

# AI & Machine Learning in Physics

PHYS3151 (6 credits)

Place: MB 122

Time : Tue 13:30-14:20, 14:30-15:20  
Fri 14:30-15:20



Teachers: Zi Yang Meng ( [zymeng@hku.hk](mailto:zymeng@hku.hk) ), HOC 231

<https://quantummc.xyz/hku-phys3151-machine-learning-in-physics-2023/>

Tutor: Ting-Tung Wang ( [leowdd@connect.hku.hk](mailto:leowdd@connect.hku.hk) ), HOC 217

# AI & Machine Learning in Physics

Teaching Materials:

<https://quantummc.xyz/hku-phys3151-machine-learning-in-physics-2023/>

Slides / Reading materials

Python notebooks

Assignments

## Assessment Methods and Weighting

- Assignments 30%
- Presentation 20%
- Project report 20%
- Exam. 30%

# AI & Machine Learning in Physics

Literature: Books

there are many, actually too many



- Ethem Alpaydin, Introduction to Machine Learning, Third Edition, MIT Press 2014
- Simon Haykin, Neural Networks and Learning Machines, Third Edition, Pearson 2009
- Stuart Russell, Peter Norvig, Artificial Intelligence, Third Edition, Pearson 2010

Literature: Online material

Andrew Ng, Stanford University

<http://www.holehouse.org/mlclass/>

[https://www.youtube.com/playlist?list=PLLssT5z\\_DsK-h9vYZkQkYNWcItqhIRJLN](https://www.youtube.com/playlist?list=PLLssT5z_DsK-h9vYZkQkYNWcItqhIRJLN)

Neuroscience For Kids

<http://faculty.washington.edu/chudler/neurok.html>

Michael Nielsen, scientist at home, the best reading material for NN

<http://neuralnetworksanddeeplearning.com>

# Ubiquitous AI and Computational Research, **from Quantum Materials to the Origin of Black Holes**

By



Dr Zi Yang MENG



Dr Hugo PFISTER



HKU  
Science



香港科學館  
HONG KONG  
SCIENCE MUSEUM

Hong Kong Science Museum 01/15/2022

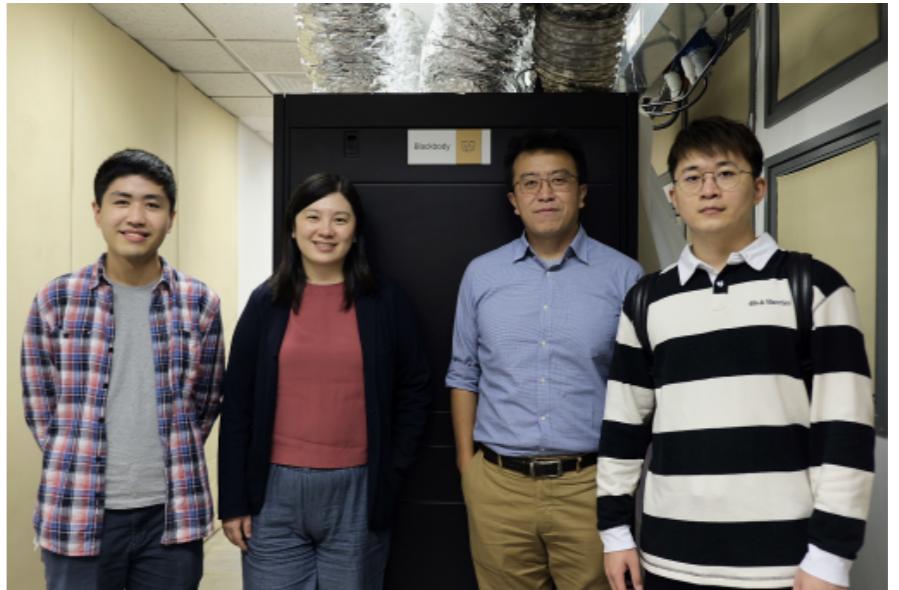
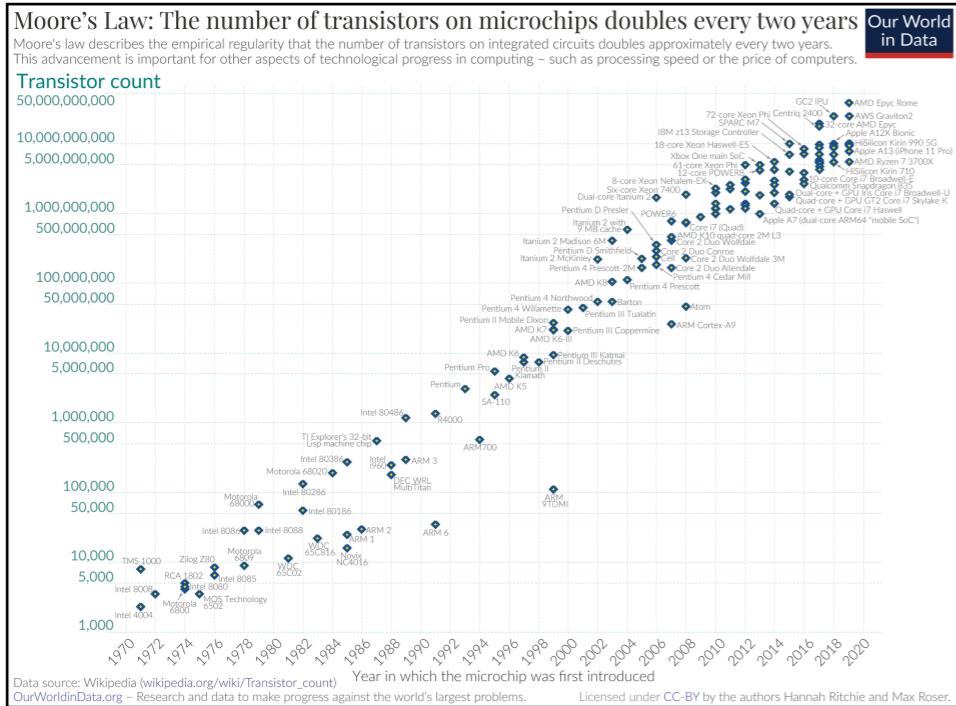
<https://www.scifac.hku.hk/events/ai-computational-research>

# Computation becomes easy

50 years of supercomputer tracks Moore's law

Our own Blackbody

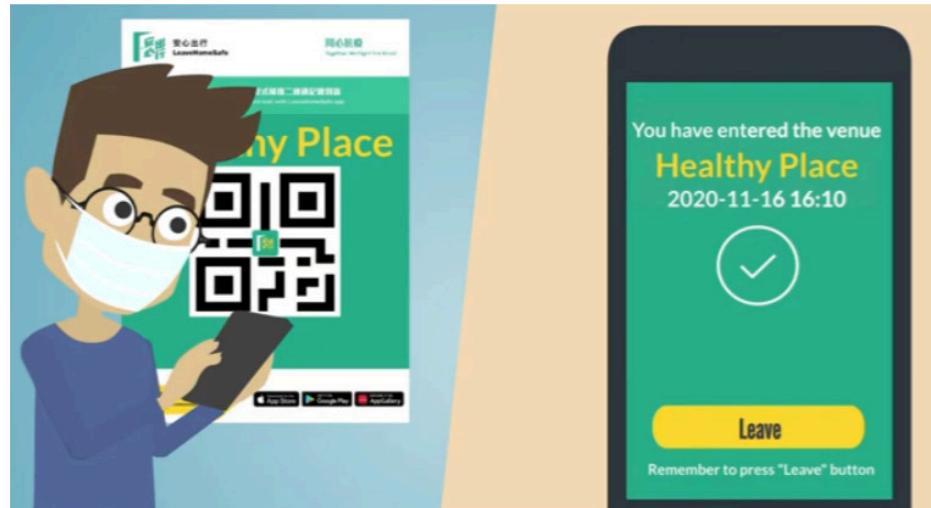
# transistors doubles every 2 years



	Supercomputer	Personal Computer	Human Brain
Computational Units	32,000 Xeon CPUs $10^{12}$ transistors	4 CPUs, $10^9$ transistors	$10^{11}$ neurons
Cycle time	$10^{-9}$ sec	$10^{-9}$ sec	$10^{-3}$ sec
Operations/sec	$10^{15}$	$10^{10}$	$10^{17}$
Memory updates/sec	$10^{14}$	$10^{10}$	$10^{14}$
Weight / Space	150 tons / Basketball court	1 Kg / A4 Paper	1.5 Kg / 1/6 basketball
Power consumption	500 megawatt	100 watt	20 watt

# AI & Machine Learning in Physics

In the era of AI & Big data



QR / Face Recognition



Smart Robots

<https://www.bostondynamics.com/>

In April 2017, AlphaGo vs. Jie Ke



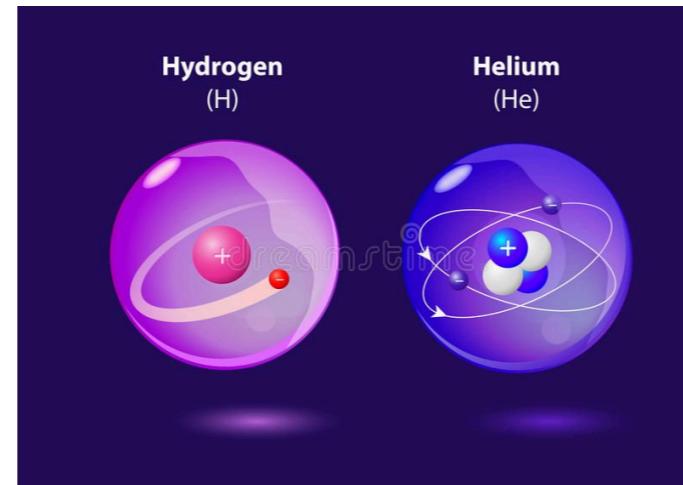
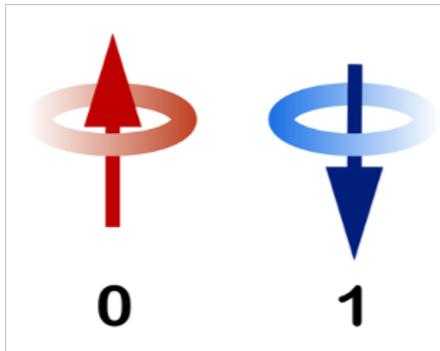
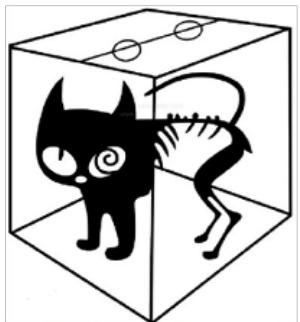
- The machine played perfect...
- I am so behind, unbelievable...
- AlphaGo is not the God, but it is a superior species than human being...

AlphaGo



Self-driving Car

# Computation on quantum bit and quantum entanglement



PERIODIC TABLE of the ELEMENTS

Lead to the famous “exponential wall”



Krishna and Radha playing chaturanga

Wheat grains on chessboard – Sissa ibn Dahir, inventor of Chaturanga

$2^{64} - 1 = 18,446,744,073,709,551,615$  grains of wheat, weighing about 1,199,000,000,000 tons.

About 1,645 times the global production of wheat.

$2^N$  Multi-electron atoms, cannot be solved exactly

$$N = 10 \quad 2^{10} = 1,024 \sim 10^3$$

$$N = 20 \quad 2^{20} = 1,048,576 \sim 10^6$$

$$N = 30 \quad 2^{30} = 1,073,741,824 \sim 10^9$$

$$N = 40 \quad 2^{40} = 1,099,511,627,776 \sim 10^{12}$$

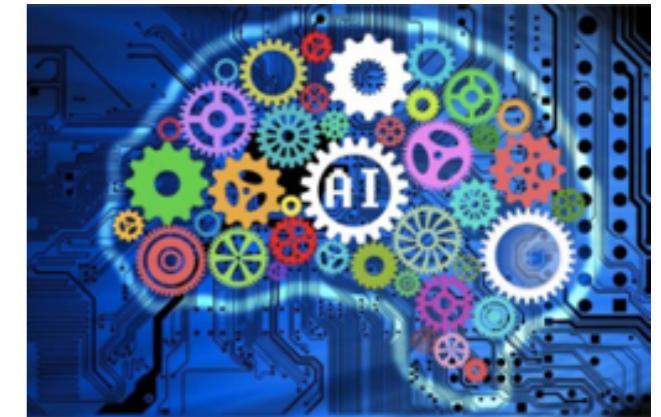
$$N = 50 \quad 2^{50} = 1,125,899,906,842,624 \sim 10^{15}$$

Solving exponentially complex problem in polynomial time

# AI & Machine Learning in Physics

## AlphaGo-1

In March 2016, AlphaGo played with Lee Sedol in Seoul.

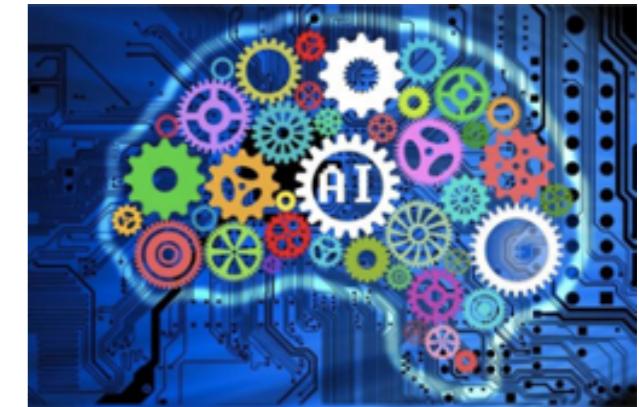


- its psychological aspects (its relentless concentration) no human can match it...
- The Go skill has improved surprisingly ...

# AI & Machine Learning in Physics

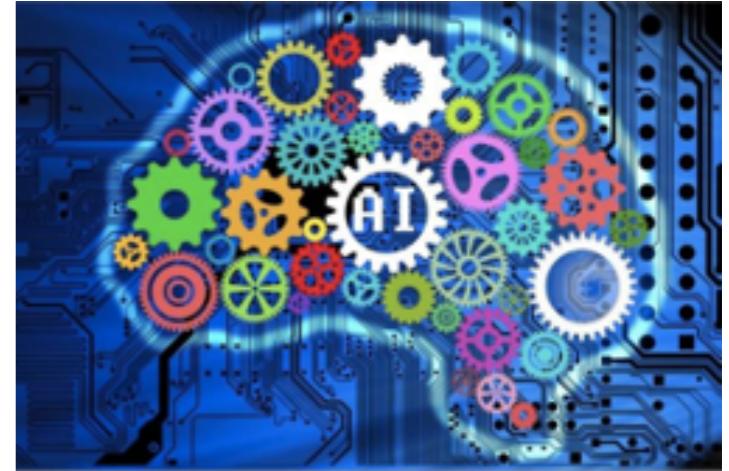
## AlphaGo-2

In April 2017, AlphaGo vs. Jie Ke



- The machine played perfect...
- I am so behind, unbelievable...
- AlphaGo is not the God, but it is a superior species than human being...

# AI & Machine Learning in Physics



## What happened to AlphaGo?

1. In January 2016, researchers revealed that AlphaGo could play 1 million games in 4 weeks. This means that AlphaGo could play 30,000 games per day. How many games could Lee and Ke play? 10 games/day, 82 years (30,000 days), only 300,000 games only. AlphaGo has played 300,000,000 games after march 2016!!! → **Big Data!!!**
2. Till now, only ~2 million games in total. AlphaGo actually learned from data generated by itself → **Self learning!!!**
3. Software/Algorithm: 12 layers → 40 layers. **Make it more complicated!!!**
4. Hardware: TPU1 → TPU2. **New Hardware is necessary!!!**

Tensor processing unit (TPU) by google

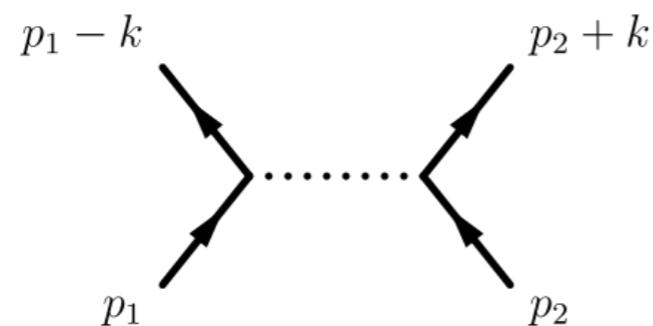
# Deep Learning And Physics

## DLAP2019

> Yukawa Institute for Theoretical Physics  
> Kyoto, Japan  
> 31 Oct - 2 Nov 2019 ■



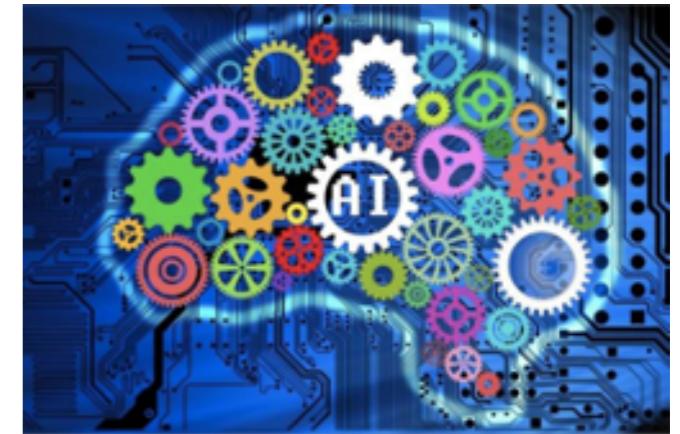
Hideki Yukawa 汤川秀树



# AI & Machine Learning in Physics

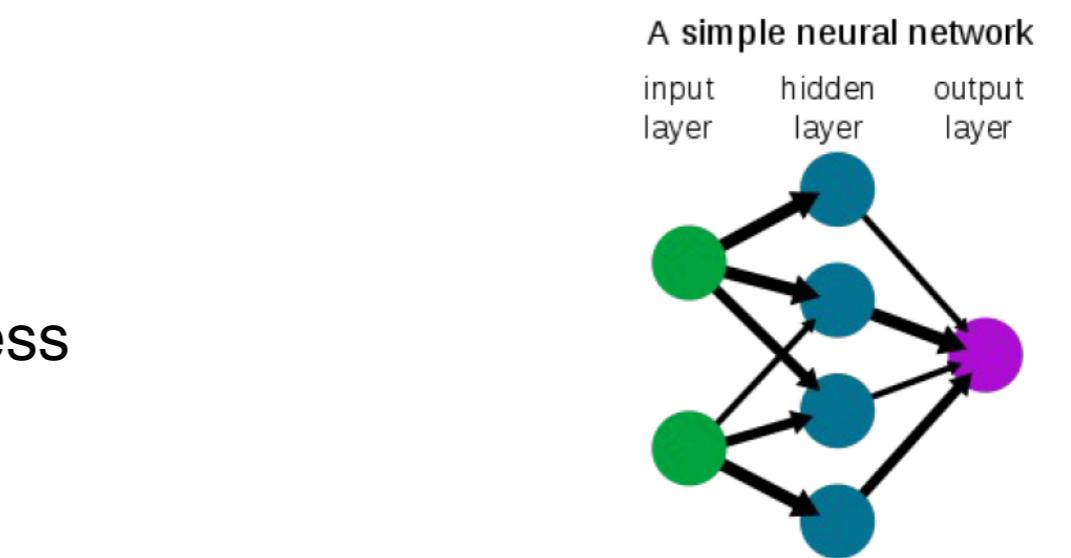
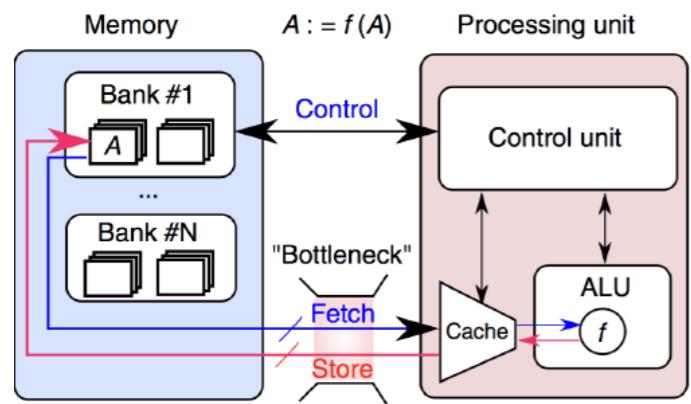
Challenges 1: models are more complicated

~ 100 layers, ~  $10^6$  weights/parameters

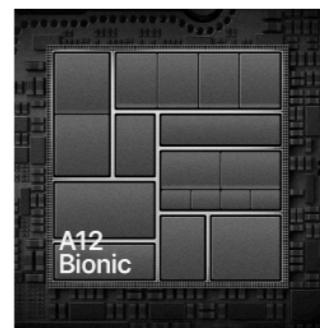


Challenges 2: memory bottleneck

Data fetch is much expensive than data process



Neural Processing unit (NPU) for AI computing



Large on-chip memory, bring computing and memory closer, using low precision computing.

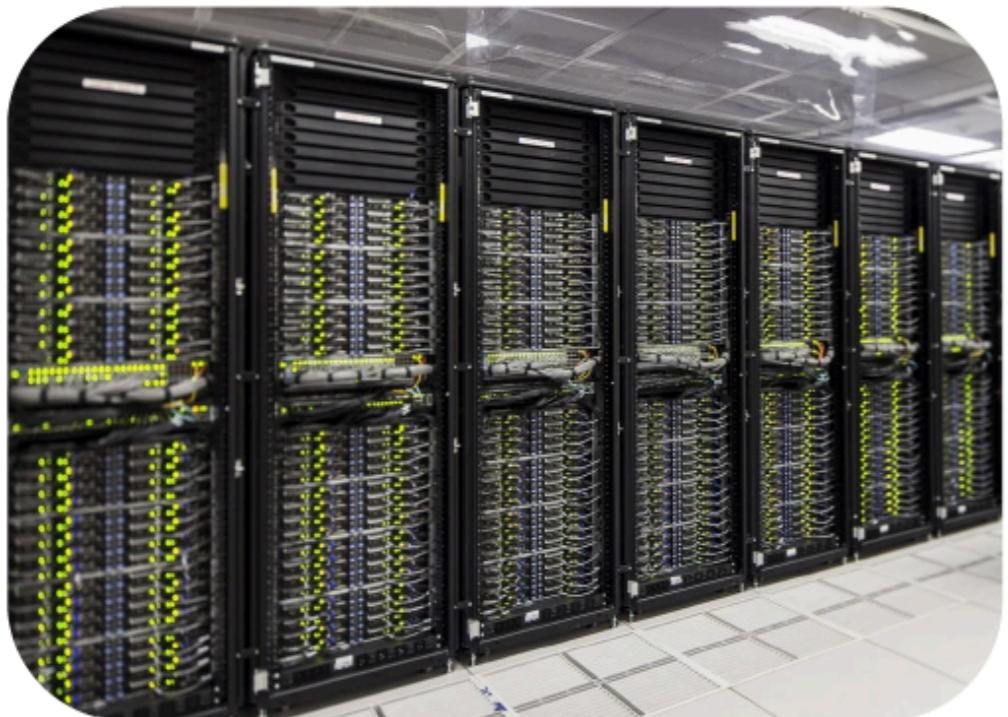
TPU by Google

A12 Bionic by Apple

Kirin 980 by Huawei

# AI & Machine Learning in Physics

## Challenges 3: energy consumption



AlphaGo:

- 176 GPUs, 1202 CPUs
- 150, 000 Watts

Jie Ke:

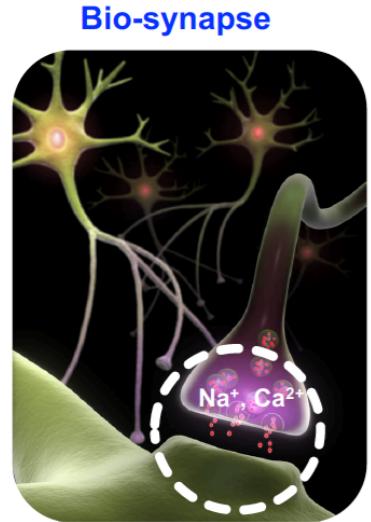
- 1.2L Human Brain
- ~20 Watts

Huge power gap between human brain and CMOS-based AI system

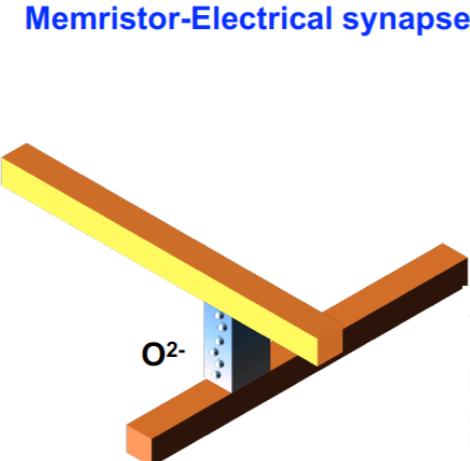
- It is much needed to develop **new hardware** with **new device** and **new architecture and new algorithm.**

# AI & Machine Learning in Physics

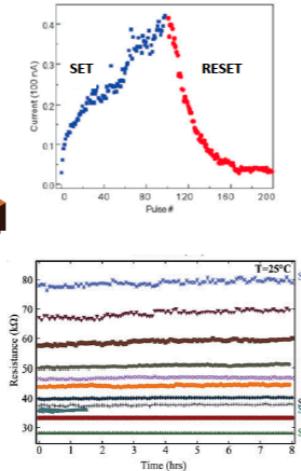
## Bio-synapse vs. Electrical synapse



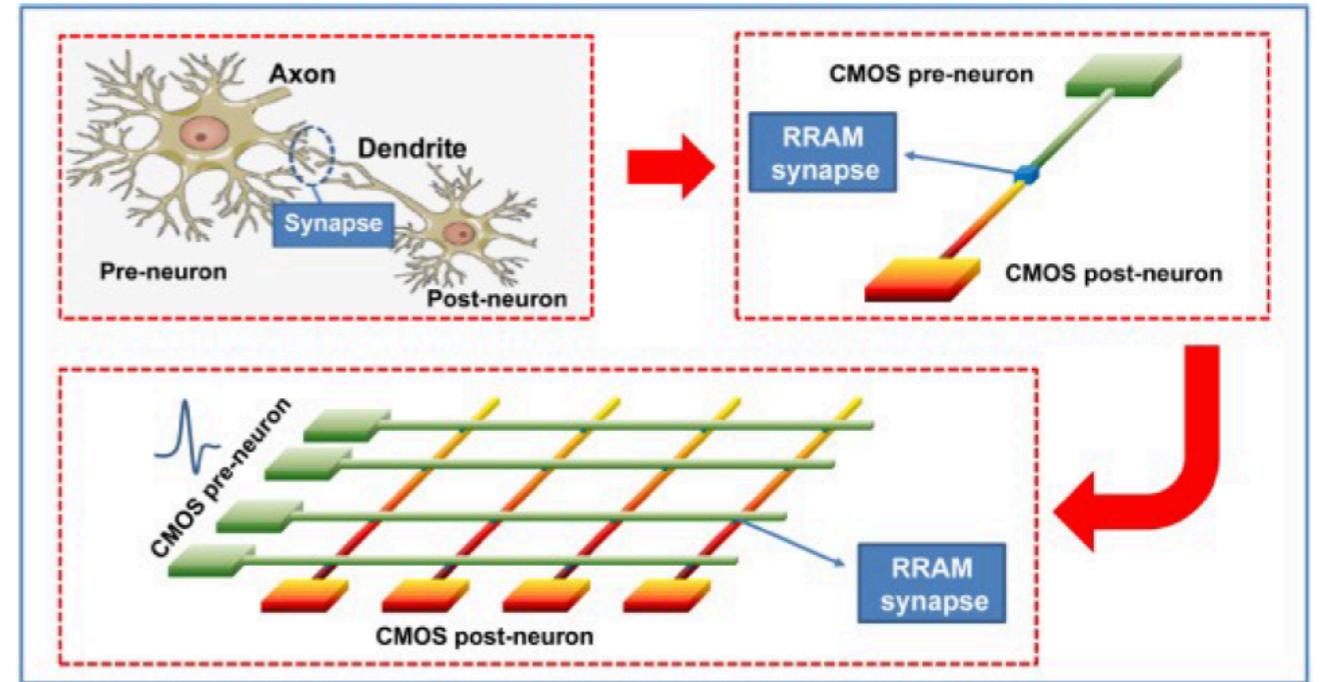
Bio-synapse conductance change through the  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  ions movement



Memristor device conductance change through  $\text{O}^{2-}$  ions movement



## Neuromorphic computing



$$\mathbb{R}^{N \times N}$$

$$Ax = b$$

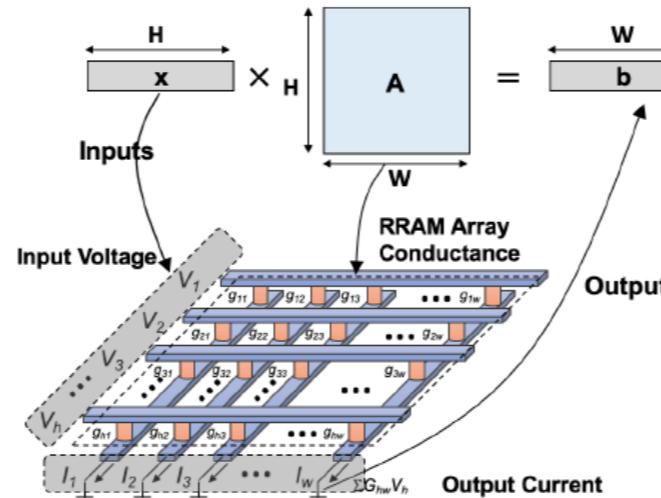
$$O(N^2)$$

$$x = A^{-1}b$$

$$O(N^3)$$

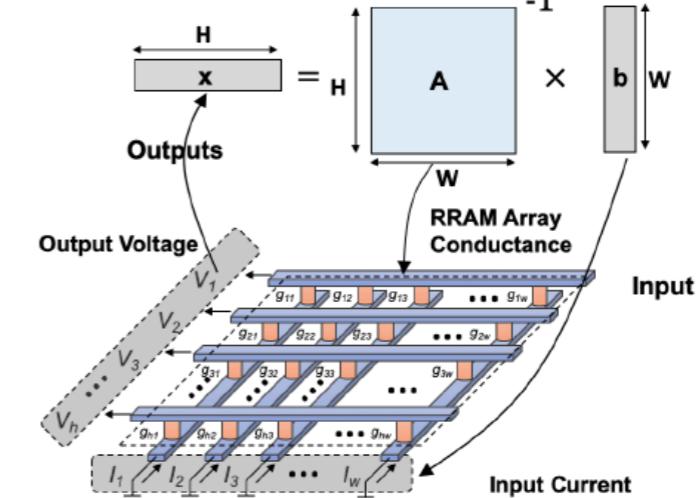
- Vector-Matrix-Multiplication (VMM)

$$Ax = b$$



- Linear equation solver

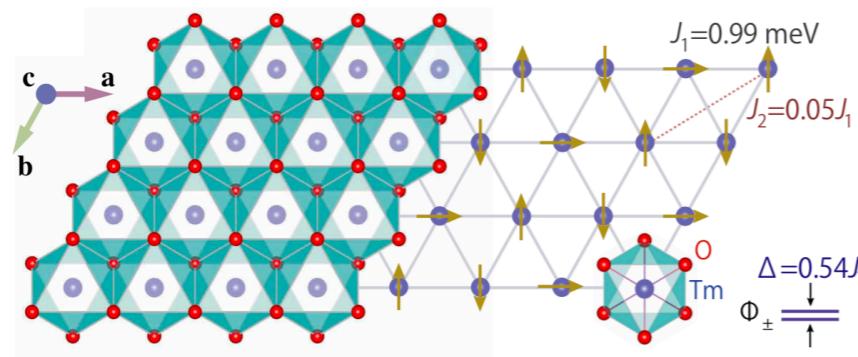
$$Ax = b \rightarrow x = A^{-1}b$$



Computing with physical law in memory: Ohm's Law and Kirchhoff's Law

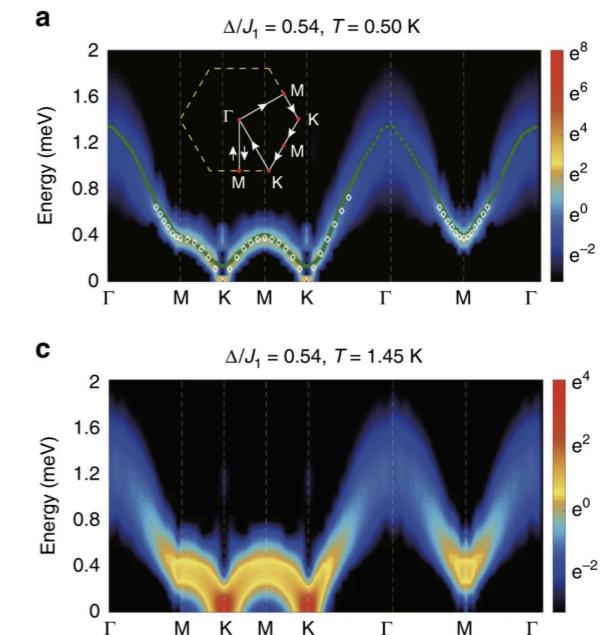
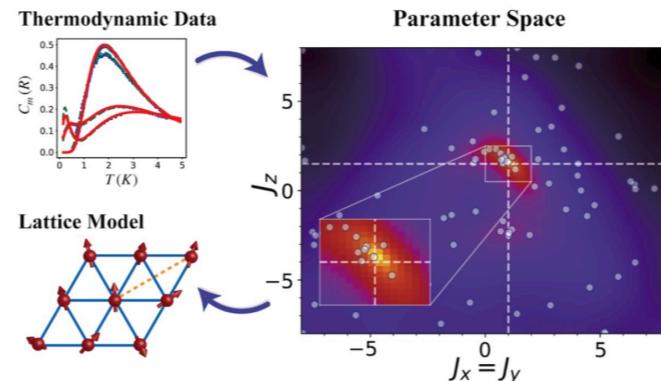
# AI & Machine Learning in Physics

❖ Quantum material research connecting physicists in Hong Kong, Beijing and Shanghai



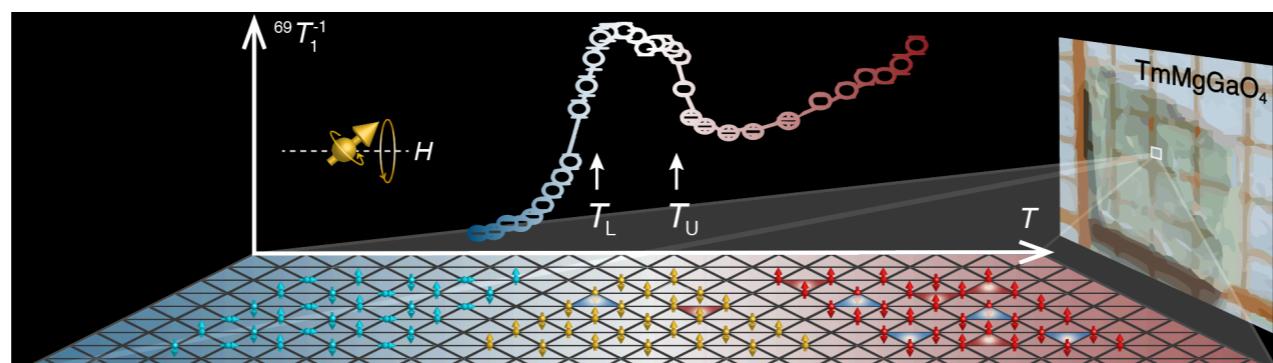
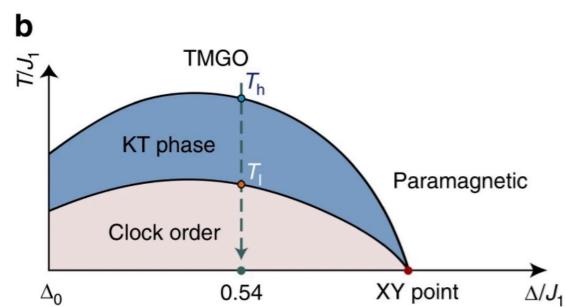
[Nature Communications 11, 1111 \(2020\)](#)

auto-gradient, Bayesian optimization

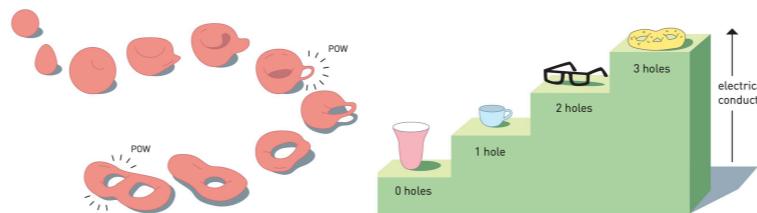


❖ Confirming simulated calculations, quantum material research reveals topological KT phase

Nuclear magnetic resonance — MRI / CT scans



[Nature Communications 11, 5631 \(2020\)](#)



The Nobel Prize in Physics 2016



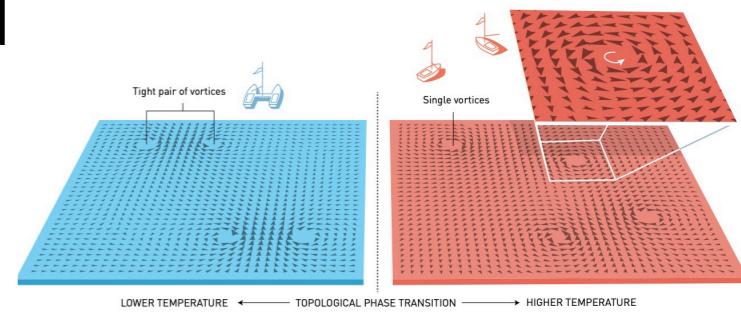
III: N. Elmehed. © Nobel Media 2016  
David J. Thouless  
Prize share: 1/2



III: N. Elmehed. © Nobel Media 2016  
F. Duncan M.  
Haldane  
Prize share: 1/4



III: N. Elmehed. © Nobel Media 2016  
J. Michael Kosterlitz  
Prize share: 1/4



# Computation and AI could solve the energy crisis



electric power transmission at high voltage



High speed rail

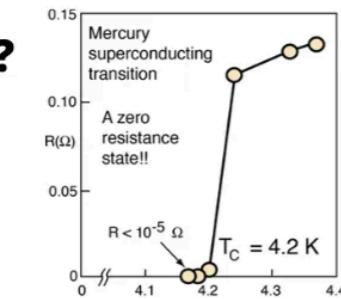
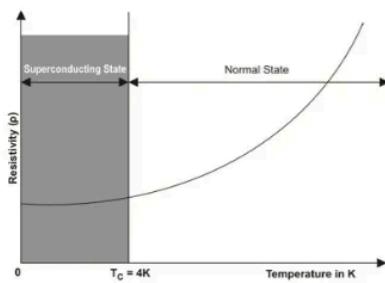


Maglev (magnetic levitation) bullet train with 600 k/h



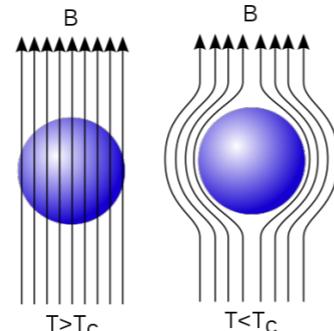
## Understanding Quantum Metals

### What is Superconductivity?

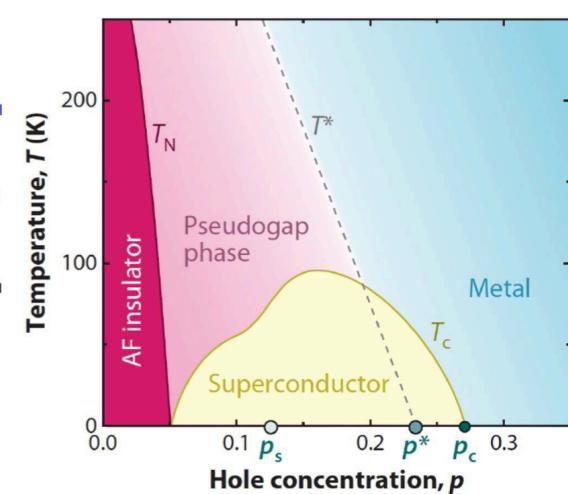
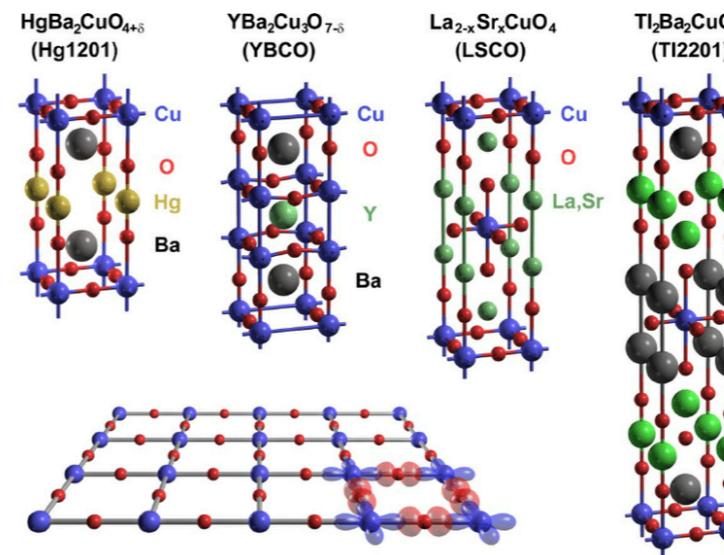


Dutch Physicist  
Heike Kamerlingh Onnes  
in 1911

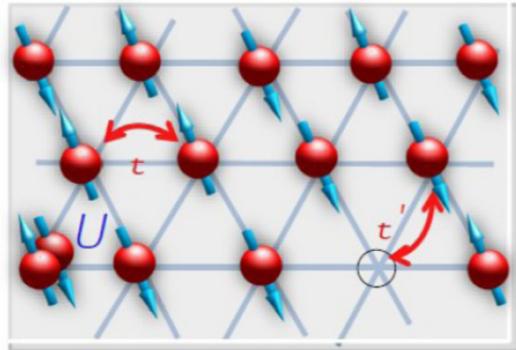
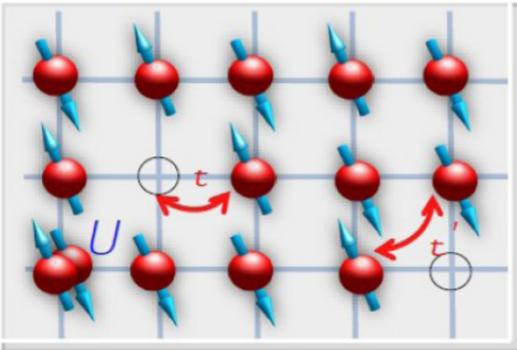
superconductors at ~ -270°C



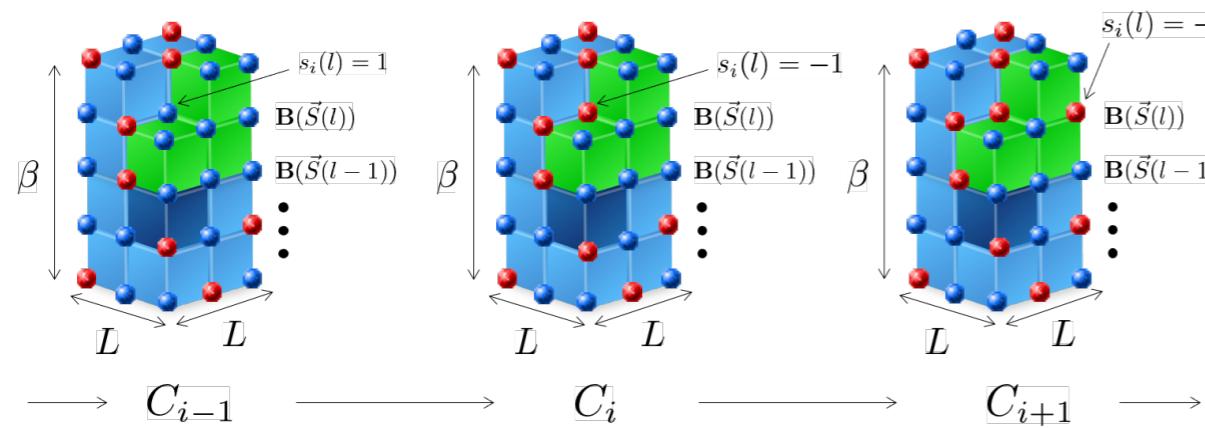
High-temperature superconductors at ~ -100°C



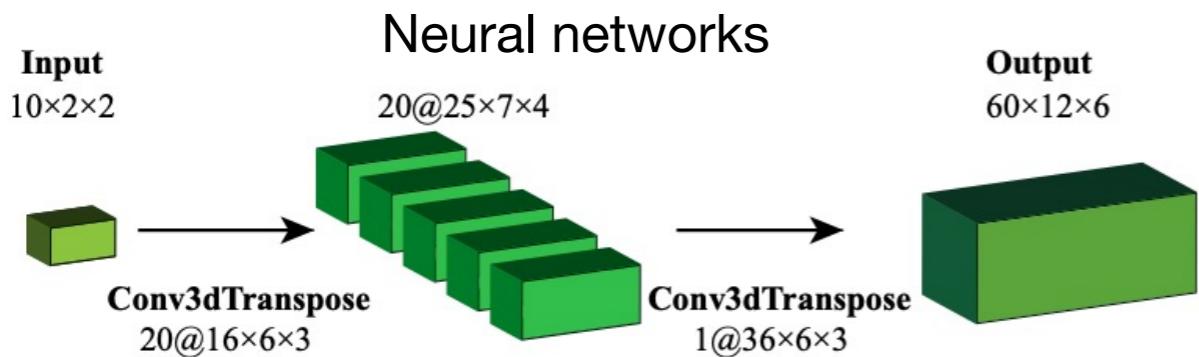
# AI & Machine Learning in Physics



**Self-learning Monte Carlo is explainable-AI**



$$Z = \sum_{\{\mathcal{C}\}} e^{-\beta \hat{H}} = \sum_{\{\mathcal{C}\}} \underbrace{\phi[\mathcal{C}] \det(1 + \mathbf{B}_\beta \cdots \mathbf{B}_1 \mathbf{B}_0)}_{\omega[\mathcal{C}]} [\mathcal{C}] = \sum_{\{\mathcal{C}\}} e^{-\beta H_{eff}[\mathcal{C}]}$$

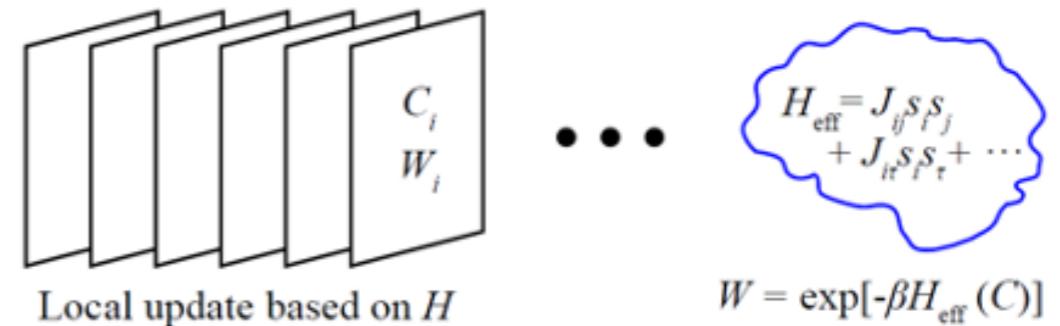


[Chin. Phys. Lett. 39, 050701 \(2022\)](#)

one of the Delphic maxims and was inscribed in the pronaos (forecourt) of the Temple of Apollo at Delphi

Trilogy of self-learning

[Phys. Rev. B 96, 041119\(R\) \(2017\)](#)  
[Phys. Rev. B 95, 241104\(R\) \(2017\)](#)  
[Phys. Rev. B. 95. 041101\(R\) \(2017\)](#)

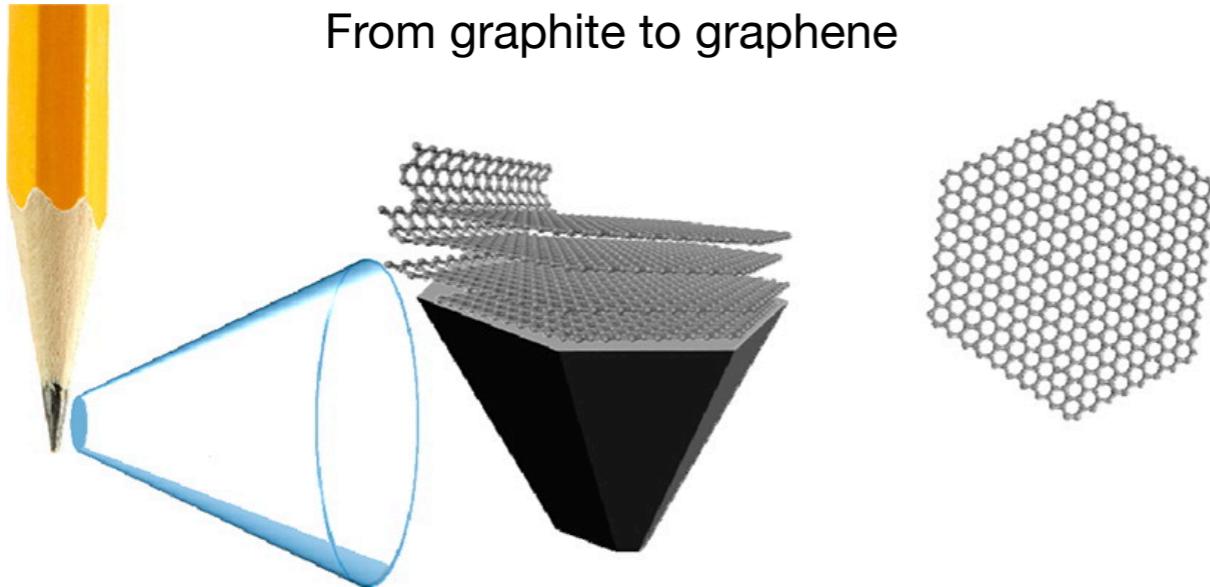


$$H_{eff}[\mathcal{C}] = E_0 + \sum_{(i,\tau);(j,\tau')} J_{(i,\tau);(j,\tau')}^{eff} S_{i,\tau}^z S_{j,\tau'}^z + \dots$$

- ▷ Reduce the computation complexity from  $O(N^4)$  to  $O(N^3)$
- ▷ Increase the acceptance rate of Monte Carlo
- ▷ Learn the effective model for quantum materials
- ▷ Save the planet

# Computation and AI in graphene face masks

From graphite to graphene



The Nobel Prize in Physics 2010

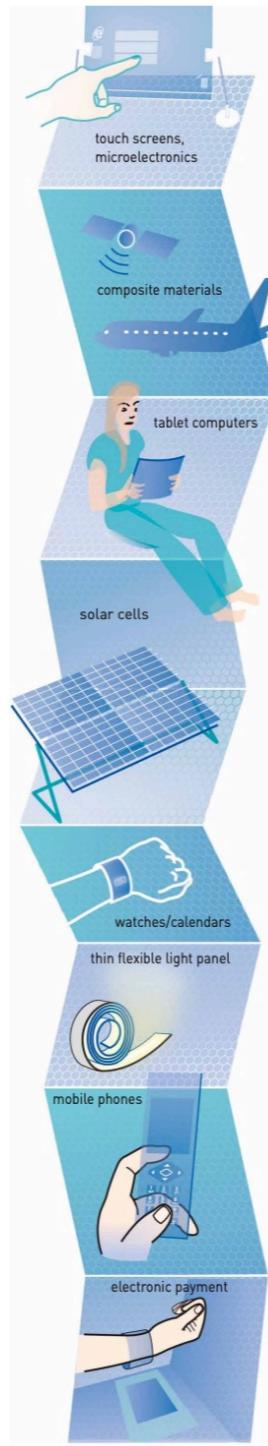


© The Nobel Foundation. Photo:  
U. Montan  
Andre Geim  
Prize share: 1/2

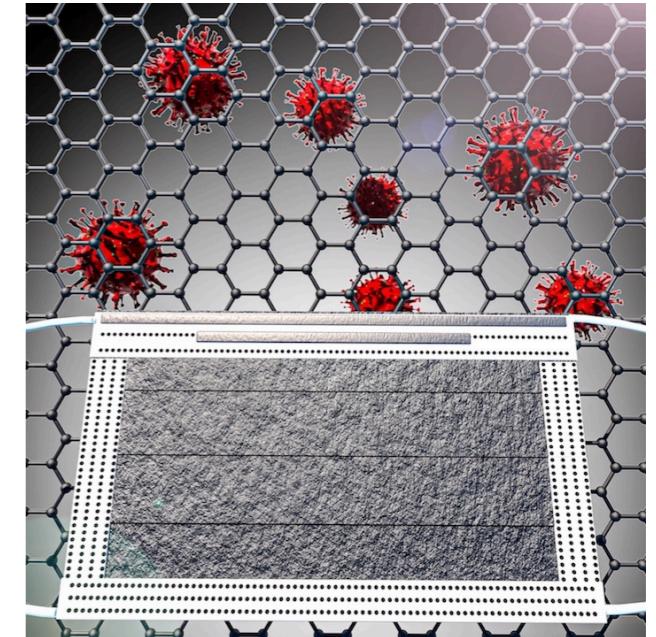


© The Nobel Foundation. Photo:  
U. Montan  
Konstantin Novoselov  
Prize share: 1/2

The 2000 Ig Nobel Prize in physics was awarded to [Andre Geim](#), Radboud University Nijmegen, and [Michael Berry](#), University of Bristol, UK, for the [magnetic levitation](#) of a live frog.



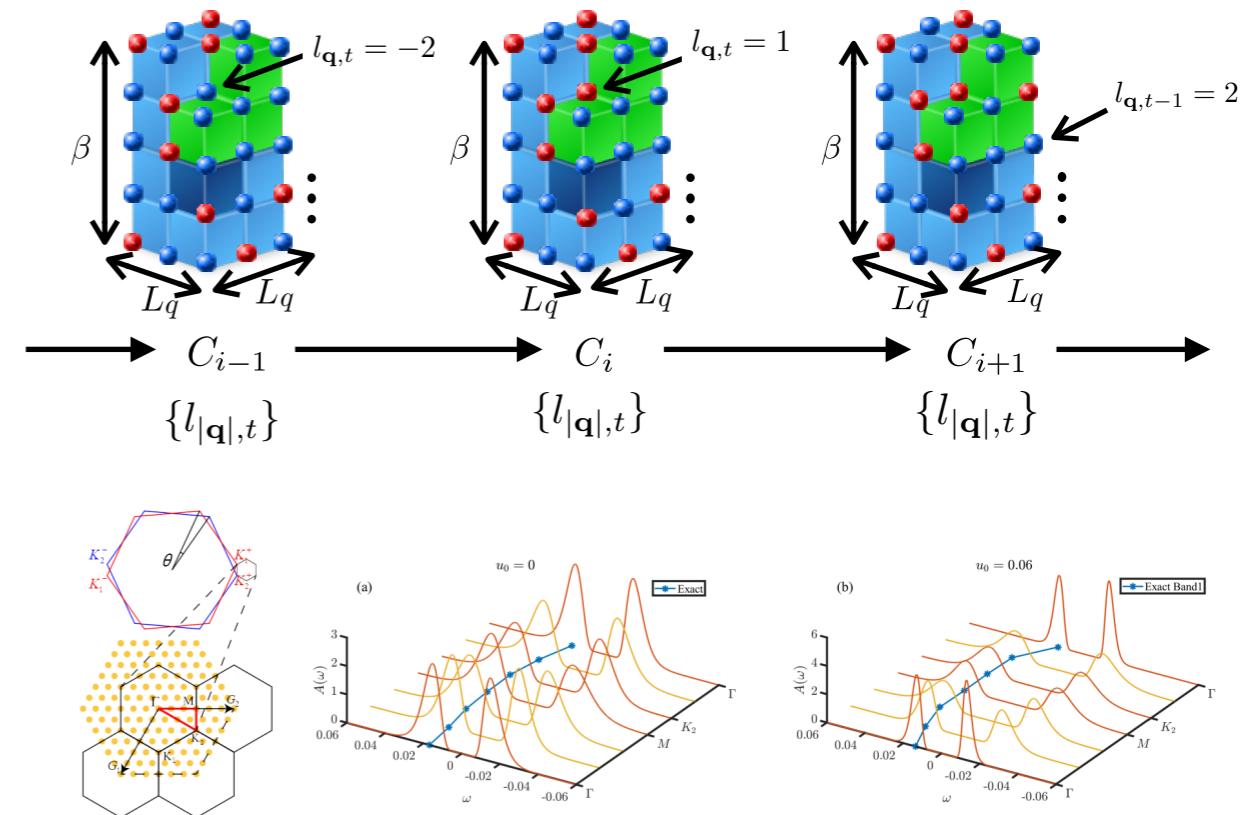
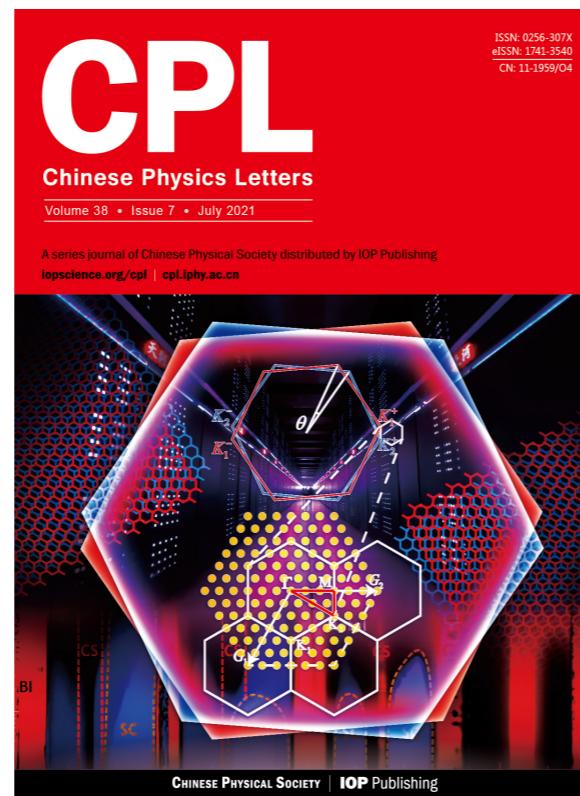
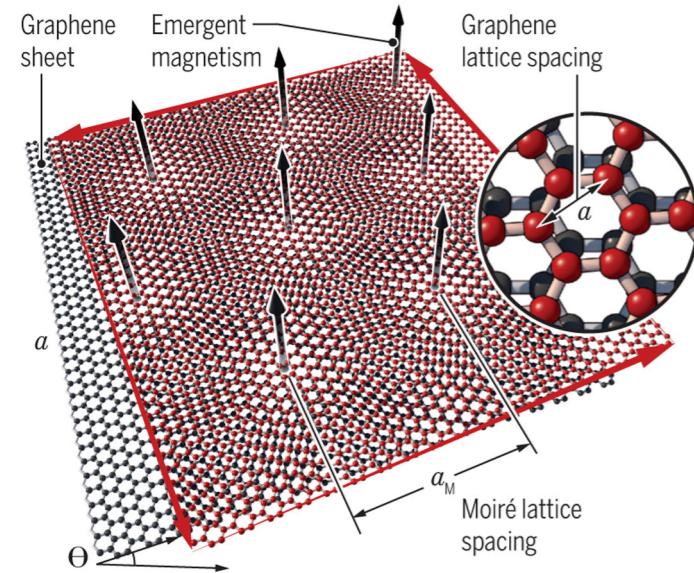
Graphene joins the fight against COVID-19



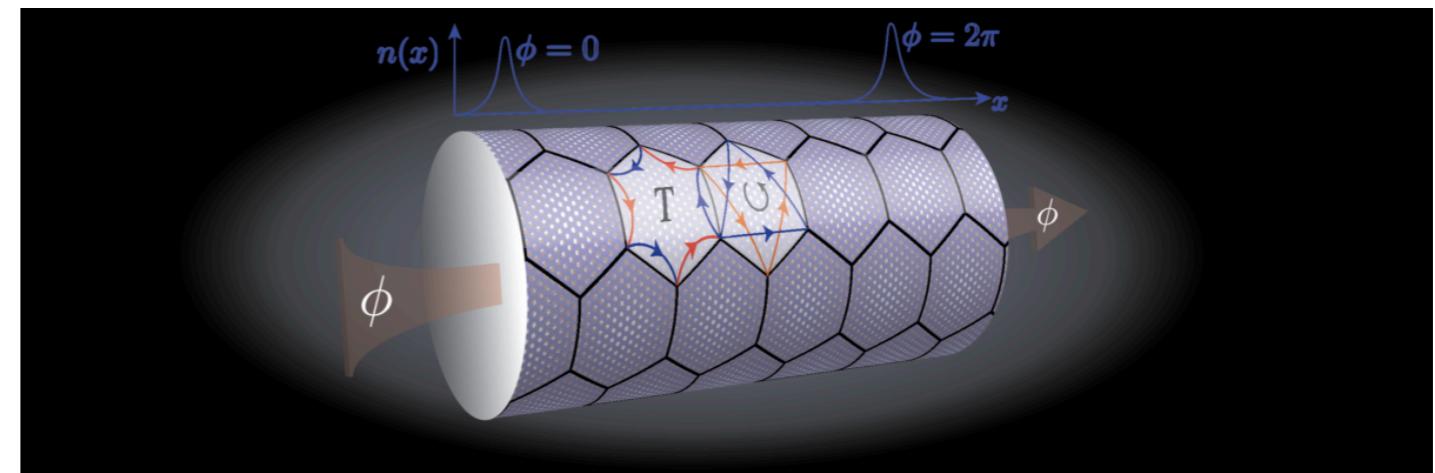
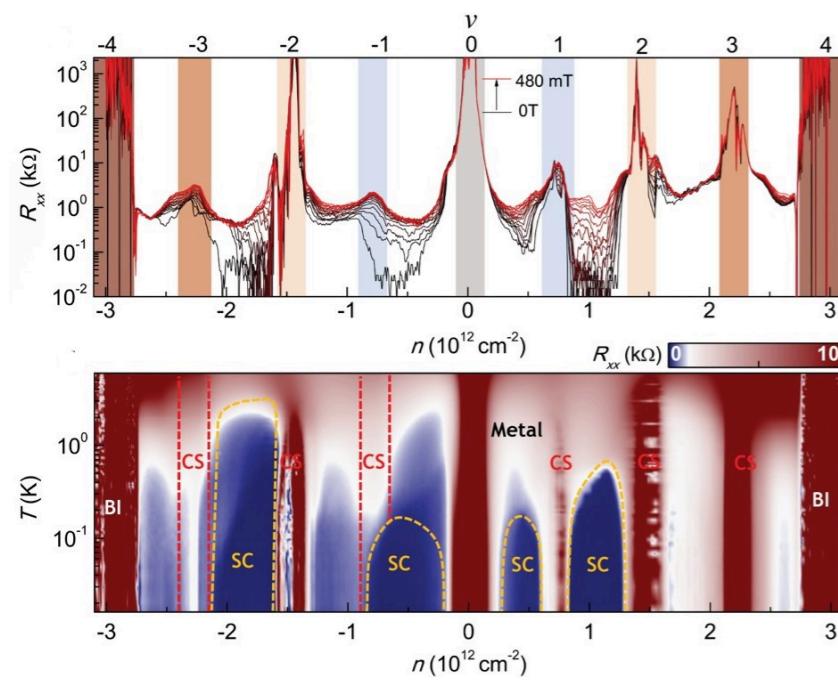
# Computation and AI in graphene face masks

## Twisted bilayer graphene

The two sheets are twisted by a small angle ( $\Theta$ ), creating a Moiré pattern that makes the bilayer both electrically insulating, with conducting edge states (red arrows), and magnetic.



[Chin. Phys. Lett. 38, 077305 \(2021\)](#) Cover story



[Nature Communications 12, 5480 \(2021\)](#)

# Momentum space quantum Monte Carlo algorithm

## Long-range Coulomb + fragile topology



$$H = H_0 + H_{int}$$

$$H_0 = \sum_m \sum_{\mathbf{k}, \tau, s} \epsilon(\mathbf{k}) d_{\mathbf{k}, m, \tau, s}^\dagger d_{\mathbf{k}, m, \tau, s}$$

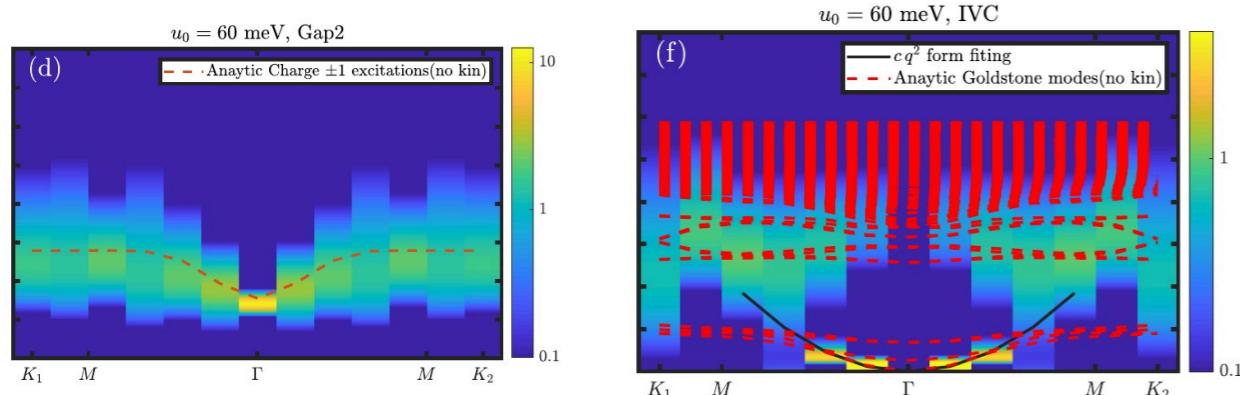
$$V(q) \sim \frac{1}{q}$$

$$\delta\rho_{\mathbf{q}+\mathbf{G}} \sim \sum_{\mathbf{k} \in mBZ, m_1, m_2, \tau, s} \lambda_{m_1, m_2, \tau}(\mathbf{k}, \mathbf{k} + \mathbf{q} + \mathbf{G}) d_{\mathbf{k}, m_1, \tau, s}^\dagger d_{\mathbf{k} + \mathbf{q}, m_2, \tau, s}$$

$$H_{int} = \frac{1}{2\Omega} \sum_{\mathbf{q}, \mathbf{G}} V(\mathbf{q} + \mathbf{G}) \delta\rho_{\mathbf{q}+\mathbf{G}} \delta_{-\mathbf{q}-\mathbf{G}}$$

(a) Hexagonal Brillouin zone with high-symmetry points  $\Gamma$ ,  $K_1$ ,  $K_2$ ,  $M$ ,  $G_1$ ,  $G_2$ .  
(b) Dispersion relations for  $L=6$  at  $u_0 = 33$  meV. Blue lines represent GapI, red dashed lines represent Bare Dispersion.  
(c) Dispersion relations for  $L=6$  at  $u_0 = 60$  meV. Blue lines represent GapI, red dashed lines represent Bare Dispersion.

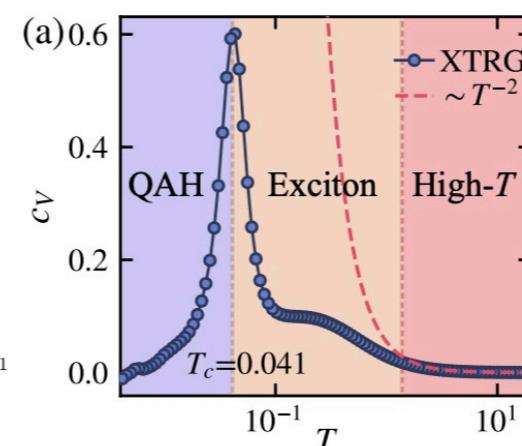
## Collective excitations



- [Phys. Rev. Lett. 130, 016401 \(2023\)](#)  
[Phys. Rev. Lett. 128, 157201 \(2022\)](#)  
[Phys. Rev. B 105, L121110 \(2022\)](#)  
[Nature Communications 12, 5480 \(2021\)](#)

<https://youtu.be/c5-bFYEL028>

## Tensor-network, thermodynamic computation



**nature** awards science in shorts

### Science in Shorts playlist 2022

Watch and share the best Shorts from 2022. Which is your favourite?



### The quantum teleportation pencil

Are you sitting comfortably? And concentrating hard? We're entering the realms of superfast computing based on quantum teleportation via twisted graphene lattices. But don't worry, you'll be fine as long as you have a pencil handy.

# AI & Machine Learning Basics

**Tradition:**      **Task**                  —>    **algorithm (algorithm for loop, sorting)**

**Big data era: Don't have algorithm**   —>    **lack in knowledge, make up for in data**

Approximation detect certain patterns or regularities, Data Mining

- Model with some parameters, model can be predictive or descriptive.
- Learning is the execution of a computer program to optimise the parameters of the model using the training data or past experience.
- Using theory of statistics, math and physics: building mathematical models, making inference from a sample
- Using computer science: efficient algorithm to solve the optimisation problem, store and process big data; representation and algorithmic solution for inference needs to be efficient
- The computational efficiency may be as important as predictive accuracy

# AI & Machine Learning Basics

## Infer hidden association rule from observed data In the era of “big data”

### Basket analysis

- In retail, associations between products bought by the customers
- People do not buy at random
- There are certain patterns (association rule) in the data, machine extract them

Conditional probability  $P(Y|X)$ , Y is the product one would like to condition on X,

$$P(\text{chips} | \text{beer}) = 0.8 \quad P(\text{bread} | \text{milk}) = 0.71$$

$P(Y|X, D)$  where D is the set of customs attributes, gender, age, martial ...

Books / Music / Shows :

$$P(\text{Game of Thrones} | \text{Fantasy, male}) =$$

$$P(\text{The daily show} | \text{Comedy, layman}) =$$

$$P(\text{Last Week Tonight with John Oliver} | \text{Comedy, sophisticated}) =$$

$$P(\text{Late-night with Seth Meyers} | \text{Comedy, politics}) = \dots$$

Webpages:

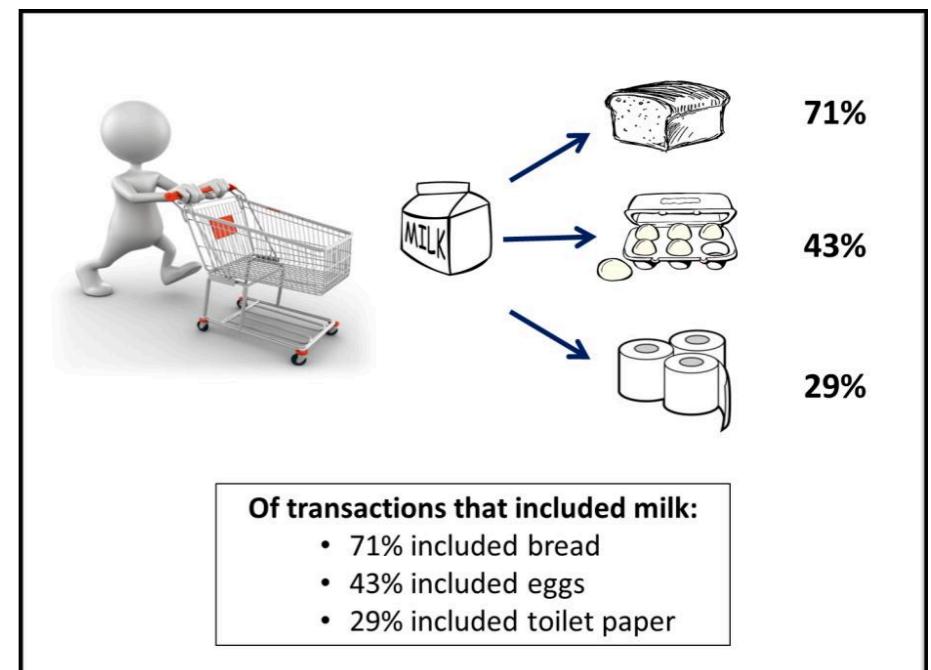
Social medias:

In spam email detection

In Fintech, credit application, stock market

In medical diagnosis, COVID-19

In Science, physics, astronomy, biology, ...



# AI & Machine Learning Basics

**Classification: input → classifier → output**

**Discriminant (two classes):**

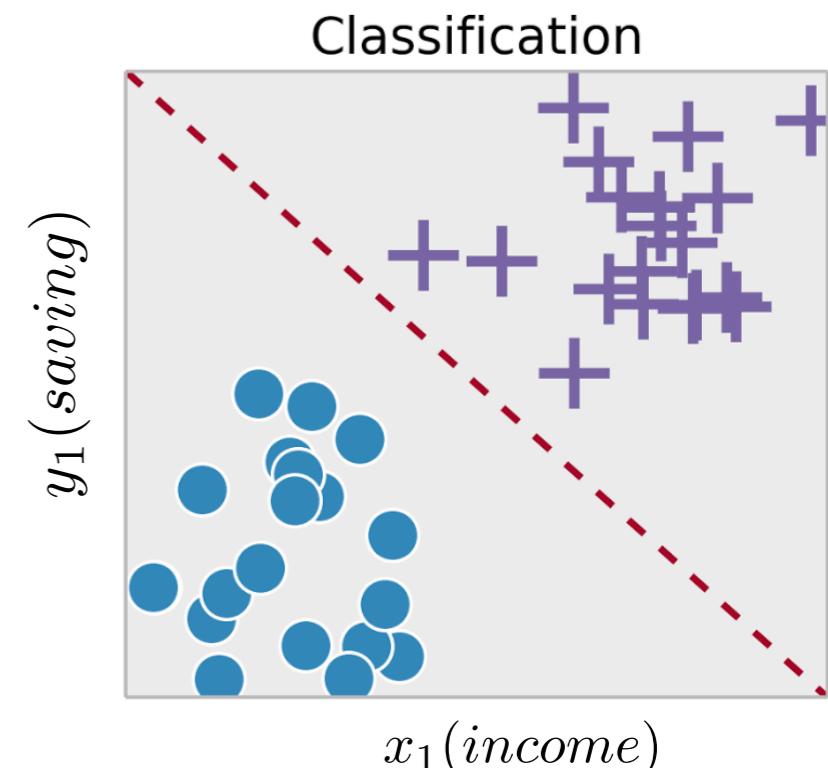
Banks classify credit for low-risk and high-risk customers

income, saving, profession, age, past financial history, ...

Machine learning fits a model to the past data,  
calculate the risk for a new application,  
decide to accept or refuse

We have a rule that fits the past data, if the future is similar to the past

Predictions: decide new customer is low-risk and high-risk



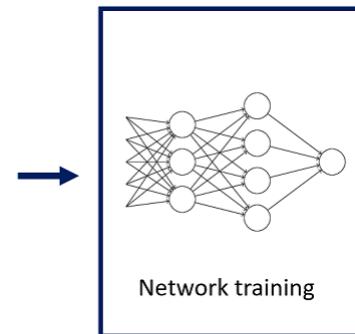
**Pattern recognition (multiple classes):**

Handwritten character recognition, MNIST database

Face recognition, medical diagnosis,

Speech recognition, time series, machine translation,  
natural language processing

A 4x10 grid of handwritten digits from 0 to 9, representing the "Data & Labels" for training a neural network.



0  
1  
2  
3  
4  
5  
6  
7  
8  
9

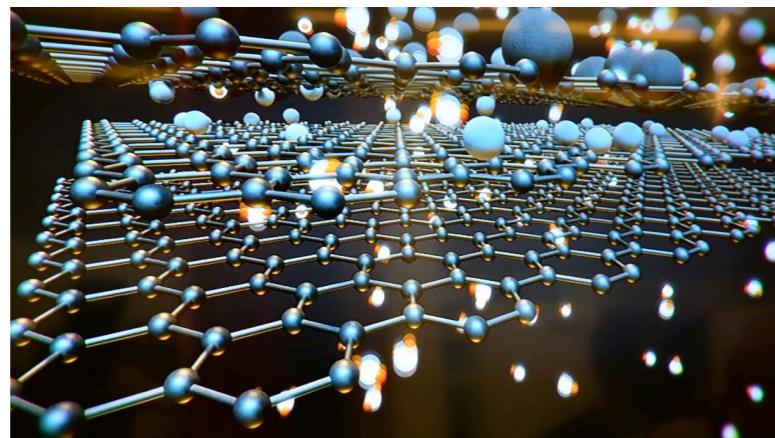
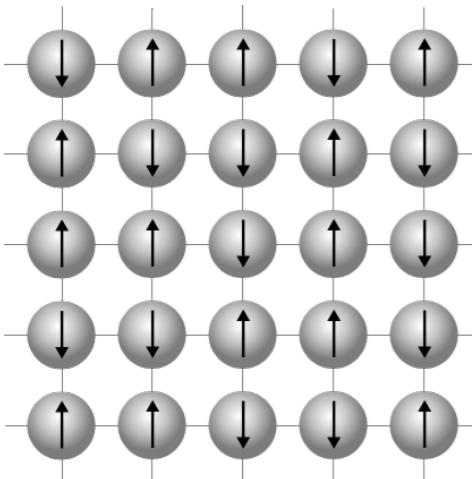
**Knowledge extraction: rule is simpler than data**

Phase transition, Landau-Ginzburg paradigm, order parameters

# AI & Machine Learning Basics

Ising model, continuous phase transition  
workhorse for statistical physics

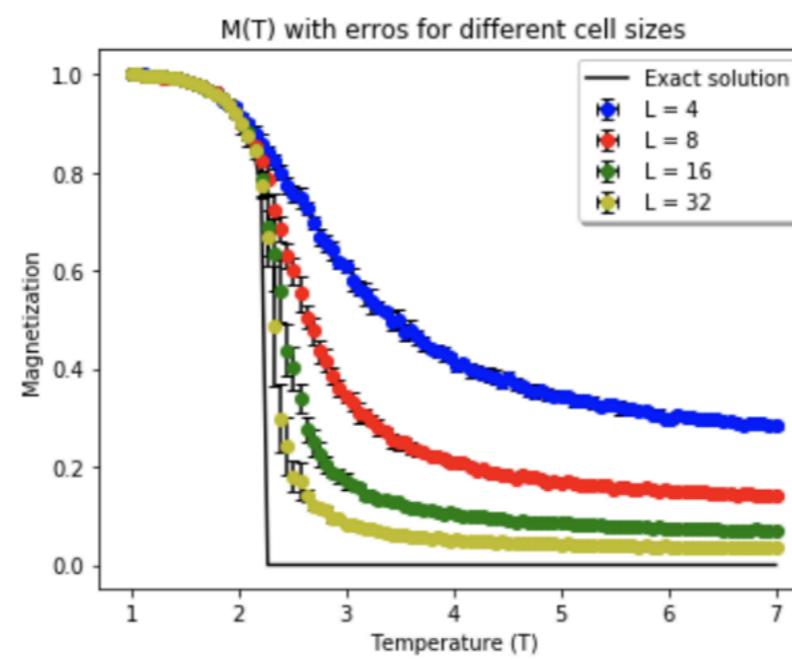
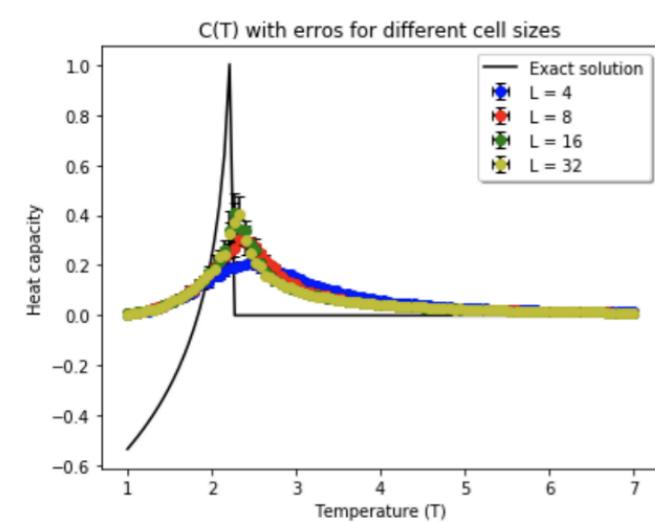
<https://mattbierbaum.github.io/isng.js/>



$$H = -J \sum_{\langle i,j \rangle} S_i^z S_j^z \quad S_i^z = \pm 1$$

Configuration space:  $2^N$

[https://en.wikipedia.org/wiki/Ising\\_model#/media/File:Ising\\_quench\\_b10.gif](https://en.wikipedia.org/wiki/Ising_model#/media/File:Ising_quench_b10.gif)



$$m = \frac{1}{N} \left| \sum_{i=1}^N S_i^z \right|$$

$$m(T) = |T - T_c|^\beta \text{ with } \beta = 1/8 \text{ in 2D}$$

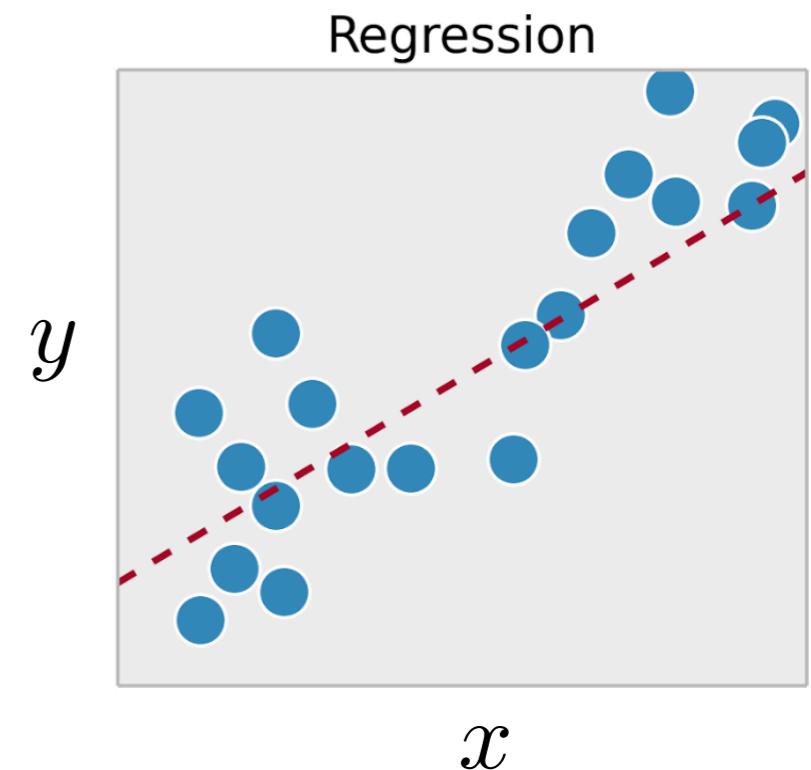
# AI & Machine Learning Basics

**Regression:**  $y = h_{\Theta}(x) = \Theta \cdot x$

$$\{(x_j^{(i)}, y^{(i)}), \theta_j\}; j = 1, 2, \dots, N; i = 1, 2, \dots, M; N < M$$

$$y^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \dots + \theta_N x_N^{(i)}$$

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_N^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1^{(M)} & x_2^{(M)} & \dots & x_N^{(M)} \end{bmatrix} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(M)} \end{bmatrix}$$



**Prediction & forecasting:**

$$\underline{\underline{X}} \cdot \underline{\Theta} = \underline{Y}$$

Least squares by Legendre 1805 and Gauss 1809

normal equation, Gradient descent and Conjugate Gradients, Lagrange multiplier

- Model / hypothesis

$$h_{\Theta}(x)$$

- Loss / cost function

$$\mathcal{L}(\Theta|X) = \sum_{i=1}^M L(y^{(i)}, h_{\Theta}(x^{(i)}))$$

- Optimisation procedure

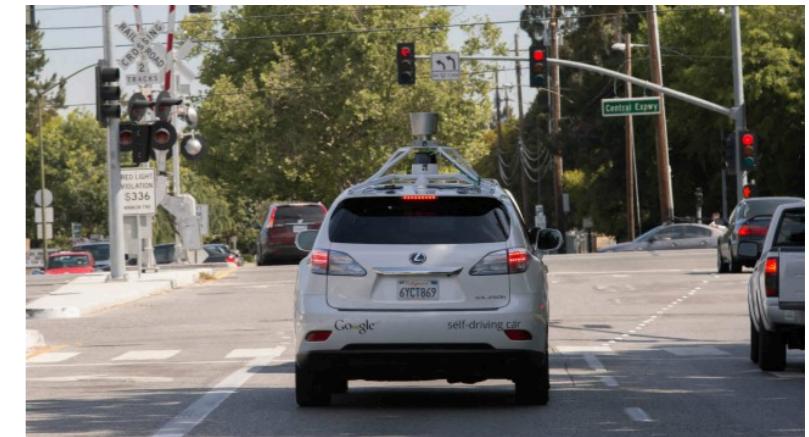
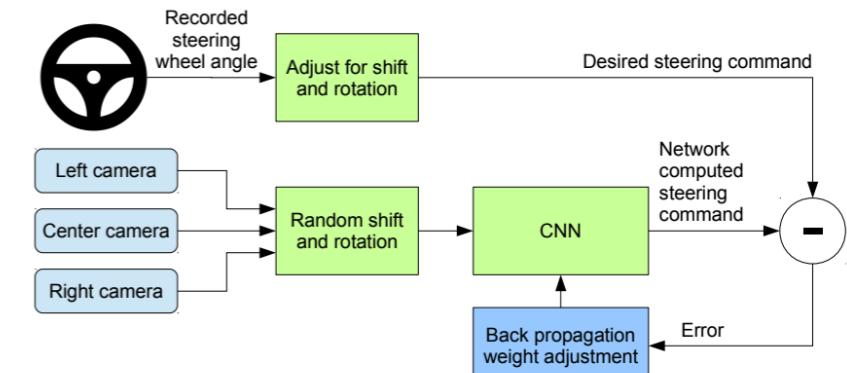
$$\Theta^* = \arg \min_{\Theta} \mathcal{L}(\Theta|X)$$

## Example: Self-Driving car

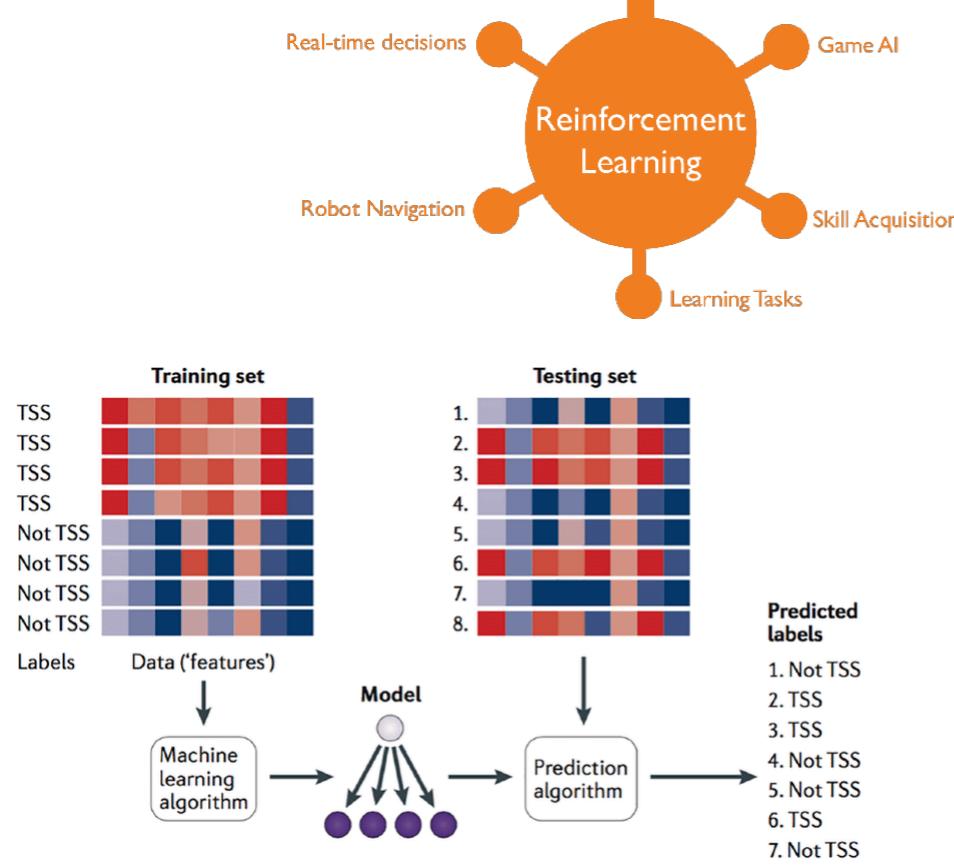
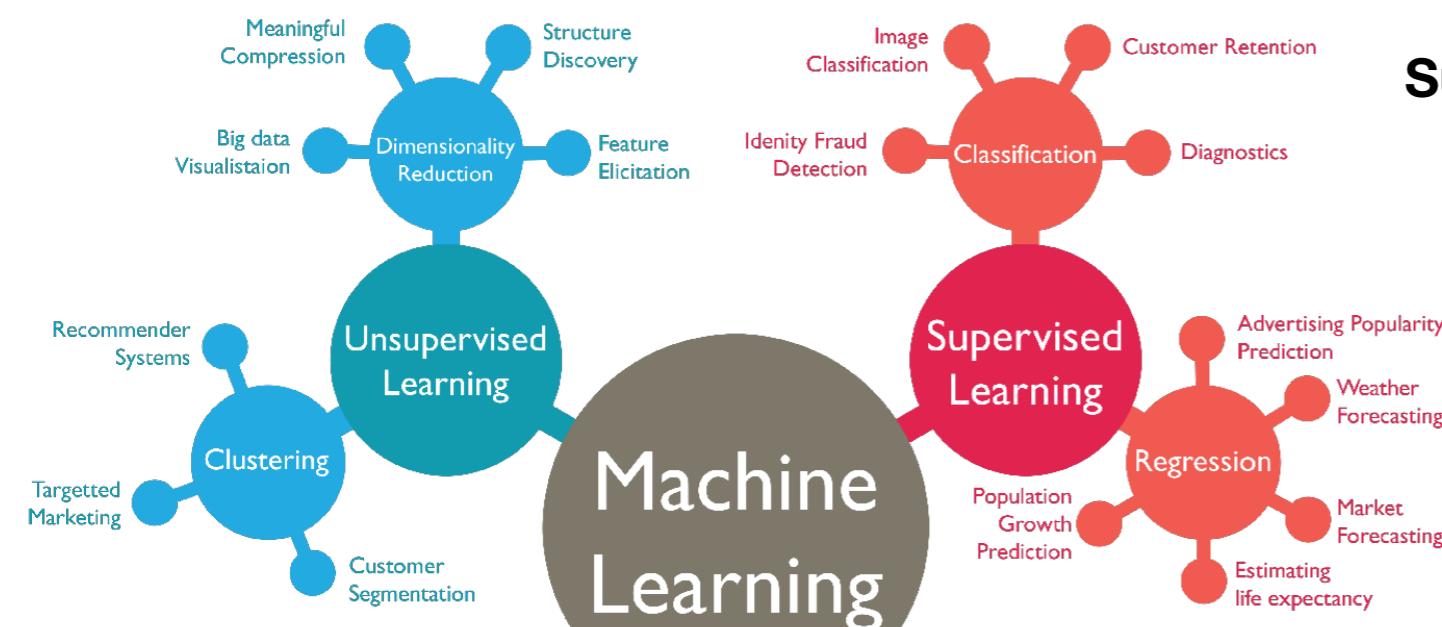
Input: sensors on the car, video camera, GPS, ...;

Output: steering wheel;

Training data: monitoring and recording the action of human driver



# AI & Machine Learning Basics



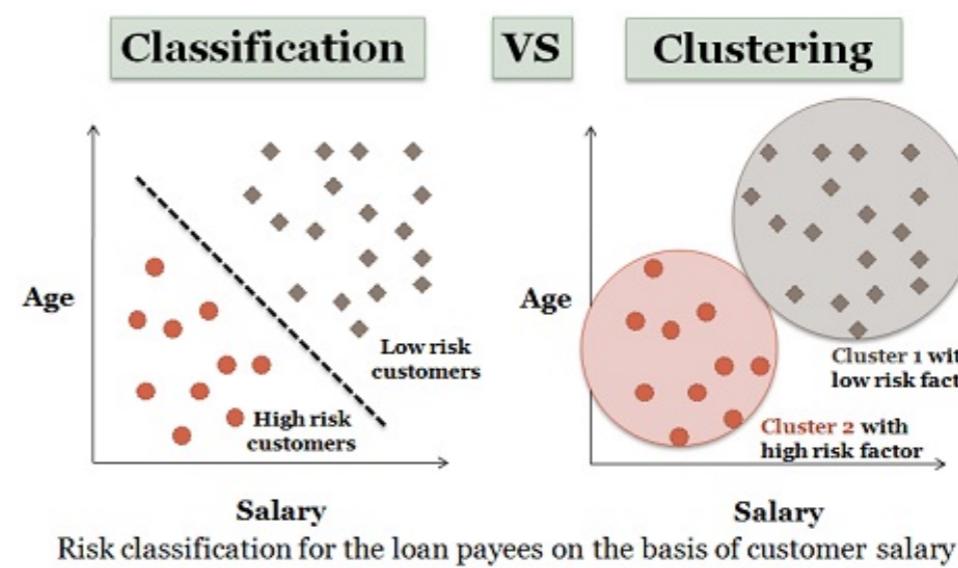
## Supervised Learning: Classification & Regression

Input → machine/model → Output  
Correct outputs are provided by the supervisor

## Unsupervised Learning: only have input data

Find regularities from the input

**Clustering:**  
customer segmentation, customer relationship management, outlier detection; Image compression  
bioinformatics: DNA, RNA, amino acids, Motif, Proteins, sequence alignments



# AI & Machine Learning Basics

**Reinforcement learning (policy generation):  
Single action is not important, good policy is the sequence of correct actions.**

**Game playing:**



AlphaGo is CNN with 12 convolution layers

**Robot navigation:**



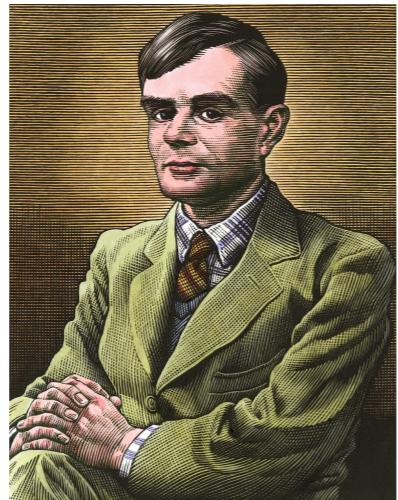
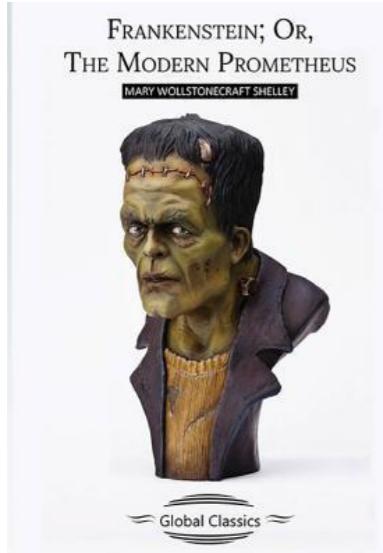
Correct sequence of action to reach the goal state from an initial state

**Watch this!**

<https://www.bostondynamics.com/spot>

# AI & Machine Learning Basics

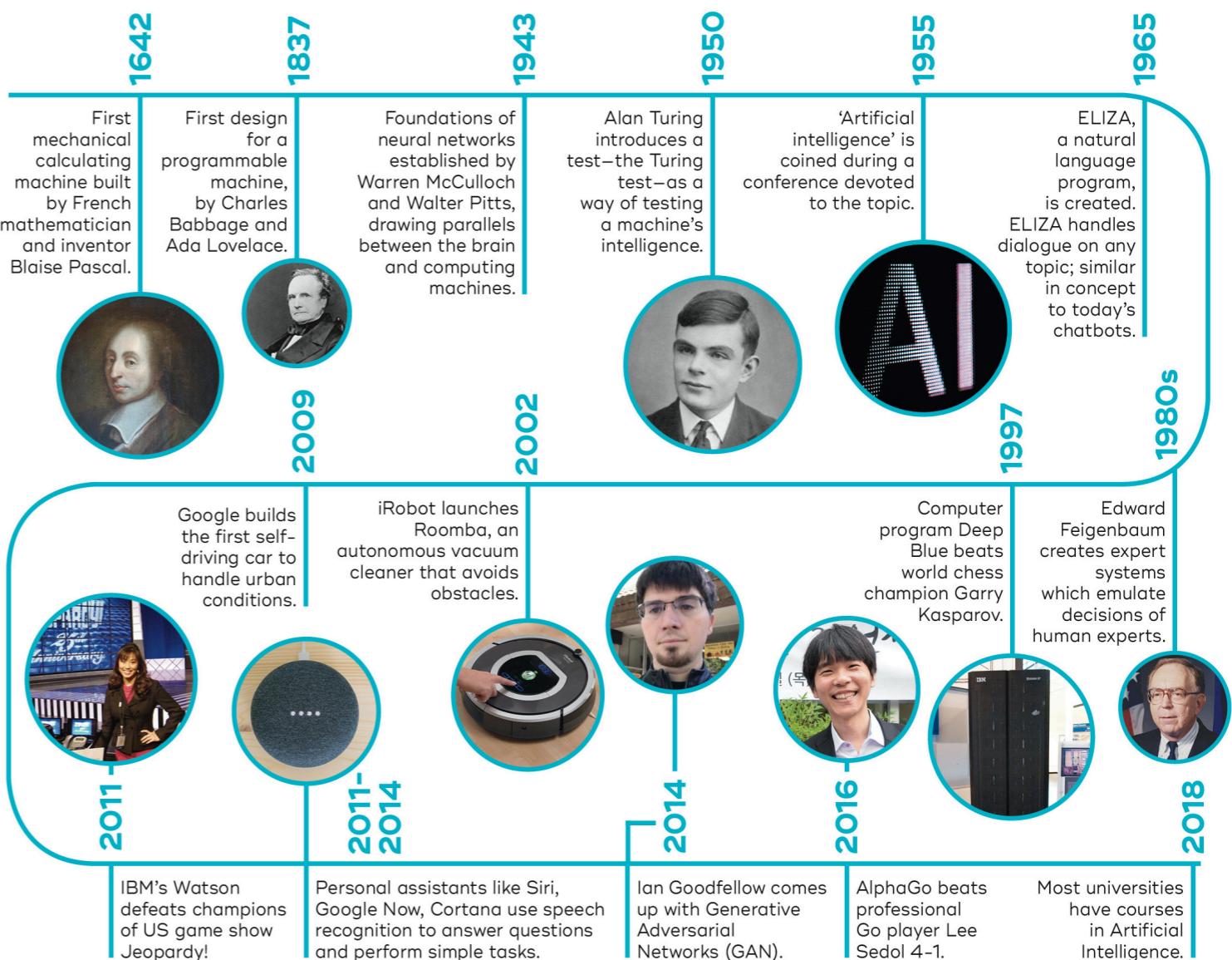
## A bit of history



- 1943, McCullouch and Pitts, artificial neurons
- 1955, workshop at Dartmouth College, Allen Newell (CMU), Herbert Simon (CMU), John McCarthy (MIT), Marvin Minsky (MIT), ...
- 1958, Rosenblatt, perceptron
- 1974, first AI winter
- 1987, second AI winter
- 1997, IBM Deepblue vs Kasparov, logistics, data mining, medical diagnosis, ...
- 2016, Alpha Go vs Lee Sedol
- 2017, Alpha Go vs Ke Jie
- Xbox, Smartphone, affordable neural networks, cloud computing, internet of things ...
- 2020, COVID-19

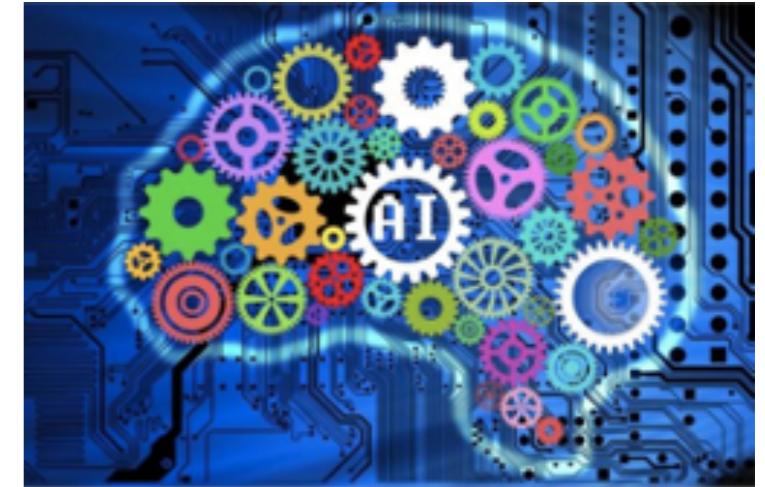
Church-Turing thesis:

If a human could not distinguish between responses from a machine and a human, the machine could be considered “intelligent”.



# Content

## 0. Introduction



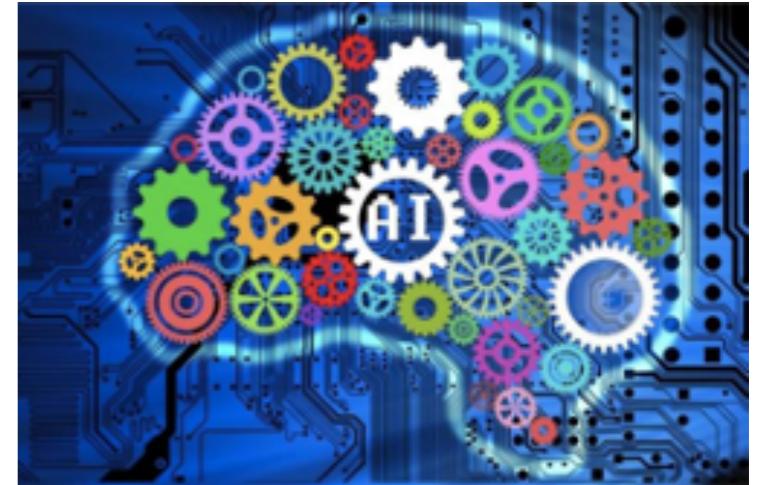
## 1. Regression

- 1.1 Multivariate Linear Regression (curve fitting)
- 1.2 Regularization (Lagrange multiplier)
- 1.3 Logistic Regression (Fermi-Dirac distribution)
- 1.4 Support Vector Machine (high-school geometry)

## 2. Dimensionality Reduction/feature extraction

- 2.1 Principal Component Analysis (order parameters)
- 2.2 Recommender Systems
- 2.3 Clustering (phase transition)

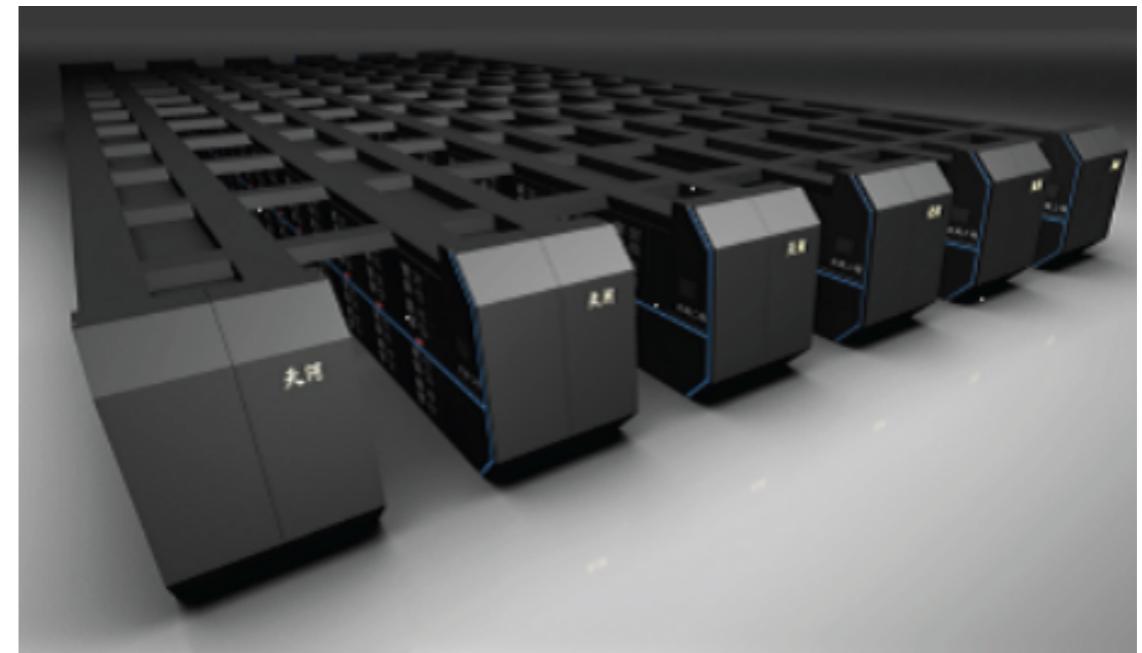
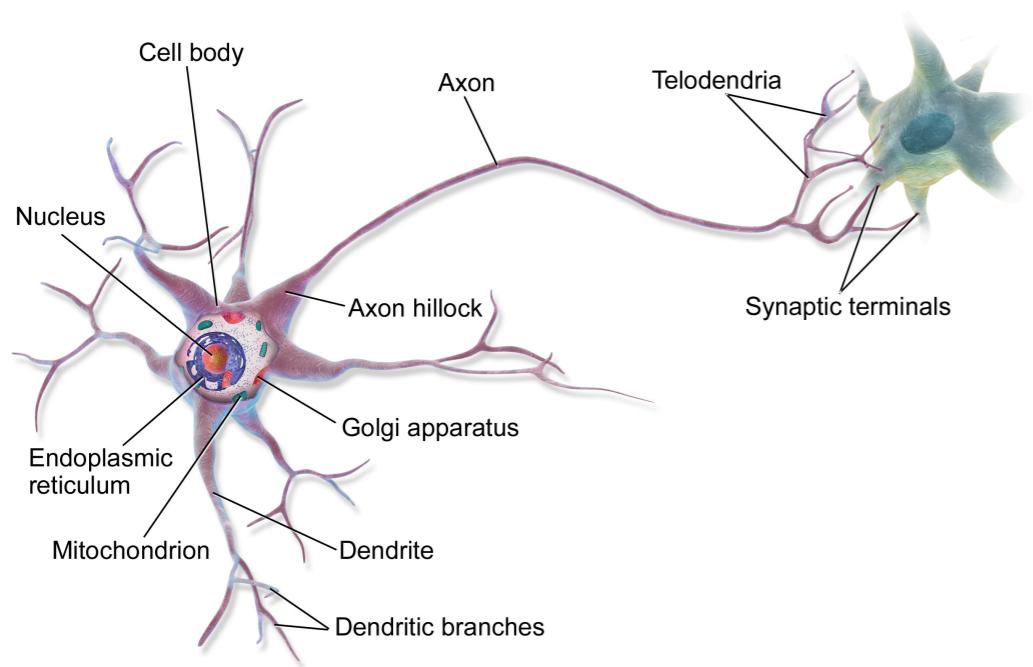
# **Content**



## **3. Neural Networks**

- 3.1 Biological neural networks**
- 3.2 Mathematical representation**
- 3.3 Factoring biological ingredient**
- 3.4 Feed-forward neural networks**
- 3.5 Learning algorithm**
- 3.6 Universal Approximation Theorem**

# AI & Machine Learning Basics



	Supercomputer	Personal Computer	Human Brain
Computational Units	32,000 Xeon CPUs $10^{12}$ transistors	4 CPUs, $10^9$ transistors	$10^{11}$ neurons
Storage units	$10^{14}$ bits RAM $10^{15}$ bits Storage	$10^{11}$ bit RAM $10^{13}$ bit Storage	$10^{11}$ neurons $10^{14}$ synapses
Cycle time	$10^{-9}$ sec	$10^{-9}$ sec	$10^{-3}$ sec
Operations/sec	$10^{15}$	$10^{10}$	$10^{17}$
Memory updates/sec	$10^{14}$	$10^{10}$	$10^{14}$
Weight / Space	150 tons / Basketball court	1 Kg / A4 Paper	1.5 Kg / 1/6 basketball
Power consumption	500 megawatt	100 watt	20 watt

# AI & Machine Learning Basics

Evolution gives human a large brain and a mechanism to learn, such that we could update ourselves with experience and adapt to different environments.

All sciences are fitting models to data.

Induction is a process of extracting general rules from a set of particular cases.

But we are at the point such analysis can no longer done by people, we need computer (machine) to learn for us.

AI is multidisciplinary

Physics, mathematics, chemistry, ...

Cognitive science

Statistics: association is inference, learning is estimation, classification is discriminate analysis.

Engineering: classification is pattern recognition

Vision, speech and robotics, learnt from sample data

Signal processing in vision and speech recognition

Neural networks, kernel-based algorithms (SVMs) in bioinformatics and language processing

Generative models, explain data through interaction of hidden factors

Lots of example data and sufficient computing power, reduced cost of storage, larger dataset over internet and cheaper computation, COVID19

Made possible now but not in the 1950s and 60s

Intelligence seems not to originate from outlandish formula, but rather from patient, sometimes brute-force use of a simple, straightforward algorithm.