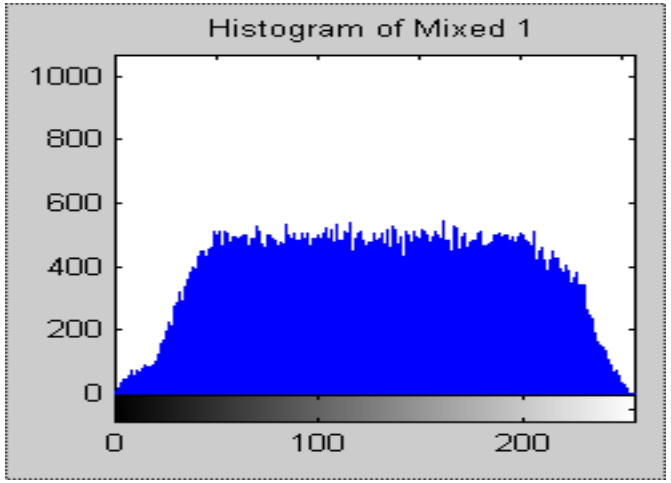
Two eyes could see through the fog---

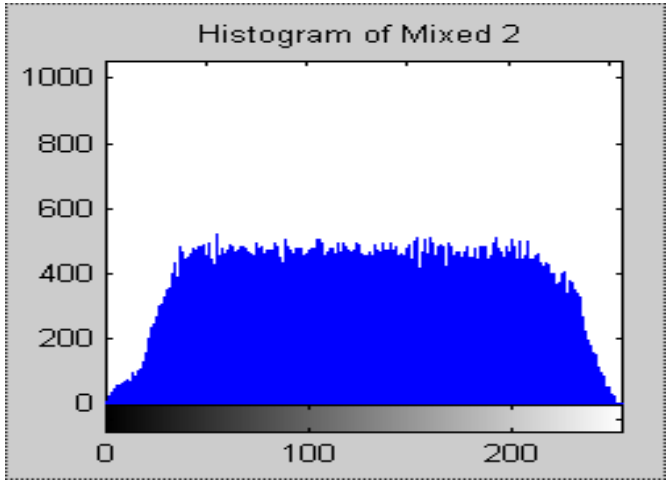
a perfect restoration among two eyes as opposed to one

eye image processing is merely re-shoveling foggy weather or snow!

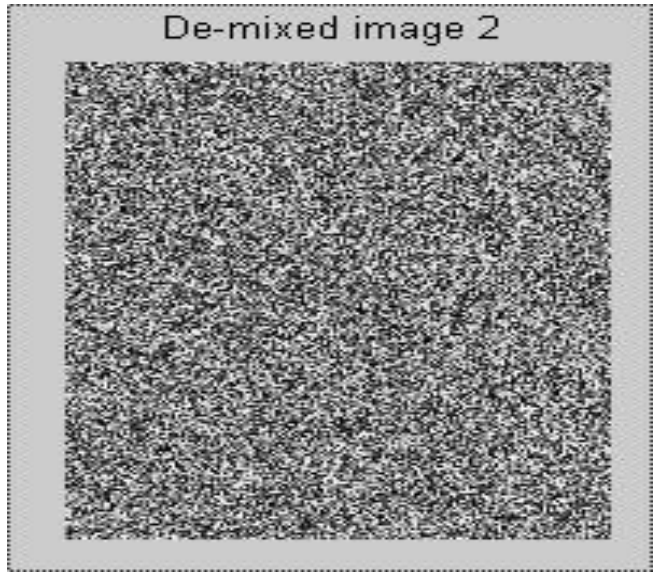


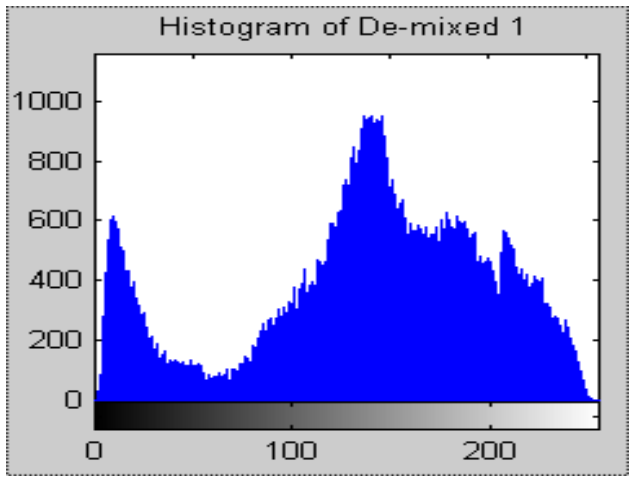


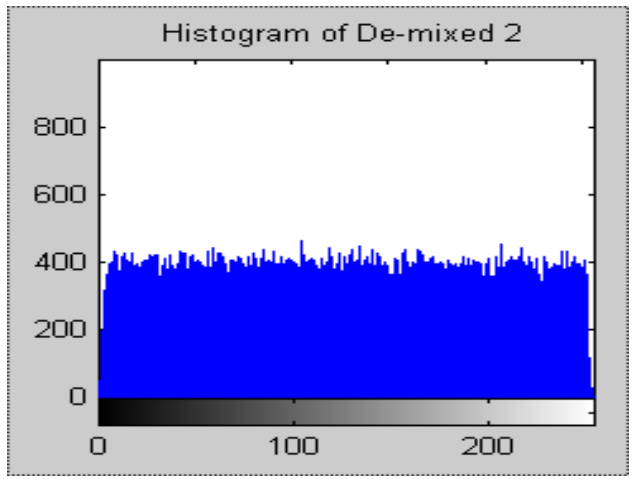




“ simple & fast image processing is “While the agreements must be the signals; the disagreements, the noises, that are universal and need no teacher!





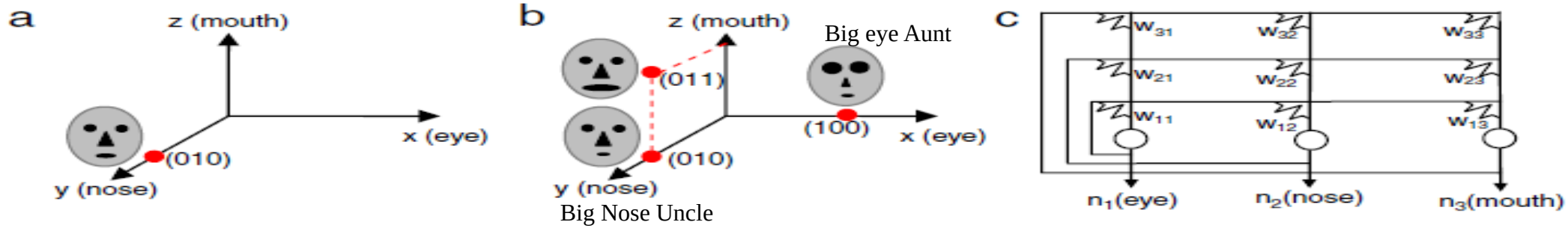


Tutorial: Computational Modeling for BNN

Architecture illustrates how BNN apply (1) **Power of Sensor Pairs** HVS V1~V7 Deep Learning **Feature Extraction (FE)** **by relaxing toward Homeostasis**; Then, (2) the features are stored into **Hippocampus Associative Memory [HAM]** defined by "**Write** by outer products" and "Read by **inner** product." Neuromorphic Chip implemented **static architecture** Massive Parallel Distributive (MPD) enjoying at least **Fault Tolerance Abstraction Generalization (FTAG).** **Then, Neurodynamics of Natural Intelligence for the survival follows the necessarily and sufficiently conditions (1) power of sensor pairs in (2) homeostasis brain.**

When a young girl met Uncle & Aunt, how does she recognize them; she will tell you it's the big eyes or big nose, explainable NI

Big Nose man open Big Mouth Laughs , Is he Uncle? *H. Szu et al. / Neural Networks 29–30 (2012) 1–7*



WRITE:

$$[AM]_{big\ eye\ aunt} = \overrightarrow{output} \otimes \overrightarrow{input} = \begin{bmatrix}1\\0\\0\end{bmatrix} \begin{bmatrix}1 & 0 & 0\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}$$

$$[AM]_{big\ nose\ unde} = \overrightarrow{output} \otimes \overrightarrow{input} = \begin{bmatrix}0\\1\\0\end{bmatrix} \begin{bmatrix}0 & 1 & 0\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$$

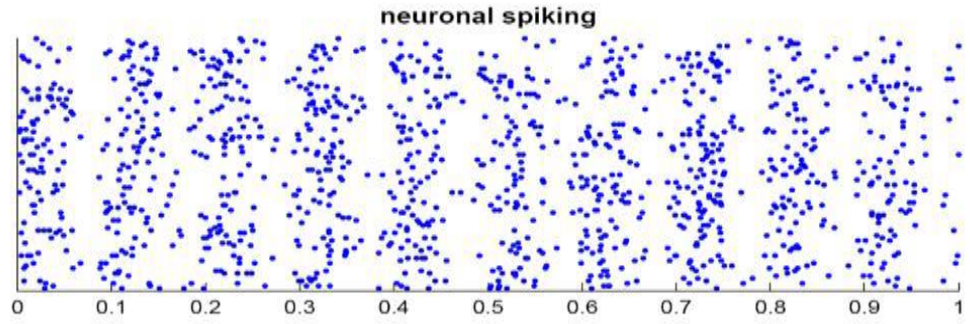
$$[AM]_{big\ nose\ unde} + [AM]_{big\ eye\ aunt} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}.$$

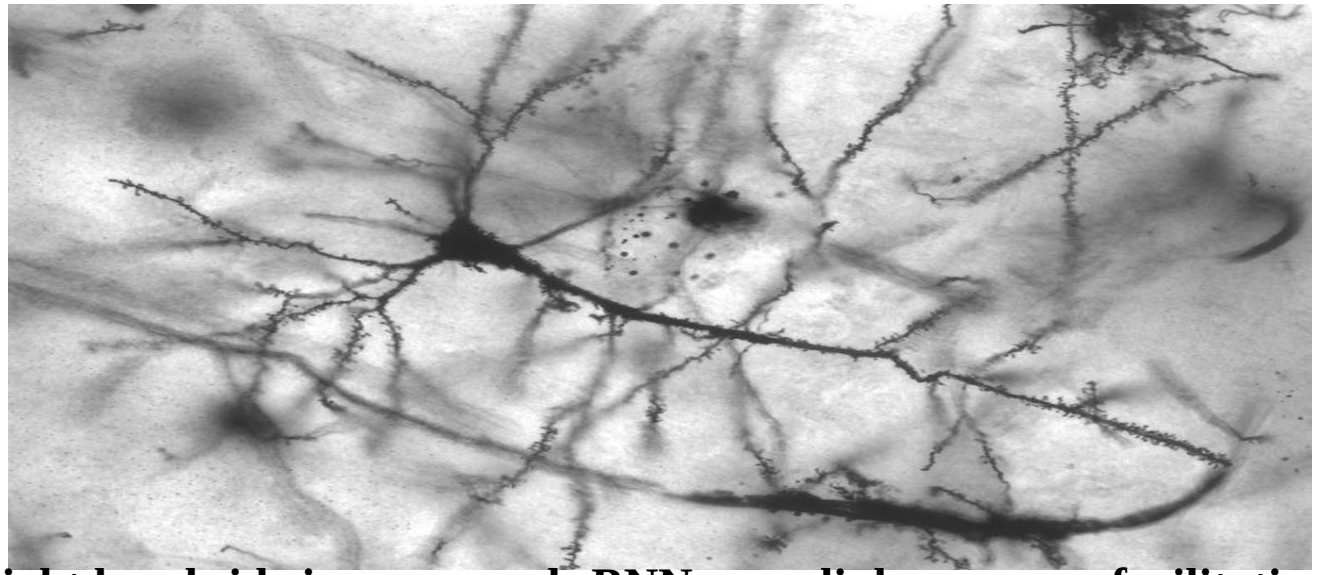
READ:

$$Recall\ Vector = [AM][error\ transmitted] \cong \sigma\left(\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\1\\0\end{bmatrix}_{uncle\ smiles}.$$

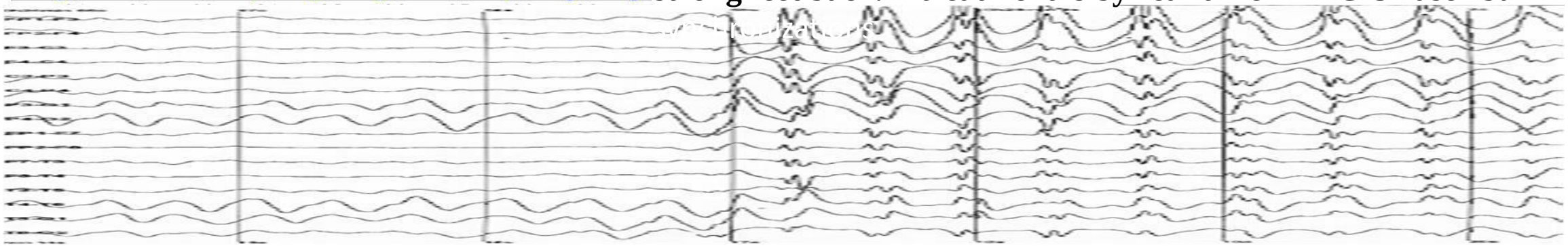
Caveat: Degree of Freedom (DoF) must be larger than Degree of information. Thus human brain is the least developed on the Earth only used 15% DoF . for yes or no neurons DoF is $O(2^{10^{10}})$ $1 \cong 100Hz$ firing; $0 \leq 50Hz\ firing$

Tutorial: Look deeper to learn deeper: What is Biological Neural Networks(BNN)? It takes two to dance tango. What are two? Elucidating EEG brainwaves at the onset of Epileptic Seizure may shed some light on BNN, due to neuron firing spiking population & local field potential for the 2nd order phase transition of Helmholtz Free Energy wearing f-EEG hat for always-on monitoring mental health, either a fast epileptic seizure or a slow Alzheimer dementia without Astrocytes Neuroglial cells work hard during good night sleep.





Right hand side image reveals BNN cross linkage axons facilitating too strong feedback. A treatment is by real-time f-EEG & laser burning



Epileptic seizures result from abnormal, excessive feedback gain instability of neuronal feedback

Missing half of Albert Einstein brain ? Ans: Neuroglial cells!

Gray matter (of 10^{11}neurons) and White matter (e.g. myelin sheath supported 10^{12} neuroglial cells about one tenth smaller in the size)

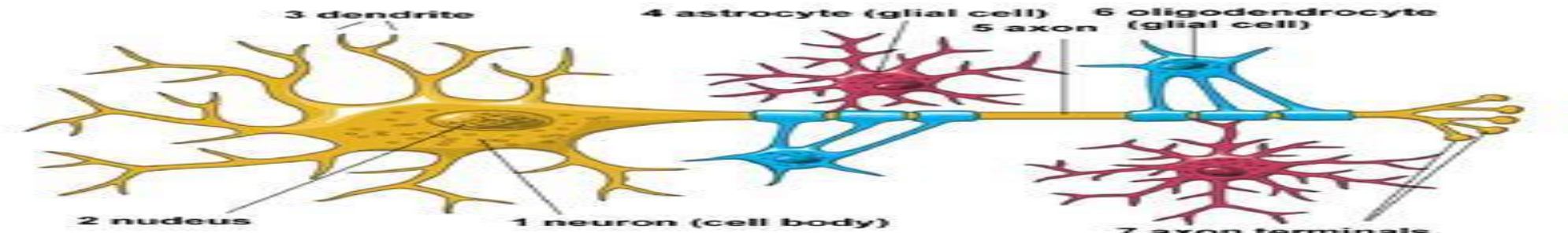
When Einstein passed away, his brain has been weighted to be 3 lb, like an ordinary human. Then, razor thin sections have been taken to investigate electrical conductivity & missed the electrically insulated neuroglial about 100 billions of neuroglial cells. Each glial is about 0.1 size of a neuron; electrical-silent, diligently house repairing and cleaning intercellular energy by-products. (NI involves glial cells, as mathematically derived first by Szu using power of pairs & Boltzmann heat death $\because \Delta S_{entropy} > 0$. supporting by the homeostasis at Herman Helmholtz **Min. Free Energy (MFE)**. Because the Ernst law of 3rd Thermodynamics, Kelvin temperature can never be zero (vacuum fluctuations). The incessant molecular collisions can homogenize the medium to more uniform composition (as mountain top rocks are eroded to beach sands with increasing entropy value without paleontology information).

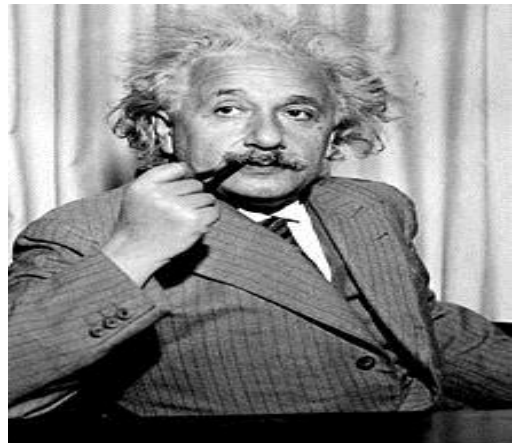
$$< T_o >= 37^o C = 310^o_K = constant$$

$\therefore \Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{entropy} \leq 0$; $\because$ D. O. Hebb bi-linear I/O learning rule;

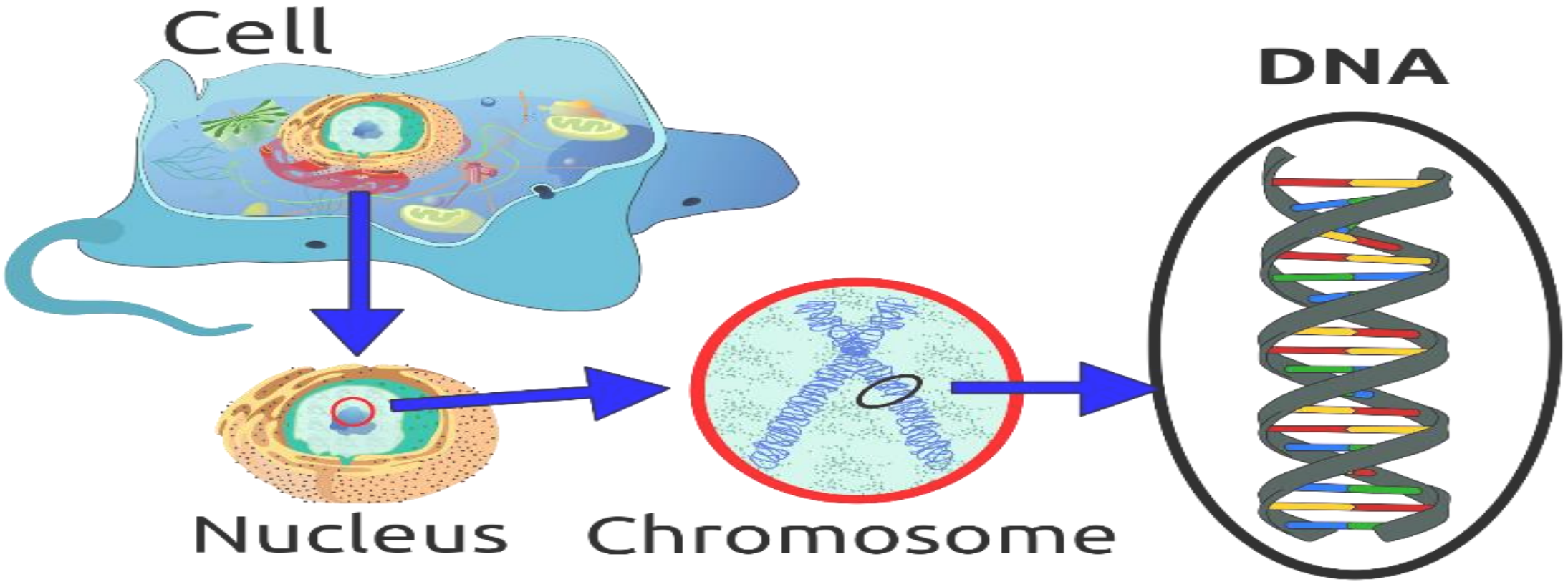
$$\therefore \overrightarrow{g_i} \equiv -\frac{\Delta H_{brain}}{\Delta \vec{D}endrite_i} \geq 0 \text{ ;where } \vec{D}endrite_{\ i} \equiv \sum_k [W_{i,k}] S_k$$

Quiz#1 Why Homosapiens $37^o C$; Quiz#2 Why Chicken $40^o C$



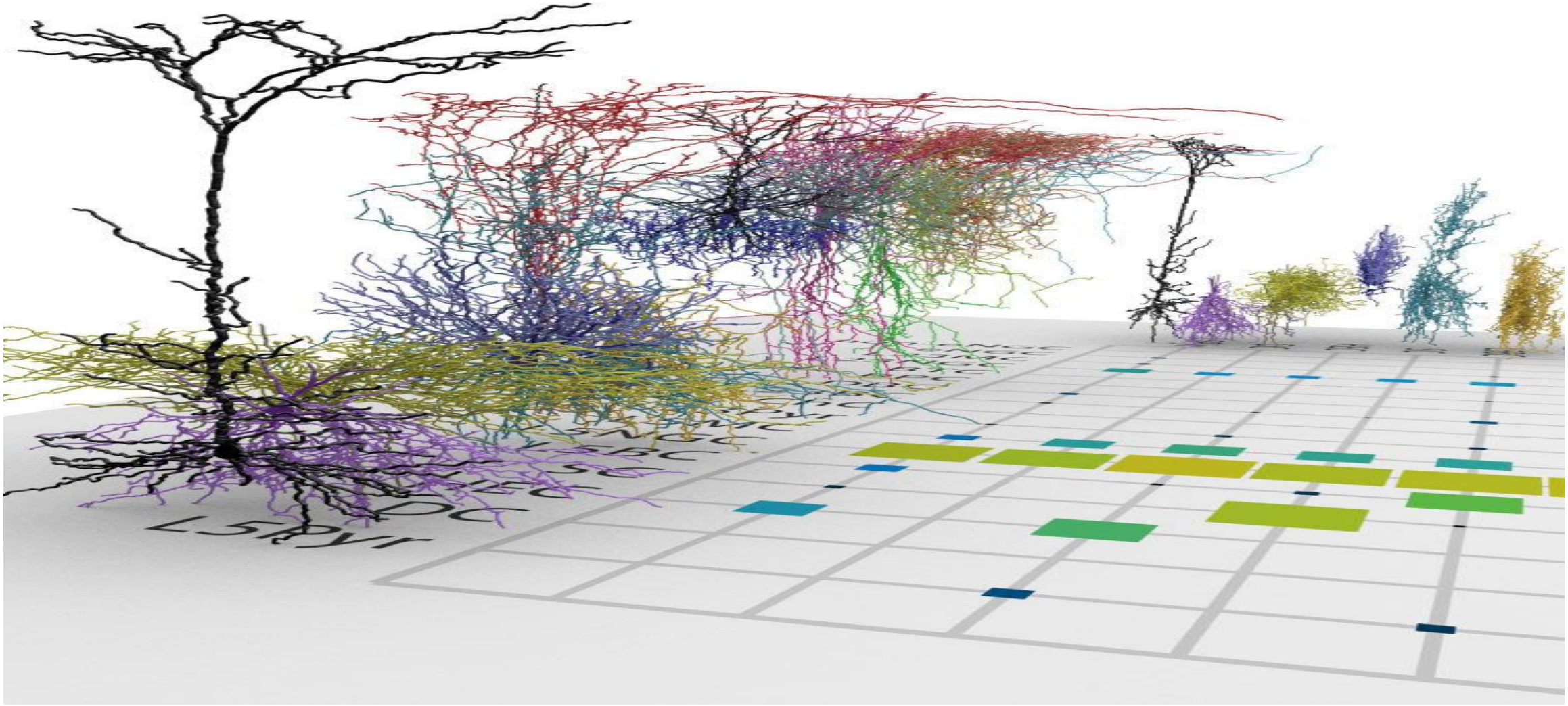


Glial cells are cells having all the necessary genetic epigenetic
Every neuron has 2 to 4 silent servant glial cells. Both are cell defined as

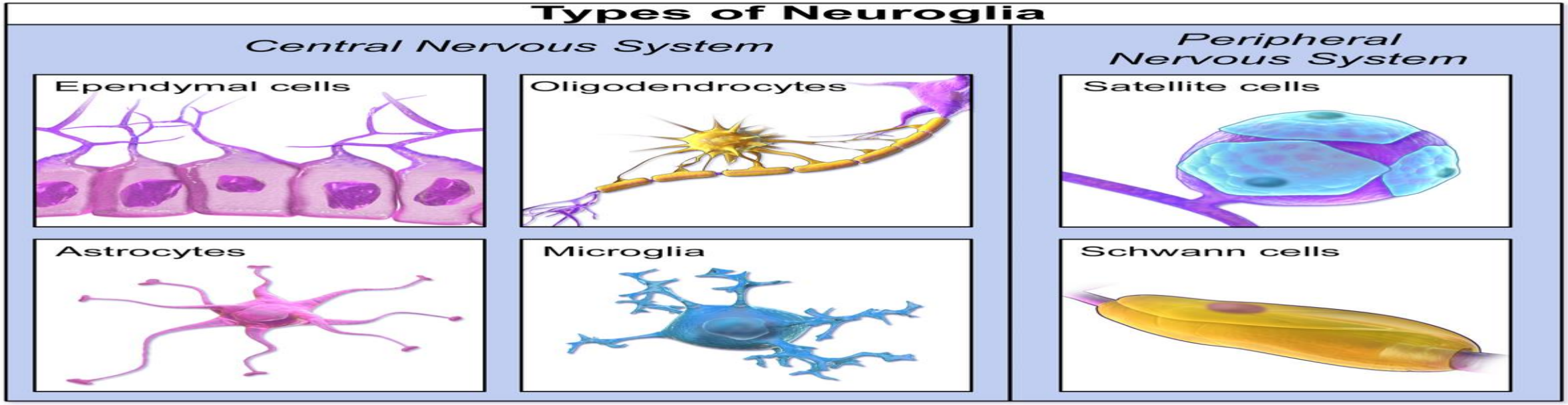


Gray Matter Neurons

Xiaolong Jiang and Andreas Tolias at Baylor College of Medicine in Houston announced six new types of 15 adult mice brain cells by the method of slicing razor-thin slices of mature brain. (William Herkewitz, Science Nov 26, 2015) that establishing a complete census of all neuron cell types is of great importance in moving the field of neuroscience forward," says Tolias, at Baylor College of Medicine



4+2=6 neuroglial cells:
Missing half of Einstein brain are major Six Glial (Greek: Glue) Cells.
(4 CNS: Ependymal, Oligodendrocytes, Astrocytes, Microglia; 2 PNS: Satellite, Schwann cells)

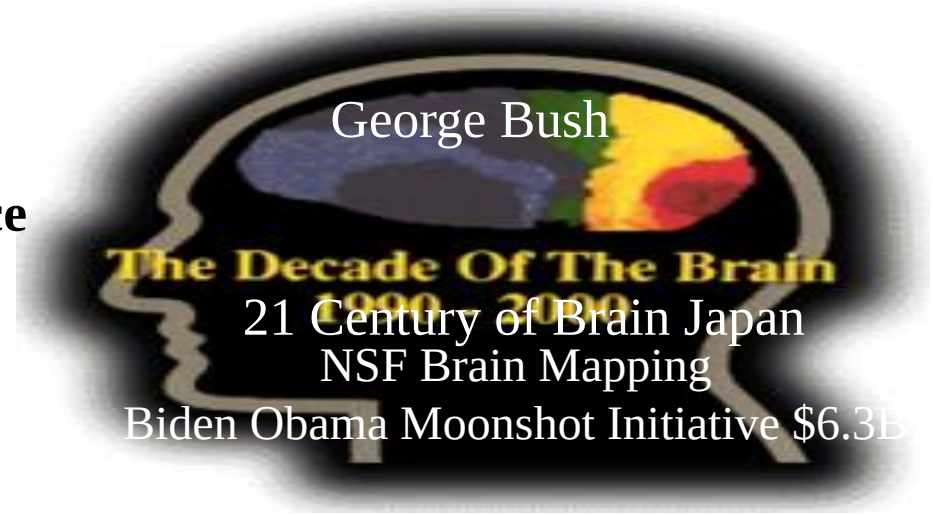


This was around same time that cell counts in the brain revealed glial cells to be nearly 90% of the brain including the white matter my sheath (we only use 10% of our brain comes from communication mediated by electromagnetic pulses. **Santiago Ramon Cajal**, Neuron Doctrine received the **Nobel Prize 1906**, while his brother **Pedro Cajal** supported Astrocytes type of **glial cells** supporting neuro transmitters that can ignite astrocyte calcium release.

The necessary and sufficient conditions of Natural Intelligence (NI),

iff 1. Power of Pairs $\vec{X}(t)$ irreversible dynamics "**Agree, Signal; Disagree, Noise;" relaxing naturally increasing the Entropy S as a measure of uniformity of firing rates,**
2. Homeostasis at Minimum Free (Helmholtz)Energy (MFE)
$\Delta H_{Brain} = \Delta E_{Brain} - T_o \Delta S_{Brain} \leq 0$ **(derived in the next page)**

5 sensor pairs co-incidence account: "Agree, Signal; Disagree, Noise;" ABSTREACTION

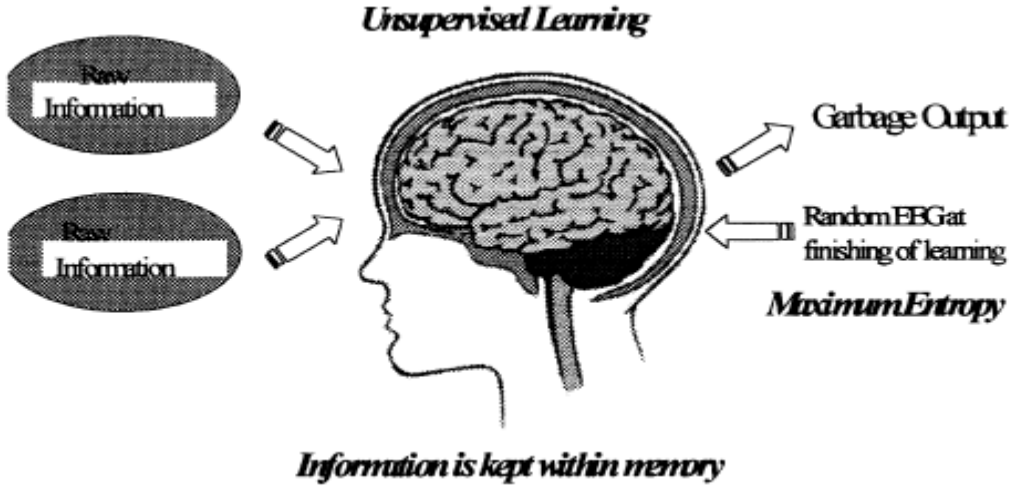


Cortex 17 (V1~V6 deep layers learning for Feature Extraction (color, edge, contour, texture, etc.) followed with Hippocampus Associative Memory recalls GENERALIZATION

Biological Neural Networks (BNN) of Homosapiens consisting of 10^{th} billions of neurons & 100^{th} billions of glial cells building the Natural Intelligence(NI) for unsupervised deep learning (UDL) spontaneity & de-confliction by elimination to focus on survival options

Unsupervised Learning Cost Function MFE

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \leq 0$$



Boltzmann Equilibrium Thermodynamics

$$S_{tot} \equiv k_B \, Log \, W_{tot}; W_{tot} \equiv \exp\left(\frac{S_{tot}}{k_B}\right) = \exp\left(\frac{(S_{brain} + S_{env.})T_o}{k_B T_o}\right) = \exp\left(\frac{S_{brain} T_o - E_{brain}}{k_B T_o}\right)$$

$$= \exp\left(-\frac{H_{brain}}{k_B T_o}\right) \equiv W_{MB}$$

Maxwell Boltzmann Canonical Probability is the statistical equilibrium is known as Homeostasis

$$H_{brain} \equiv E_{brain} - T_o S_{brain}$$

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain}$$

Boltzmann heat death $\Delta S_{brain} > 0$ due to incessant collisional mixing towards more uniformity (Inequilibrium Thermodynamics)

$$\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \leq 0$$

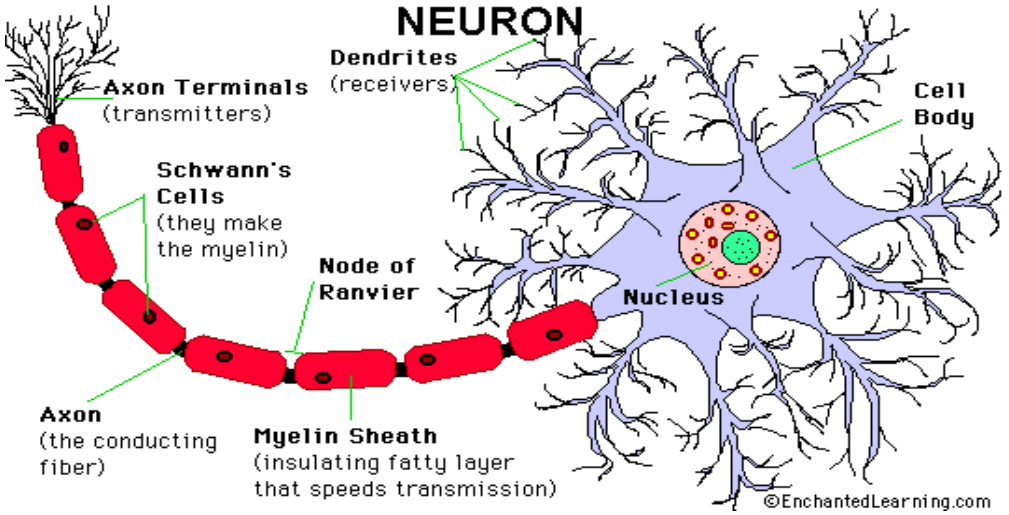
Stable Brain Dynamics is necessary for robustness.

MFE: $\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain} \leq 0$

Aleksandr Lyapunov

https://en.wikipedia.org/wiki/File:Alexander_Ljapunow_jung.jpg

June 6, 1857,
November 3, 1918,
St. Petersburg State University



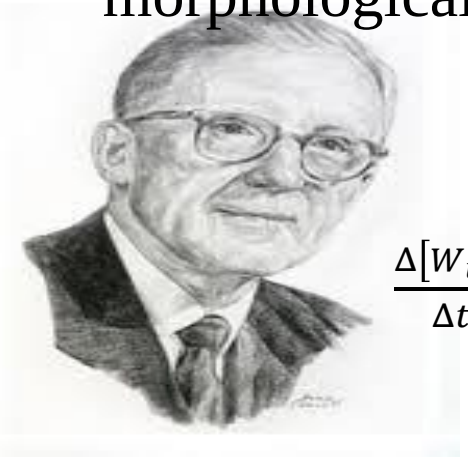
$$\frac{\Delta H_{brain}}{\Delta t} = \left(\frac{\Delta H_{brain}}{\Delta [W_{i,j}]}\right)\frac{\Delta [W_{i,j}]}{\Delta t} = -\frac{\Delta [W_{i,j}]}{\Delta t}\frac{\Delta [W_{i,j}]}{\Delta t} = -\left(\frac{\Delta [W_{i,j}]}{\Delta t}\right)^2 \leq 0$$

we have derived the MFE Gradient Force in consistent with the Neurodynamics

$$\frac{\Delta [W_{i,j}]}{\Delta t} = -\frac{\Delta H_{brain}}{\Delta [W_{i,j}]}$$

A half century ago, Canadian Biologist Donald. O. Hebb (1904-1985) observed the synaptic weight Learning Rule :

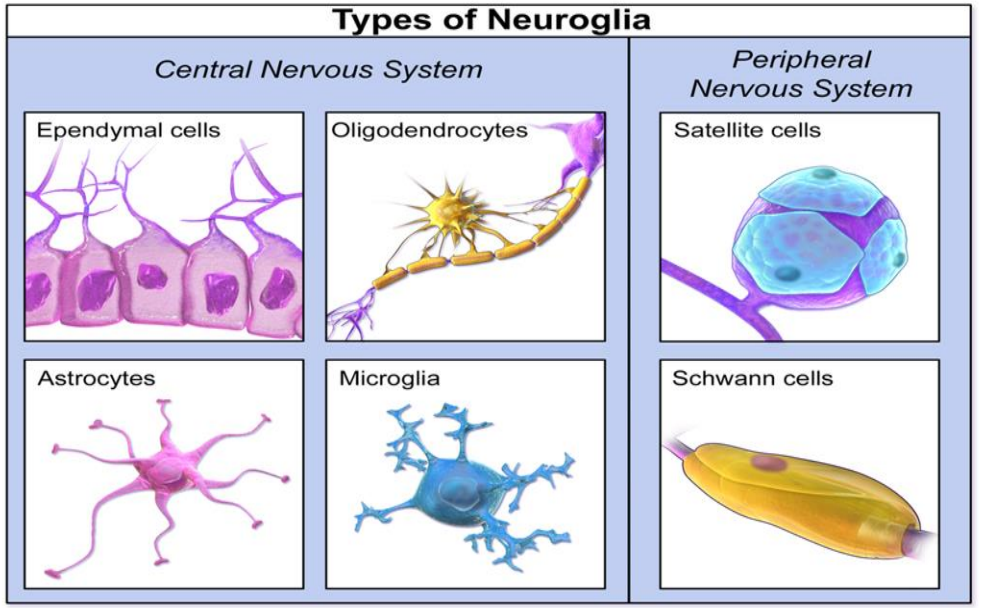
Dendrite net defines input vector glial cells that are morphologically dependent on cell typing



$$\overrightarrow{Dendrite_j} \equiv \sum_i [W_{i,j}]\vec{S'}_i$$

$$\frac{\Delta[W_{i,j}]}{\Delta t} \equiv \left(-\frac{\Delta \mathrm{H}_{brain}}{\Delta[W_{i,j}]}\right) = \left(-\frac{\Delta \mathrm{H}_{brain}}{\Delta \overrightarrow{Dendrite_j}}\right)\frac{\Delta \overrightarrow{Dendrite_j}}{\Delta[W_{i,j}]} \approx \vec{\mathrm{g}}_j \vec{S'}_i$$

$$\vec{\mathrm{g}}_j \equiv -\frac{\Delta \mathrm{H}_{brain}}{\Delta \overrightarrow{Dendrite_j}}$$ (Szu 2017 US PTO)



We have derived for the first time the mathematical definition of glial cell (discovered by pathologist **Rudolf Virchow *1856).*** This math definition of glial cells explains the unified theory of all different neuron typing about 15 in mice, morphology predicts different Dendrite vectors defining different kinds of Glial Cells, led to all kinds of Glial cells (4 in CNS: Ependymal, Oligodendrocytes, Astrocytes, Microglia; 2 in PNS: Satellite, Schwann cells). $\Delta[W_{i,j}] = [W_{i,j}(t+1)] - [W_{i,j}(t)] = \vec{\mathrm{g}}_j \vec{S'}_i \eta$

Rosette Stone 2016
Math Theory of Deep Learning in BNN NI or ANN AI

$$\text{NI glial } g_i \equiv -\frac{\partial \mathrm{H}}{\partial Dendrite_i} \qquad \text{AI delta } \delta_i \equiv -\frac{\partial \mathrm{LMS}}{\partial net_i}$$

$$Dendrite_j \equiv \sum_i [W_{j,i}] S_i; \qquad net_j \equiv \sum_i [W_{j,i}] O_i;$$

$$S_i = \sigma(Dendrite) \qquad O_i = \sigma(net_i)$$

Einstein at Princeto

Supervised & Unsupervised Deep Learning

What is the Cost Functions for supervised and unsupervised DL ?

- ***Supervised DL*** utilizes **the *Least Means Squares (LMS)*** errors for ***Artificial Intelligence (AI), ANN learnable relational databases***;

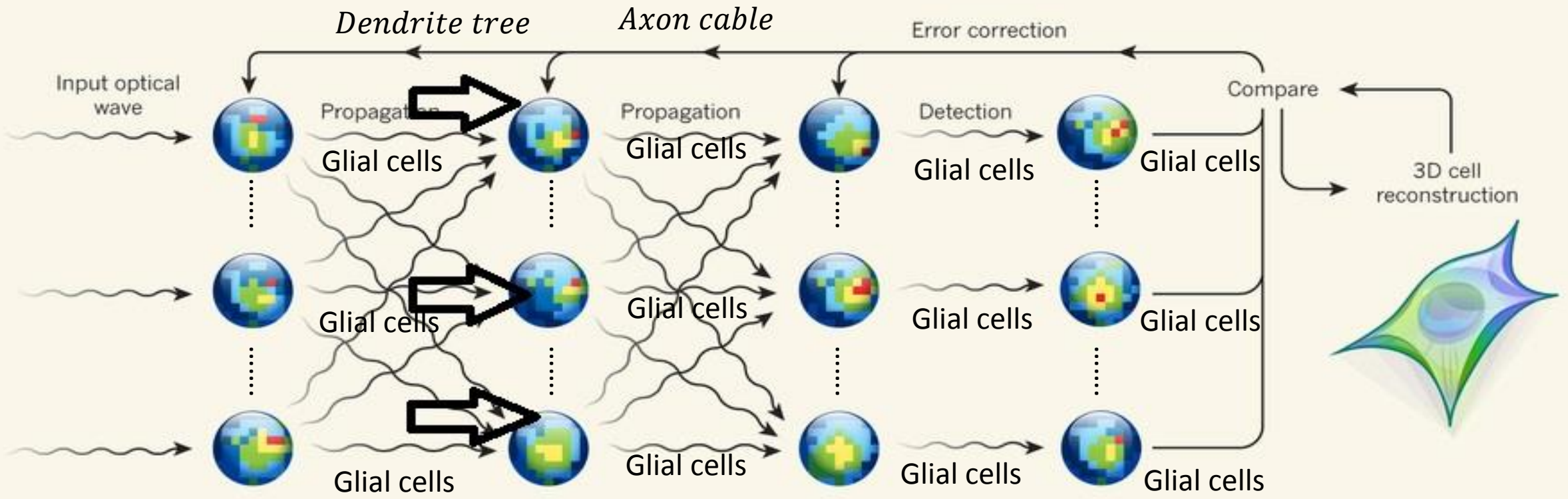
- ***Unsupervised DL*** utilizes ***the Minimum Free Energy (MFE)*** at ***BNN at*** Helmholtz ***MFE*** for ***Natural Intelligence (NI), if and only if***
- ***(i) Isothermal Brain (homosapiens at 37^0C (optimum for hemoglobin elasticity); while Chicken 40^oC (for egg hatching); but lacking of big thumb for holding tools becomes less intelligent than homosapiens (we ate them, not vice versa, Q.E.D.).***
- ***(ii) Power of Pairs (eyes, ears, nostrils, taste buds, tactile pre-processing "agree, signal; disagree, noise"): BNN Learning*** $[W_{i,j}]\vec{X}_{in,pair}(t) = \vec{S}_{out,fusion}(t)$

AI LMS Error cannot separate figure from the ground

NI Human Target Recognition must be able separate binary Figure and Ground under the dusk dim light far away. This could be any simple Ambiguity Figures for computational simplicity. The idea of NI in BNN for the survival is manifested clear in "$\vec{F}$ =Tigress" & Ground "$\vec{G}$ =Tree". In contrast, the supervised cost function is LMS AI based on ANN becomes ambiguity of binary figure & ground Least Mean Squares (LMS) cost function $|(\vec{F}-\vec{G})^2|=|(\vec{G}-\vec{F})^2|$could not separate to run away for the survival of the species due to the switch of the algebra sign. However, higher order of moment expansion of MFE can separate the tiger and tree in the remote dime light for the survival of Homosapiens.

The linear term can already tell the difference between the target lion versus the background tree, $(0-1)\neq(1-0)$without suffering the LMS parity invariance: $(0-1)^2=(1-0)^2$

Biological Neural Net (BNN), weighted dendritic sum (bold arrow vector input to a neuron $Dendrite_i = \sum_j [\mathrm{W}(\epsilon t)_{i,j}] \vec{S}_{j,out}(t))$ and electrically neural Glial Cells housekeeping servant cells supported real-time co-axial Axon cable output
e.g. HVS Cortex 17 has lavers V1~V4



Analytic Proof of Glial Cells

Taylor Expansion defines the negative slope as the glia cells due to Biologist D.O. Hebb learning rule circa 1950

$$H_{brain} = H_o\left(\vec{X}(t_o)\right) + \sum_{i=1}^{N'}\left(\frac{\partial H_{brain}}{\partial S_i}\right)_o\{S_i - S_i(t_o)\} + \sum_{j}\sum_{i=1}^{N'}\left(\frac{\partial^2 H_{brain}}{\partial S_j \partial S_i}\right)_o\{S_i - S_i(t_o)\}\{S_j - S_j(t_o)\} + O(|\Delta\vec{S}|^3)$$

$$H_{brain} = H_o + \vec{g}_{glial}\{[W]\vec{X}_o - \vec{S}\} + \frac{\partial \vec{g}_{glial}}{\partial \vec{S}} * \{[W]\vec{X}_o - \vec{S}\}\{[W]\vec{X}_o - \vec{S}\}$$

- **D.O. Hebb observed bi-linear learning rule (dendritic Input glial, g_j, -MFE slope, output Axon firing rate S_i** MFE itself):

$$\frac{\partial[W_{ji}]}{\partial t} = -\frac{\partial \mathrm{H}}{\partial[W_{ji}]} = -\frac{\partial \mathrm{H}}{\partial Dendrite_j}\frac{\partial Dendrite_j}{\partial[W_{ji}]} = = g_j\, S'_i$$; where *bi-linear Hebb rule follows:*

$$\text{Input } g_{glialj} = -\frac{\partial \mathrm{H}}{\partial Dendrite_j}; output \quad \frac{\partial Dendrite_j}{\partial[W_{ji}]} = \frac{\partial}{\partial[W_{ji}]}\sum_i [W_{j,i}]S'_i = S'_i\,.$$

Sigmoid Logic & Exact solution of Control Ricatti Equation

Single layer Input $\left[\mathrm{W}(\epsilon \mathrm{t})_{i,j}\right]\vec{X}_{j,in}(t) = \vec{S'}_{i,out}(t)$

$Multiple\ layer\ output\ \vec{S'}_{i,out}(t) \Longrightarrow Dendrite_i = \sum_j \left[\mathrm{W}(\epsilon \mathrm{t})_{i,j}\right]\vec{S}_{j,out}(t)$

$$\boldsymbol{Sigmoid\ Logic}\ \boldsymbol{\sigma(H^*_{brain} - H_{brain})} = \frac{\boldsymbol{exp\left(\frac{-H^*_{brain}}{k_B T_o}\right)}}{\boldsymbol{exp\left(\frac{-H^*_{brain}}{k_B T_o}\right) + exp\left(\frac{-H_{brain}}{k_B T_o}\right)}} = \frac{\boldsymbol{1}}{\boldsymbol{1 + exp\left(\frac{H^*_{brain} - H_{brain}}{k_B T_o}\right)}}$$

$$\boldsymbol{H^*_{brain} - H_{brain}} = Dendrite_i = \sum_j \left[\mathrm{W}(\epsilon \mathrm{t})_{i,j}\right]\vec{S}_{j,out}(t)$$

$$\sigma(Dendrite_j) = \overrightarrow{S'}_j;$$

$$\boldsymbol{\frac{dy}{dx} = \frac{d}{dx}(1 + exp(x))^{-1} = -1((1 + exp(x))^{-2}([exp(x) + 1] - 1) = y^2 - y = y(y - 1)};$$

$$\sigma'_j(Dendrite_j) = \overrightarrow{S'}_j\left(\vec{S}'_j - 1\right)$$

is a limiting window function [0, 1] 1 for 100 Hz

Back Prop of MFE slopes glial cells

$$g_j \equiv -\frac{\partial H}{\partial Dendrite_j} = -\frac{\partial H}{\partial S_j}\frac{\partial S_j}{\partial Dendrite_j} = -\frac{\partial H}{\partial S_j}\sigma_j'(Dendrite_j)$$

$$-\frac{\partial H}{\partial S_j} = -\sum_k \frac{\partial H}{\partial Dendrite_k}\frac{\partial Dendrite_k}{\partial S_j}$$

$$= -\sum_k \frac{\partial H}{\partial Dendrite_k}\frac{\partial}{\partial S_j}\sum_i [W_{k,i}]S_i = -\sum_k \frac{\partial H}{\partial net_k}[W_{k,j}] = \sum_k g_k [W_{k,j}]$$

$$g_j = \sigma_j'(net_j)\sum_k g_k [W_{k,j}]$$

Q.E.D.

MFE $\Delta H_{brain} \equiv \Delta E_{brain} - T_o \Delta S_{brain}$; & Irreversible $\Delta S_{brain} > 0$

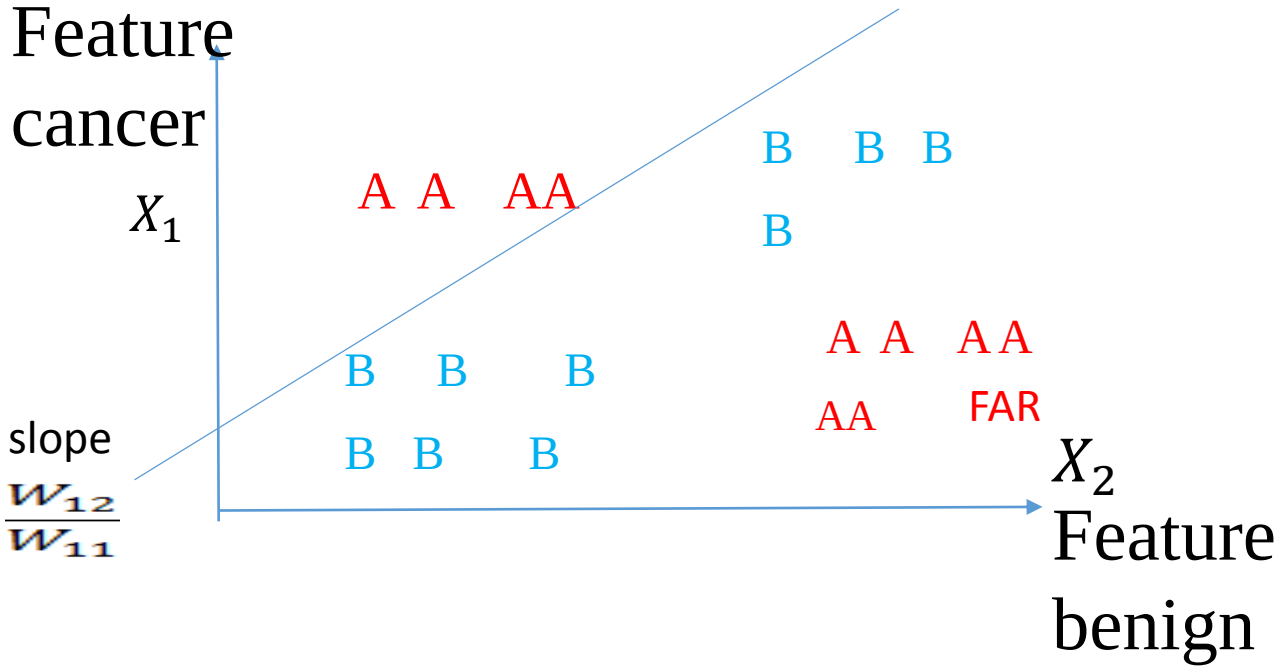
BNN, Natural Intelligence (NI) & Deep Learning (Multiple Layers) Back Prop Learning Rule:

$$\left[W_{ji}(t+1)\right] = \left[W_{ji}(t)\right] + \eta\, \vec{\mathrm{g}}_j \vec{S}_i' + \alpha_{momtum} \left[W_{ji}(t) - \left[W_{ji}(t-1)\right]\right]$$

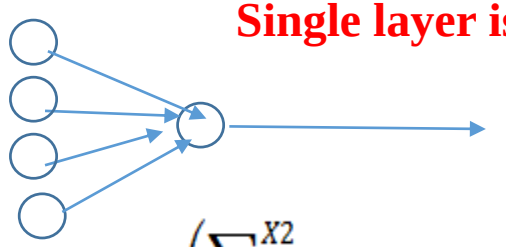
Q.E.D.

where ad hoc momentum term is to bypass the Mexican standoff local minimum

Why Deep (Layers) Learning (DL)? We begin with a single layer ANN:



Single layer is Linear classifier:



$$S_i = \sigma\left(\sum_{j=X1}^{X2} W_{ij}\, X_j - \theta_i\right) \geq 0$$

$$X_1 W_{11} + X_2 W_{12} \geq \theta_1$$

$$X_1 W_{21} + X_2 W_{22} \geq \theta_2$$

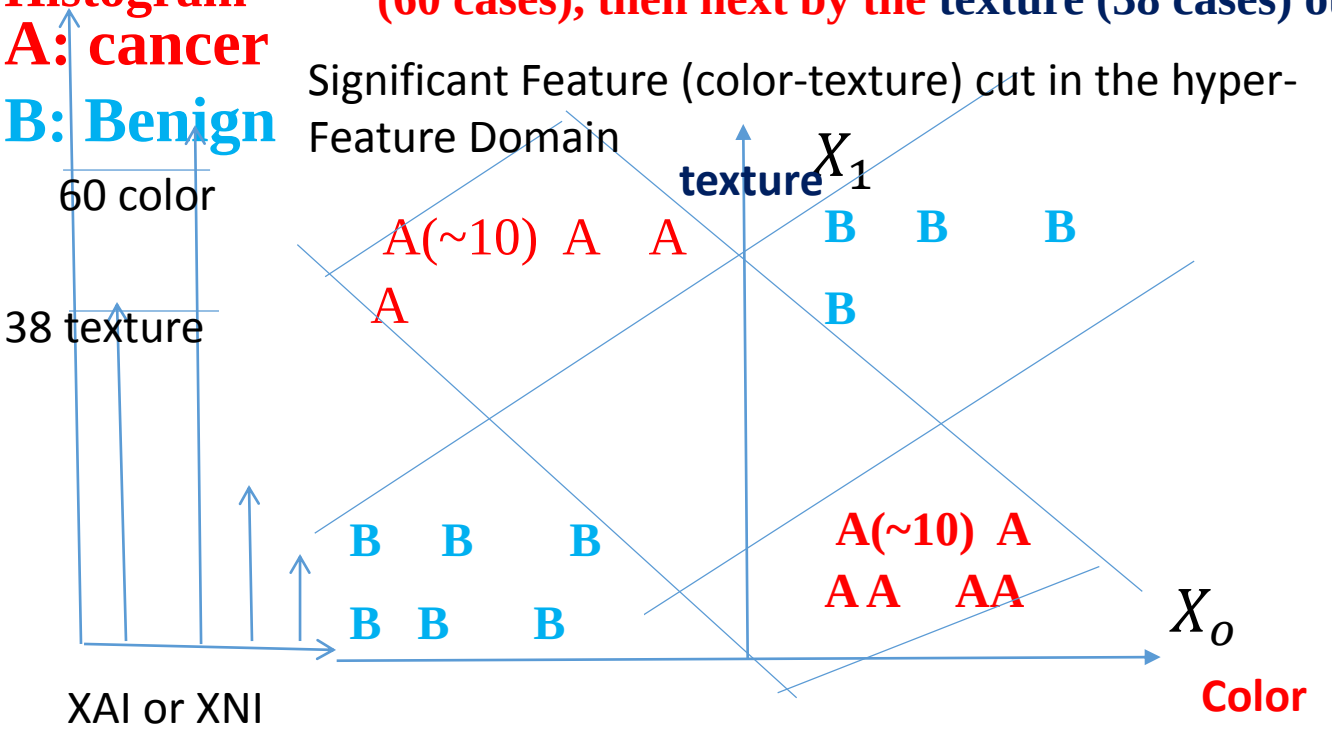
$$X_1 + \frac{W_{12}}{W_{11}} X_2 = \Delta\theta$$

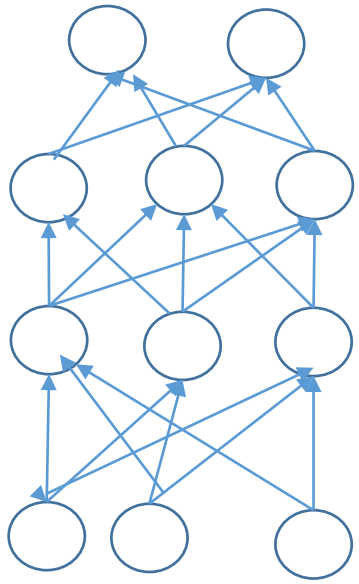
1-layer Artificial Neural Network (ANN)

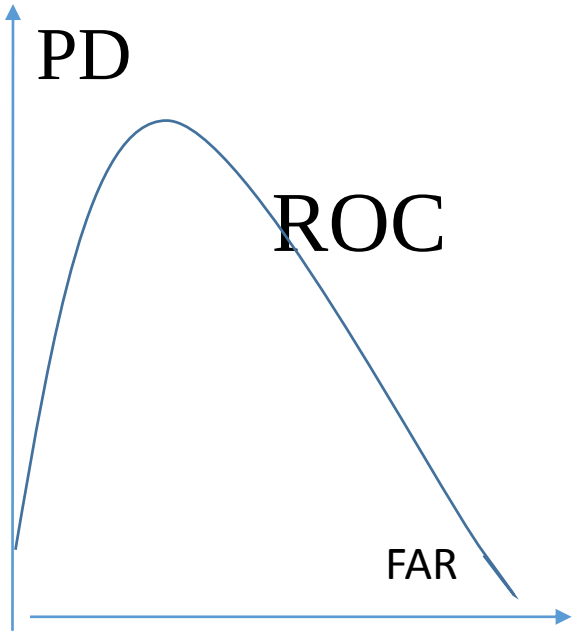
is a **Linear Classifier** for target A **(e.g.** malignant cancer) & non-target **B (e.g.** benign tumor), with high **False Alarm (Cancer A) Rate (FAR) causing delay Multiple Layers namely deep (layer-wise convex hull) learning will be better (next).**

We need Multiple Spectral Layers (MSL) separates the Cortex 17 back of the head Feature Extraction (SFE) from Hippocampus Associative Memory (HAM) under two hemisphere of brain (Logical LHS & Emotional RHS) where Multiple Layers (ML) Machine (Convex Hulls) Classifier (MC) to explainable NI with Histogram.

FE histogram yields XAI or XNI that Melanomas are significantly determined by the multispectral color (60 cases), then next by the texture (38 cases) out of 100, as % of Significant Features Domains(SFD).







- 3x7-layers ***BNN UDL*** are for a Convex Hull **Machine Classifier (MC)**
- Enhance **Receiver Operation Characteristics (ROC),**
- Learning to Increase the **Probability of Detection (PD)** reduce the **False Alarm Rate (FAR)**,
- This goal ***can be directly mapped*** into **Unsupervised Deep Learning**. (cf. **NI & BNN**; Ph. D. Thesis in Data Mining, Intelligent Robots, Autopilots, XAI, ECI, NI, etc.)

Synopsis& Conclusion

- **NI** is defined to be **Unsupervised/Supervised Deep Learning (U/S DL) taking two (Neurons & Glial Cells) to Tango** for Darwinian Survival if only if. **2 conditions:** (1) **Homeostasis** via **Min Free Energy (MFE)/Least Mean Squares (LMS) ala Herman Helmholtz** at Const. Temp. T_o (general than **SDL LMS)** (2) **Power of Pair**. When sensors stimulus brainwaves and pair **coincidence** MFE: "Agree is the signal & kept, and disagree are the noises thus rejected, relaxing brainwaves towards MFE (Homospiens at 37oC for optimum elasticity of hemoglobium; chicken @40oC hatching egg) $\Delta H_{brain} = \Delta E_{brain} - T_o \Delta S_{brain} \leq 0 \quad irreversible\ \Delta S_{brain} > 0$
-
- If we know the shortfall, we know how to deal it better. If we follow the principle of **Physics (Entropy as a measure of uniformity for MFE) & Physiology (Glial Biology for unsupervised)**, the breakthrough of 6 Glial cells from brain tumor **Glioma** to Alzheimer, Epileptic Seizure etc. disorders, to understand missing half of Einstein's, because Homosapiens brains have 3 lb weight with 10 B gray matter neurons & 100 B white matter Glial Cells. By Lyaponov convergence, D.O. Hebb learning $\frac{\Delta[W_{i,j}]}{\Delta t} = -\frac{\Delta H}{\Delta[W_{i,j}]} = g_i S_j \quad g_i \equiv -\frac{\Delta H_{brain}}{\Delta D_i}$
- **where Glial Cells** are defined as gradient of MFE over Dendrite sum
BNN take 10 B neurons &100 B Glial cells to dance Tangle unsupervised . $\vec{D}endrite_{\ i} \equiv \sum_k [W_{i,k}] S_k$
- **What do you expect to learn** from this **Seminar** (if 1-way) or **Colloquium** (if 2-way dialogue):

- **0. Courage (answer your own question), Comprehensive (Maxwell displacement current in capacitor), Complementarity (Q.M. commutator), Conscientious (Edison hard working); 10 group creativity rule how to be 1+1 =11, otherwise 1+1<1. ("Lesson in Creativity," Harold Szu, Ron Driggers, Appl. Opt. V.25, Aug. 2015)**
- **1. To Be Graduated**: Plenty jobs available at Internet Giants, Google Brain Alpha-Go (winning chess games), YouTube (discovery of the most favorite video is cat), as well as Face book (independent of age, expression). NI is beyond **Alan Turing definition of Artificial Intelligence(AI)for the fittest of survival.**
- **2. For Ph. D. Thesis topics**: We can compare AI Supervised Deep Learning (SDL) using ANN versus NI Unsupervised Deep Learning (UDL) BNN(**100 ways to skin a Cat)** ;
- **3. For faculty**: Seeking shortfalls to matching federal funding (beyond CR): **(DoC) Consumer Index Machine IQ defines from Zoombot to House Maid; (DoT) Automated pilot** i-Cars; **(DoD) UXS swarming; (DoE) Robotic Rescue Missions in space, undersea, reactors.**

Appendix A Modern Brain Physiology in Life & Health

rom death we learn life, from disorder we learn orderings, as the truth complement to the false, to one another:

(i) Biological Neural Net based on Thermodynamics Brain at Minimum Free Energy (MFE);

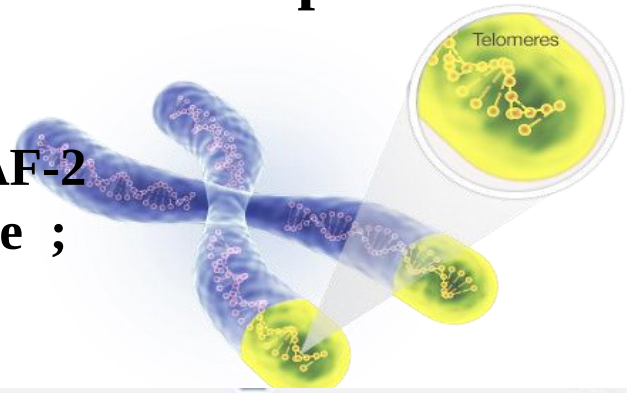
(ii) Smartness inheritance? Genome by DNA & Phenome by Epigenetic;

(iii) Heathy Longevity by Telomerase Enzyme.

(iv) Brain Tumor is known as **Glioma**. A **glioma** is a type of **tumor** that starts in the **brain** or spine. It is called a **glioma** because it arises from glial cells. The most common site of **gliomas** is the **brain**. **Gliomas** make up about 30% of all **brain** and central nervous system **tumors** and 80% of all malignant **brain tumors**. (https://en.wikipedia.org/wiki/**Glioma)**

Appendix B More about BNN heat death. Is it inevitable? Calico (California Life Co.) Inc. said maybe. To prevent the heat death , we need all scale attack on BNN systems in 5 pages as follows some may apply to ECI. (One may skip) The longevity requires **Yamanaka Induced Pluripotent Stem cells** 4 genes, **Blackburn Telomere Length &** Hayflick turns, epigenetic Methylation Histone marker, Mattson Restrictive Calorie;;

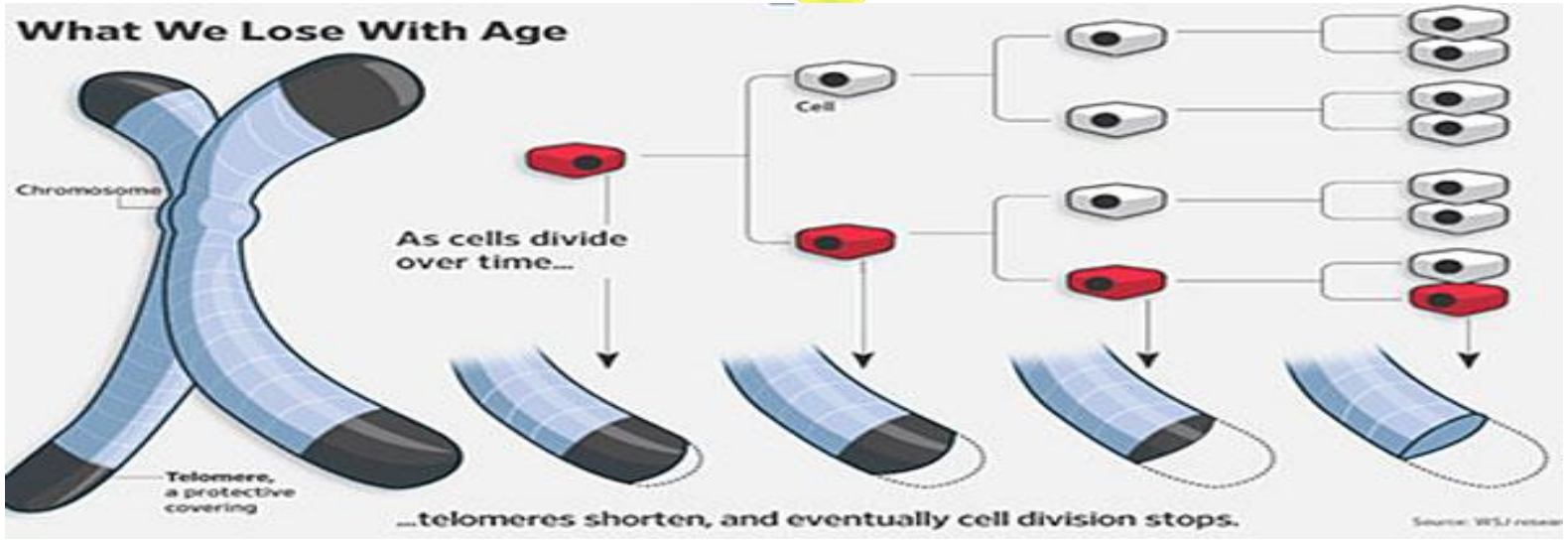
3 billion pairs of A-T C-G codons of DNA packed in 23 pairs in a total of 46 chromosomes un-winded linearly in 3m long



2 mm to 20 mm
X-X female, X-Y male

Kenyon of UCSF discovered worm DAF-2 aging gene; DAF-16 for Longevity gene ; but we're complex

We lose telomere size i.e. 50 times in 9 months known as Hayflick limit

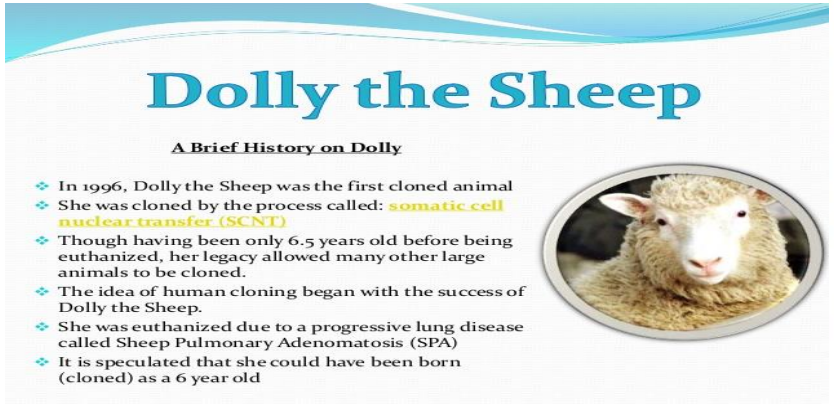


Eukaryote ("Eu=well" has a nuclei & Mitochondria foreign energy production cells inside our cell) billions years ago; Prokaryote ("Pro=pre") primitive cell

Longevity attribute to Nonstop Production of Telomerase Enzyme; Mattson Calorie Restriction; Yamanaka induced Pluripotent Stem (iPS) 4 Genes. We anticipate Scotland Dolly the Sheep cloned at mother 6 years old **somatic cells nuclear tfr, should be winding the clock back to embryonic state using iPS (TBD).**

Cf. W. Duan, et. al. M. P. Mattson (Dir. NIA), "Dietary (Calorie) Restriction Normalizes Glucose Metabolism and Brain-Derived Neurotrophic Factor Levels, Slows Disease Progression and Increases Survival in Huntington Mutant Mice" PNAS Feb. 10 2003; Nobel 2012 Shinya Yamanaka 4 Genes induced Pluripotent Stem Cells of mice & human.

- **Lobsters (non-stop growth of telomerase keeping longevity) can live 150 yrs. On the contrary, Mammals have high growth rates in embryonic and juvenile phases and no growth in adult old phases.**



"**Longevity of lobsters is linked to ubiquitous telomerase expression**," Klapper, et al. U. *Kiel*. FEBS Letters, V. 439, pp.143–146 (1998)

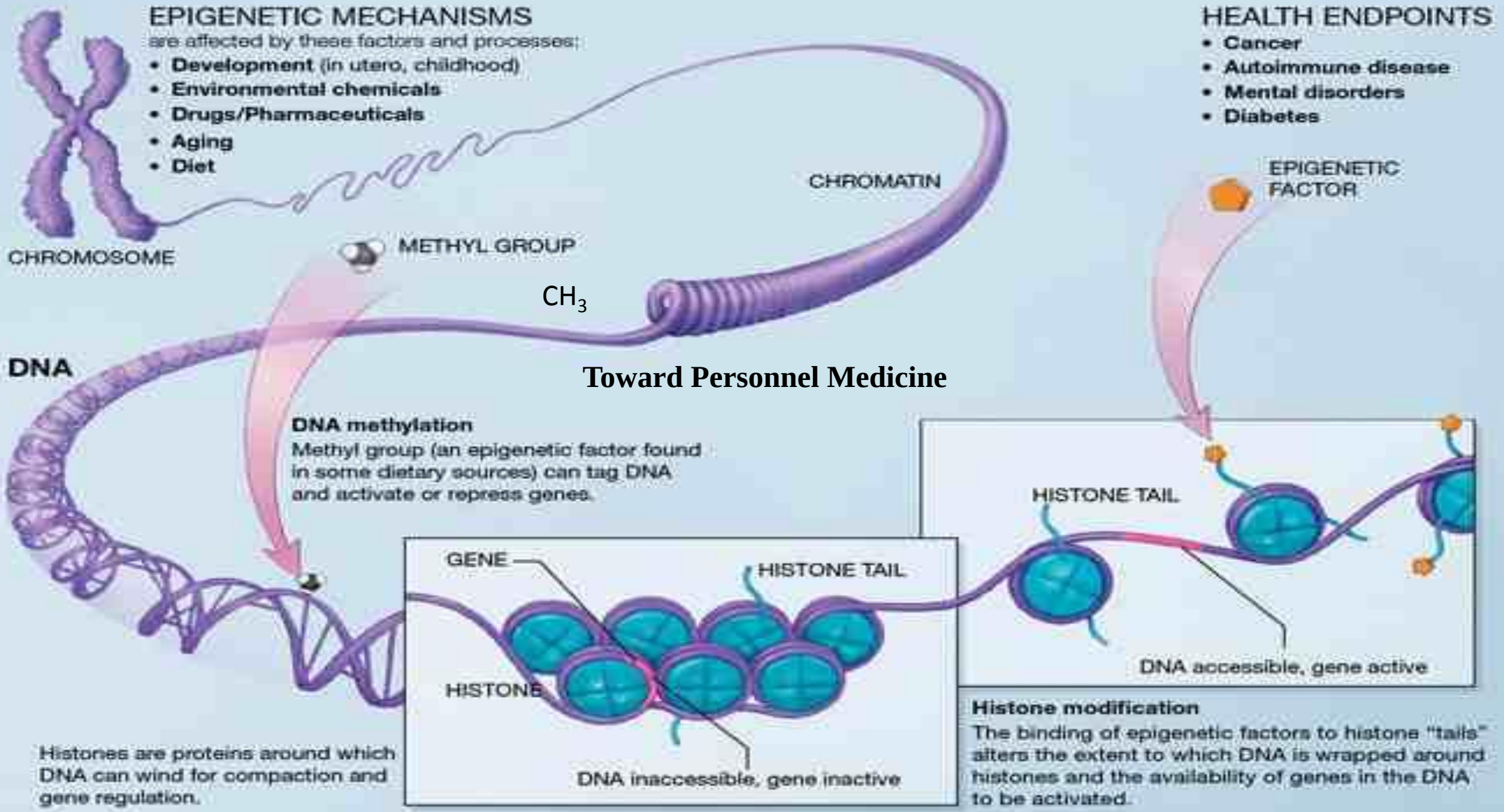
Galápagos giant tortoise of **190 years**; **Arctic Clam of 400 years**

UN/WHO/NIH/NAAS

Le télomère protège le chromosome contre les dégradations

Elizabeth H. Blackburn Carol W. Greider Jack W. Szostak

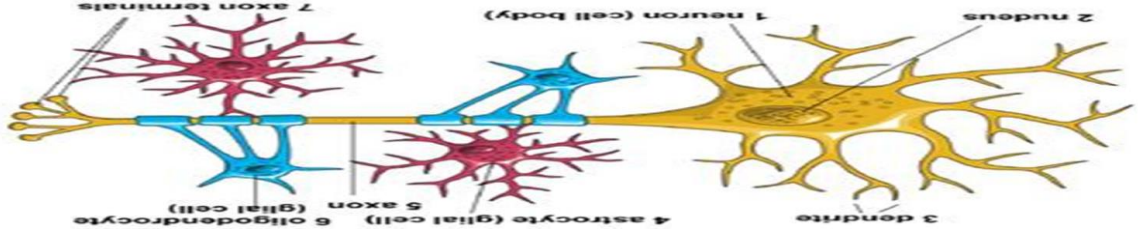
Prix Nobel de physiologie ou médecine, 2009


EPIGENETIC MECHANISMS
are affected by these factors and processes:
• Development (in utero, childhood)
• Environmental chemicals
• Drugs/Pharmaceuticals
• Aging
• Diet
HEALTH ENDPOINTS
• Cancer
• Autoimmune disease
• Mental disorders
• Diabetes
CHROMOSOME
CHROMATIN
METHYL GROUP
CH3
EPIGENETIC FACTOR
DNA
Toward Personnel Medicine
DNA methylation
Methyl group (an epigenetic factor found in some dietary sources) can tag DNA and activate or repress genes.
GENE
HISTONE TAIL
HISTONE
DNA inaccessible, gene inactive
HISTONE TAIL
DNA accessible, gene active
Histones are proteins around which DNA can wind for compaction and gene regulation.
Histone modification
The binding of epigenetic factors to histone "tails" alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.


FEBRUARY 23 / MARCH 2, 2015
TIME
SPECIAL
HEALTH
DOUBLE
ISSUE
THIS BABY COULD LIVE TO BE
142 YEARS OLD
Dispatches From the Frontiers of Longevity
time.com

Appendix C for Cancer Neuroglial Biology insures 4 functionalities: (1) real time communication; (2) convex hull classifier; (3)multiple morphology multiple neuroglial (4) disorder at the singularity

(1) Real Time (RT) Communication is possible because Axon ion vesicles are confined and aligned up in **Axon Cable surrounded by electrically insulated Glial Cells** known as **White Matter Myelin Sheath** making axon insulator an co-axial cable, becoming "how the duck cross the road?" ***"While one slow duck pops in; the other ducks pops out"*** **as ion charge carrier are slow but the propagation speedway is fast as** $O(\Delta t)$ =10-th mille-sec. One meter long from the tail end of the spinal cord to the big toe running away for the survival of the species



(2) Multiple layer Convex Hull Classifier reduce False Alarm Rate

$output\ \vec{S}'_{i,out}(t) \Longrightarrow Dendrite_i = \sum_j [W(\epsilon t)_{i,j}]\vec{S}_{j,out}(t)$, Single layer Input $[W(\epsilon t)_{i,j}]\vec{X}_{j,in}(t) = \vec{S}'_{i,out}(t)$

(3) Unified Neuroglial Theory: $\vec{g}_j \equiv -\frac{\Delta H_{brain}}{\Delta \overrightarrow{Dendrite_j}}$ Multiple Dendrite Morphologies

insure Multiple Neuroglial (pp.13-14): 4+2 kinds Glial, glue, cells composed of white matter.

(4) Divergence of the gradient may define the brain tumor Glioma.

Appendix D for Data Big Data Analysis:

DoC/DoT/DHS: Intelligent Robots/Cops require Explainable & Quantifiable Computational Intelligence (EQCI) and **Consumer Index Machines IQ: MIQ = AI/NI**

1. MIQ=10% is loyal to its human master and its own survivability to differentiate electric power plug having a two-porn's of 110 Volts or three porn's of 220 Volts.2.
2. MIQ=20% is able to understanding human conversation in a fixed semantic network for a closed domain dialogue.
3.**MIQ=30%** is able to read facial expression and voice tone for e-IQ to understanding the emotion need of human being.
4. **MIQ=40%** is able to command and control a small team of other robots.
5. **MIQ=50%** is able to "explore the tolerance of imprecision," e.g. using fuzzy logic to negotiate a single precision path finding in an open save terrain.

This goal is possible unless we emulate BNN NI to separate Feature Extraction from Machine Classification for BDA in drug discovery, in law enforcement, in order to accountable of salient features for Machine Classifier

- We keep i-Robot without e-IQ loving feeling toward master & jeopardies spouse. Then, no Terminator III danger of AI control the world. In general, ***MIQ*** $\subseteq$ **50%** should be in a supervised learning category using large data basis training by the lookup table setup. Since no e-IQ in MIQ.
- Machine intelligence about 20% to 30% should understand simple language & recognize human emotion and some extrapolation and interpolation capability.
- Other than those robots working on the factory floor, the futurist intelligent robots happen in **Deep Ocean, Outer space, Melt down Reactor** with an **unforeseeable NL dynamics interwoven with non-stationary complexity**.
- They may need ***MIQ*** $\supseteq$ **50%** When **MIQ** $\supseteq$ **50%** a machine shall behave human-like with some intrinsic *e-IQ* for better machine-human interface, machine sensory are equipped with learning without the supervision. For example, the team **MIQ** of *UXV (X=air, ground, marine)* shall be equipped with the *Swarming Intelligence* for unsupervised self-organization team formation.

BDA Drug Discovery: FDA Application of Explainable Computational Intelligence to :

Is the Herbal Mushroom G Lucidum, Lingzhi (that 2000 Nobel Laureate Literature Mr. Gao Xingjian recovered in cancer) similar to Merck immunotherapy Keytruda (Pembrolizumab) drug (that President Jimmy Carter Liver and Brain Metastasis cancer: Aug. 2015 ~Feb. 2016)?

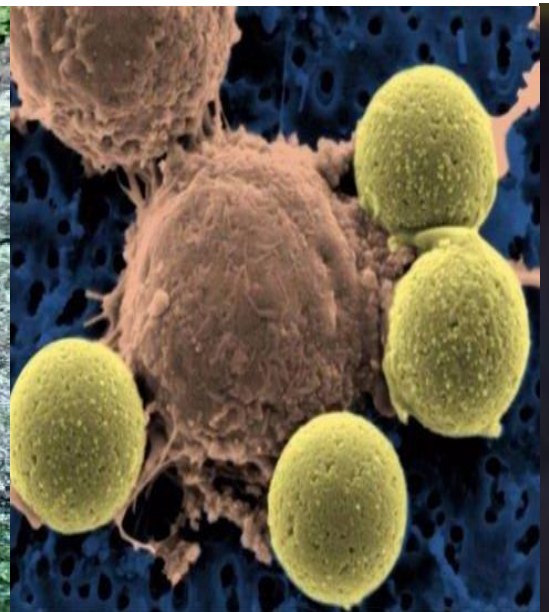
While Merck drug (Yellow balls) are targeted at the Programmed cell Death 1 (PD-1) receptor and allows the body's own immune system go after the cancer cells. While they are all worked on human immune systems, the key difference between Oriental Herbal Medicine and Western Molecular personalized precision targeted drug is mainly in that the holistic is slow in nature of herbal drug for years versus fast drug in half a year.

2000 Literature Nobel Laureate China Mr. Gao Xingjian recovered from cancer

NIH/CAM: Herbal Medicine G Lucidum, Lingzhi

Merck Keytruda (Pembrolizumab)

Jimmy Carter Liver and brain Metastasis advanced Melanoma Aug. 2015 ~Feb. 2016



Appendix E for Entropy Boltzmann Paradox in Thermodynamics:

Is there an Arrow in Time ? Yes, if absolute temperature 0 is never achievable.

(a) Ludwig Boltzmann; $\boldsymbol{S_{tot} = k_B\, Log\, W_{tot}}$; circa 1890 for the irreversible thermodynamics $\Delta\vec{S} > 0$ heat death

(b) Henri Poincare the dynamics is time reversible ,$m_o \frac{d^2\vec{X}}{dt^2} = m_o \frac{d^2\vec{X}}{d(-t)^2}$;

$$\pm i\hbar \frac{\partial \Psi}{\partial(\pm t)} = -\frac{\hbar^2}{2m}\nabla^2\Psi$$

(c) We now know Boltzmann is right, the trajectory is more than dynamics but initial boundary conditions which are irreproducible.

Quiz: What is the entropy? Which has a larger entropy? Sands or Rocks? Boltzmann said the Entropy is a measure of the degree of uniformity.

Let's consider a beautiful beach White Sands (e.g. Australia) versus Mountain Top Rocks, assuming that both composed of similar atoms silicon oxide and impurities.

Answer: Beach white sand has obviously more degree of uniformity, versus the original before eroded at mountain top rocks that have stronger variation of molecular *binding energy* known *archeologically* to the *paleontologist* as the information. Since Neal's third law of thermodynamics: Kelvin temperature can never be absolute zero, incessant molecular collisions that will mix toward maximum uniformity, irreversible increase of the entropy. In other words, the inter-molecular collision will decay gradually the molecular binging energy, namely the archeology information dear to paleontologist at hearts.

Appendix B **"Science has nothing to do with the truth, but the consistency," Albert Einstein circa 1910**

we can unified Sources of Attractive Field Theory: electron radius, gravitational diameter, and glial cell size.

$$\frac{ee}{r_o} = E = m_e C_o^2; \quad r_o = \frac{e^2}{m_e C_o^2} = 2.8\ 10^{-13}\text{cm}$$

$$F = -m_o g = -G\frac{m_o M}{|d_o|^2}; \quad E = m_o C_o^2; \qquad d_o = \frac{1}{C_o}\sqrt{\frac{GME}{F}} =$$

$$\frac{2F}{\pi GM<\rho_o>} \qquad \vec{\mathrm{g}}_j \equiv -\frac{\partial \mathrm{H}}{\partial \overrightarrow{Dentritic_j}} = O(\frac{neuron}{10\ glials})$$

$\left|\overrightarrow{Dendritic_j}\right| \equiv \left|\sum_{i=1}^{N}\left[W_{i,j}\right]\vec{S}_i'^{(t)}\right| = size\ of\ glial$= 0.1 size of neuron; the size of H is the diameter of neurons in sub-millimeters. Inversely, we determine from observation of 0.1 neuron size, the Glue force is 0.1 chemical affinity of neuron layers

Whether the smartness of NI is endowed or cultured?

Epigenetic Methylation (Phenome) versus Genetic (**A-T C-G codons** Genome)?

- Older & Wiser (Longevity gain more experience & more judgment.)
- Is NI coded in Genome DNA A-C, G-T pairs or Phenome Methylation (**3 billion pairs of A-T C-G codons of DNA packed in 23 pairs in 46 Chromosomes un-winded linearly in 3m long)**
- European (EU) **Human Epigenetic Programs (HEP)** have investigated many identical twins that their lifestyles may have influenced the epigenetics that pass down to influence the next gen.
- The United Sates (US) **Human Genome Program (HGP)** has decoded the full human genome.
- If Genetic DNA like a **hardware**; then Epigenetic is like a **software**. We need both the US/HGP & the EU/HEP.

Appendix D: Morphological Learning

(1) Beer Belly architecture for a large internal representation degree of freedom (d. o. f.) for easy Generalization of numerous representations capability in compression to the output class.

(2) Hourly glass architecture with less d. o. f. for Abstraction

(3) Input data-driven determines connectivity morphology by glial cells two ad hoc principles

(i) Use it or Lose it, Pruning Node or Growth Connect Node may be decided by difference in MFE's with it or without

(ii) Hot spot traffic jam recruit more neurons

Appendix E: Hopfield Error Correction Dynamic Code: Nothing but living style can beat a good gene for health longevity: Error Correction in walking meditation can keep the Telomerase enzyme in dynamic balance of telomeres to prevent a premature aging &/or carcinoma.

- In contrary to popular belief, the hydrogen bonds do not stabilize the DNA, and stabilization is mainly due to 3-D stacking (***epigenetic phenome*** helps the ***DNA genome***). The ***back-chaining proofreading*** can reduce error rate below Maxwell-Boltzmann Probability (MBP), 1 error in 10,000 ten thousands

$$MBP_{CG} = \exp -\left(\frac{\Delta H_{CG}}{k_B T}\right) \leq MBP_{AT} = \exp -(\frac{\Delta H_{AT}}{k_B T})$$

ΔH_{CG} is slightly higher than ΔH_{AT}, thus MBP_{AT} is more stable than MBP_{CG}

- BSS $\vec{X} = [A]\vec{S}\,;\ \vec{S} = [W]\vec{X}.\ e.g.$ Given Beethoven first 3 notes: " 5, 5, 1…." :X=5 = (0+5; 1+4; 2+3; 3+2; 4+1; 5+0) in MBP unit at K_BT=1/40eV for T=300°; Find hidden sources tones 2=3 and 3+2 occurring twice that have the highest MBP 2 $\exp(-2/K_BT)\exp(-3/K_BT)$ after ruling out rare *high energy* cases: 0+5 and 1+4, in favor of lower energy: twice 2+3.
- Given multiple spectral bands, MFE to find hot spot cancer or not by 1-layer Lagrange Constraint Neural Net (LCNN), where Lagrange coeff. are Glial cells ("Unsup. Learning at MFE," Szu, Miao, Qi SPIE Vol. 6576, 657605, (2007).

More on Hopfield back−chaining kinetic proofreading for $DNA\ Codec\ \ P = \exp(-H/k_B T_o)$ *John Hopfield "Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity". Proc. Natl. Acad. Sci. U.S.A. **71** (10): 4135–9. Oct 1974*

- Kinetic proofreading allows enzymes to discriminate between two possible reaction pathways leading to correct or incorrect products with an accuracy higher than what one would predict based on the difference in the activation energy between these two pathways.
- Similar DNA reading by tRNA of A-T & C-G pairs the energy difference is minutia. Backward chaining help increase the success reading.
- To achieve an error rate of e^{-10} requires several comparison steps. Hopfield predicted on the basis of this theory that there is a multistage ratchet in the ribosome which tests the match several times before incorporating the next amino acid into the protein.

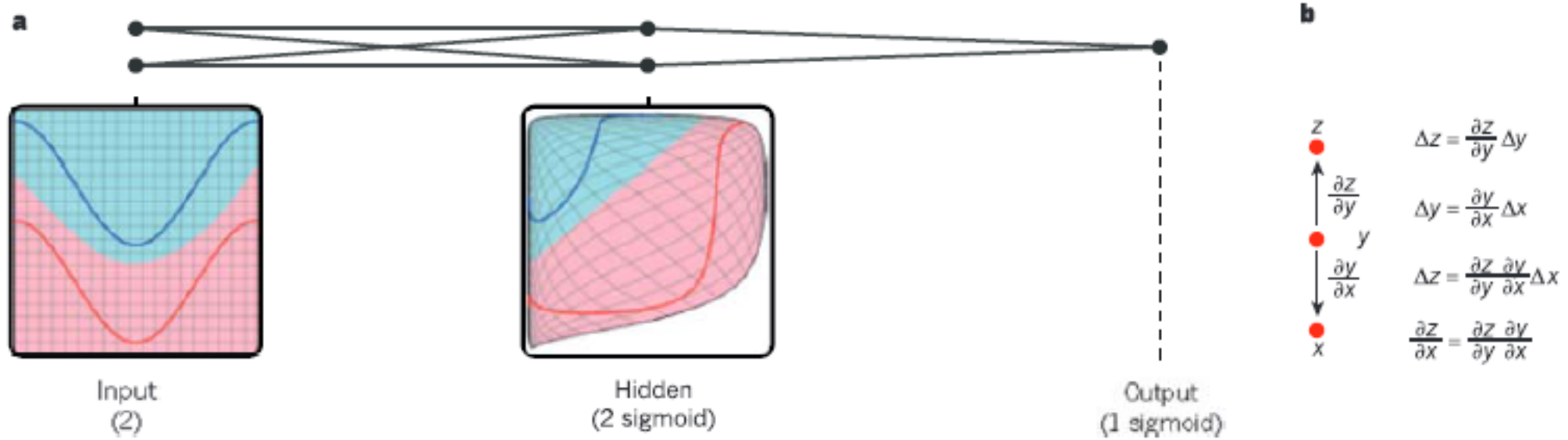
Appendix F: Deep Learning Machine Learning Tensor flow Python Language or Math Lab by Math Works: Dr. Joanna Pingel, "Object Recognition Deep Machine Learning for Computer Vision"

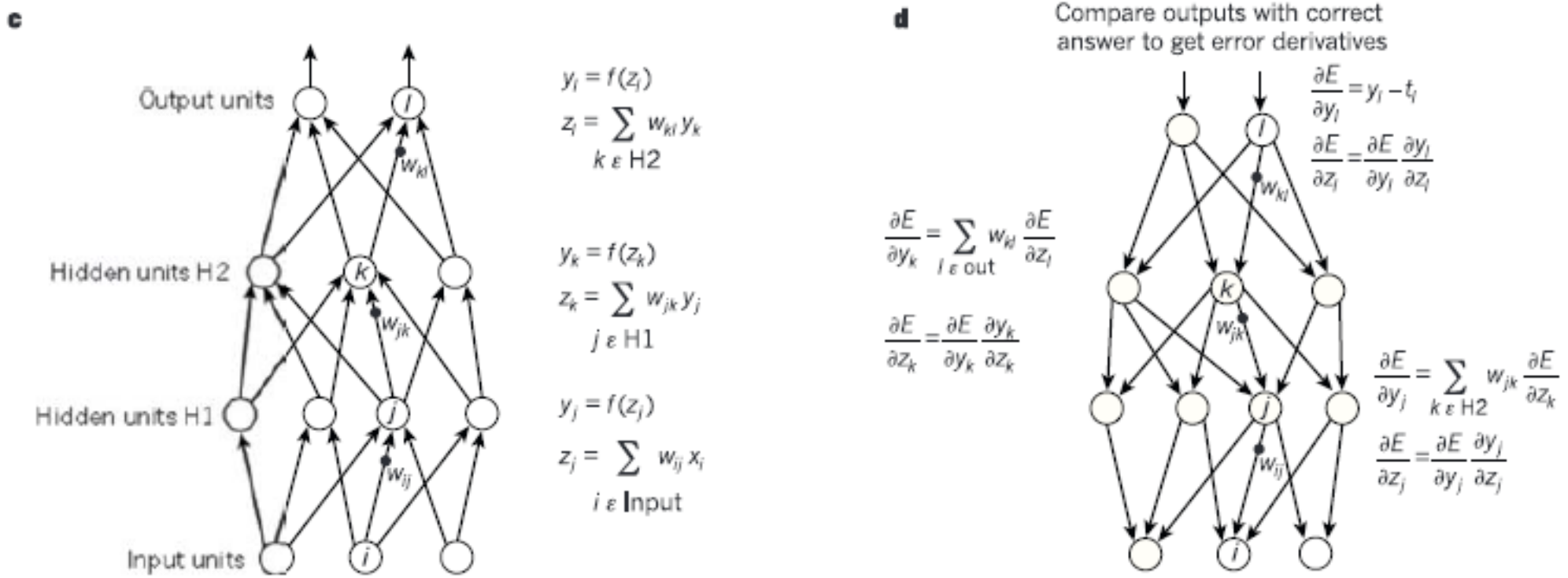
https://www.tensorflow.org/ Python Language for GPU

another video from MATHWORKS - szuharoldh@gmail.com - Gmail.html

Appendix B: Dr. Hesham M. Eraqi; hesham.eraqi@gmail.com Matlab Code (8 pages with comments)

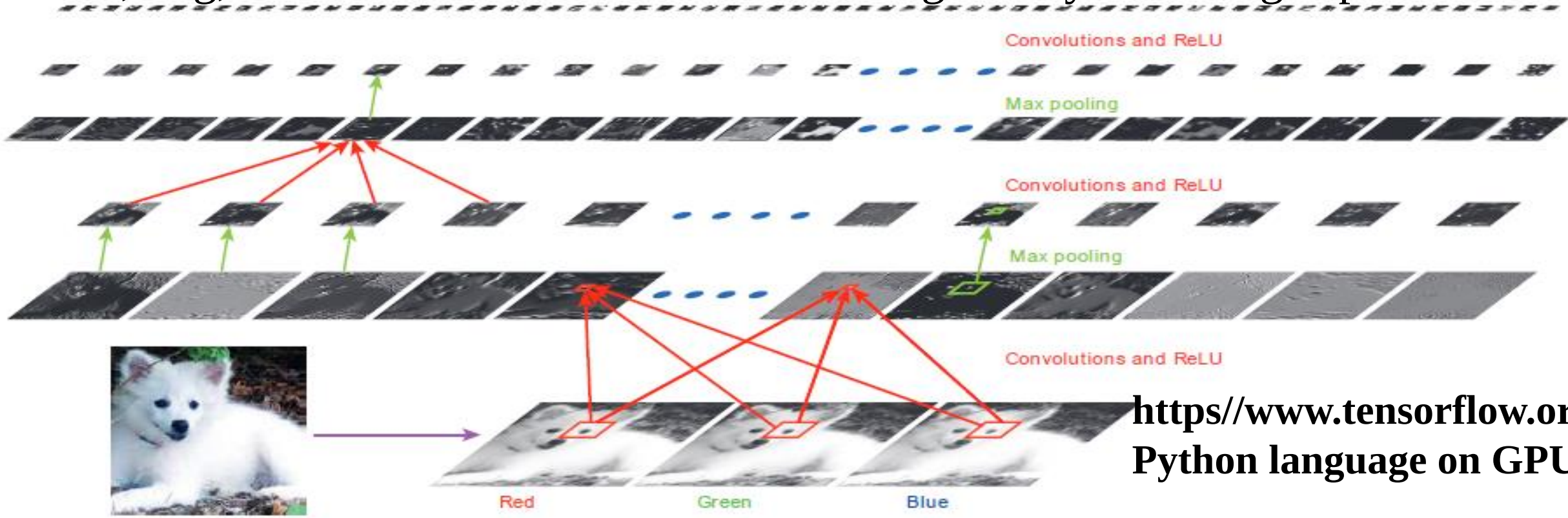
http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html


a
Input
(2)
Hidden
(2 sigmoid)
Output
(1 sigmoid)
b
z
y
x
$\Delta z = \frac{\partial z}{\partial y}\Delta y$
$\Delta y = \frac{\partial y}{\partial x}\Delta x$
$\Delta z = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}\Delta x$
$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$


c
Output units
Hidden units H2
Hidden units H1
Input units
$y_i = f(z_i)$
$z_i = \sum_{k \varepsilon H2} w_{ki} y_k$
$y_k = f(z_k)$
$z_k = \sum_{j \varepsilon H1} w_{jk} y_j$
$y_j = f(z_j)$
$z_j = \sum_{i \varepsilon Input} w_{ij} x_i$
d
Compare outputs with correct answer to get error derivatives
$\frac{\partial E}{\partial y_i} = y_i - t_i$
$\frac{\partial E}{\partial z_i} = \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial z_i}$
$\frac{\partial E}{\partial y_k} = \sum_{l \varepsilon out} w_{kl}\frac{\partial E}{\partial z_l}$
$\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k}$
$\frac{\partial E}{\partial y_j} = \sum_{k \varepsilon H2} w_{jk}\frac{\partial E}{\partial z_k}$
$\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial z_j}$

Training Procedure:

- **Input** has 4 layers edge at layer V1, curvature layer V2, texture layer v3, RGB color pixels per neurons at V4 , hidden layers Internal knowledge representation.
- **Hidden layers** has 10 to 100 layers.
- **Output** is a category vector has components, one for each class of object: cat, fox, dog, etc. whose numerical value will be given by a training supervisor.



https//www.tensorflow.org
Python language on GPU

Appendix B: Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function[1]

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Multilayer Perceptron (MLP) Neural Network Function using MATLAB:      %
%  An implementation for Multilayer Perceptron Feed Forward Fully        %
%  Connected Neural Network with a sigmoid activation function. The      %
%  training is done using the Backpropagation algorithm with options for  %
%  Resilient Gradient Descent, Momentum Backpropagation, and Learning     %
%  Rate Decrease. The training stops when the Mean Square Error (MSE)     %
%  reaches zero or a predefined maximum number of epochs is reached.      %
%                                                                         %
%  Four example data for training and testing are included with the       %
%  project. They are generated by SharkTime Sharky Neural Network         %
%   (http://sharktime.com/us_SharkyNeuralNetwork.html)                    %
%                                                                         %
% Copyright (C) 9-2015 Hesham M. Eraqi. All rights reserved.              %
%                hesham.eraqi@gmail.com                                   %
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% Clear Variables, Close Current Figures, and Create Results Directory
clc;
clear all;
close all;
mkdir('Results//'); %Directory for Storing Results
%% Configurations/Parameters
dataFileName = 'sharky.spirals.points'; %sharky.linear.points - sharky.circle.points - sharky.wave.points -
sharky.spirals.points
% nbrOfNeuronsInEachHiddenLayer = [10 10]; %linear:[4] - circle:[10] - wave,spirals:[10 10]
nbrOfNeuronsInEachHiddenLayer = [4 4]; %linear:[4] - circle:[10] - wave,spirals:[10 10]
nbrOfOutUnits = 2;
unipolarBipolarSelector = 0; %0 for Unipolar, -1 for Bipolar

learningRate = 0.15;
nbrOfEpochs_max = 500000;
```

```
enable_resilient_gradient_descent = 1; %1 for enable, 0 for disable
learningRate_plus = 1.2;
learningRate_negative = 0.5;
deltas_start = 0.9;
deltas_min = 10^-6;
deltas_max = 50;

enable_decrease_learningRate = 0; %1 for enable decreasing, 0 for disable
learningRate_decreaseValue = 0.0001;
min_learningRate = 0.05;

enable_learningRate_momentum = 0; %1 for enable, 0 for disable
momentum_alpha = 0.05;

draw_each_nbrOfEpochs = 100;

%% Read Data
importedData = importdata(dataFileName, '\t', 6);
Samples = importedData.data(:, 1:length(importedData.data(1,:))-1);
TargetClasses = importedData.data(:, length(importedData.data(1,:)));
TargetClasses = TargetClasses - min(TargetClasses);
ActualClasses = -1*ones(size(TargetClasses));

%% Calculate Number of Input and Output NodesActivations
nbrOfInputNodes = length(Samples(1,:)); %=Dimention of Any Input Samples
% nbrOfOutUnits = ceil(log2(length(unique(TargetClasses)))) + !; %Ceil(Log2( Number of Classes ))

nbrOfLayers = 2 + length(nbrOfNeuronsInEachHiddenLayer);
nbrOfNodesPerLayer = [nbrOfInputNodes nbrOfNeuronsInEachHiddenLayer nbrOfOutUnits];

%% Adding the Bias as Nodes with a fixed Activation of 1
nbrOfNodesPerLayer(1:end-1) = nbrOfNodesPerLayer(1:end-1) + 1;
Samples = [ones(length(Samples(:,1)),1) Samples];
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_2$

```
%% Calculate TargetOutputs %TODO needs to be general for any nbrOfOutUnits
TargetOutputs = zeros(length(TargetClasses), nbrOfOutUnits);
for i=1:length(TargetClasses)
   if (TargetClasses(i) == 1)
      TargetOutputs(i,:) = [1 unipolarBipolarSelector];
   else
      TargetOutputs(i,:) = [unipolarBipolarSelector 1];
   end
end

%% Initialize Random Wieghts Matrices
Weights = cell(1, nbrOfLayers); %Weights connecting bias nodes with previous layer are useless, but to
make code simpler and faster
Delta_Weights = cell(1, nbrOfLayers);
ResilientDeltas = Delta_Weights; % Needed in case that Resilient Gradient Descent is used
for i = 1:length(Weights)-1
   Weights{i} = 2*rand(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1))-1; %RowIndex: From Node
Number, ColumnIndex: To Node Number
   Weights{i}(:,1) = 0; %Bias nodes weights with previous layer (Redundant step)
   Delta_Weights{i} = zeros(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1));
   ResilientDeltas{i} = deltas_start*ones(nbrOfNodesPerLayer(i), nbrOfNodesPerLayer(i+1));
end
Weights{end} = ones(nbrOfNodesPerLayer(end), 1); %Virtual Weights for Output Nodes
Old_Delta_Weights_for_Momentum = Delta_Weights;
Old_Delta_Weights_for_Resilient = Delta_Weights;

NodesActivations = cell(1, nbrOfLayers);
for i = 1:length(NodesActivations)
   NodesActivations{i} = zeros(1, nbrOfNodesPerLayer(i));
end
NodesBackPropagatedErrors = NodesActivations; %Needed for Backpropagation Training Backward Pass

zeroRMSReached = 0;
nbrOfEpochs_done = 0;
```

```
%% Iterating all the Data
MSE = -1 * ones(1,nbrOfEpochs_max);
for Epoch = 1:nbrOfEpochs_max

   for Sample = 1:length(Samples(:,1))
      %% Backpropagation Training
      %Forward Pass
      NodesActivations{1} = Samples(Sample,:);
      for Layer = 2:nbrOfLayers
         NodesActivations{Layer} = NodesActivations{Layer-1}*Weights{Layer-1};
         NodesActivations{Layer} = Activation_func(NodesActivations{Layer},
unipolarBipolarSelector);
         if (Layer ~= nbrOfLayers) %Because bias nodes don't have weights connected to
previous layer
            NodesActivations{Layer}(1) = 1;
         end
      end
      % Backward Pass Errors Storage
      % (As gradient of the bias nodes are zeros, they won't contribute to previous layer errors
nor delta_weights)
      NodesBackPropagatedErrors{nbrOfLayers} =  TargetOutputs(Sample,:)-
NodesActivations{nbrOfLayers};
      for Layer = nbrOfLayers-1:-1:1
         gradient = Activation_func_drev(NodesActivations{Layer+1}, unipolarBipolarSelector);
         for node=1:length(NodesBackPropagatedErrors{Layer}) % For all the Nodes in current
Layer
            NodesBackPropagatedErrors{Layer}(node) =  sum(
NodesBackPropagatedErrors{Layer+1} .* gradient .* Weights{Layer}(node,:) );
         end
      end
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_3$

```
% Backward Pass Delta Weights Calculation (Before multiplying by learningRate)
      for Layer = nbrOfLayers:-1:2
         derivative = Activation_func_drev(NodesActivations{Layer}, unipolarBipolarSelector);
         Delta_Weights{Layer-1} = Delta_Weights{Layer-1} + NodesActivations{Layer-1}' *
(NodesBackPropagatedErrors{Layer} .* derivative);
      end
   end

   %% Apply resilient gradient descent or/and momentum to the delta_weights
   if (enable_resilient_gradient_descent) % Handle Resilient Gradient Descent
      if (mod(Epoch,200)==0) %Reset Deltas
         for Layer = 1:nbrOfLayers
            ResilientDeltas{Layer} = learningRate*Delta_Weights{Layer};
         end
      end
      for Layer = 1:nbrOfLayers-1
         mult = Old_Delta_Weights_for_Resilient{Layer} .* Delta_Weights{Layer};
         ResilientDeltas{Layer}(mult > 0) = ResilientDeltas{Layer}(mult > 0) * learningRate_plus; % Sign
didn't change
         ResilientDeltas{Layer}(mult < 0) = ResilientDeltas{Layer}(mult < 0) * learningRate_negative; % Sign
changed
         ResilientDeltas{Layer} = max(deltas_min, ResilientDeltas{Layer});
         ResilientDeltas{Layer} = min(deltas_max, ResilientDeltas{Layer});
         Old_Delta_Weights_for_Resilient{Layer} = Delta_Weights{Layer}
         Delta_Weights{Layer} = sign(Delta_Weights{Layer}) .* ResilientDeltas{Layer};
      end
   end
   if (enable_learningRate_momentum) %Apply Momentum
      for Layer = 1:nbrOfLayers
         Delta_Weights{Layer} = learningRate*Delta_Weights{Layer} +
momentum_alpha*Old_Delta_Weights_for_Momentum{Layer};
      end
      Old_Delta_Weights_for_Momentum = Delta_Weights;
   end
```

```
   if (~enable_learningRate_momentum && ~enable_resilient_gradient_descent)
      for Layer = 1:nbrOfLayers,       Delta_Weights{Layer} = learningRate * Delta_Weights{Layer};
      end
   end

   %% Backward Pass Weights Update
   for Layer = 1:nbrOfLayers-1
      Weights{Layer} = Weights{Layer} + Delta_Weights{Layer};
   end

   % Resetting Delta_Weights to Zeros
   for Layer = 1:length(Delta_Weights),       Delta_Weights{Layer} = 0 * Delta_Weights{Layer};
   end
    %% Decrease Learning Rate
   if (enable_decrease_learningRate)
      new_learningRate = learningRate - learningRate_decreaseValue;
      learningRate = max(min_learningRate, new_learningRate);
   end
 %% Evaluation
   for Sample = 1:length(Samples(:,1))
      outputs = EvaluateNetwork(Samples(Sample,:), NodesActivations, Weights, unipolarBipolarSelector);
      bound = (1+unipolarBipolarSelector)/2;
      if (outputs(1) >= bound && outputs(2) < bound) %TODO: Not generic role for any number of output
nodes
         ActualClasses(Sample) = 1;
      elseif (outputs(1) < bound && outputs(2) >= bound)
         ActualClasses(Sample) = 0;
      else
         if (outputs(1) >= outputs(2)),             ActualClasses(Sample) = 1;
         else
            ActualClasses(Sample) = 0;
         end
      end
   end
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

Multilayer Perceptron Feed Forward Fully Connected Neural Network with a Sigmoid activation function$_4$

```
MSE(Epoch) = sum((ActualClasses-TargetClasses).^2)/(length(Samples(:,1)));
  if (MSE(Epoch) == 0)
    zeroRMSReached = 1;
  end

  %% Visualization
  if (zeroRMSReached || mod(Epoch,draw_each_nbrOfEpochs)==0)
    % Draw Decision Boundary
    unique_TargetClasses = unique(TargetClasses);
    training_colors = {'y.', 'b.'};
    separation_colors = {'g.', 'r.'};
    subplot(2,1,1);
    cla;
    hold on;
    title(['Decision Boundary at Epoch Number ' int2str(Epoch) '. The max number of Epochs is '
int2str(nbrOfEpochs_max) '.']);

    margin = 0.05; step = 0.05;
    xlim([min(Samples(:,2))-margin max(Samples(:,2))+margin]);
    ylim([min(Samples(:,3))-margin max(Samples(:,3))+margin]);
    for x = min(Samples(:,2))-margin : step : max(Samples(:,2))+margin
      for y = min(Samples(:,3))-margin : step : max(Samples(:,3))+margin
        outputs = EvaluateNetwork([1 x y], NodesActivations, Weights, unipolarBipolarSelector);
        bound = (1+unipolarBipolarSelector)/2;
        if (outputs(1) >= bound && outputs(2) < bound) %TODO: Not generic role for any number of
output nodes
          plot(x, y, separation_colors{1}, 'markersize', 18);
        elseif (outputs(1) < bound && outputs(2) >= bound)
          plot(x, y, separation_colors{2}, 'markersize', 18);
        else
          if (outputs(1) >= outputs(2))
            plot(x, y, separation_colors{1}, 'markersize', 18);
          else
            plot(x, y, separation_colors{2}, 'markersize', 18);
          end
        end
      end
    end
    for i = 1:length(unique_TargetClasses)
      points = Samples(TargetClasses==unique_TargetClasses(i), 2:end);
      plot(points(:,1), points(:,2), training_colors{i}, 'markersize', 10);
    end
    axis equal;

    % Draw Mean Square Error
    subplot(2,1,2);
    MSE(MSE==-1) = [];
    plot([MSE(1:Epoch)]);
    ylim([-0.1 0.6]);
    title('Mean Square Error');
    xlabel('Epochs');
    ylabel('MSE');
    grid on;

    saveas(gcf, sprintf('Results//fig%i.png', Epoch),'jpg');
    pause(0.05);
  end
  display([int2str(Epoch) ' Epochs done out of ' int2str(nbrOfEpochs_max) ' Epochs. MSE = '
num2str(MSE(Epoch)) ' Learning Rate = ' ...
    num2str(learningRate) '.']);
  nbrOfEpochs_done = Epoch;
  if (zeroRMSReached)
    saveas(gcf, sprintf('Results//Final Result for %s.png', dataFileName),'jpg');
    break;
  end
end
display(['Mean Square Error = ' num2str(MSE(nbrOfEpochs_done)) '.']);
```

http://heraqi.blogspot.com.eg/2015/11/mlp-neural-network-with-backpropagation.html

References & Notes

Nature Publication has an impact factor 40 points (Science 35; IEEE 1.5; Opt. 3); but the most impact papers among Nature/Sci. may be delayed (~10 years esp. in physiology & physics).

1. "***Learning Machine***," Nicola Jones **V. 505**, pp146-148, **2014**;
2. "***Deep Learning***," Yann LeCun, Yoshui Bengio, Geoffrey Hinton, **V. 521**, pp. 436-440, **2015**.
3. "**Natural Intelligence Neuromorphic Engineering**," Harold Szu, Elsevier 2017, pp.1-350.
4. " **ANN, Deep Learning & Apps**," Harold Szu, Henry Chu & Simon Foo (**Gulf Mexico Spring School** April 16-19, 2017 Tallahassee FL, Elsevier Book Publisher)
5. "**Unsupervised Learning at MFE" (**single layer LCNN for one class breast cancer or not), appeared in Harold Szu, Lidan Miao, Hairong Qi, Proc. SPIE Vol. 6576, p. 657605, (2007)
6. **Multiple Layer Deep Learning** appeared in "Introduction to Computing with Neural Nets," Richard **Lipmann**, IEEE ASSP Magazine April 1987 & PDP MIT book by David **Rumelhart** , James **McCelland**; P. **Werbos** PhD Thesis.
7. **Mary Nedergaad**, **Steve Goldman (Sweden & Rochester)** "Brain Drain," Sci. Am. March 2016.
8. Nobel Prize in Medicine & Physiology has been given in 2012 to the discovery genes by Kyoto Prof. **Shinya Yamanaka**, and these **4 Yamanaka genes** can be unwind cells back to the embryonic (adult cells induced pluripotent: mice, Dolly Sheep, Homosapiens longevity)
9. **Common Sense Longevity: Sleep Tight**, **Eat Right** (calorie restriction: Luigi Fontana Wash U.**), Deep Exercise** (e.g. Yoga,Tai Chi Quan), **Be Happy**.
10. "**Natural Intelligence Neuromorphic Engineering** (NINE)" book by Harold Szu et. al. Elsevier 2017
11. "**Deep learning ANN & Appl.**" book edited by Foo, Chu, Szu et al. from GMSS Tallahassee FL April 16-18, 2018 Elsevier 2018.