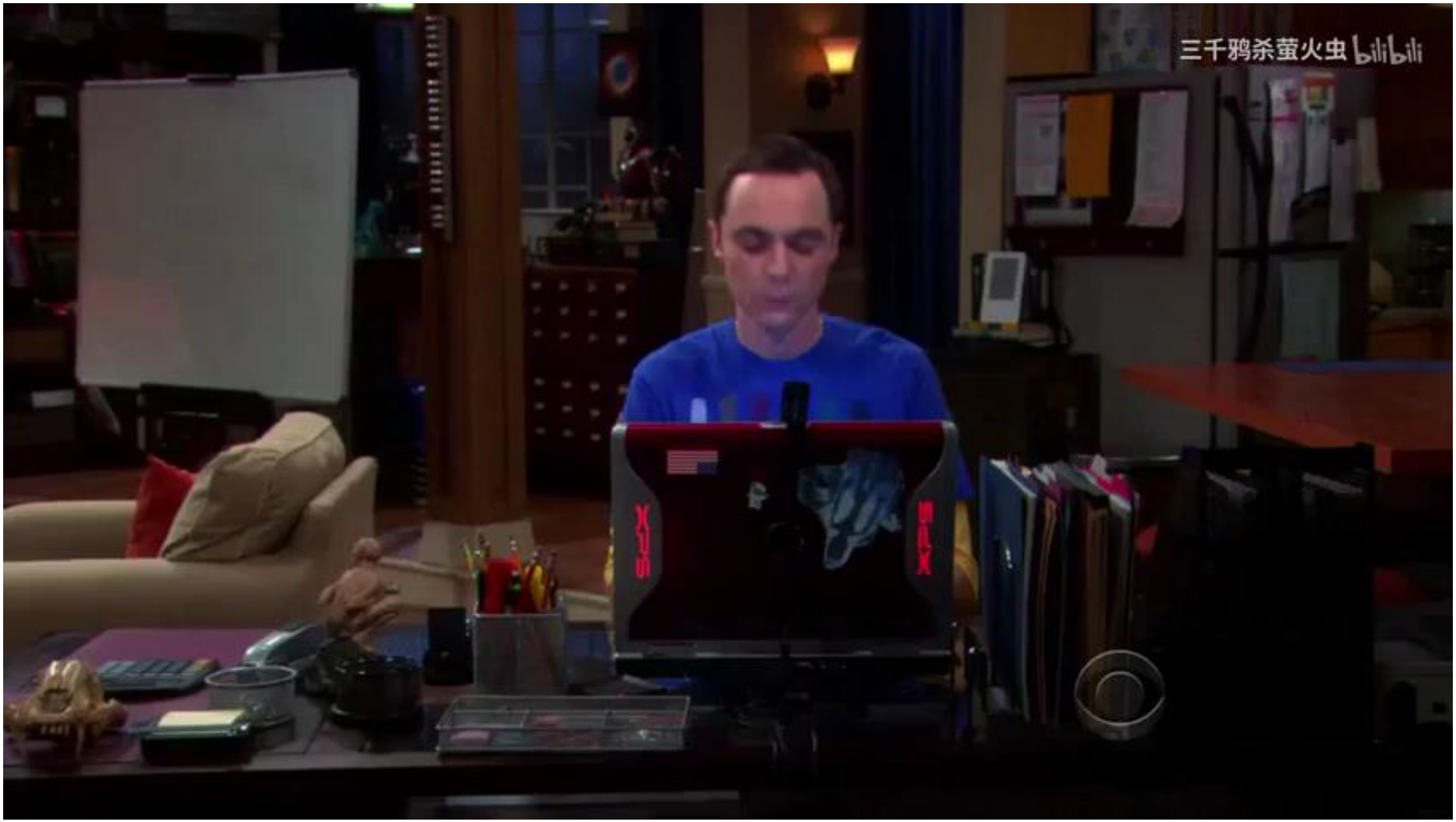
Introduction to Computational and Modeling Tools

Lec 7: Applications to Physics Problems

Instructor: Junwei Liu

10/21/2023

What is physics?



https://www.bilibili.com/video/av14218525/?spm_id_from=333.788.videocard.1

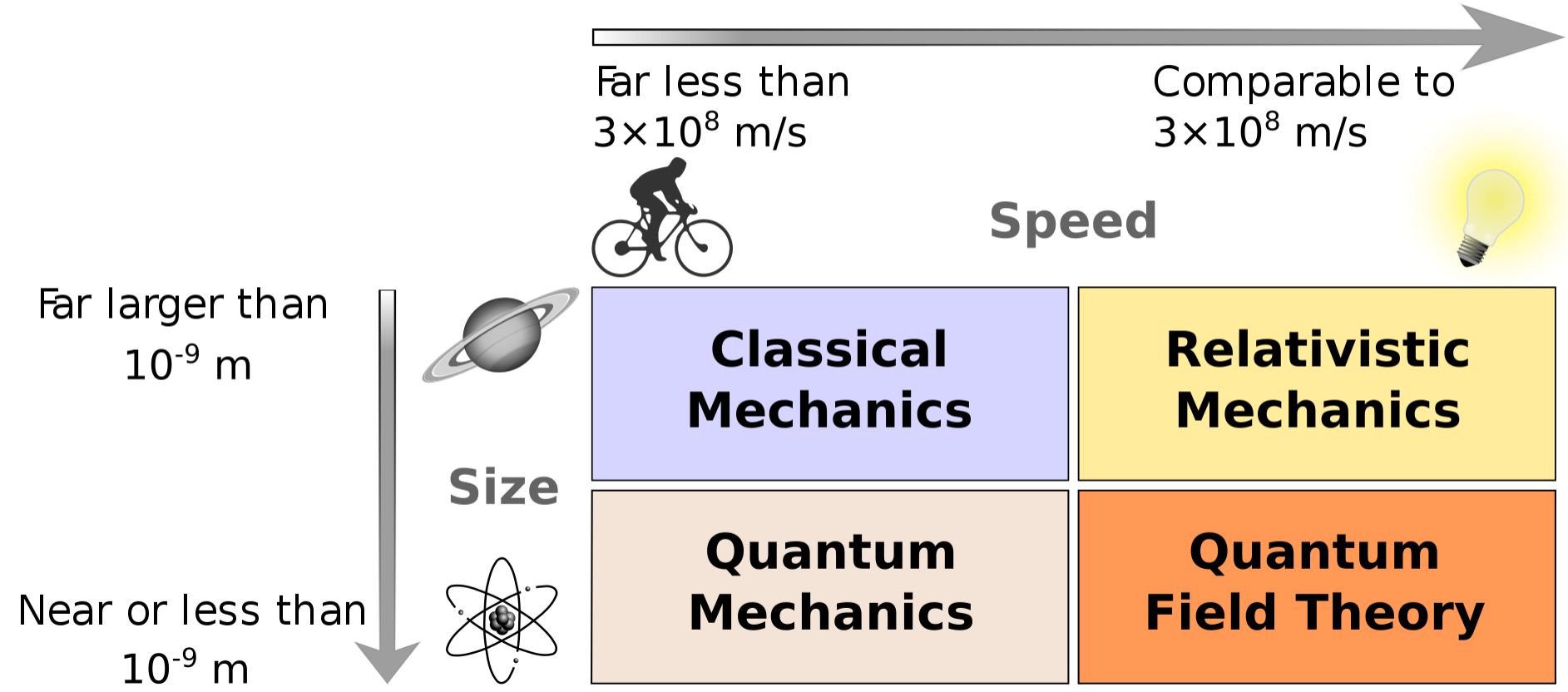
What is physics?

Ancient Greek: φυσική (ἐπιστήμη)
romanized: *physikḗ (epistḗmē)*

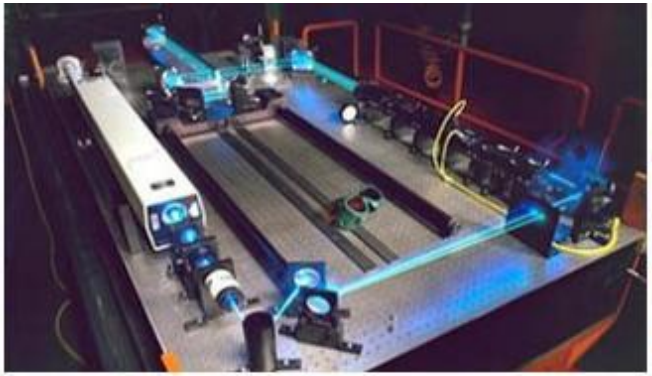
Knowledge of Nature

Physics studies matter, its motion and behavior through space and time, and its main goal is to understand how the universe behaves.

The scope of physics

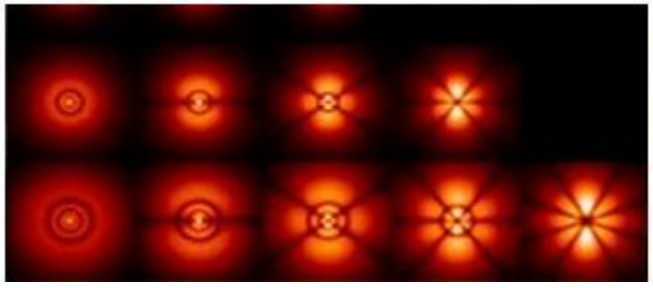
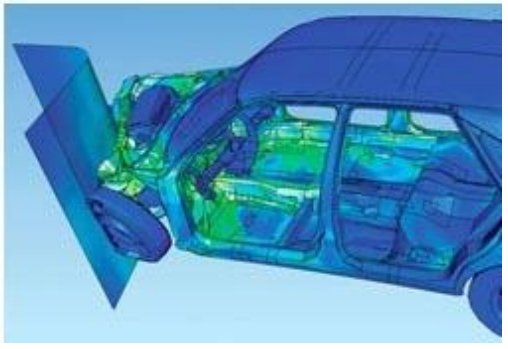


Various physics phenomena





How do we solve physics problems?

Blue ball move to the red ball from the left, and what will happen after the blue ball hit the red ball?

Experiments

https://www.youtube.com/watch?v=51IFubnEAsU

Two balls will exchange their velocities!

Analytical analysis

Momentum conservation law

$$m_b v_{b1} + m_r v_{r1} = m_b v_{b2} + m_r v_{r2}$$

Total energy conservation law

$$\frac{1}{2} m_b v_{b1}^2 + \frac{1}{2} m_r v_{r1}^2 = \frac{1}{2} m_b v_{b2}^2 + \frac{1}{2} m_r v_{r2}^2$$

When $m_b = m_r$, the solution is

$$v_{b2} = v_{r1}$$
$$v_{r2} = v_{b1}$$

A harder problem

How about the movement of stars in the sky?

Experiments

https://www.youtube.com/watch?v=tp6UkqIwVfk

Analytical analysis

Newton's law of universal gravitation

$$m_1 \frac{d^2 r_1}{dt^2} = G \sum_{i \neq 1}^{N} \frac{m_1 m_i}{|r_1 - r_i|^2}$$

$$m_2 \frac{d^2 r_2}{dt^2} = G \sum_{i \neq 2}^{N} \frac{m_2 m_i}{|r_2 - r_i|^2}$$

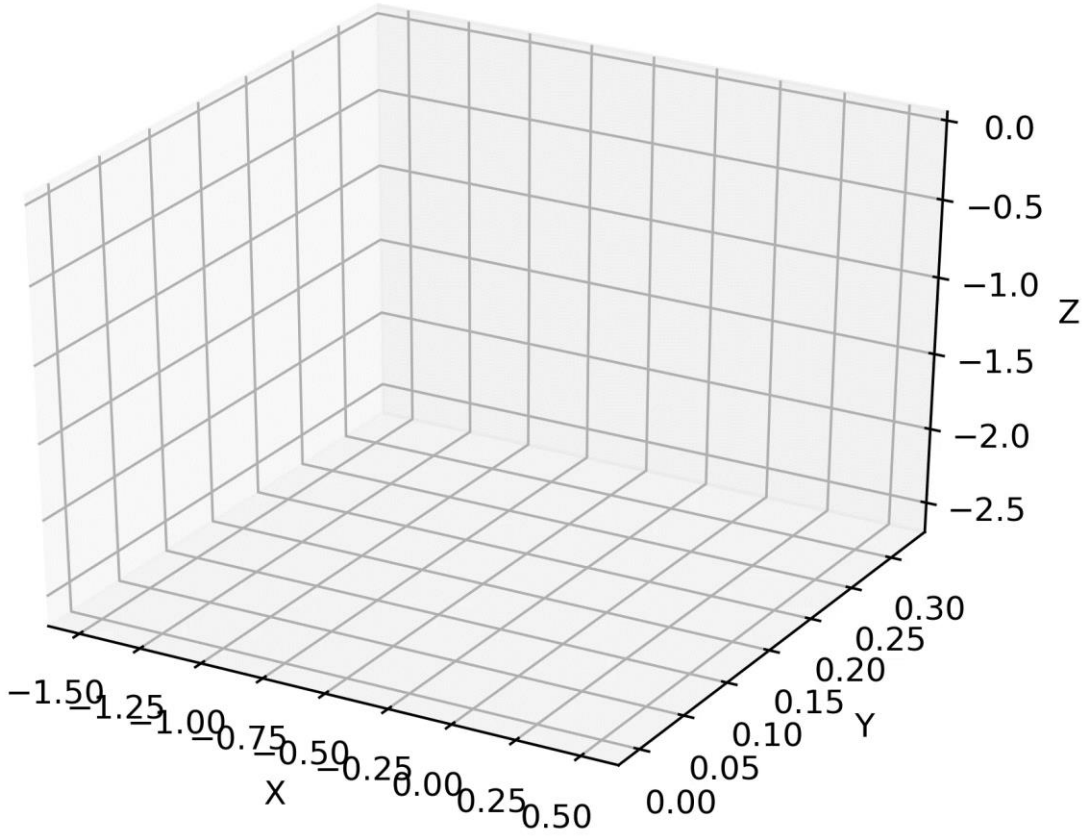
$$\dots$$

$$m_3 \frac{d^2 r_3}{dt^2} = G \sum_{i \neq 3}^{N} \frac{m_3 m_i}{|r_3 - r_i|^2}$$

Impossible to get an exact solution!!

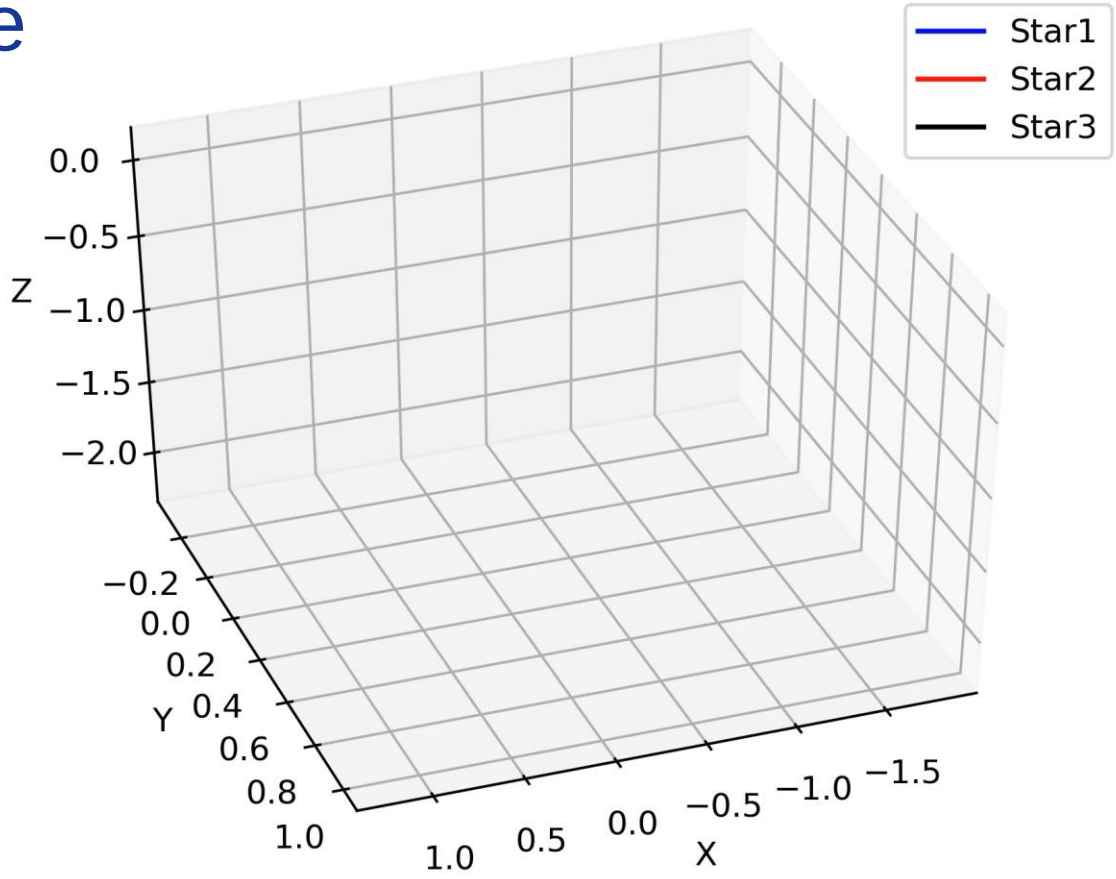
Numerical simulations for two stars

Two stars rotate with each other



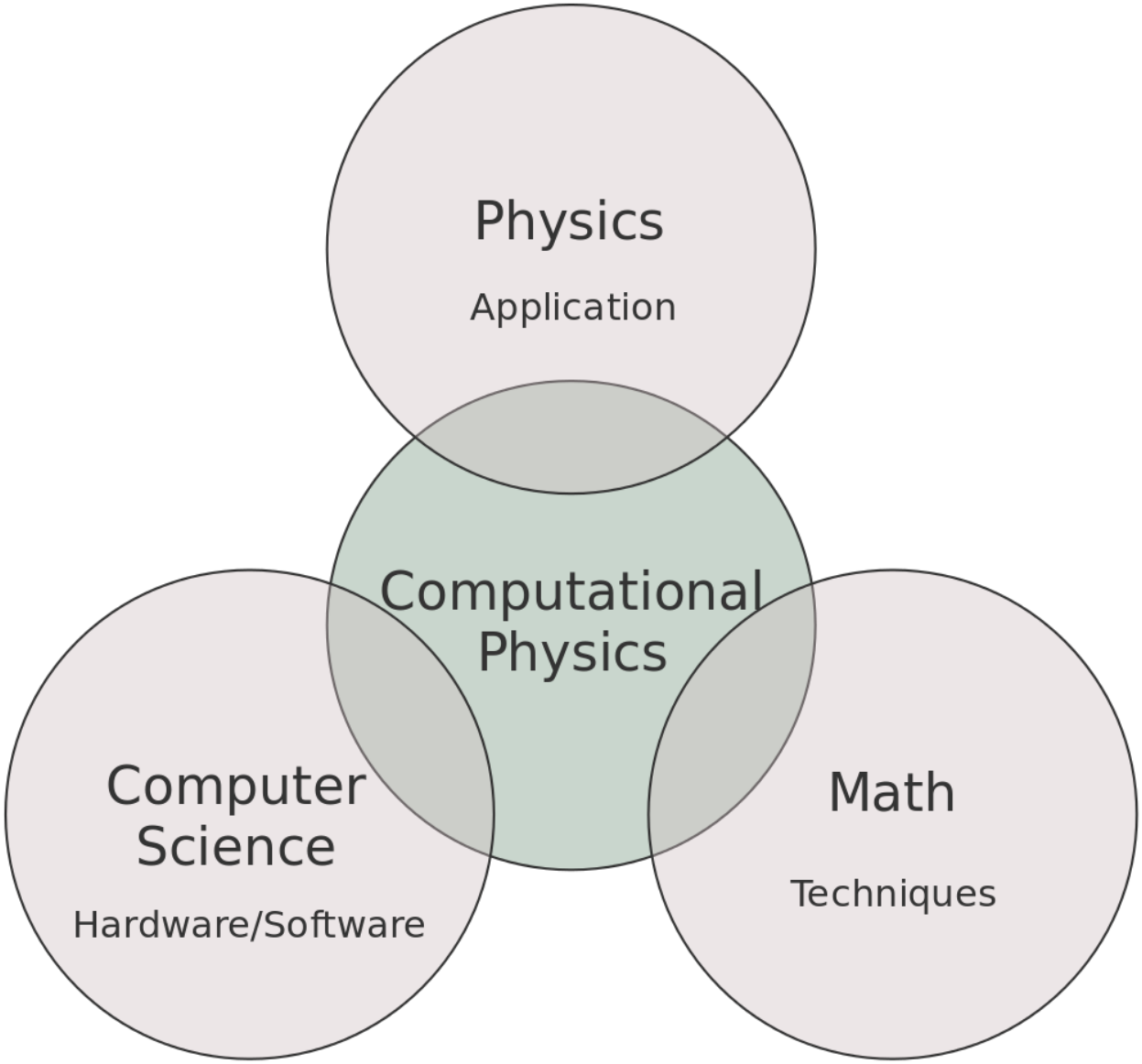
Numerical simulations for three stars

Even hard
to describe



Another example: take picture for black hole

Computational Physics

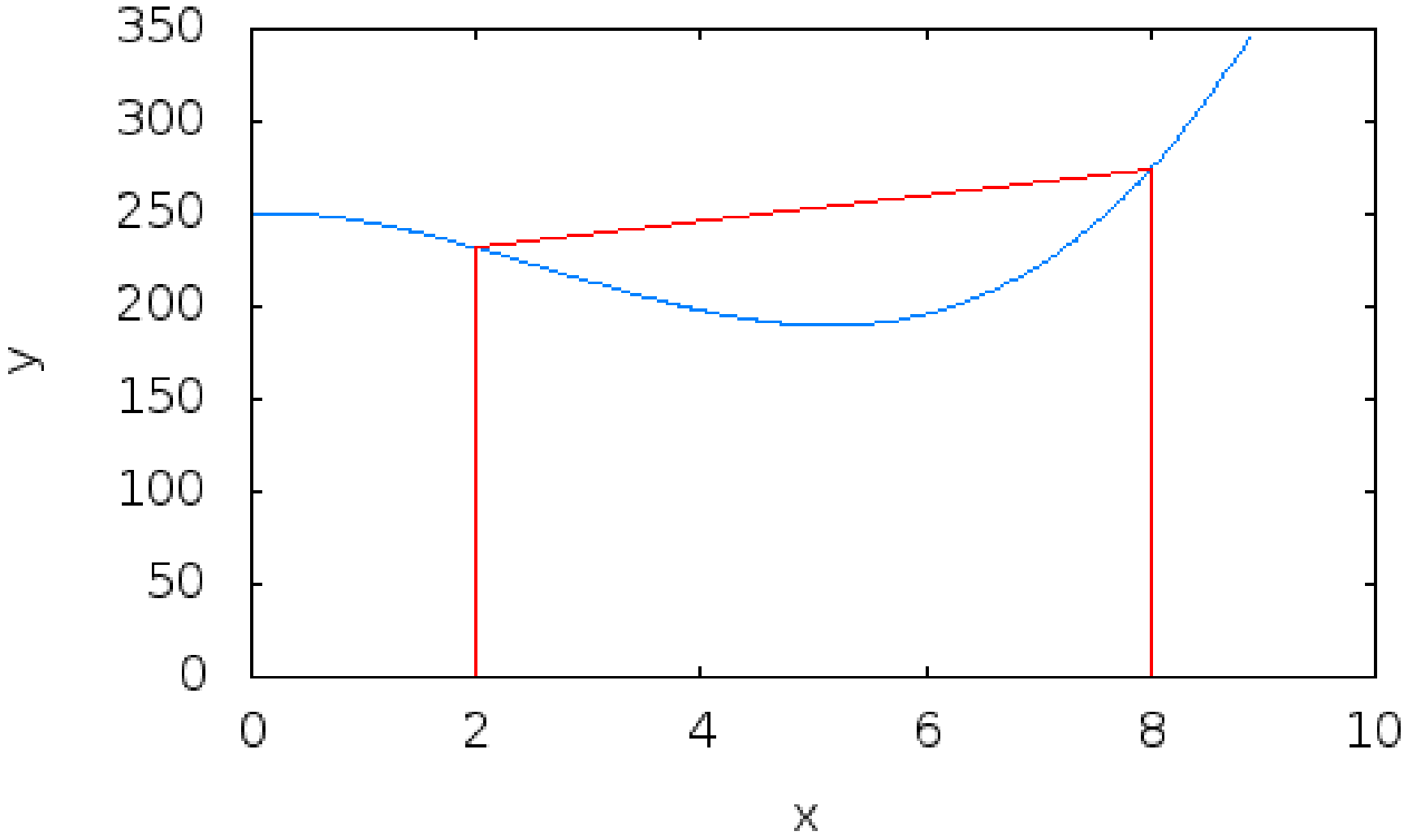


Common methods and algorithms

- Integration (using e.g. Romberg method)
- Root finding (using e.g. Newton-Raphson method)
- Linear algebra (using e.g. LU decomposition)
- Ordinary differential equations (using e.g. Runge–Kutta methods)
- Partial differential equations (using e.g. finite difference method and relaxation method)
- Random process and Monte Carlo

...

Integration



Linear equations (linear algebra)

- A single linear equation in one variable, such as $x - 1 = 0$, is trivial to solve.

- Simultaneous sets of linear equations is also not hard, but it really tedious to solve.

$$
\begin{aligned}
2w + x + 4y + z &= -4, \\
3w + 4x - y - z &= 3, \\
w - 4x + y + 5z &= 9, \\
2w - 2x + y + 3z &= 7
\end{aligned}
$$

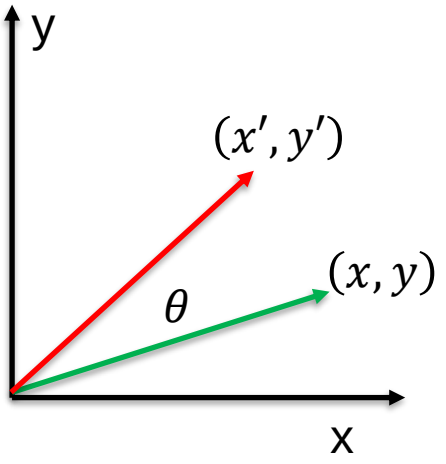
- We can write it in a more compact way by matrix

$$
\begin{pmatrix} 2 & 1 & 4 & 1 \\ 3 & 4 & -1 & -1 \\ 1 & -4 & 1 & 5 \\ 2 & -2 & 1 & 3 \end{pmatrix}
\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}
=
\begin{pmatrix} -4 \\ 3 \\ 9 \\ 7 \end{pmatrix}
$$

- Alternatively, we could write this out shorthand as $\mathbf{Ax} = \mathbf{v}$

Rotation between different vectors

- Suppose we have a vector (x, y), and we want to rotate it counterclockwise by θ and the new vector is (x', y')



$$x' = \cos(\theta)x - \sin(\theta)y$$
$$y' = \sin(\theta)x + \cos(\theta)y$$

- The formula is also several linear equations, and we can also write it in matrix form

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

- Here, the meaning of a matrix is to rotate a vector.

Nonlinear equation (root finding)

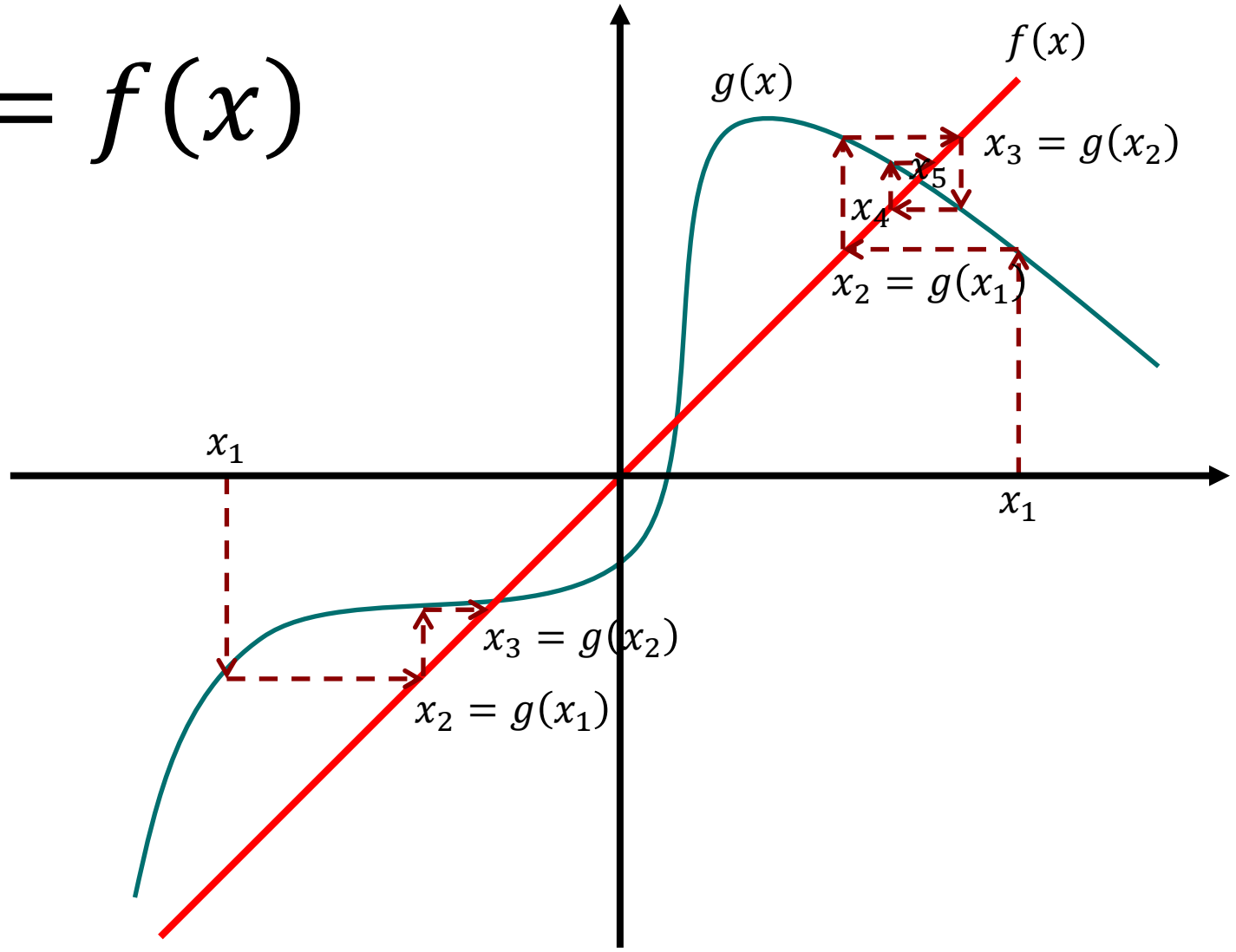
$$x = f(x)$$

$$f(x) = 0$$

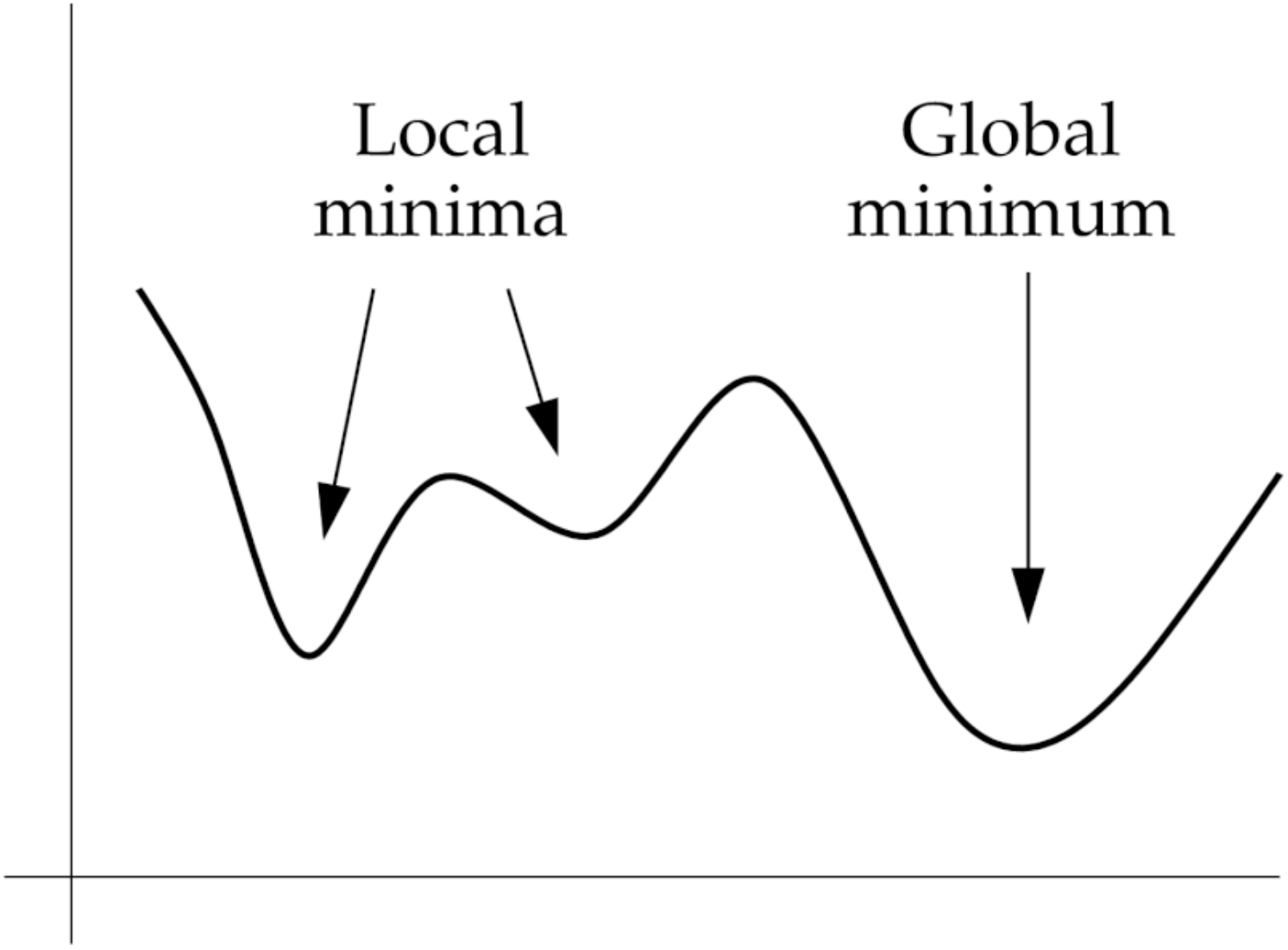
- $f(x)$ is non-linear equation.

Graphic description of relaxation method

$$x = f(x)$$

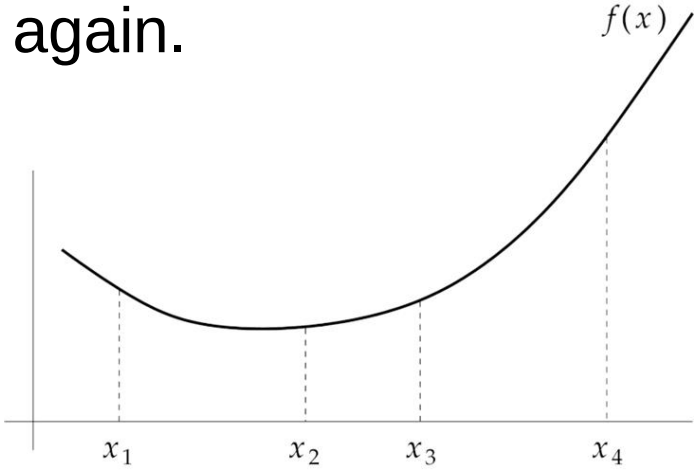


Minima (root finding)



The idea of golden ratio search

- Suppose we have four points $x_1 < x_2 < x_3 < x_4$, and $f(x_2)$ or $f(x_3)$ is smaller than both $f(x_1)$ and $f(x_4)$. In that case, we know that there must be at least one minimum between x_1 and x_4, because the function goes down and up again.



- By comparing $f(x_2)$ and $f(x_3)$, we can further narrow down the location of this minimum, e.g. $f(x_2) < f(x_3)$, know the minimum must lie between x_1 and x_3.
- It means that we can narrow down our search for the minimum to a smaller range, encompassing three of four points.
- Now we can add a new point x_5 between x_1 and x_3, and re-label all the four points x_1, x_2, x_3 and x_5 as the new $x_1 < x_2 < x_3 < x_4$ and repeat the process again. By doing this, we can narrow down the location of the minimum to any accuracy.

Ordinary differential equations

$$\frac{dx}{dt} = f(x,t)$$

Euler's method (initial condition)

- For $\frac{dx}{dt} = f(x,t)$, we have the general solution form

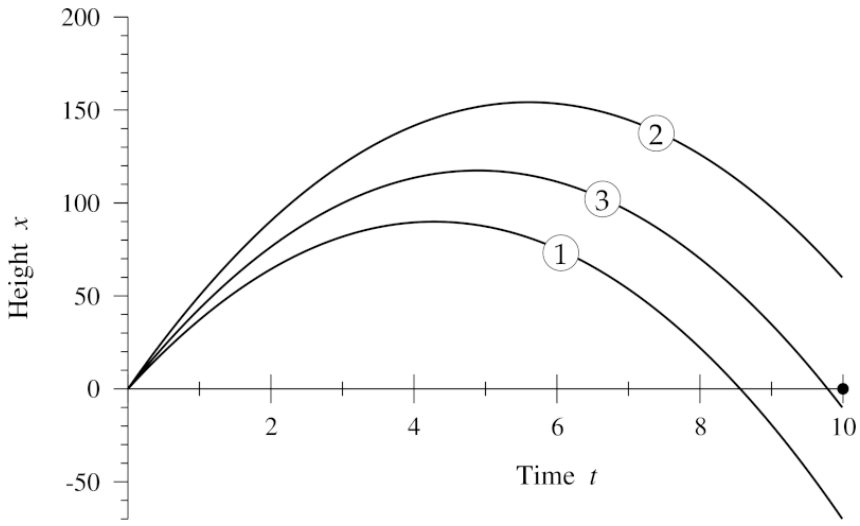
$$x(t_1) = x(t_0) + \int_{t_0}^{t_1} dx = x(t_0) + \int_{t_0}^{t_1} f(x,t)dt$$

- It seems that we can solve first-order ODE by just performing the integral, however, it is not true because we don't know $x(t)$ for a given t except for the initial time t_0.
- Even so, this give us the most important idea about solving the differential equation. If t_1 is very close to t_0, i.e. $t_1 = t_0 + h$, then we can use $x(t_0)$ to approximate all $x(t)$ for $t_0 \le t \le t_1$, then $\int_{t_0}^{t_1} f(x,t)dt \approx f(x(t_0),t_0)\int_{t_0}^{t_1} dt = f(x(t_0),t)(t_1 - t_0) = hf(x(t_0),t)$.
- Finally, we get

$$x(t_0 + h) = x(t_0) + \int_{t_0}^{t_1} f(x,t)dt \approx x(t_0) + hf(x(t_0),t_0)$$

The shooting method (boundary condition)

- We start by guessing a value of initial upward velocity and then calculate the position of the ball x at $t = t_1$which is an initial value problem and can be easily solved. If $x \neq 0$ we should make another guess and do it again.



- The obvious solution is that: if $x < 0$ (undershot), we should increase the initial velocity; if $x > 0$ (overshot), we should decrease the initial velocity.

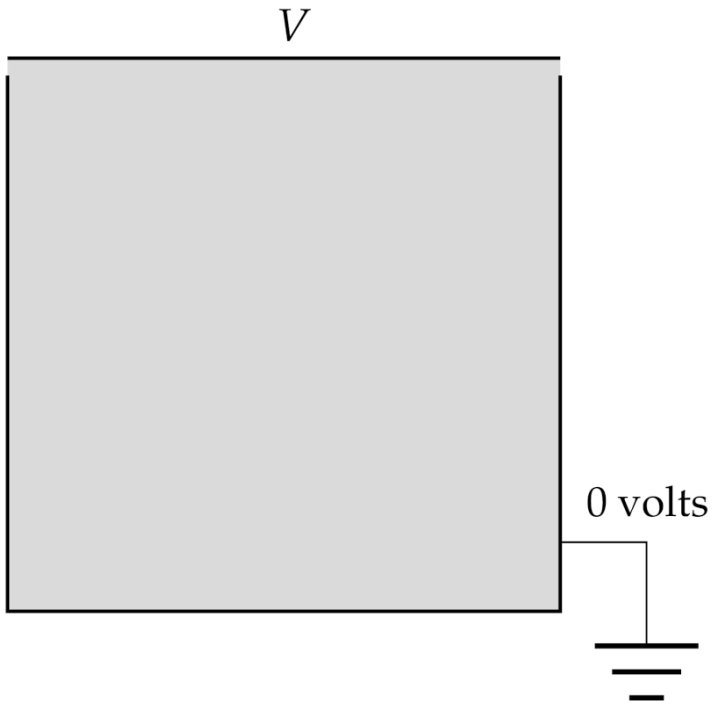
- In principle, the boundary value problems will become the root-finding problems with initial value problems for the differential equations. We can use various methods we have learned for the root-finding problems as binary search or secant method.

Partial difference equations

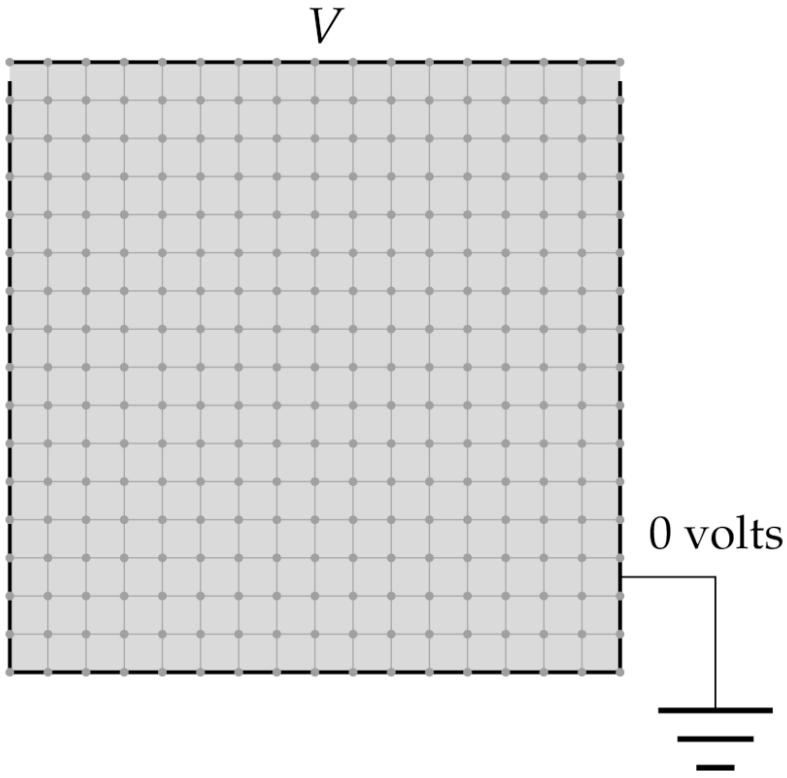
- For a general partial difference equation, it is very hard. We will mainly focus on some examples to show the spirits of solving partial difference equations.

- Consider an electrostatics example, for the electric potential subject to the appropriate boundary conditions. To be simple, we only consider the problem in 2D space, where the Laplace's equation takes the simpler form

$$\frac{\partial^2 \phi(x,y)}{\partial x^2} + \frac{\partial^2 \phi(x,y)}{\partial y^2} = 0$$



First step: discretize the variables



- To use the computer, the first step is to discretize the variables. Note that we put points on the boundaries of the space as well as in the interior.
- Then we have

$$\frac{\partial^2\phi(x,y)}{\partial x^2}+\frac{\partial^2\phi(x,y)}{\partial y^2}$$

$$=\frac{\phi(x+a,y)+\phi(x-a,y)-2\phi(x,y)}{a^2}$$

$$+\frac{\phi(x,y+a)+\phi(x,y-a)-2\phi(x,y)}{a^2}$$

$$=\frac{\phi(x+a,y)+\phi(x-a,y)+\phi(x,y+a)+\phi(x,y-a)-4\phi(x,y)}{a^2}$$

Laplace's equation in grid

- Then the Laplace's equation becomes

$$\frac{\phi(x+a,y)+\phi(x-a,y)+\phi(x,y+a)+\phi(x,y-a)-4\phi(x,y)}{a^2}=0$$

- a^2 can be cancelled, then we have

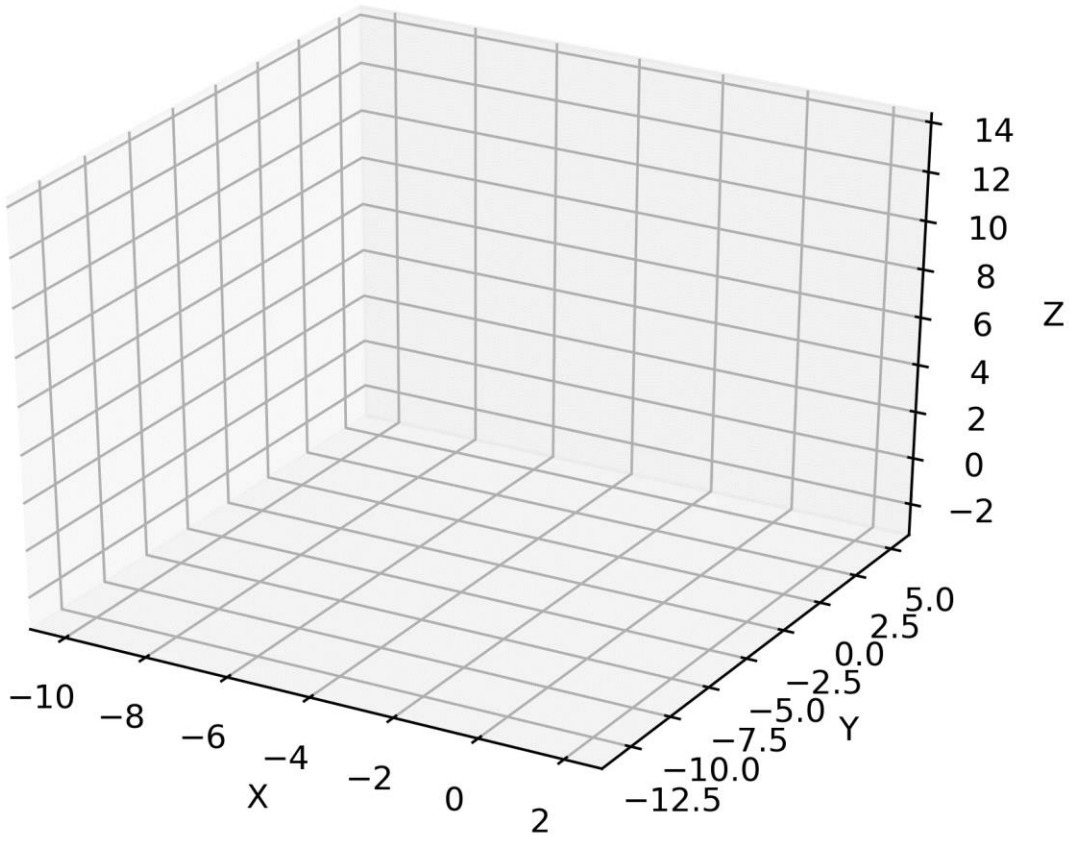
$$\phi(x+a,y)+\phi(x-a,y)+\phi(x,y+a)+\phi(x,y-a)-4\phi(x,y)=0$$

- Now we have simultaneous linear equations, and we can rearrange $\phi(x,y)$ as a vector $\boldsymbol{\phi}$ and then transform them to be linear algebra problem

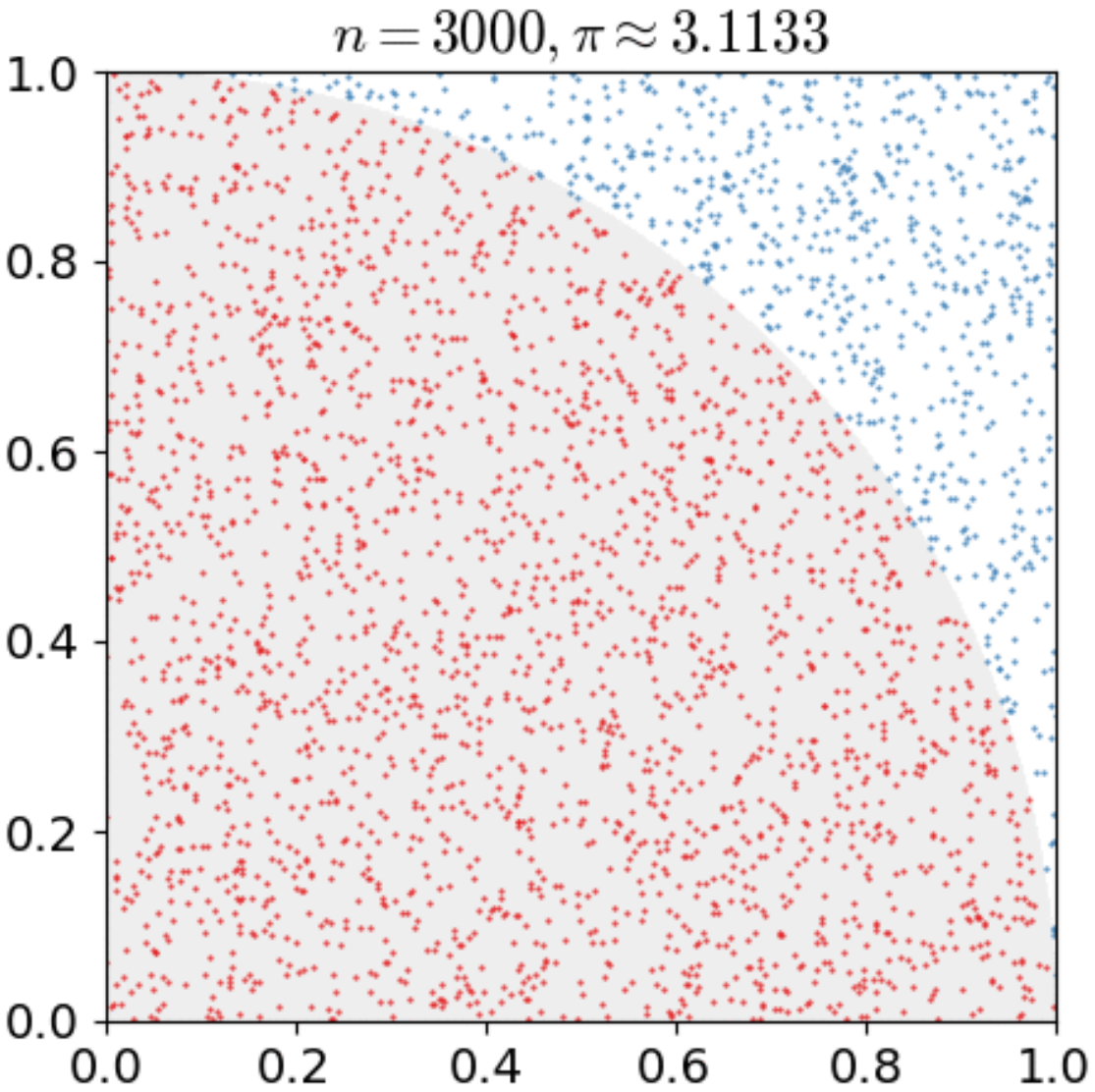
$$\boldsymbol{A\phi}=0$$

- We can solve it by standard methods in linear algebra such as Gaussian elimination or LU decomposition.

Random process and Monte Carlo



MC for Pi



A simpler summary

- Most of the computational problems can be divided into two kinds: **sum up many different terms** or **finding a root for an equation (minimum for a function)**

- The core tools we use: **Taylor expansion**, **Matrix**, **Statistics**, and **Iteration process**.

Summary of Error analysis

- Using Taylor expansion to find the error (exact error, could be negative and could be absorbed)
- Statistical error in the Monte Carlo method
- Error in the iteration process (convergence and stability)
- Use Fourier transform to do stability analysis (not exact the error)

General method for hard problems

Guess

Check

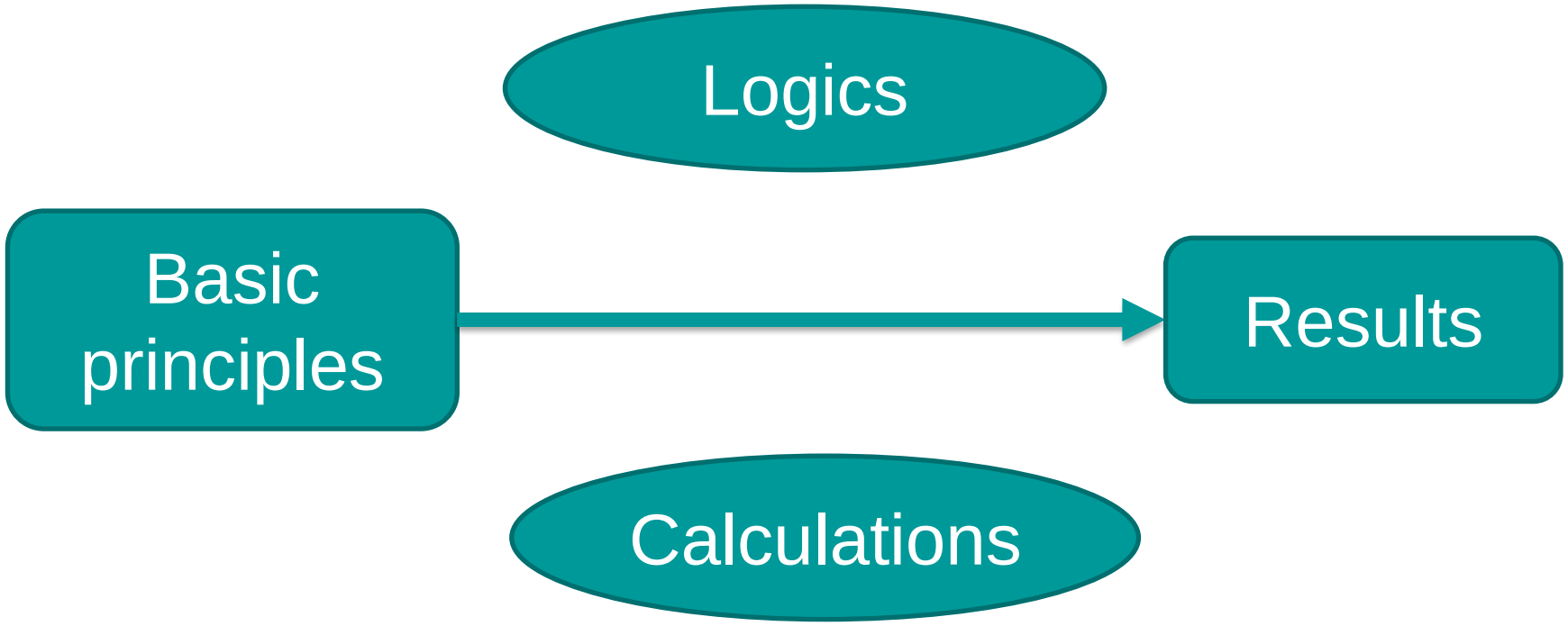
Try it again

Euclidean geometry

- Euclidean geometry is an axiomatic system, in which all theorems ("true statements") are derived from a small number of simple axioms.

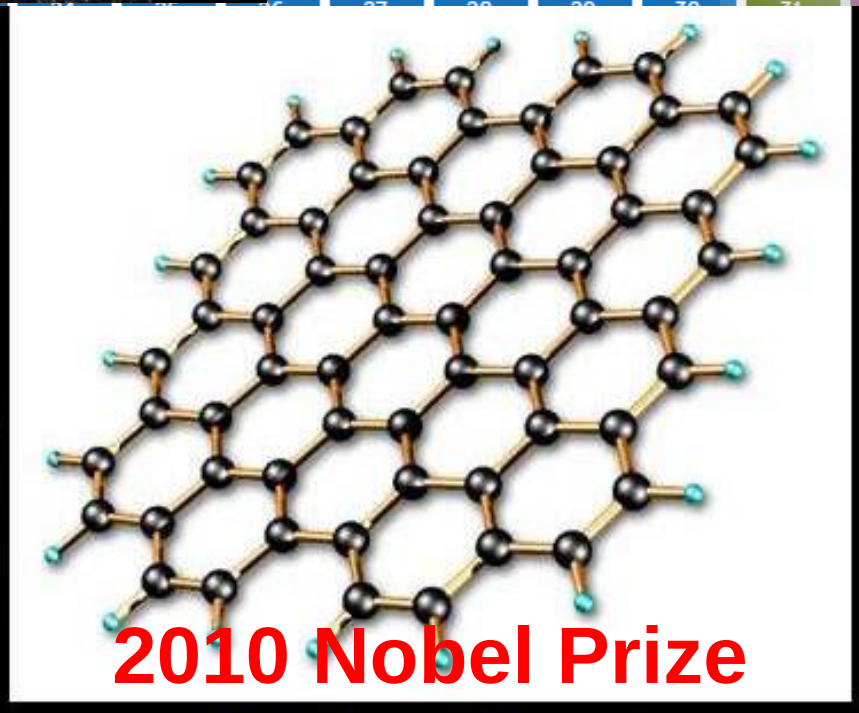
1. To draw a straight line from any point to any point.
2. To produce (extend) a finite straight line continuously in a straight line.
3. To describe a circle with any center and distance.
4. That all right angles are equal to one another.
5. [The parallel postulate]: That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

First-principle way



If the basic principles are “right”,
then all the results will be “right”!!!

Structure and properties



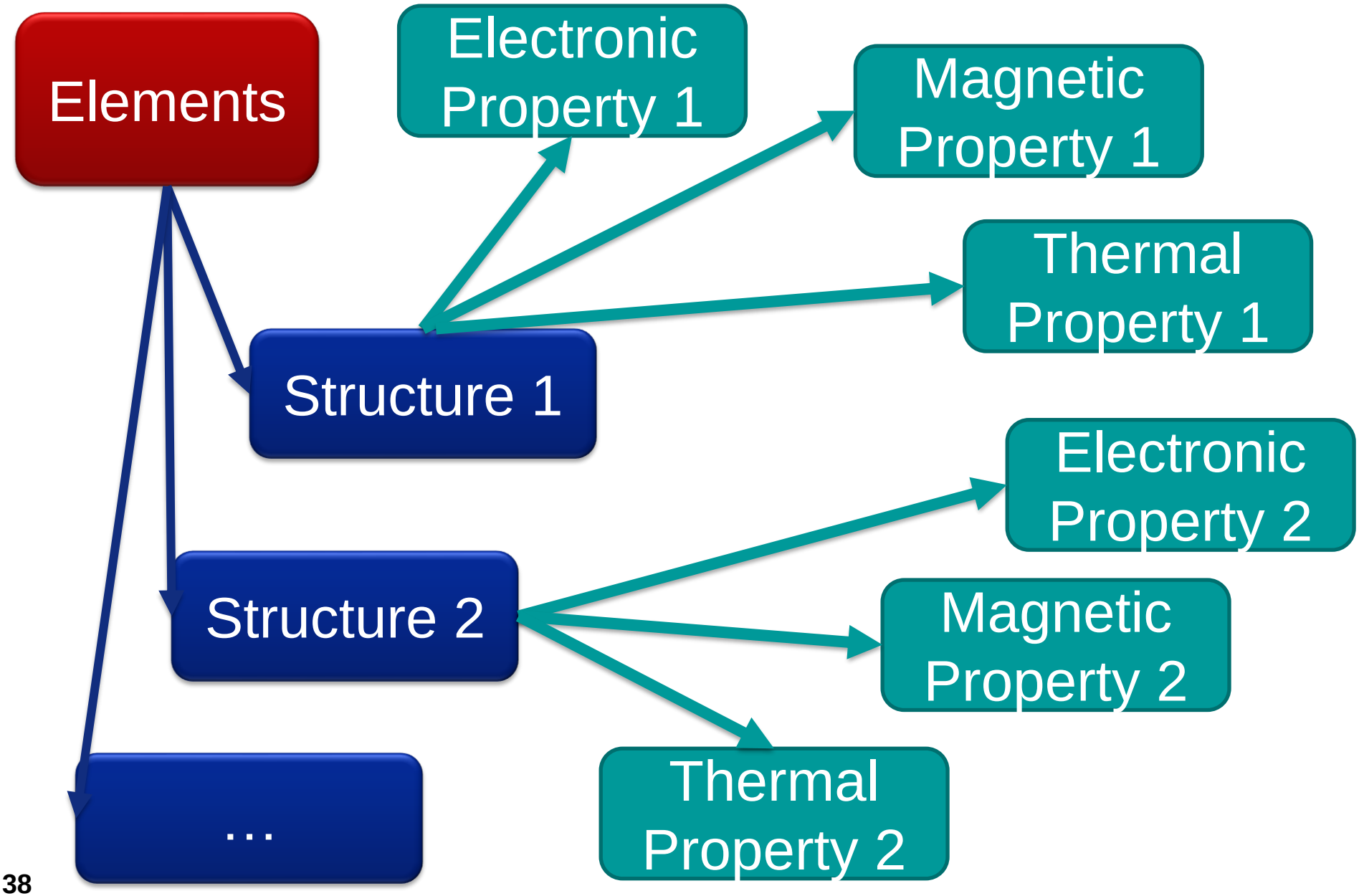
2010 Nobel Prize

A Geim

Graphene Structure

K Novoselov

First-principle calculations



Schrodinger Equation

The movement of N particles are controlled by

$$i\hbar\frac{\partial}{\partial}\Psi(r_1, r_2, \dots, r_N; t)$$
$$= \left[-\frac{\hbar^2}{2m}\nabla_i^2 + V(r_1, r_2, \dots, r_N; t)\right]\Psi(r_1, r_2, \dots, r_N; t)$$

If the state does not change with time, then

$$\left[-\frac{\hbar^2}{2m}\nabla_i^2 + V(r_1, r_2, \dots, r_N)\right]\Psi(r_1, r_2, \dots, r_N)$$
$$= E\Psi(r_1, r_2, \dots, r_N)$$

$$N \sim 10^{23}$$

Just impossible to solve directly !!!

Density functional theory

The properties of many-body systems could be decided by the charge density of the ground state.

Dimension Reduction

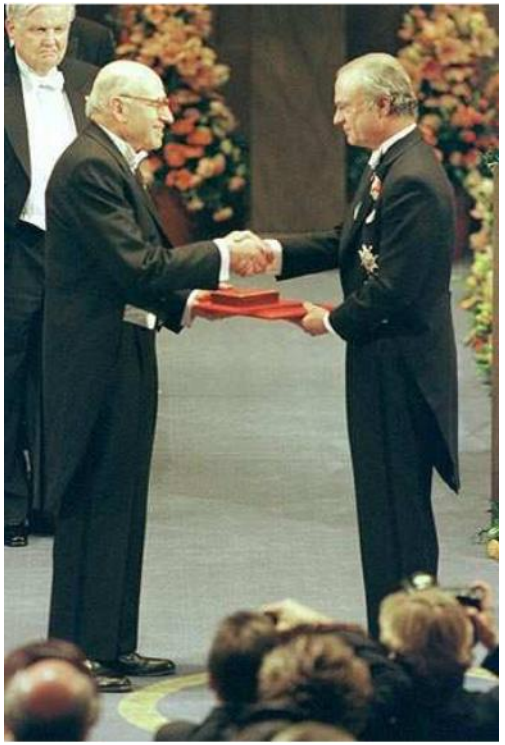
$3N \rightarrow 3$

$$\hat{H} = \hat{T}_{\boldsymbol{r}} + \hat{V}_{\boldsymbol{r}} + \hat{V}_{\boldsymbol{r};\boldsymbol{R}}$$

given an external potential $v_{\mathrm{ext}}(\boldsymbol{r})$ (f=3) $\xleftarrow{\text{HK}}$ $n_0(\boldsymbol{r})$ (f=3) the ground state density

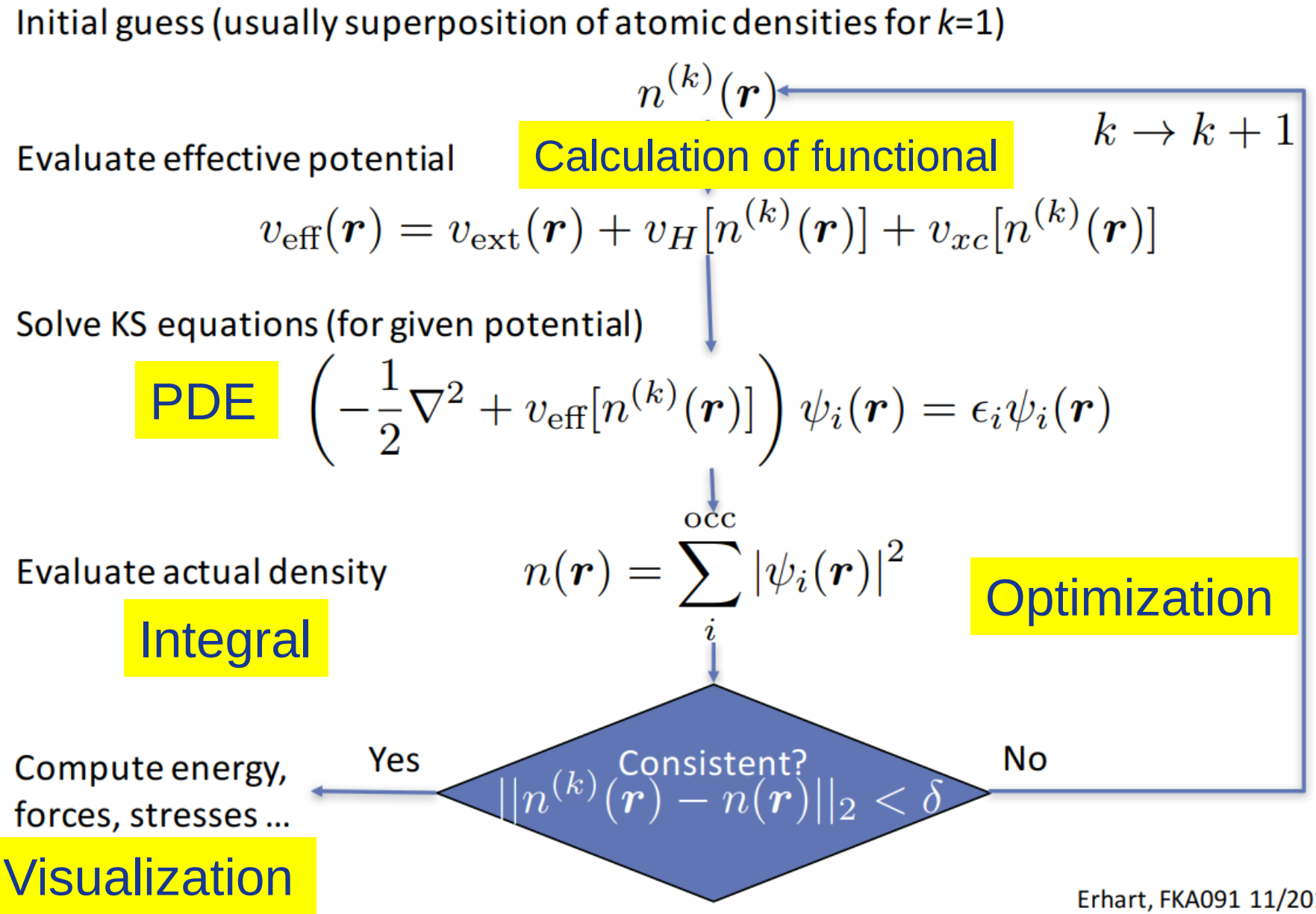
a solution to the Schrödinger solution $\Psi_n(\{\boldsymbol{r}_i\})$ (f=3N) $\rightarrow$ $\Psi_0(\{\boldsymbol{r}_i\})$ (f=3N) the ground state solution (lowest energy)

From Prof Henrik Johannesson's lecture



Walter Kohn receiving his Nobel Prize at the Stockholm Concert Hall.

Self-consistent calculations



Applications of DFT in quantum chemistry

Molecular structures: DFT gives the bond lengths of a large set of molecules with a precision of 1-2%. The hybrid functionals have improved the LDA results.

Bond lengths for different bonding situations [Å]:

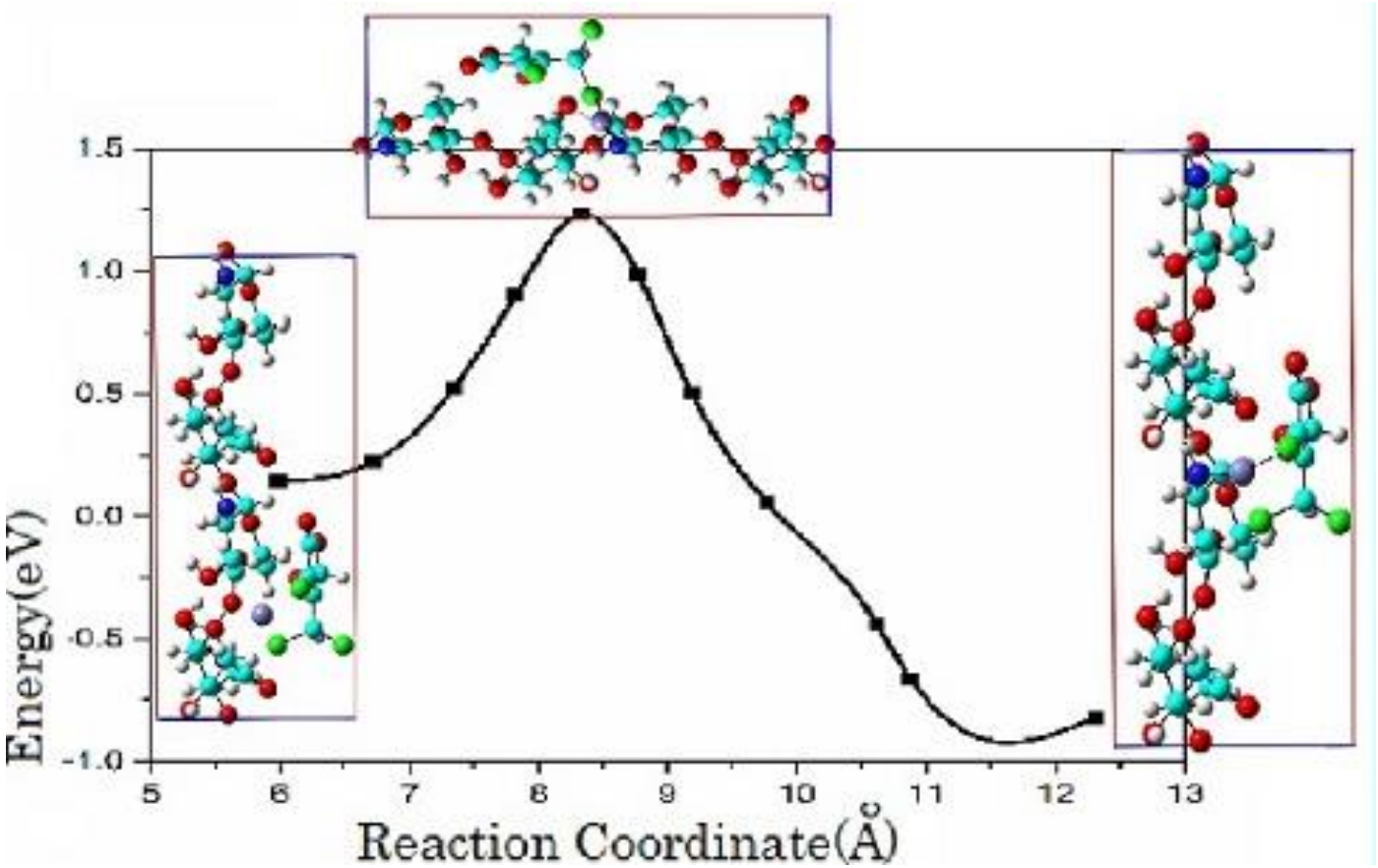
Bond		LDA	BLYP	BP86	Experiment
H-H	R_{H-H}	0.765	0.748	0.752	0.741
H_3C-CH_3	R_{C-C}	1.510	1.542	1.535	1.526
	R_{C-H}	1.101	1.100	1.102	1.088
HC≡CH	R_{C-C}	1.203	1.209	1.210	1.203
	R_{C-H}	1.073	1.068	1.072	1.061

Vibrational frequencies: DFT predicts the vibrational frequencies of a broad range of molecules within 5-10% accuracy.

Vibrational frequencies of a set of 122 molecules: method, rms deviations, proportion outside a 10% error range and listings of problematic cases (taken from Scott and Radom, 1996).

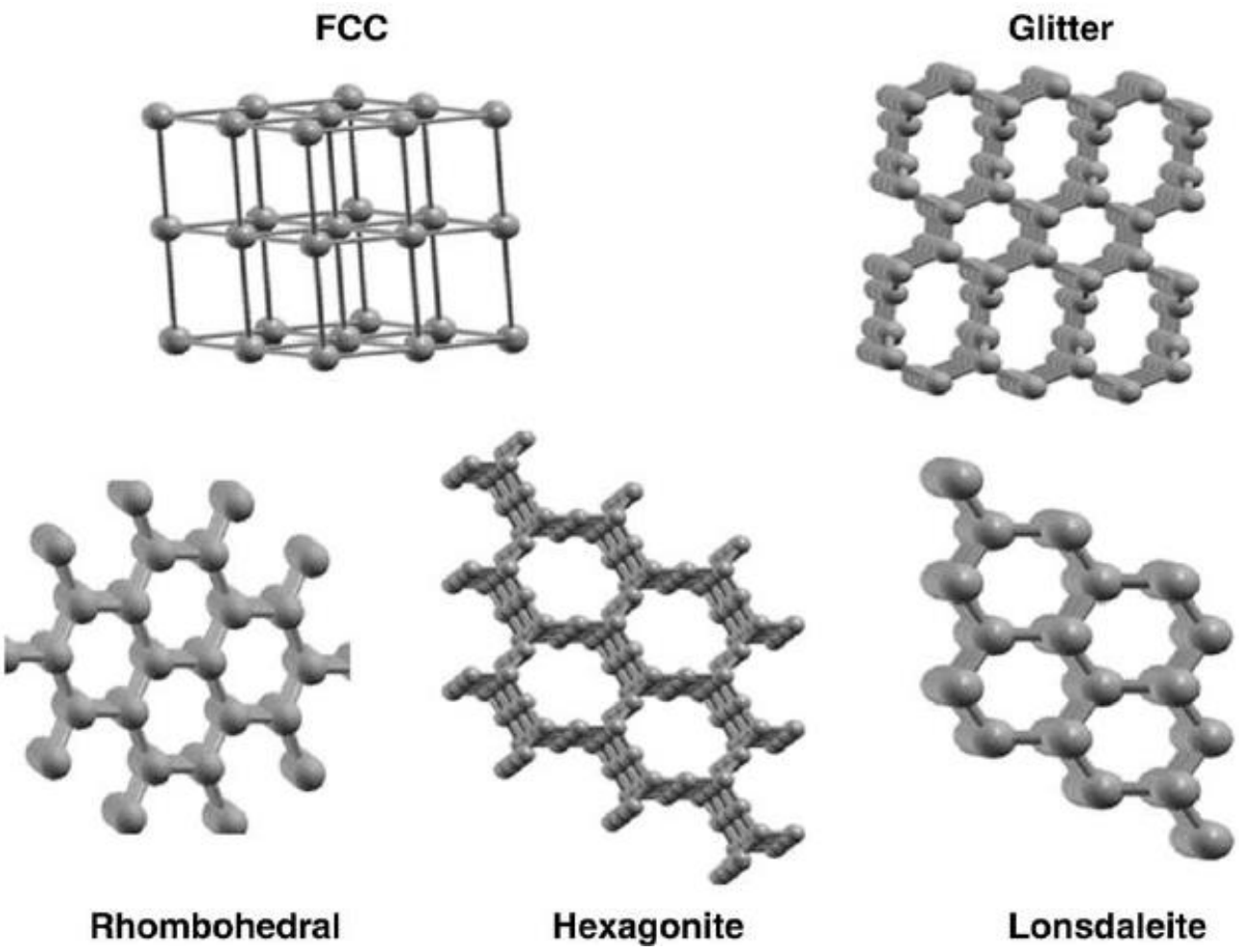
Method	RMS	10%	Problematic cases (deviations larger than 100 cm^{-1})
BP86	41	6	142(H_2), 115(HF), 106(F_2)
B3LYP	34	6	132(HF), 125(F_2), 121(H_2)

Applications of DFT in quantum chemistry



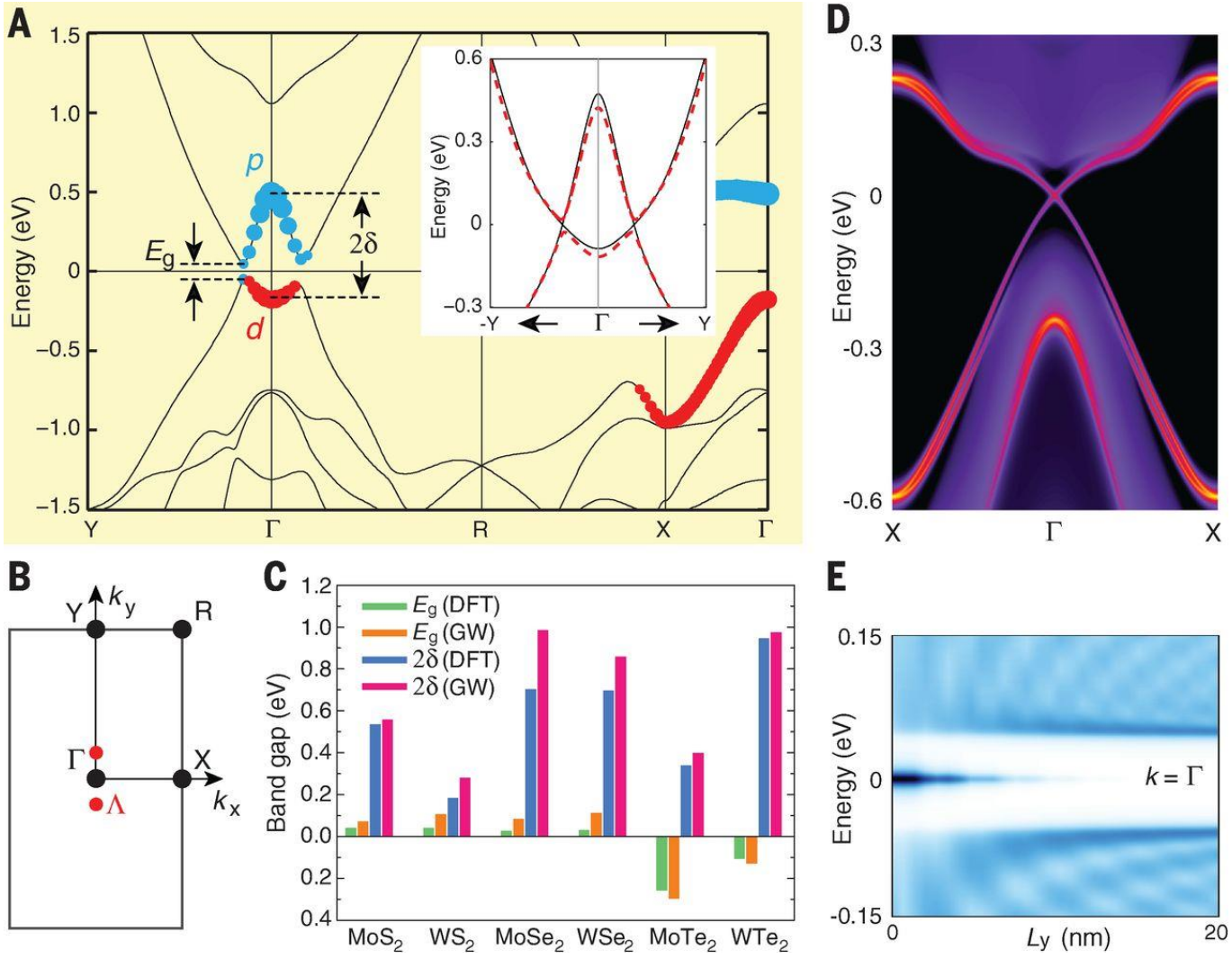
Calculate the chemical reaction

Application of DFT in condensed matter physics



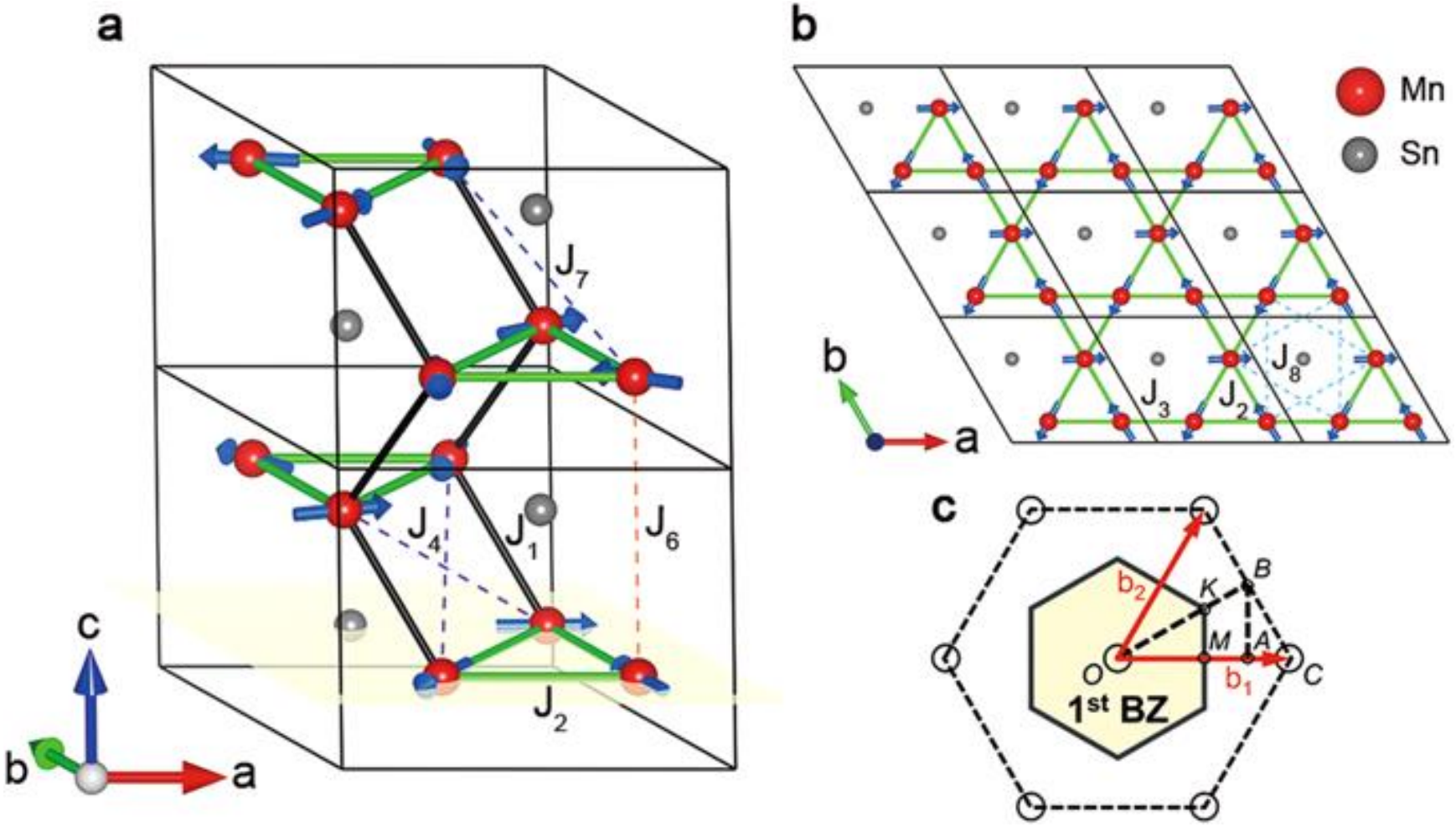
Design new structure or new materials

Application of DFT in condensed matter physics



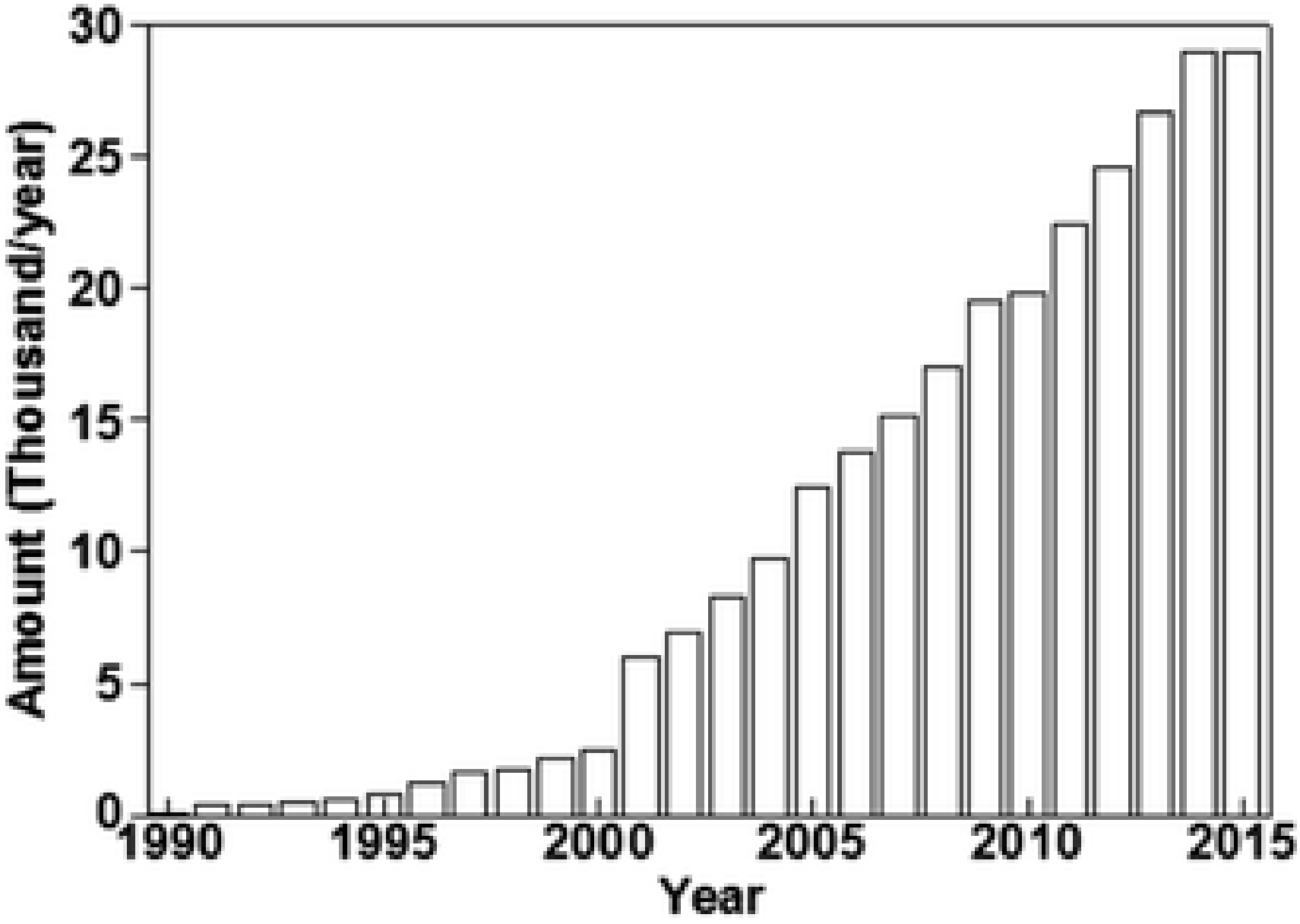
Calculate the electronic properties

Application of DFT in condensed matter physics



Calculate the magnetic properties

Amount of papers when DFT is searched as a keyword in Web of Science



Monte Carlo Simulations

Monte Carlo simulation is, in essence, the generation of random objects or processes by means of a computer. These objects could arise "naturally" as part of the modeling of a real-life system, such as a complex road network, the transport of neutrons, or the evolution of the stock market. In many cases, however, the random objects in Monte Carlo techniques are introduced "artificially" in order to solve purely deterministic problems. In this case the MCM simply involves random sampling from certain probability distributions. In either the natural or artificial setting of Monte Carlo techniques the idea is to repeat the experiment many times (or use a sufficiently long simulation run) to obtain many quantities of interest using the Law of Large Numbers and other methods of statistical inference.

Typical usage of MC

Sampling. Here the objective is to gather information about a random object by observing many realizations of it. An example is *simulation modeling*, where a random process mimics the behavior of some real-life system, such as a production line or telecommunications network. Another example is found in Bayesian statistics, where *Markov chain Monte Carlo* (MCMC) is often used to sample from a *posterior distribution*.

Estimation. In this case the emphasis is on estimating certain numerical quantities related to a simulation model. An example in the natural setting of Monte Carlo techniques is the estimation of the expected throughput in a production line. An example in the artificial context is the evaluation of multi-dimensional integrals via Monte Carlo techniques by writing the integral as the expectation of a random variable.

Optimization. The MCM is a powerful tool for the optimization of complicated objective functions. In many applications these functions are deterministic and randomness is introduced artificially in order to more efficiently search the domain of the objective function. Monte Carlo techniques are also used to optimize *noisy* functions, where the function itself is random — for example, the result of a Monte Carlo simulation.

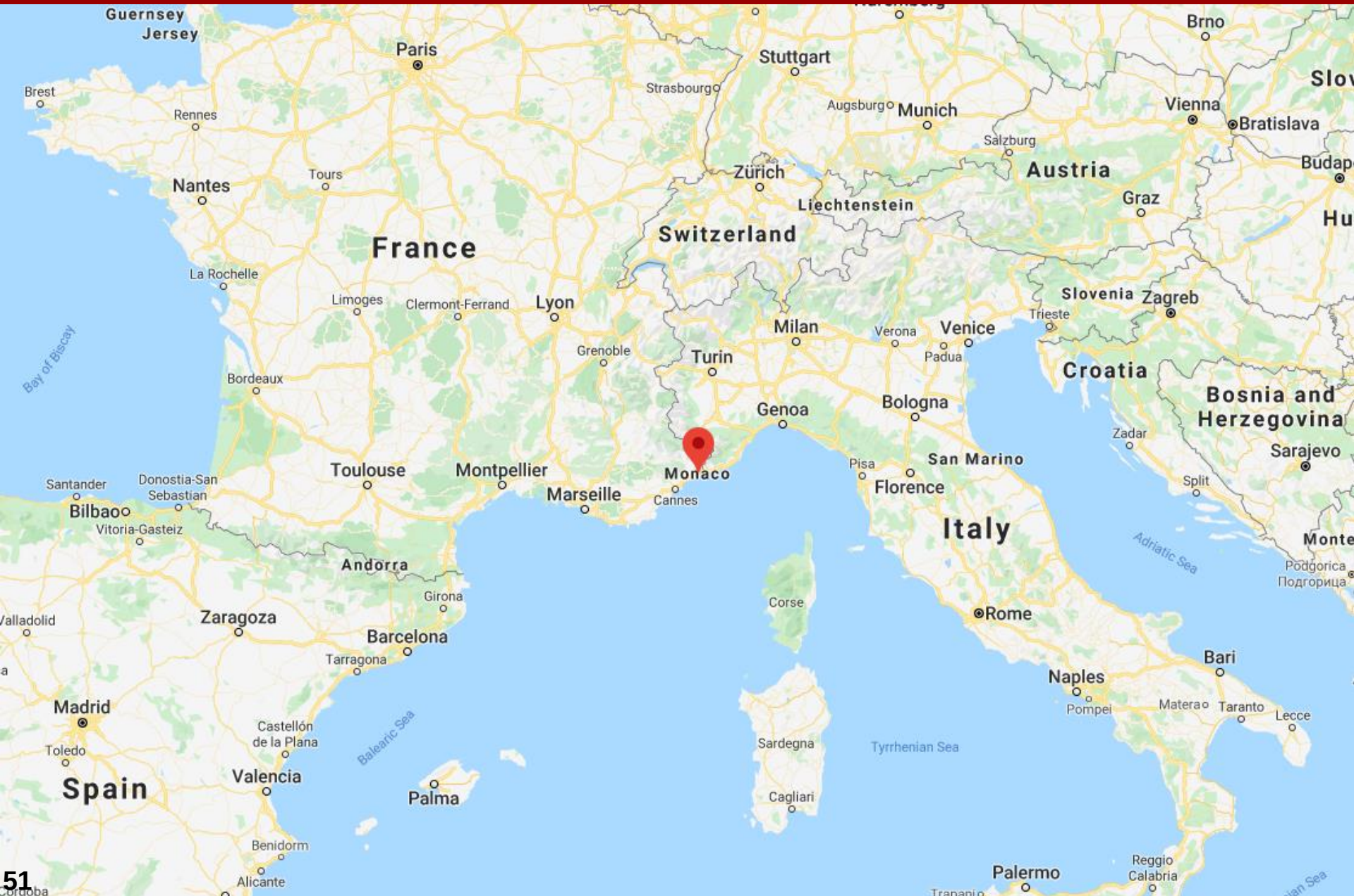
Why MC?

Easy and Efficient

Randomness as a strength

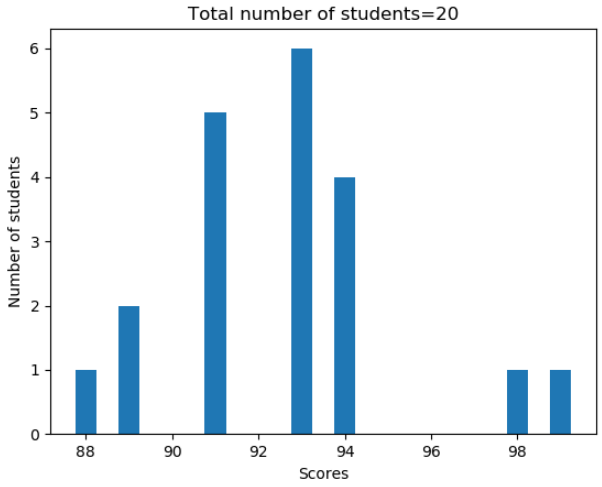
Insight into randomness

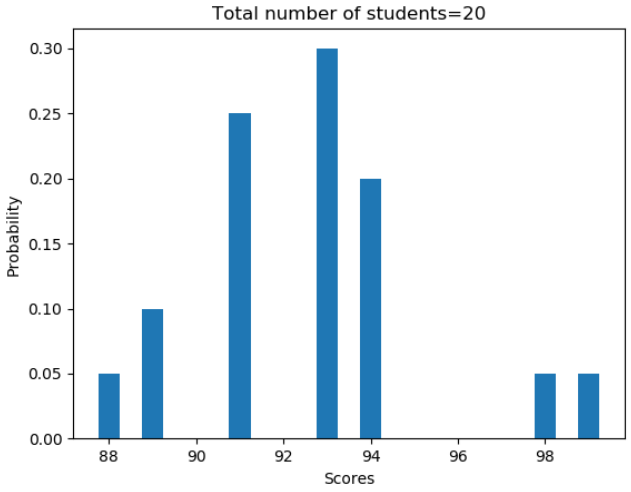
Monte-Carlo Casino in Monaco



Population(总体) and Probability Distribution

- A population means a set of data points. For example, scores of all the students form a population {88, 89, 89, 91, 91, 91, 91, 91, 93, 93, 93, 93, 93, 93, 94, 94, 94, 94, 98, 99}.
- Note different students could have the same score. It is inconvenient to store all these values if we only care about the statistic of a population.
- In probability theory and statistics, a probability distribution is a mathematical function that provides the probabilities of different possible values in a population.





Mean(平均值), Variance(方差) and Standard derivation(标准差)

- Here we use X_i to label different values with the probability $P(X_i)$ (Note: $\sum_{i=1}^{M} P(X_i) = 1$), and x_i label the value for all the possible data point, then the **mean** μ of this population is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \sum_{i=1}^{M} X_i P(X_i)$$

- **Variance** is the mean of the squared deviation of a variable from its mean. Informally, it measures how far a set of (random) numbers are spread out from their average value.

$$\mathrm{Var}(X) = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2 = \sum_{i=1}^{M} P(X_i)(X_i - \mu)^2$$

- Note that variance $\mathrm{Var}(X)$ has different unit than X. To solve this issue, we define another quantity, called the **standard derivation** (noted as σ), which is the square root of variance

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

An example: two ways of presenting data

- Directly show all the data, e.g.

1 1 1 1 2 3 3 4 5 6

The mean, $\mu = \frac{1+1+1+1+2+3+3+4+5+6}{12} = 2.7$

- Show all the different data and their counts, e.g.

X	1	2	3	4	5	6
N(X)	4	1	2	1	1	1
P(X)	0.4	0.1	0.2	0.1	0.1	0.1

The mean, $\mu = \frac{1*4+2*1+3*2+4*1+5*1+6*1}{4+1+2+1+1+1} = 2.7$

The mean, $\mu = 1*0.4+2*0.1+3*0.2+4*0.1+5*0.1+6*0.1=2.7$

P(X): the probability distribution of X

Sample and sample statistic

- To get the statistical properties of a population, we need all the data points. However, usually, it is impossible and we can only have finite number (n) of data points. This set of data point is called a **sample(样本)**, and n is the sample size. We can use the sample statistical properties as estimations of the population statistical properties.
- **Mean of the sample** is similarly defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{m} X_i p(X_i)$$

- **Variance s^2 (standard derivation s) of a sample** are

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{m} p(X_i)(X_i - \bar{x})^2$$

- Note $p(X_i)$ from points in one sample is very **close to** but usually **different from** $P(X_i)$ in the whole population and thus there will be an error.

Bessel's Correction for sample statistic

- When you work with a sample, you've only got a small fraction of the population to work with. Therefore, your answers are not as accurate as those you would have gotten, if you had the entire set of data to work with.
- In the sample variance and standard deviation, the particular statistic you are working with is the sample mean ($\bar{x}$) instead of the population mean (μ). Any x-value in your sample is closer to $\bar{x}$ than to μ. This leads to $\sum_{i=1}^{n}(x_i - \bar{x})^2 < \sum_{i=1}^{n}(x_i - \mu)^2$, thus we need to use a smaller denominator to calculate the variance. It can be proven that we can use $n - 1$ instead of n to correct this. This is called Bessel's correction. This is very important especially for small n.
- Thus, variance of a sample are

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\sum_{i=1}^{m} p(X_i)(X_i - \bar{x})^2$$

Confidence intervals(置信区间)

- We know sample statistic is only an estimation of the population statistics. Then we want to ask how good the estimation is. We take the mean as an example. However, the sample mean by itself is a single point. This can not give us the answer.
- If we want to assess the accuracy of this estimation we will use confidence intervals which provide us with information as to how good our estimation is. And the confidence intervals can be estimated by the variance σ of many independent samples.
- We usually write our final results as

$$\mu = \bar{x} \pm \delta$$

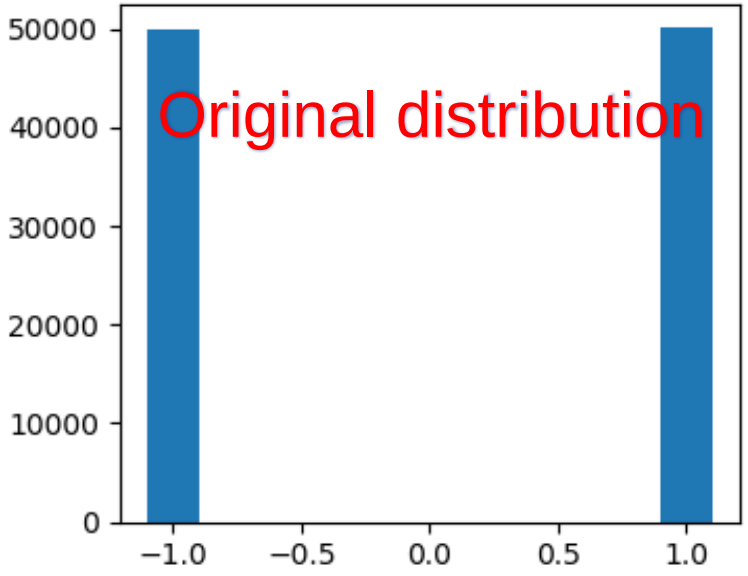
- This means that the probability of actual mean μ falled in between $\bar{x} + n\delta$ and $\bar{x} - n\delta$ is P.

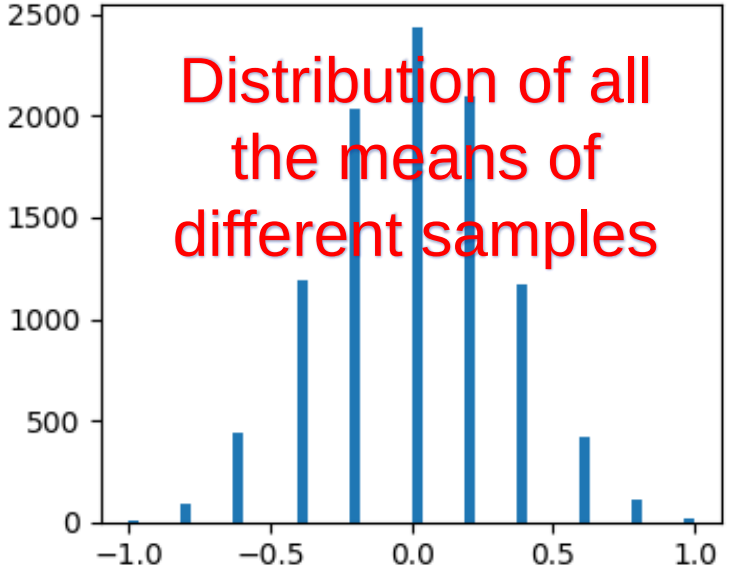
Why?

1	0.682 689 492 137
2	0.954 499 736 104
3	0.997 300 203 937
4	0.999 936 657 516
5	0.999 999 426 697
6	0.999 999 998 027

Central limit theorem

- In probability theory, the central limit theorem (CLT) establishes that, in most situations, when **independent** random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.
- In other words, the estimated means from different samples will obey a normal distribution no matter the form of distribution of the original population!!!

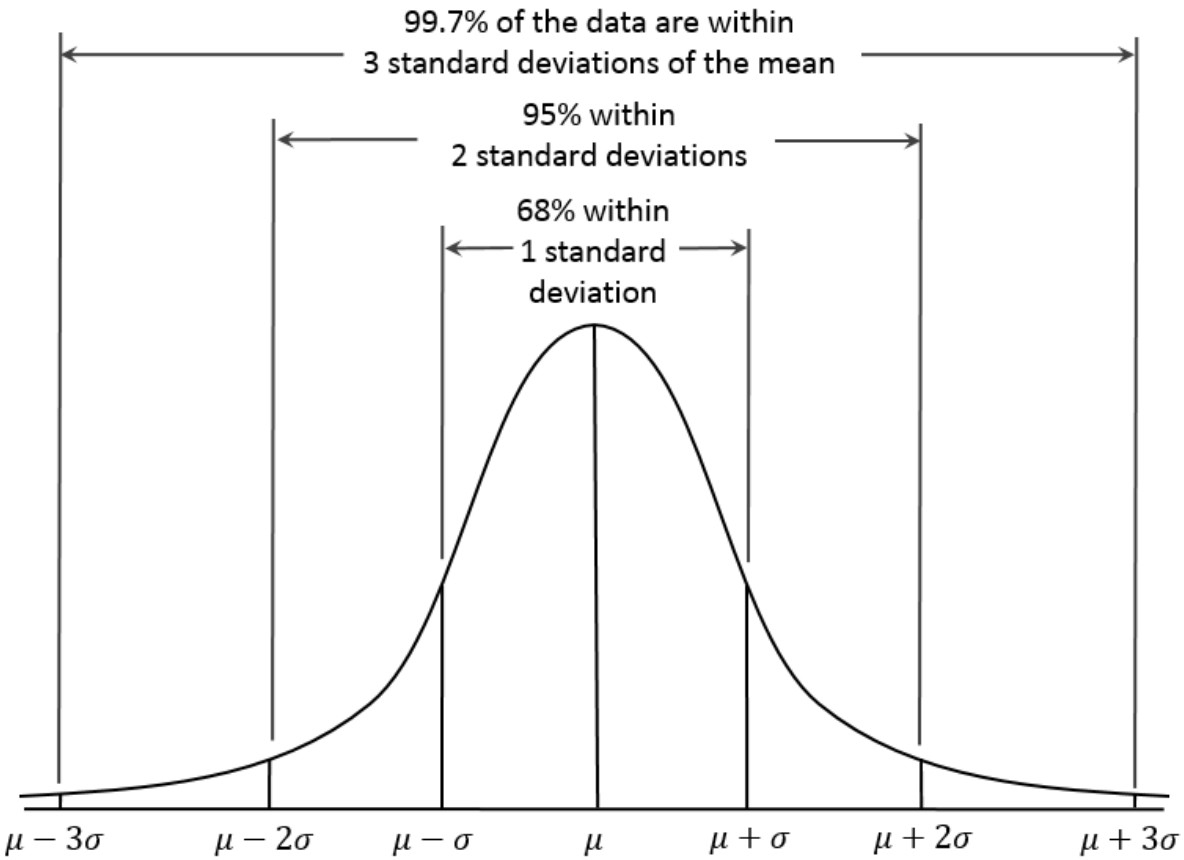




Gauss distribution

- Gauss distribution, also called the normal distribution, is a very common continuous probability distribution.

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Estimated results and standard error

- Assume we could have many samples, labelled as $\{x_i\}^1$, $\{x_i\}^2$, $\cdots$, $\{x_i\}^N$. To be simple, in each sample, we have M data points, $x_1^j, x_2^j, \cdots, x_M^j$. The means of all samples are then $\bar{x}^j = \frac{1}{M}\sum_{i=1}^{M} x_i^j$
- Then we get the final estimated value for the actual means μ is $\bar{x} \approx \frac{1}{N}\sum_{j=1}^{N} \bar{x}^j = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{M}\sum_{i=1}^{M} x_i^j$
- Moreover, all these means $\bar{x}^j$ will form a Gaussian distribution with variance $s^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(\bar{x}^j - \bar{x}\right)^2$
- Since all the data points are generated directly by random number, we can assume there is no autocorrelation, thus we can use the variance to estimate **standard error of the mean**

$$\delta \approx \frac{s}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(\bar{x}^j - \bar{x})^2}$$

Standard deviation and standard error

- The ***standard deviation*** of the sample data is a ***description of the variation in measurements***, while the ***standard error of the mean*** is a probabilistic statement about how the sample size will ***provide a better bound on estimates of the population mean***, in light of the central limit theorem.

- Put simply, the standard error of the sample mean is an estimate of how far the sample mean is likely to be from the population mean, whereas the standard deviation of the sample is the degree to which individuals within the sample differ from the sample mean. If the population standard deviation is finite, the **standard error** of the mean of the sample will **tend to zero with increasing sample size**, because the estimate of the population mean will improve, while the **standard deviation** of the sample will tend to **approximate the population standard deviation** as the sample size increases.

Autocorrelation

- To characterize the correlation in the series of numbers X_i, we use the definition of autocorrelation

$$R(\tau) = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)(X_{i+\tau} - \mu)}{\sigma^2} = \frac{\sum_{i=1}^{N}(X_i - \mu)(X_{i+\tau} - \mu)}{\sum_{i=1}^{N}(X_i - \mu)(X_i - \mu)}$$

- $R(\tau)$ is between -1 and 1. If the data is random, $R(\tau)$ should be near zero for any and all τ. If non-random, then one or more of the autocorrelations will be significantly non-zero.

- The autocorrelations between the values of successive random number will introduce errors into physics calculations that are of significant size yet hard to detect. For serious computations in physics, it is important to use high quality random numbers.

Random number

- To mimic randomness our computer programs themselves will need to have an element of randomness in them and for that we need random numbers.
- The accurate random number is usually related to some physical processes such as the rolling of dice, coin flipping, the shuffling of playing cards, or a particle following quantum mechanics. These random number generator can be called hardware random number generator. However, Because of the mechanical nature of these techniques, generating large numbers of sufficiently random numbers required a lot of work and/or time. Thus, results would sometimes be collected and distributed as random number tables.
- In real calculations, in fact, we use pseudorandom numbers, which are not really random at all. They only look random, being generated by a deterministic formula referred to (inaccurately) as a random number generator.

Random number generators

- Luckily, there are many other random number generators that give much better random numbers. Usually the generator of choice for physics calculations seems to be so-called Mersenne twister, which is a “generalized feedback shift-register generator”. It is quite complicated to program, but fortunately we don’t have to do that. Many software like python provides a version. It comes in the ***random*** package, which contains the following useful functions:

random()	Gives a random float number uniformly in [0,1), including 0 but not including 1
randrange(n)	Gives a random integer from 0 to n-1
randrange(m,n)	Gives a random integer from m to n-1
randrange(m,n,k)	Gives a random integer from m to n-1 in steps of k
Seed(n)	Set up the seed value

Monte Carlo Method: one sentence summary

$$I = \sum_{i=1}^{N} C_i$$

- In one sentence, Monte Carlo Method is used to calculate the summation of a very large number of (almost infinite) terms based on a random process.
- This method is quite impressive, because we use a random process to calculate the answer to an exact, nonrandom question!!! It is really genius and amazing!!!

Monte Carlo calculation for Pi

- The trick to use a new method is always to transfer your new problems to the problems, which can be handled by the new method, Monte Carlo method.

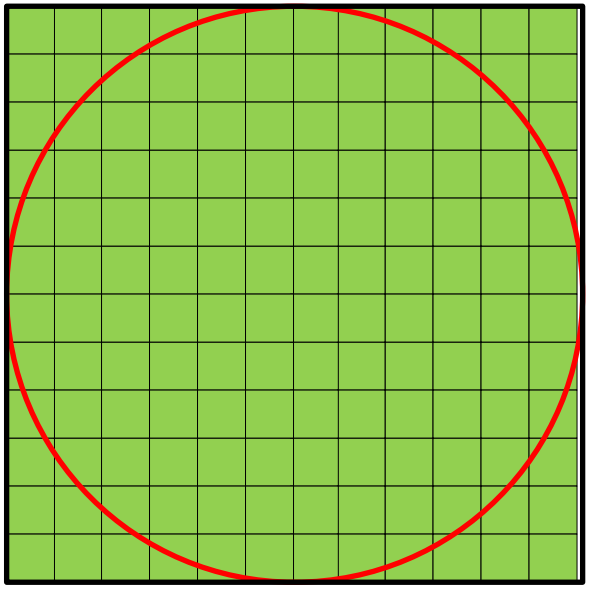
- Monte Carlo Method is used to calculate the summation of a very large number of (almost infinite) terms based on a random process.

- Another important trick is to use the definition of the concepts in your problems. Then, what's the definition of Pi?

- Pi is original defined as the ratio of a circle's circumference to its diameter. It can be defined as the ratio of a circle's area and its diameter's square. Here, we use the second definition, then to calculate Pi, we need to calculate the area very accurately!

Discretize your variable

- To calculate Pi, we need to calculate the area in the circle.
- We first discretize x and y (N points), and make a grid over the square. Then the total number of small square it N^2.
- We count the number of small square is M, then $\pi \approx M/N^2$



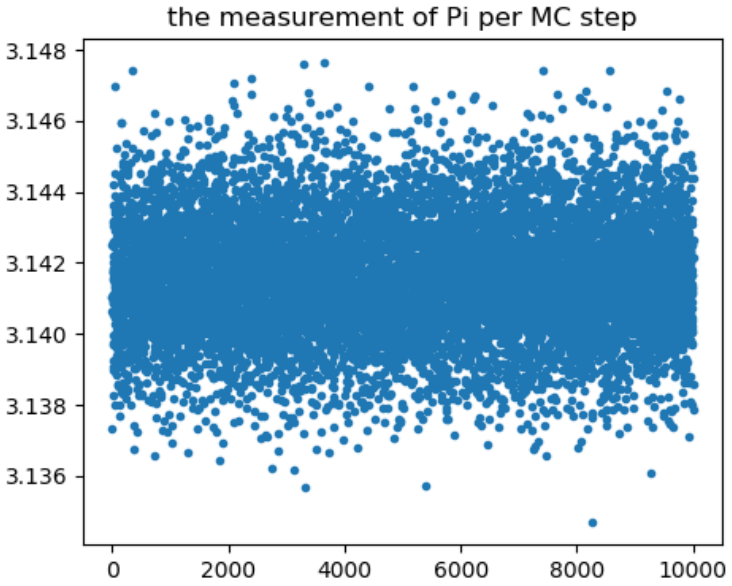
- The large number of points N, the more accurate the results. Could we choose infinite N? Yes, we can! We can use a point to replace the small square.
- First, we **randomly** choose the positions of M points. Then by counting the number of points falling in the circle m, we can estimate the area of circle and then the value of π

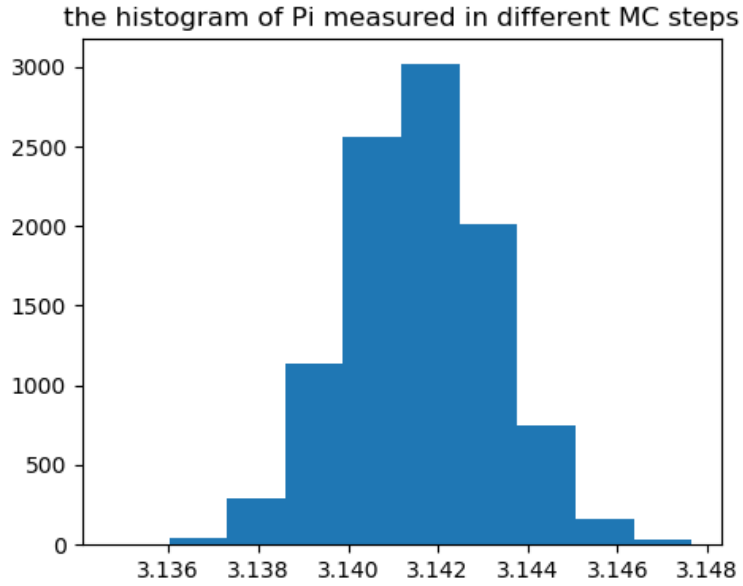
$$\pi \approx \frac{m}{M}$$

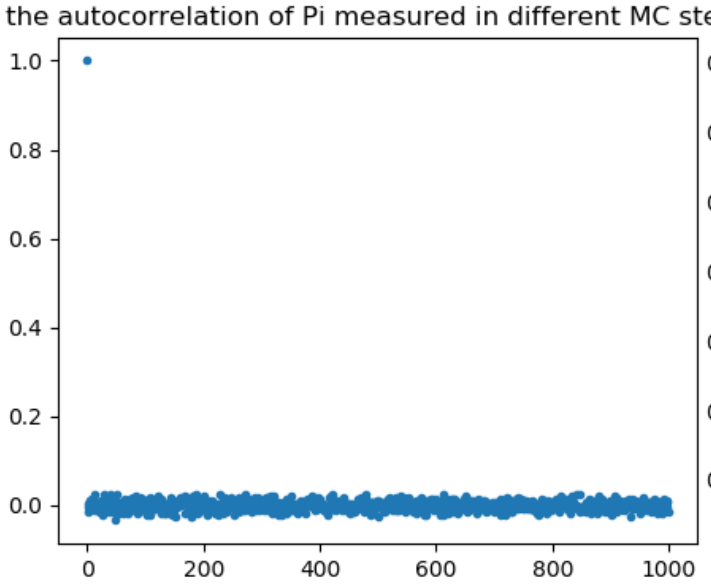
Practice: actual calculations and error estimation

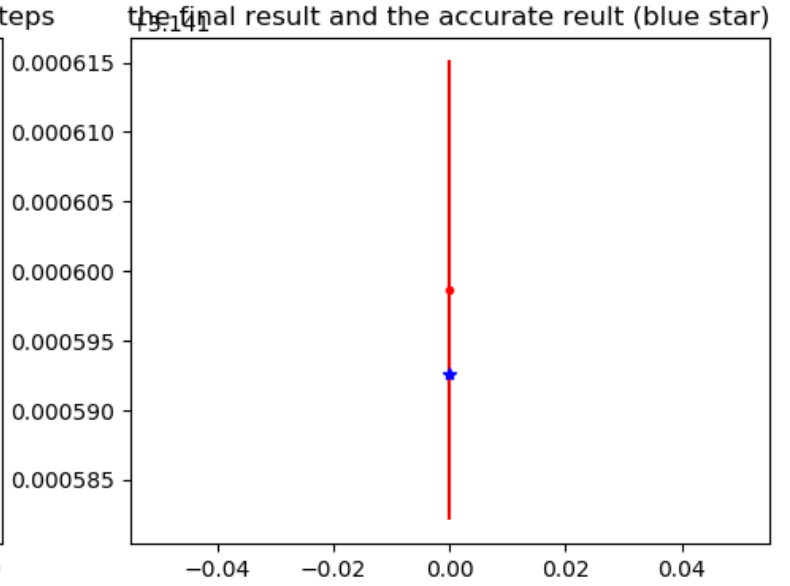
- Write a function, which could get one sample to calculate the value of $\pi \approx \frac{m}{M}$
- Get N different samples and the corresponding value of π
- the value of π for all the samples.
- Plot the histogram for all the value of π for all the samples.
- Calculate the autocorrelation function of π for all the samples.
- Estimate the confidence interval (error estimation)

Results of Pi calculations







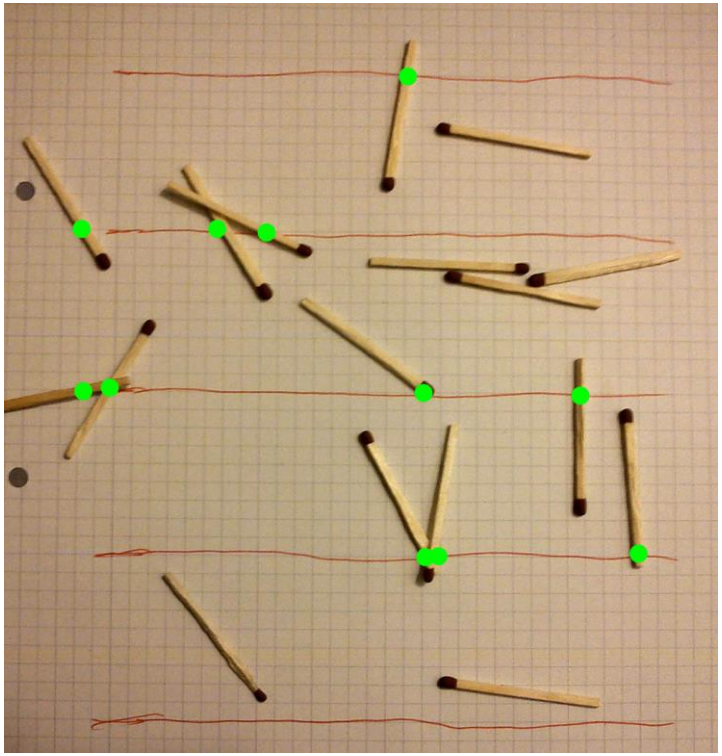


Buffon's needle for Pi

- Find a needle or stick with the length to be l.
- Find a large paper and plot multiple parallel lines with equal space $t > l$.
- Drop the stick to the paper and count the times m with the stick crossing the lines and the total number M you dropped.
- After many times, you will get

$$\pi \approx \frac{M}{m}\frac{2l}{t}$$

- You can play or write a code for this. Have Fun!



Some interesting facts or guess about Pi

- How many digits have human ever got for Pi?

22,459,157,718,361 from Peter Trueb in 2016 after 105 days' calculation by using a super computer with 72 cores.

- How many digits does a human can memorize?

70030 from Guinness Book.

- If we have Pi up to N digits, are the number of 0,1, 2…,9 almost equal? Any proof?

We believe that the answer is yes, but no one can demonstrate it for now! You will be famous if you can demonstrate it.

- All the number in Pi appears in random order? If so, can we use the calculations of Pi as pseudo random number generator?

I think the answer is also yes, but I did not find a solid proof.

Last remark about Pi

https://www.youtube.com/watch?v=a6bT_DVwo7M

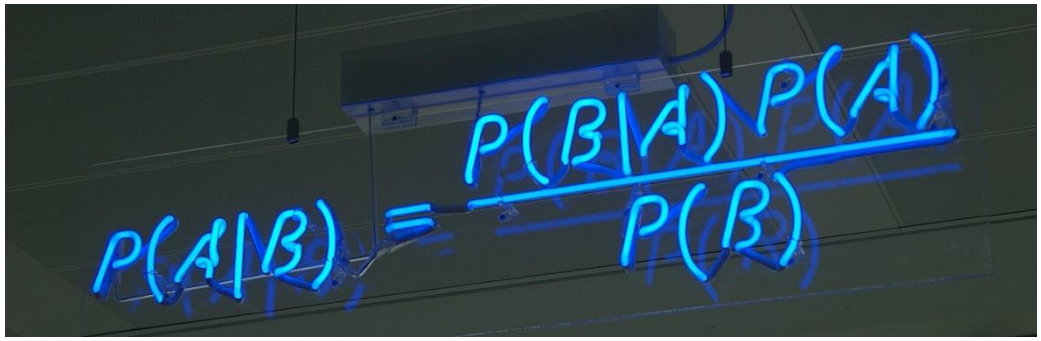
Bayesian statistics

- Bayesian statistics is a theory in the field of statistics based on the **Bayesian interpretation of probability** where **probability expresses a degree of belief in an event**.

- The degree of belief may be based on prior knowledge about the event, such as the results of previous experiments, or on personal beliefs about the event. This differs from a number of other interpretations of probability, such as the **frequentist interpretation** that views **probability as the limit of the relative frequency of an event after many trials**.

- Bayesian statistical methods use Bayes' theorem to **compute** and **update** probabilities after obtaining new data. Bayes' theorem describes **the conditional probability** of an event based on **data** as well **as prior information or beliefs** about the event or conditions related to the event.

Bayes' theorem



- A and B are different events and $P(B) \neq 0$.
- $P(A|B)$ is a conditional probability: the likelihood of event A occurring given that B is true.
- $P(B|A)$ is a conditional probability: the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B respectively; they are known as the marginal probability.
- We can write it in more symmetric form

$$P(A|B)P(B) = P(B|A)P(A)$$

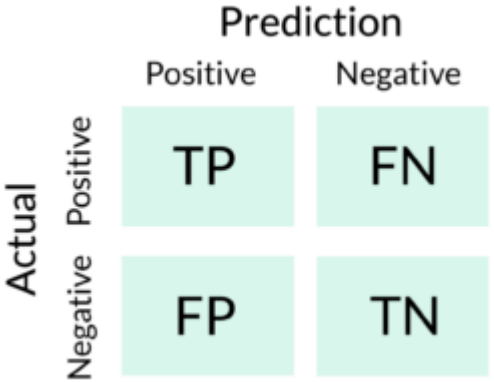
Example: understanding the test for covid-19

Q1: If one randomly chosen person go to a hospital to have a PCR test and get a positive results, is this person really infected? If not, what is the actual probability of this person got infected?

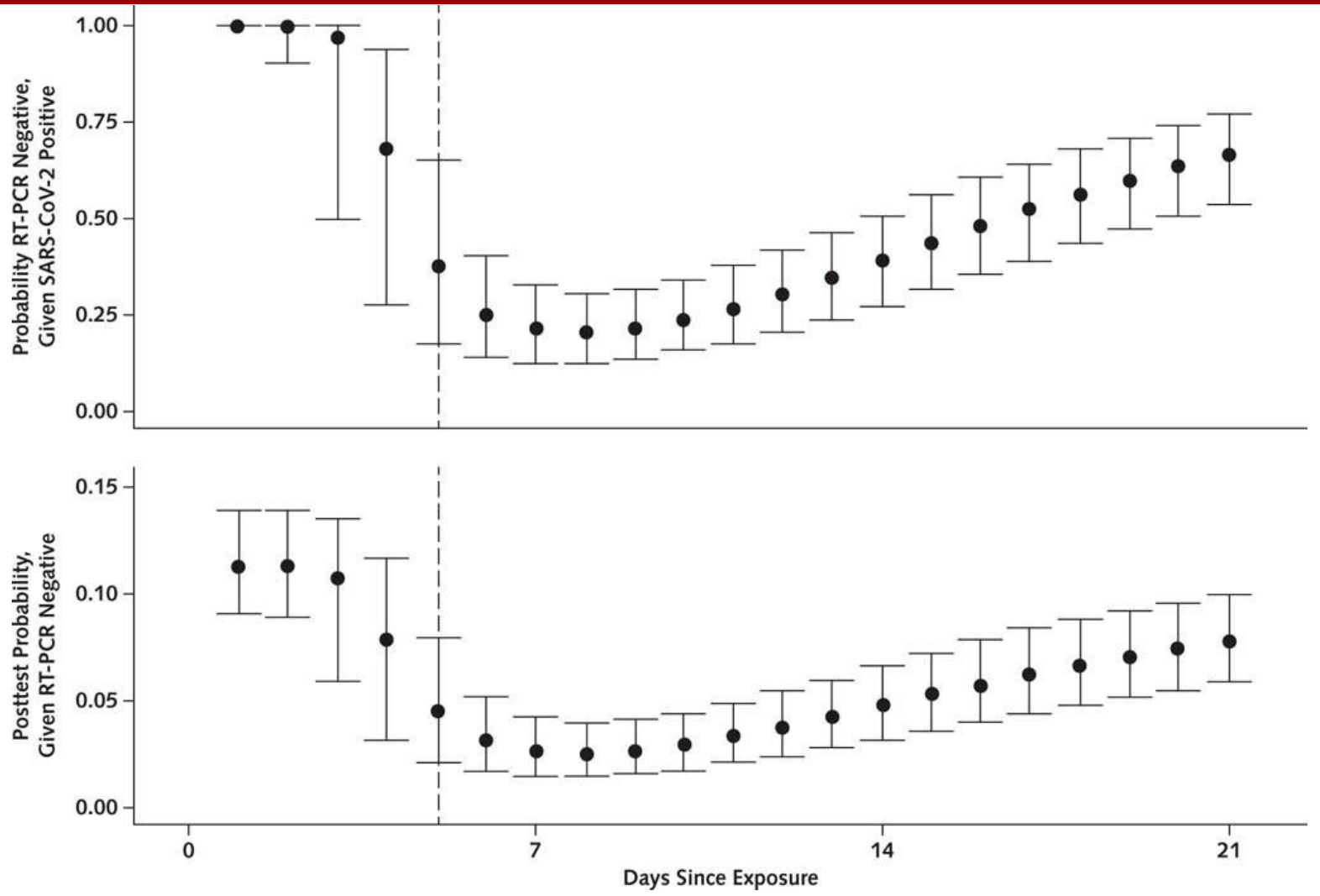
Q2: It was reported that Pfizer and Moderna's vaccine's efficacy is more than 95%. Do you believe it and why?

False positives and false negatives

- A **false positive** is an error in binary classification in which a test result incorrectly indicates the presence of a condition such as a disease when the disease is not present.
- A **false negative** is the opposite error where the test result incorrectly fails to indicate the presence of a condition when it is present.
- These are the two kinds of errors in a binary test, in contrast to the two kinds of correct result (a **true positive** and a **true negative**.)
- They are also known in medicine as a false positive (or false negative) diagnosis, and in statistical classification as a false positive (or false negative) error.



PCR test accuracy



Lauren M. Kucirka, MD, PhD*, et al, Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. https://www.acpjournals.org/doi/10.7326/m20-1495

For question 1

- The Hong Kong population: 7.451 M
- The total number of confirmed cases in Hong Kong:

  10884 (updated on 2021/02/23)
- The number of confirmed cases in Hong Kong on 2021/02/23: 16
- For a random person, the probability of got infected is around 10,884/7,451,000~0.00146
- We assume that the PRC false positive rate is 0.03, and true positive 1

Answer: For a random person got positive result in PCR test, the actual probability of get infected is

$$\frac{0.00146 \times 1}{0.00146 \times 1 + (1 - 0.00146) \times 0.03} = 4.64\%$$

For question 2

- The test is composed of a total of 43,448 (21720 vaccine, 21728 placebo) participants >16 years of age, regardless of duration of follow-up;
- For participants without evidence of SARS-CoV-2 infection prior to 7 days after Dose 2, VE against confirmed COVID-19 occurring at least 7 days after Dose 2 was 95.0%. The case split was 8 COVID-19 cases in the BNT162b2 group compared to 162 COVID-19 cases in the placebo group.
- Data is from https://www.fda.gov/media/144416/download
- Vaccine efficacy formula:

Vaccine efficacy=(Attack rate of unvaccinated people-Attack rate of vaccinated people)/Attack rate of unvaccinated people

$$VE = \left(\frac{162}{21728} - \frac{8}{21720}\right) \Big/ \frac{162}{21728} \approx 95.1\%$$

However!!!

- Among 3, 410 total cases of suspected but unconfirmed COVID-19 in the overall study population, 1, 594 occurred in the vaccine group vs. 1816 in the placebo group.
- Data is from https://www.fda.gov/media/144416/download
- We assume that the PRC false negative rate is $\alpha = 0.2$
- The answer

$$VE = \left(\frac{162 + 1816 \times \alpha}{21728} - \frac{8 + 1594 \times \alpha}{21720}\right) \Big/ \frac{(162 + 1816 \times \alpha)}{21728}$$

$$\approx 37.8\%$$

- How about other PRC false negative rate α?

Lessons we could learn

In Math We Trust!

By Physics We Understand!

One Example: AlphaGO

Deep learning→
Action-value
function

Monte Carlo Tree
Search →
Optimized move

The other example: encrypted communication