



Machine Learning and the Physical World

Lecture 2 : Simulation

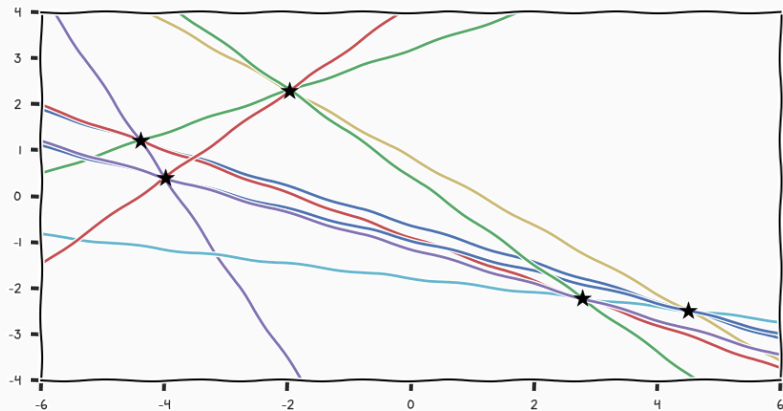
Carl Henrik Ek - che29@cam.ac.uk

14th of October, 2026

<http://carlhenrik.com>



An Over-determined System





We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.

All these efforts in the search for truth tend to lead [the human mind] back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed.

– Pierre Simon Laplace, *A Philosophical Essay on Probabilities*
Laplace, 1814

"The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance. Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision it is impossible for us to announce their occurrence with certainty. It is, however, probable that one of these events, chosen at will, will not occur because we see several cases equally possible which exclude its occurrence, while only a single one favors it."

An Over-determined Sysatem

- Over determined system

$$y_i = [w_1, w_2] \begin{bmatrix} x_i \\ 1 \end{bmatrix}$$

An Over-determined System

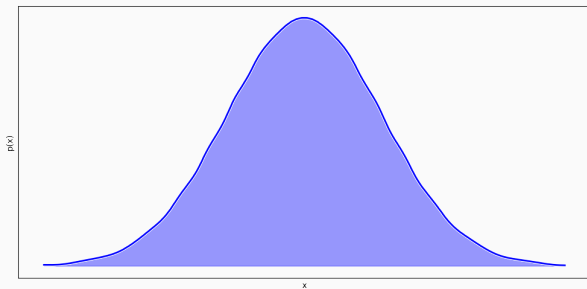
- Over determined system

$$y_i = [w_1, w_2] \begin{bmatrix} x_i \\ 1 \end{bmatrix}$$

- Parametrisation of ignorance

$$y_i = [a, b] \begin{bmatrix} x_i \\ 1 \end{bmatrix} + \epsilon_i$$

The Gaussian Distribution



$$y_i = [a, b] \begin{bmatrix} x_i \\ 1 \end{bmatrix} + \epsilon_i$$
$$\epsilon_i \sim \mathcal{N}(0, \cdot)$$

- Model

$$p(y_i \mid \mathbf{x}_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mathbf{w}^T \mathbf{x}_i)^2}{2\sigma^2}\right)$$

- Optimisation

$$\hat{\mathbf{w}} = \operatorname{argmax}_w \prod_{i=1}^N p(y_i \mid \mathbf{x}_i)$$

- We will never have complete knowledge in the real world

- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it

- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it
- The notion of probabilities allows us parametrise our ignorance

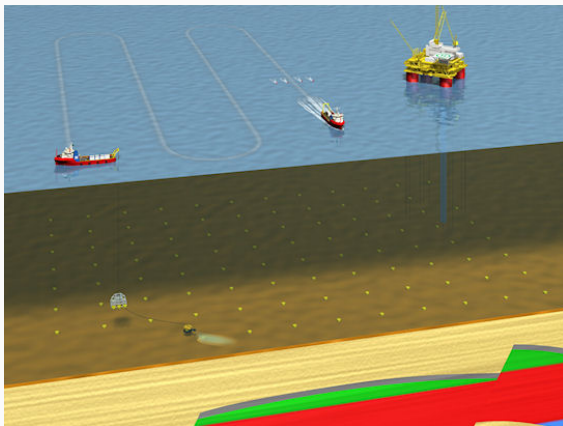
- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it
- The notion of probabilities allows us parametrise our ignorance
- Our ignorance comes from many sources

- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it
- The notion of probabilities allows us parametrise our ignorance
- Our ignorance comes from many sources
 - model

- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it
- The notion of probabilities allows us parametrise our ignorance
- Our ignorance comes from many sources
 - model
 - data

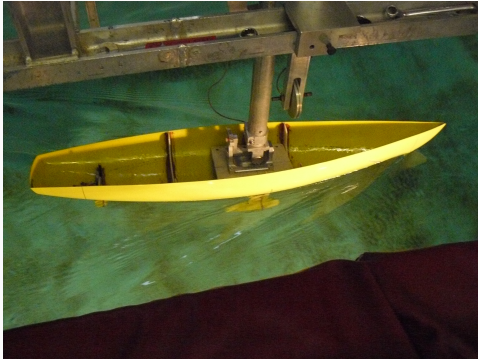
- We will never have complete knowledge in the real world
- In order to reduce our ignorance and learn we need to parametrise it
- The notion of probabilities allows us parametrise our ignorance
- Our ignorance comes from many sources
 - model
 - data
 - compute

Simulation

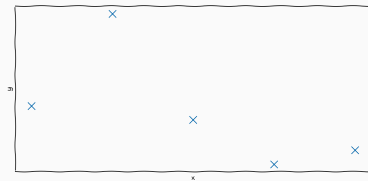




Data collection



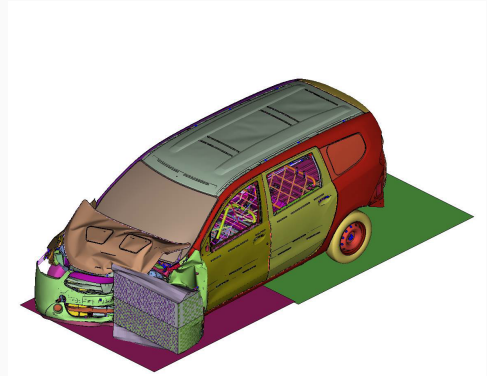
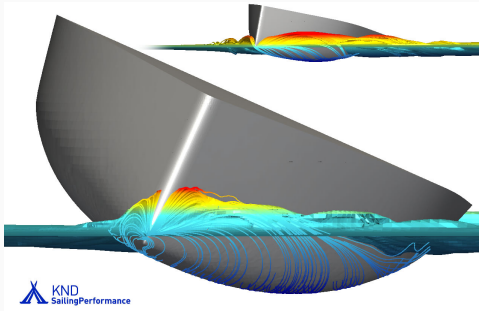
- Computing integrals
 - what is the reservoir oil content
- Uncertainty propagation
 - what is the impact score distribution
- Optimisation
 - what is the best hull design



- In a supervised learning setting, data comes as pairs $\{x_i, y_i\}$
 - Mining: x is the location, y is the oil quantity
 - Crash test: x is the car design, y is the impact score
 - Boat shape: x is the hull design, y is the drag force

- In a supervised learning setting, data comes as pairs $\{x_i, y_i\}$
 - Mining: x is the location, y is the oil quantity
 - Crash test: x is the car design, y is the impact score
 - Boat shape: x is the hull design, y is the drag force
- Commonalities,
 - y is costly to obtain (weeks to months) $\rightarrow n$ is small
 - x is high-dimensional
 - y is noise (measurement errors, variability of the physical system)

Numerical experiments



The principle behind simulation is to introduce a mathematical model of a physical phenomenon, and to use computation power to solve it

Two extremely common types of models are:

- Physics models based on partial differential equations (PDEs) (e.g. fluid dynamics, elasticity, electromagnetism).
- Discrete event models: The physical system is described by discrete events (e.g. queuing systems, network traffic).

There is often a trade-off between the accuracy of the model and its computational cost.

Partial Differential Equations

$$\begin{aligned}\frac{\partial u}{\partial t} + \mu(x, t) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 u}{\partial x^2} &= 0 & \forall x \in \Omega \\ u(x, T) &= g(x) & \forall x \in \partial\Omega\end{aligned}$$

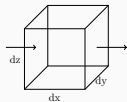
- multivariate function and one or more of its partial derivatives
- the function u is often unknown and considered its solution
- very important both in math¹ and Physics
- often not possible to explicitly solve

¹https://en.wikipedia.org/wiki/Ricci_flow

$$\rho\left(\frac{\partial v}{\partial t} + v \cdot \nabla v\right) = -\nabla p + \nabla \cdot T + f$$

Alexander Bastounis et al. (2026). “Navier-Stokes Lost in Translation: Why Lean Verification of Ai Autoformalisation Does Not Guarantee Correct Natural Language Proofs”. In: *CoRR*

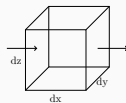
The Heat Equation (Fourier)



- Flux in through left-side

$$\frac{\partial u}{\partial x} dt dy dz$$

The Heat Equation (Fourier)



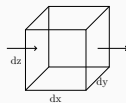
- Flux in through left-side

$$\frac{\partial u}{\partial x} dt dy dz$$

- Flux out through right-side

$$\frac{\partial u}{\partial x} dt dy dz - \partial \frac{\partial u}{\partial x} dt dy dz \approx \frac{\partial u}{\partial x} dx dt dy dz - \frac{\partial^2 u}{\partial x^2} dt dy dz$$

The Heat Equation (Fourier)



- Flux in through left-side

$$\frac{\partial u}{\partial x} dt dy dz$$

- Flux out through right-side

$$\frac{\partial u}{\partial x} dt dy dz - \frac{\partial u}{\partial x} dt dy dz \approx \frac{\partial u}{\partial x} dx dt dy dz - \frac{\partial^2 u}{\partial x^2} dt dy dz$$

- Flux through left and right

$$\frac{\partial u}{\partial x} dt dy dz - \left(\frac{\partial u}{\partial x} dt dy dz - \frac{\partial^2 u}{\partial x^2} dx dt dy dz \right) = \frac{\partial^2 u}{\partial x^2} dx dt dy dz$$

- Conservation (the heat inside the cube)

$$\begin{aligned}\partial u dV &= \frac{\partial^2 u}{\partial x^2} dx dt dy dz + \frac{\partial^2 u}{\partial y^2} dx dt dy dz + \frac{\partial^2 u}{\partial z^2} dx dt dy dz \\ &= \underbrace{dx dy dz}_{dV} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) dt\end{aligned}$$

- Conservation (the heat inside the cube)

$$\begin{aligned}\partial u dV &= \frac{\partial^2 u}{\partial x^2} dx dt dy dz + \frac{\partial^2 u}{\partial y^2} dx dt dy dz + \frac{\partial^2 u}{\partial z^2} dx dt dy dz \\ &= \underbrace{dx dy dz}_{dV} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) dt\end{aligned}$$

- Re-arrange terms

$$\frac{\partial u}{\partial t} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right)$$

Heath Equation (Fourier)

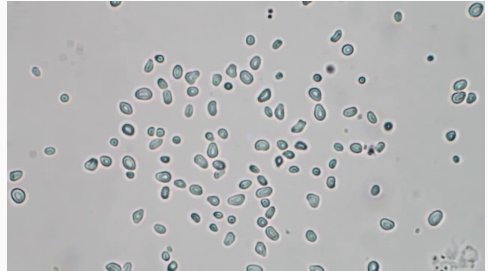
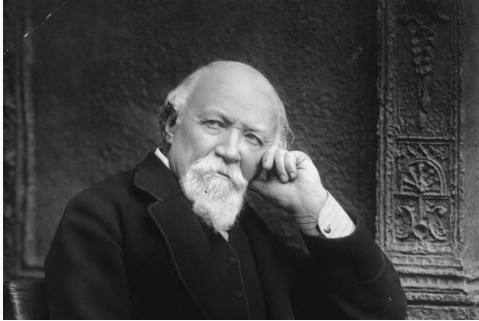
- Conservation (the heat inside the cube)

$$\begin{aligned}\partial u dV &= \frac{\partial^2 u}{\partial x^2} dx dt dy dz + \frac{\partial^2 u}{\partial y^2} dx dt dy dz + \frac{\partial^2 u}{\partial z^2} dx dt dy dz \\ &= \underbrace{dx dy dz}_{dV} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) dt\end{aligned}$$

- Re-arrange terms

$$\frac{\partial u}{\partial t} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right)$$

- α is a heat conductivity term that scales the relationship



- We have K particles

- We have K particles
- Each particle is displaced $x + d\epsilon$ during time T

- We have K particles
- Each particle is displaced $x + d\epsilon$ during time T
- The displacement follows a probability $p(\epsilon)$,

- We have K particles
- Each particle is displaced $x + d\epsilon$ during time T
- The displacement follows a probability $p(\epsilon)$,
- Medium is homogenous

$$p(\epsilon) = p(-\epsilon)$$

- We have K particles
- Each particle is displaced $x + d\epsilon$ during time T
- The displacement follows a probability $p(\epsilon)$,
- Medium is homogenous

$$p(\epsilon) = p(-\epsilon)$$

- Expected displacement

$$\mathbb{E}[\epsilon]_{p(\epsilon)} = \int_{-\infty}^{\infty} \epsilon p(\epsilon) d\epsilon = 0$$

- Number of particles at time t and location x is $u(x, t)$

$$\int_{-\infty}^{\infty} u(x, t) dx = K$$

- Number of particles at time t and location x is $u(x, t)$

$$\int_{-\infty}^{\infty} u(x, t) dx = K$$

- We want to compute the distribution of particles after a time interval T

$$u(x, t + T) = \int_{-\infty}^{\infty} u(x + \epsilon, t) p(\epsilon) d\epsilon$$

- Time

$$u(x, t + T) \approx u(x, t) + \frac{\partial u}{\partial t} T$$

- Space

$$u(x + \epsilon, t) \approx u(x, t) + \frac{\partial u}{\partial x} \epsilon + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \epsilon^2$$

$$\begin{aligned}u(x, t + T) &= \int_{-\infty}^{\infty} u(x + \epsilon, t) p(\epsilon) d\epsilon \\&= \int_{-\infty}^{\infty} \left(u(x, t) + \frac{\partial u}{\partial x} \epsilon + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \epsilon^2 \right) p(\epsilon) d\epsilon \\&= u(x, t) \underbrace{\int_{-\infty}^{\infty} p(\epsilon) d\epsilon}_1 + \frac{\partial u}{\partial x} \underbrace{\int_{-\infty}^{\infty} \epsilon p(\epsilon) d\epsilon}_0 + \frac{\partial^2 u}{\partial x^2} \frac{1}{2} \underbrace{\int_{-\infty}^{\infty} \epsilon^2 p(\epsilon) d\epsilon}_{\alpha T} \\&= u(x, t) + \alpha \frac{\partial^2 u}{\partial x^2} T\end{aligned}$$

- Now we have two expressions for the change in particle density

$$u(x, t) + \frac{\partial u}{\partial t} T = u(x, t) + \alpha \frac{\partial^2 u}{\partial x^2} T$$
$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}$$

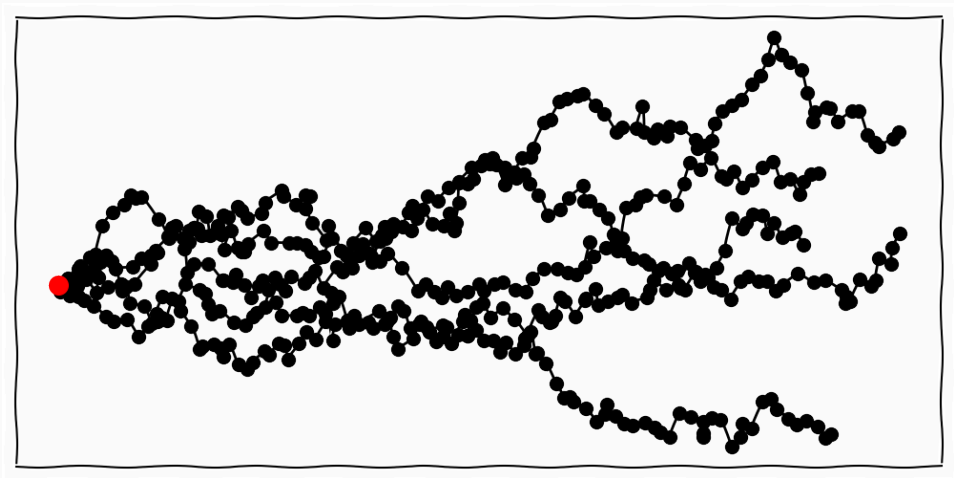
- So the distribution of particles in a homogenous fluid follows the same PDE as heat diffusion

- $\{B_t, t \geq 0\}$ is a Brownian Motion with variance σ^2 if,

$$B_t = W_t + x_0$$

$$W_{t_0}, W_{t_1} - W_{t_0}, \dots, W_{t_n} - W_{t_{n-1}} \sim \mathcal{N}(0, \sigma^2 t)$$

Simulation of Brownian Motion



- Transition Density

$$p(x \mid t, x_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-x_0)^2}{2\sigma^2 t}}$$

²take log of both sides

- Transition Density

$$p(x \mid t, x_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-x_0)^2}{2\sigma^2 t}}$$

- Partial derivatives²

$$\frac{\partial}{\partial t} p(x \mid t, x_0) = \frac{1}{2} \frac{\partial^2}{\partial x^2} p(x \mid t, x_0)$$

²take log of both sides

- Transition Density

$$p(x \mid t, x_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-x_0)^2}{2\sigma^2 t}}$$

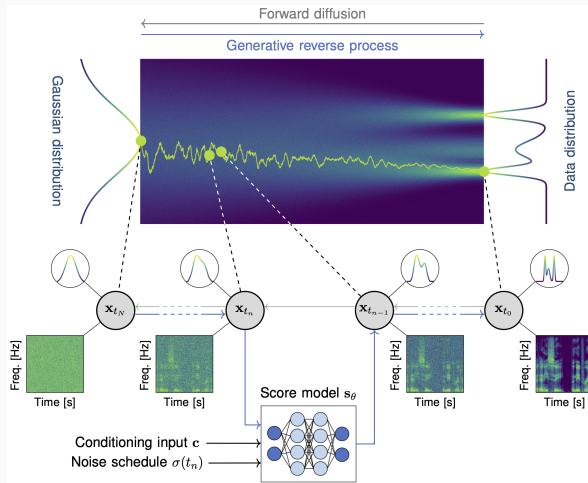
- Partial derivatives²

$$\frac{\partial}{\partial t} p(x \mid t, x_0) = \frac{1}{2} \frac{\partial^2}{\partial x^2} p(x \mid t, x_0)$$

- *Brownian motion solves the heat equation*

²take log of both sides

Diffusion Models ML



- Partial Differential Equation

$$\frac{\partial u}{\partial t} + \mu(x, t) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 u}{\partial x^2} = 0 \quad \forall x \in \Omega$$

$$u(x, T) = g(x) \quad \forall x \in \partial\Omega$$

- Partial Differential Equation

$$\begin{aligned}\frac{\partial u}{\partial t} + \mu(x, t) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 u}{\partial x^2} &= 0 & \forall x \in \Omega \\ u(x, T) &= g(x) & \forall x \in \partial\Omega\end{aligned}$$

- Stochastic Process

$$u(x, t) = \mathbb{E}[g(X_T) \mid X_t \in \partial\Omega] \approx \frac{1}{N} \sum_i^N g(x_T^{(n)})$$

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t)$$

$$W(t + dt) - W(t) \sim \mathcal{N}(0, dt)$$

- Partial Differential Equation

$$\begin{aligned}\frac{\partial u}{\partial t} + \mu(x, t) \frac{\partial u}{\partial x} + \frac{1}{2} \sigma^2(x, t) \frac{\partial^2 u}{\partial x^2} &= 0 & \forall x \in \Omega \\ u(x, T) &= g(x) & \forall x \in \partial\Omega\end{aligned}$$

- Stochastic Process

$$u(x, t) = \mathbb{E}[g(X_T) \mid X_t = x] \approx \frac{1}{N} \sum_i^N g(x_T^{(n)})$$

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t)$$

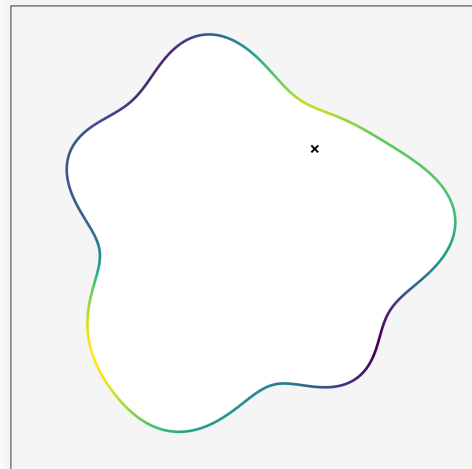
$$W(t + dt) - W(t) \sim \mathcal{N}(0, dt)$$

- We can simulate the PDE using an approximation of the stochastic process using a Euler–Maruyama approximation

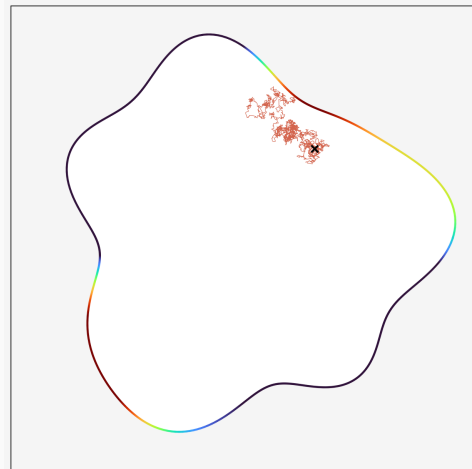
Steady state heat equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0 \quad \forall x \in \Omega$$

$$u(x) = g(x) \quad \forall x \in \Omega$$



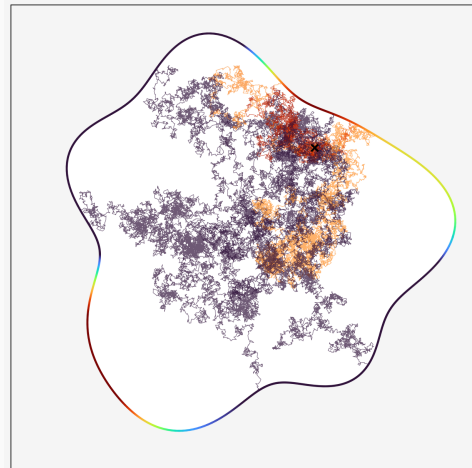
Single Brownian Motion Path



Example Feynman-Kac

10 paths

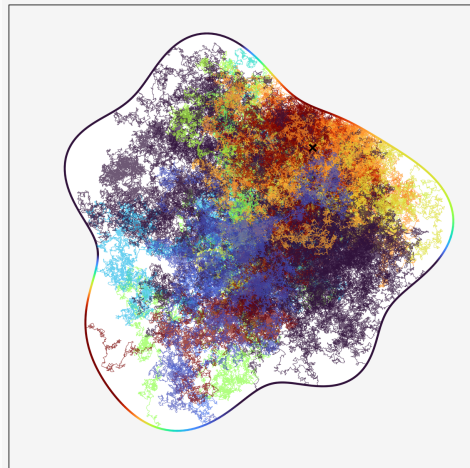
- $u(x) \approx 0.21$
- standard deviation 0.20



Example Feynman-Kac

100 paths

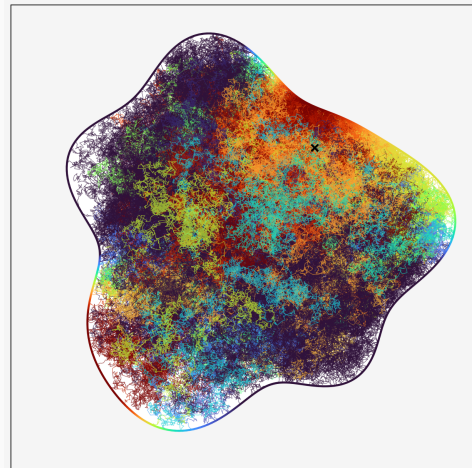
- $u(x) \approx 0.47$
- standard deviation 0.06



Example Feynman-Kac

1000 paths

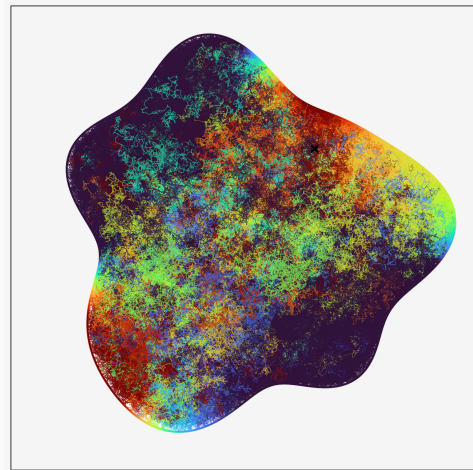
- $u(x) \approx 0.55$
- standard deviation 0.02



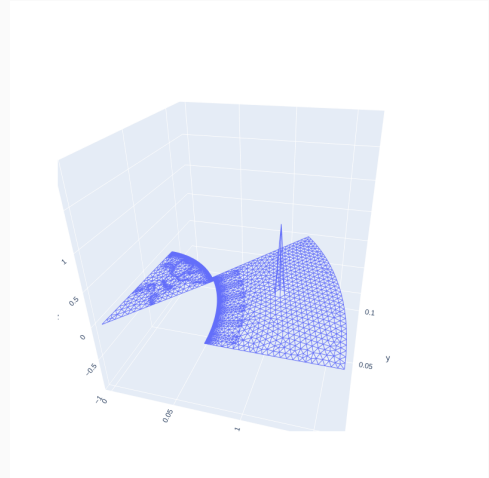
Example Feynman-Kac

10000 paths

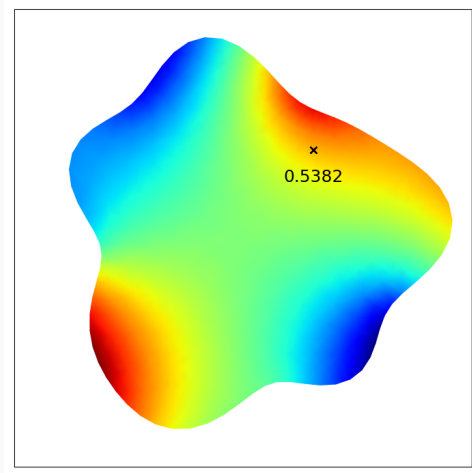
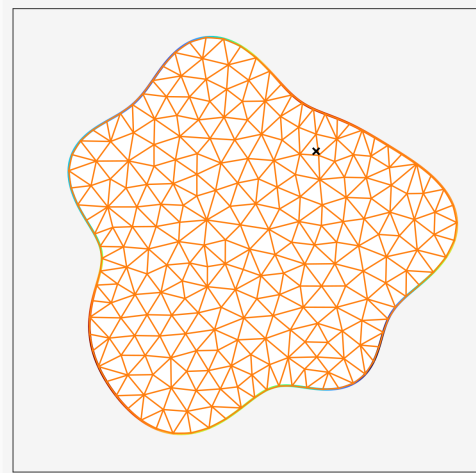
- $u(x) \approx 0.53$
- standard deviation 0.00



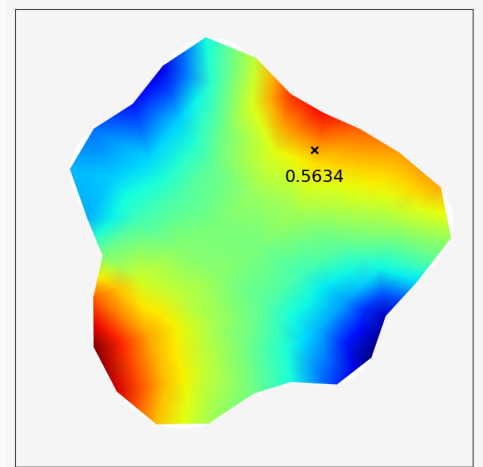
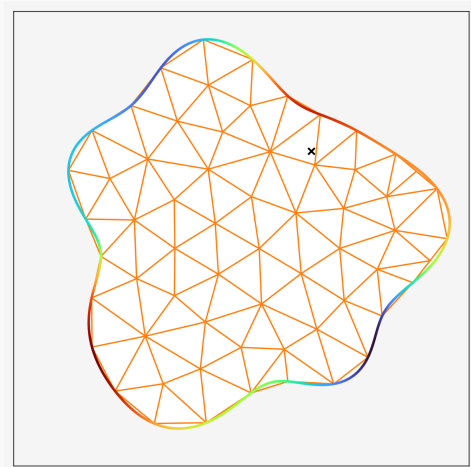
- Approximate domain with a set of finite simple elements
- Approximate integral across domain as a sum
- Leads to system of linear equations
- Interpolate out solution



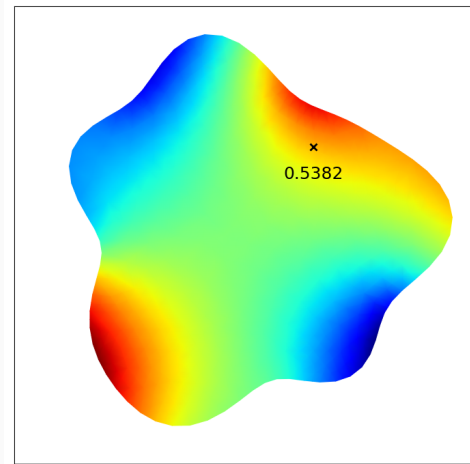
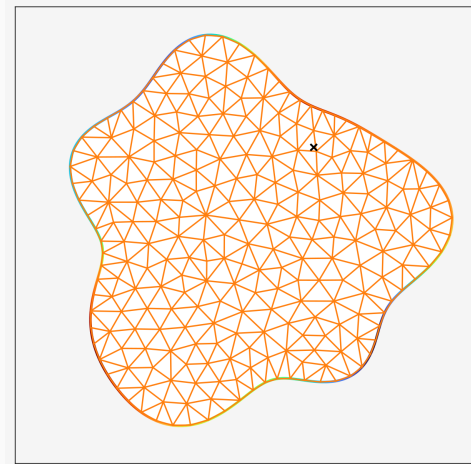
Example FEM



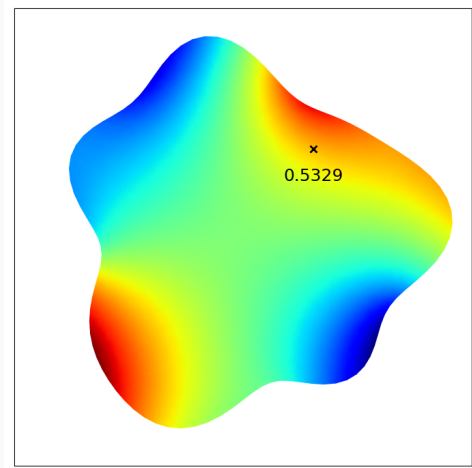
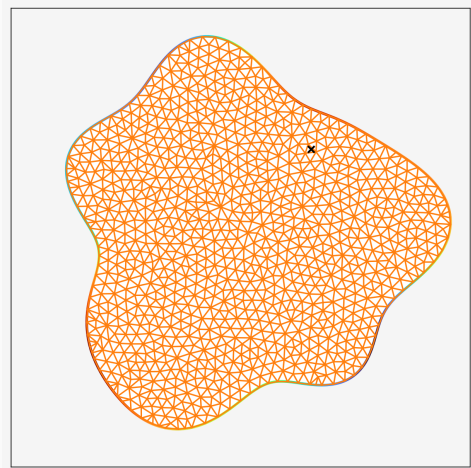
Example FEM



Example FEM



Example FEM

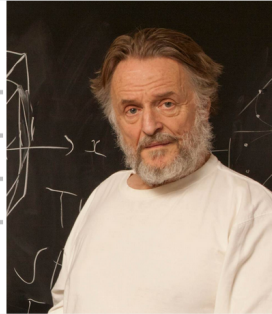
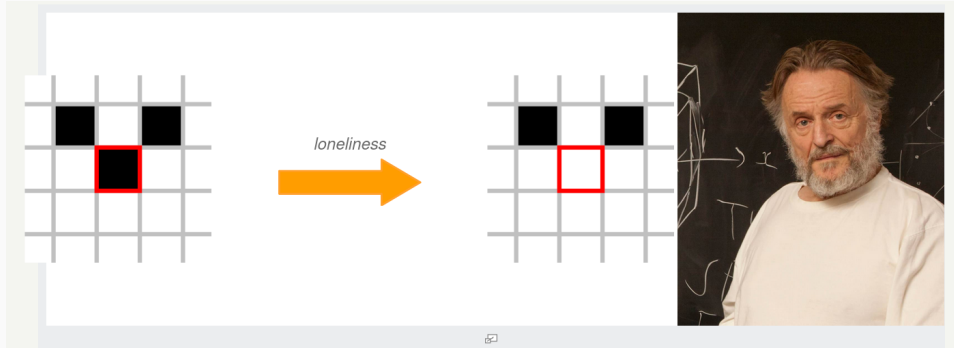


- FEM Solver
 - Computes solution "everywhere"
 - Makes quantification of error challenging
 - Suffers from curse of dimensionality

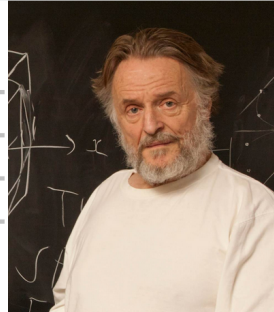
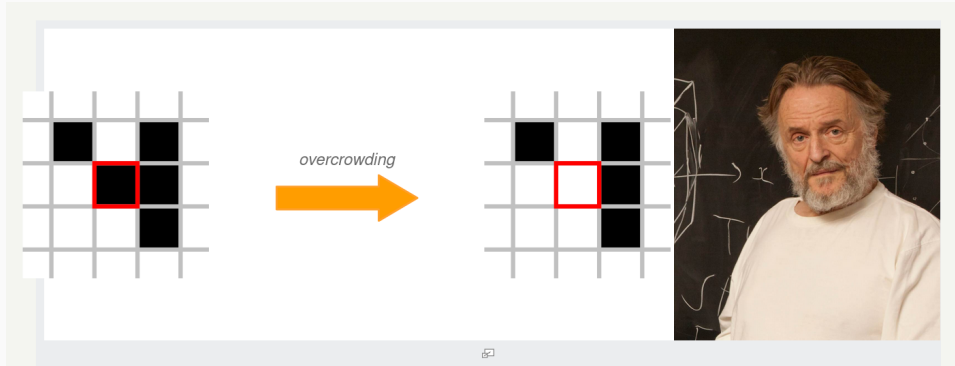
- FEM Solver
 - Computes solution "everywhere"
 - Makes quantification of error challenging
 - Suffers from curse of dimensionality
- Monte Carlo Solver
 - Computes the solution at selected point
 - Error quantification straight forward
 - Convergence is slow (but independent of dimension)

Emergent Behaviours

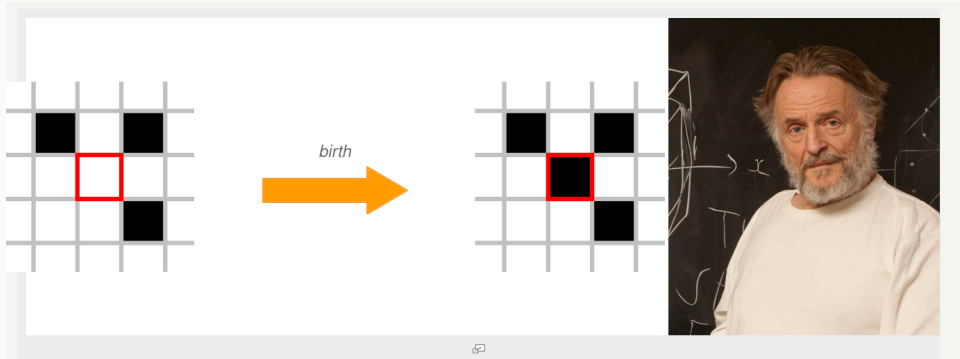
Game of Life

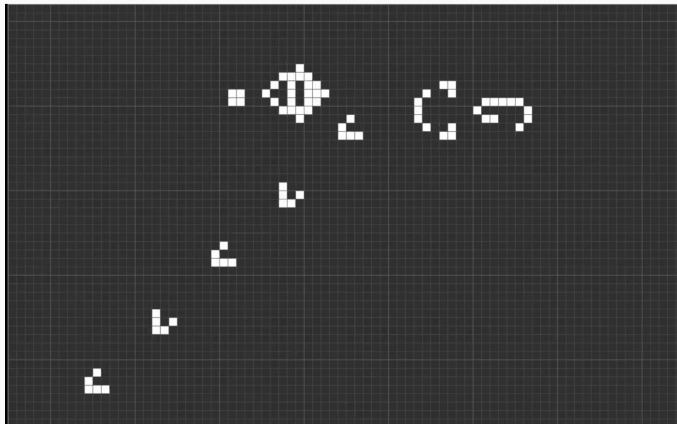


Game of Life



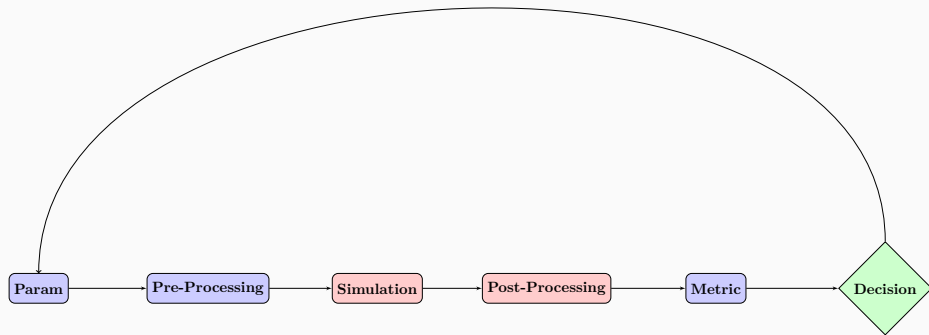
Game of Life

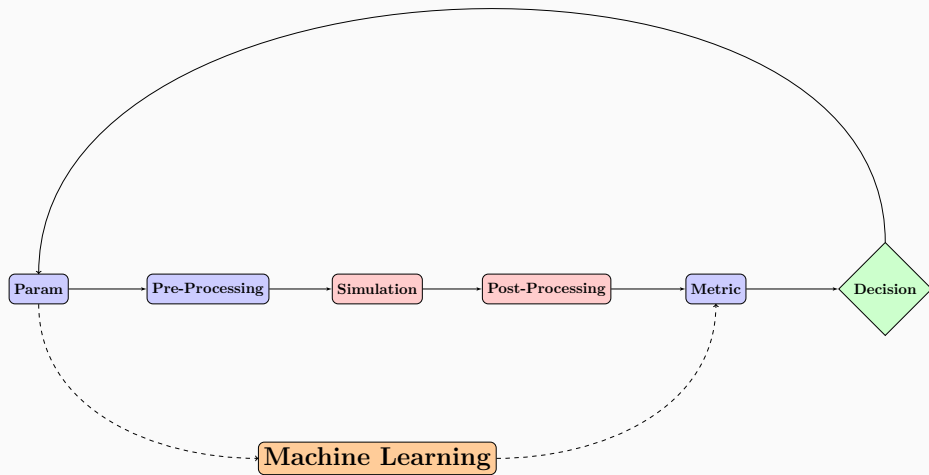


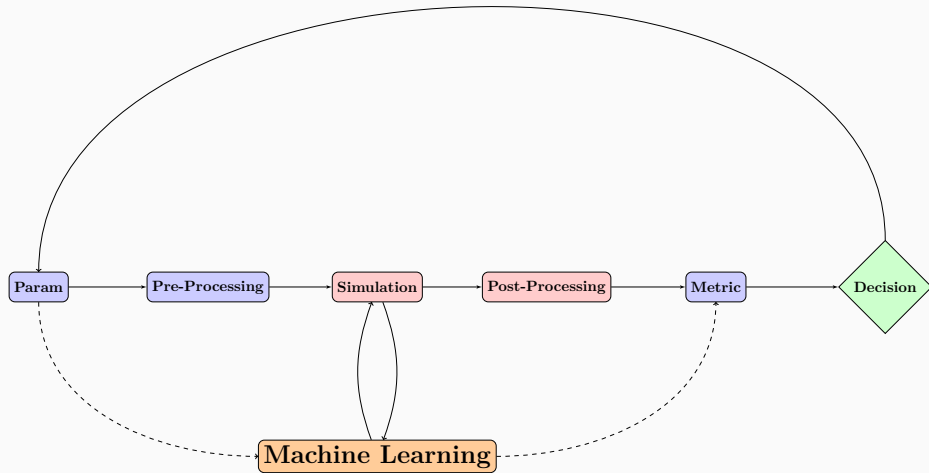


<https://youtu.be/C2vgICfQawE?t=77>

Machine Learning and Simulations







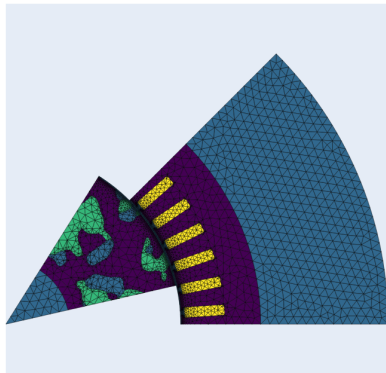
Model

- Operate on mesh/graph
- Predict magnetic field value

Benefits

- No data imbalance
- same granularity as FEM
- Gradients
- Simpler tasks?

³Nicholas Durrande Monumo



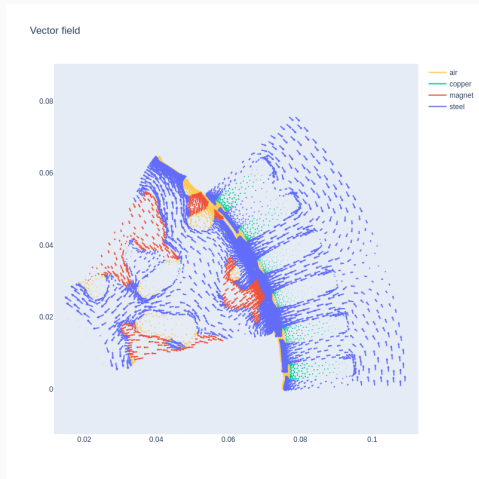
Model

- Operate on mesh/graph
- Predict magnetic field value

Benefits

- No data imbalance
- same granularity as FEM
- Gradients
- Simpler tasks?

³Nicholas Durrande Monumo

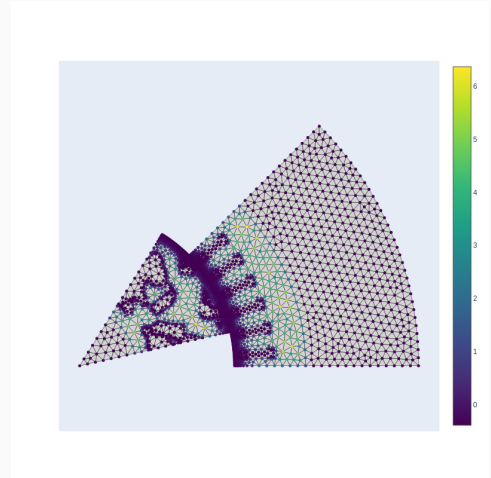


Input

- vertex location
- material
- magnetic polarisation
- current

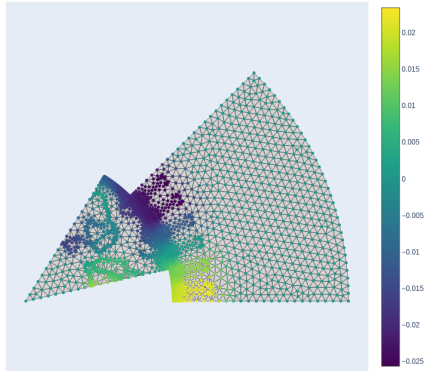
Output

- A and B field

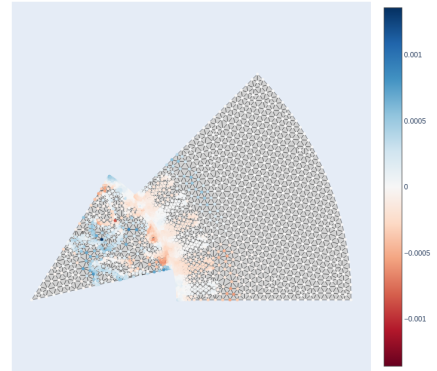


FEM example

Predicted a-field

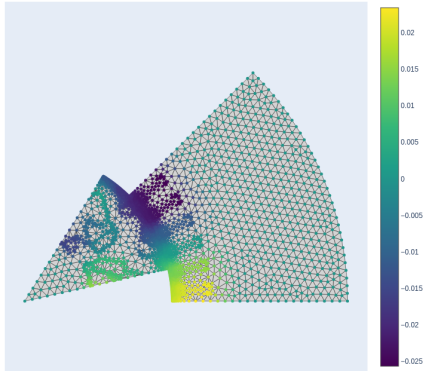


a-field prediction error (actual - predicted)

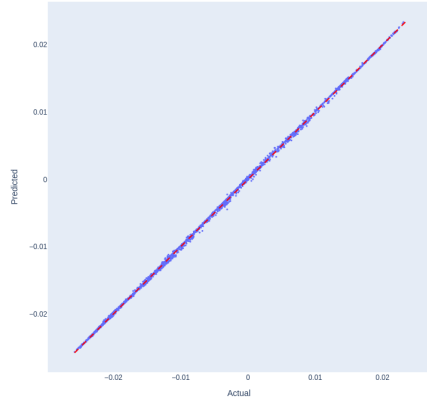


FEM example

Predicted a-field



True vs Pred



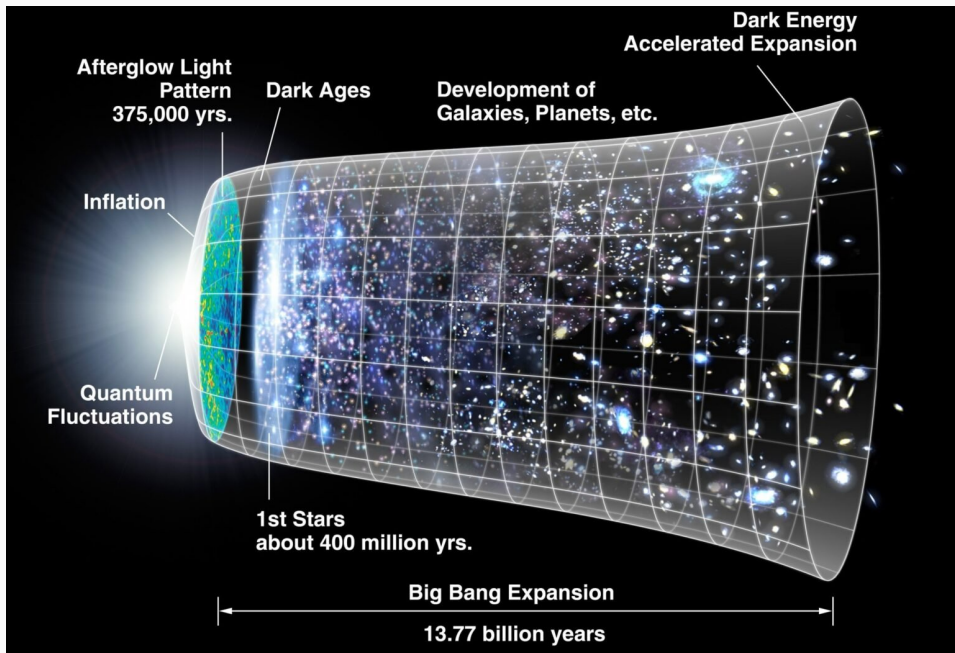
Summary

- Real world data is typically extremely costly to obtain
- Simulation is a key component of the science workflow to reduce cost based on first principles
- Statistical Modelling can act as a simplifier

- Real world data is typically extremely costly to obtain
- Simulation is a key component of the science workflow to reduce cost based on first principles
- Statistical Modelling can act as a simplifier
 - Focus on Quantity of interest

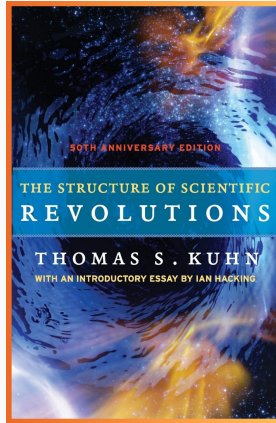
- Real world data is typically extremely costly to obtain
- Simulation is a key component of the science workflow to reduce cost based on first principles
- Statistical Modelling can act as a simplifier
 - Focus on Quantity of interest
 - Multi-fidelity models: use real data, simulated data, multiple simulators

- Real world data is typically extremely costly to obtain
- Simulation is a key component of the science workflow to reduce cost based on first principles
- Statistical Modelling can act as a simplifier
 - Focus on Quantity of interest
 - Multi-fidelity models: use real data, simulated data, multiple simulators
 - Experimental design



- Quantification of Beliefs
- Gaussian Processes

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.



eof