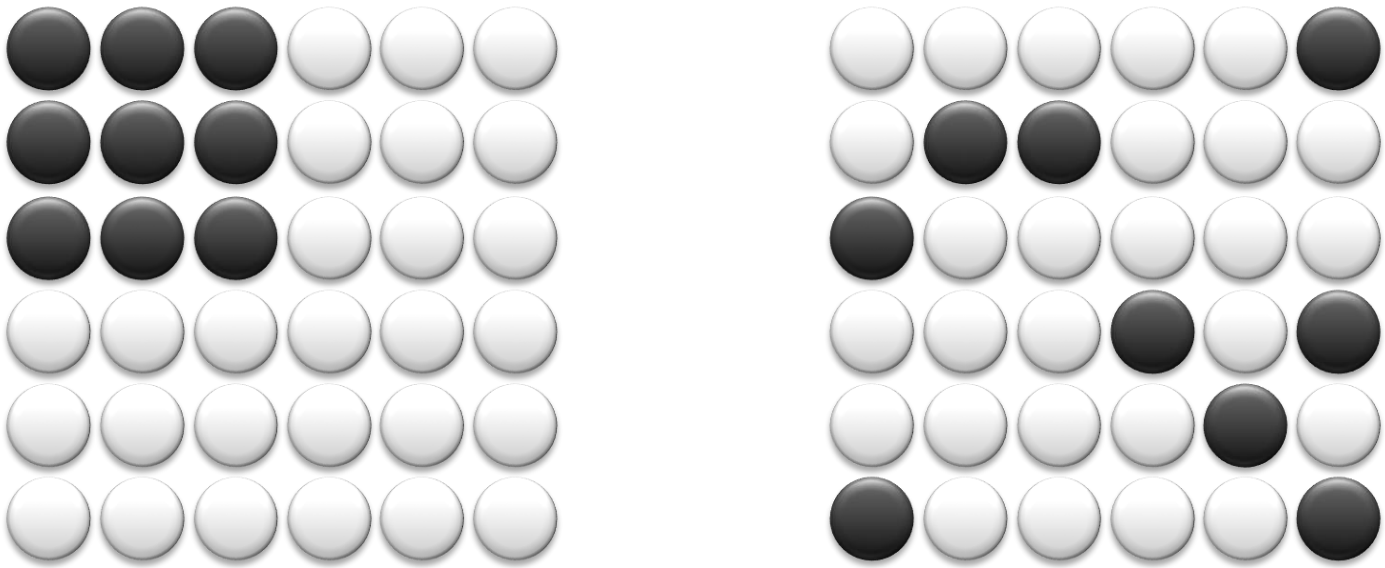
Modelling 2

STATISTICAL DATA MODELLING



Chapter 12

Physics and Self-Organization

Michael Wand · Institut für Informatik, JGU Mainz · michael.wand@uni-mainz.de

Video #12

Physics & Self-Organization

- **Physics**
- **Self-Organization**

Introduction

Physics & Machine Learning?

- Why does would this matter?

Big problems

- "*The*" research question: Solve AI
- In other words: Universal machine learning

Big question

- Does universal learning exist?

Introduction

Meta-Priors

- **Math:** No free lunch!
- **Physics:** ? ← our focus
- **Biology:** Sure (if you are optimistic)

Computer Science

- < 2010: We have none, but good luck.
- 2010-2020: Ups, maybe
  - Deep networks solve very limited tasks
  - But they seem disturbingly universal at that

Introduction

Perspective

- Taking fundamental physics as we know it as model
- Does this tell us
  how to build a universal learning machine?
- We will not be able to answer this question.

Three steps

1. All of physics in 45min (from a CS perspective)
2. Methods / results on self-organizing systems
3. Do your own research (beyond this lecture)

Disclaimer

I am not a physicist

- Educated in computer science
- This lecture gives an overview / starting point
- This is not a physics lecture
- Take everything with a grain of scepticism

Physical Dynamics

Physics: Dynamical Systems

Dynamical systems

- State Space ("microstates")
  - Set Ω of possible system states
- Examples
  - 8 planets orbiting the sun

$$\Omega = \mathbb{R}^{2\times 3\times 9} = \mathbb{R}^{54}$$

  - Cellular automata
    - Discrete state space on infinite grid

$$\Omega = \{\omega_1, \dots, \omega_N\}^{\mathbb{Z}^2}$$

- Temporal evolution

$$\text{function } s: \mathbb{R} \to \Omega$$

Dynamical Systems

Dynamics

- State evolves over time
  - The future only depends on the “last time step”
- Continuous case:

$$f: \mathbb{R} \to \Omega, t \mapsto f(t)$$

$$\frac{d}{dt} f(t) = F(f(t), t)$$

- Discrete case:

$$f: \mathbb{Z} \to \Omega, t \mapsto s(t)$$

$$f(t+1) = F(f(t), t)$$

Dynamical Systems

Dynamics

- State evolves over time
  - The future only depends on the "last time step"
- Continuous case

$$f: \mathbb{R} \to \Omega, t \mapsto f(t)$$

$$\frac{d}{dt} f(t) = F(f(t), t) \text{ (Markov process)}$$

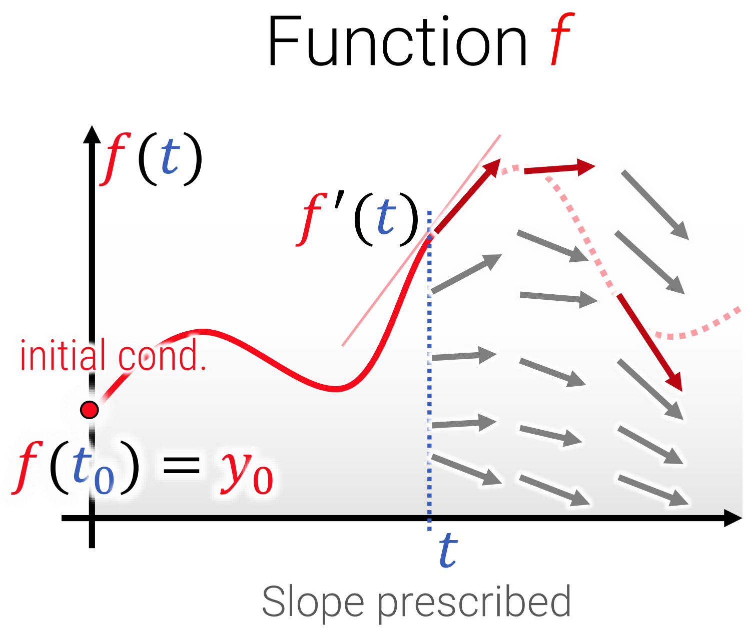
- Discrete case

$$f: \mathbb{Z} \to \Omega, t \mapsto s(t)$$

$$f(t+1) = F(f(t), t) \text{ (Markov chain)}$$

- "New information": F can be a random variable
  - Markovian dynamics

Integration



$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$f'(t) = F(f(t), t)$$

Newtonian Physics

Newtonian Physics

Newtonian Physics

$$\text{“}F = m \cdot a\text{”}$$

Which means:

$$\mathbf{F}\big(t, \mathbf{s}(t)\big) = m \cdot \mathbf{a}(t) = m \cdot \ddot{\mathbf{s}}(t)$$

In other words…

$$\frac{d^2}{dt^2}\mathbf{s}(t) = \frac{1}{m}\mathbf{F}\big(t, \mathbf{s}(t)\big)$$

Typical Forces

force vector

$$\mathbf{F}_{1,2} = \|\mathbf{F}\| \frac{(\mathbf{s}_i - \mathbf{s}_j)}{\|\mathbf{s}_i - \mathbf{s}_j\|}$$

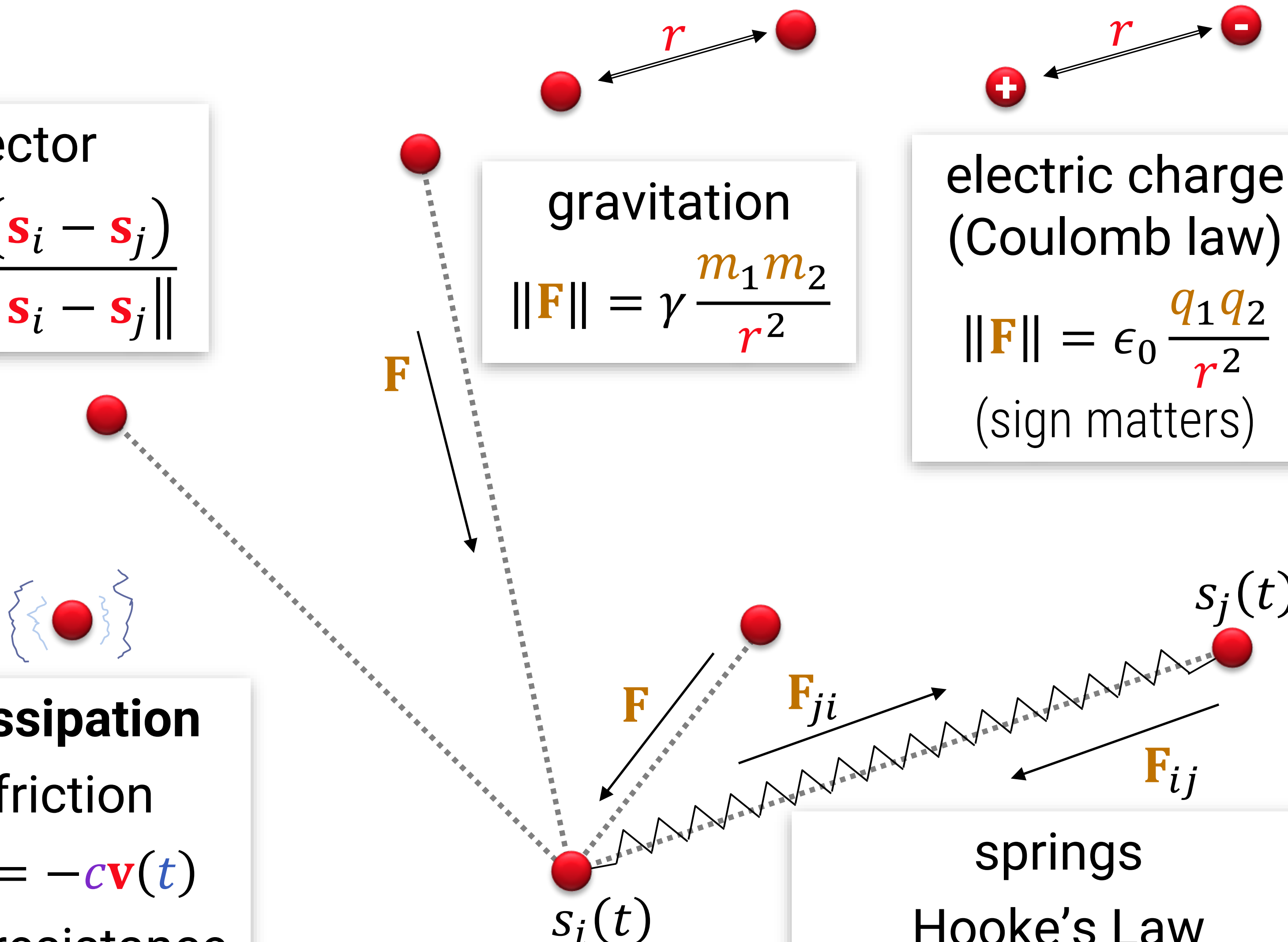
gravitation

$$\|\mathbf{F}\| = \gamma \frac{m_1 m_2}{r^2}$$

electric charge
(Coulomb law)

$$\|\mathbf{F}\| = \epsilon_0 \frac{q_1 q_2}{r^2}$$

(sign matters)



dissipation

friction

$$\mathbf{F} = -c\mathbf{v}(t)$$

air resistance

$$\mathbf{F} = -c\mathbf{v}(t)^2$$

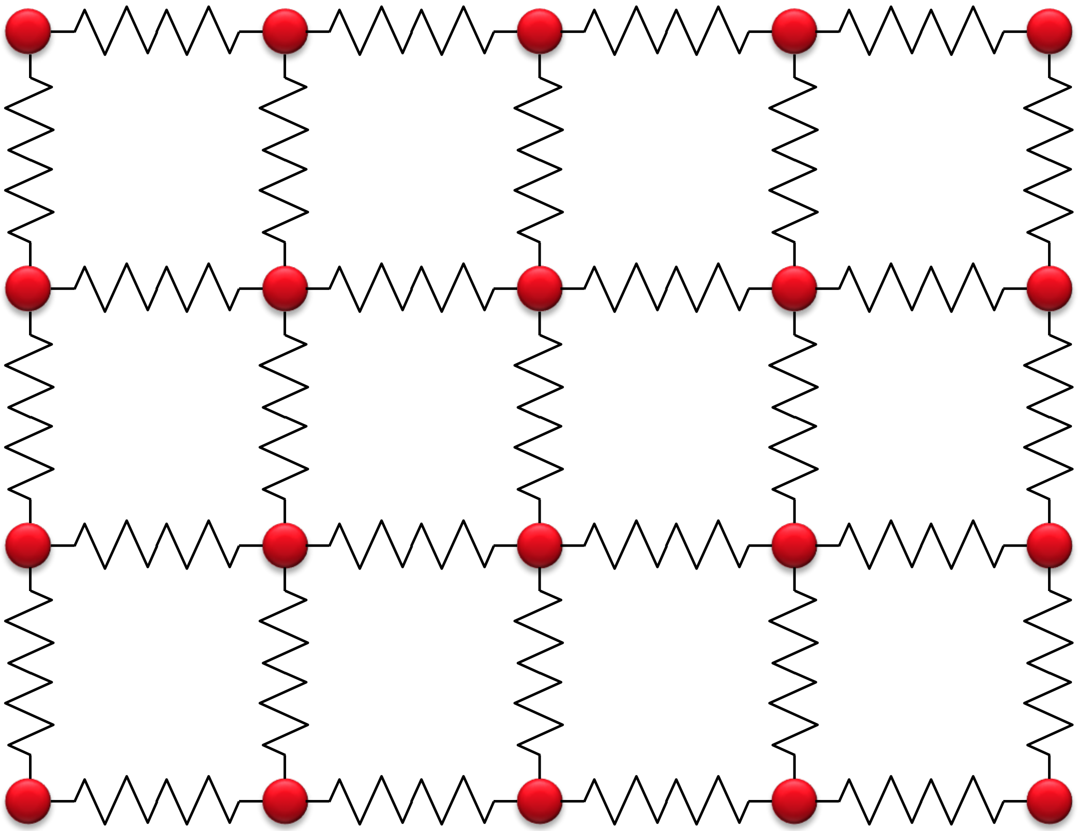
springs

Hooke's Law

$$\mathbf{F}_{ij} = k\|\mathbf{s}_j(t) - \mathbf{s}_i(t)\|$$

Classical
Fields & Waves

We Like the Springs



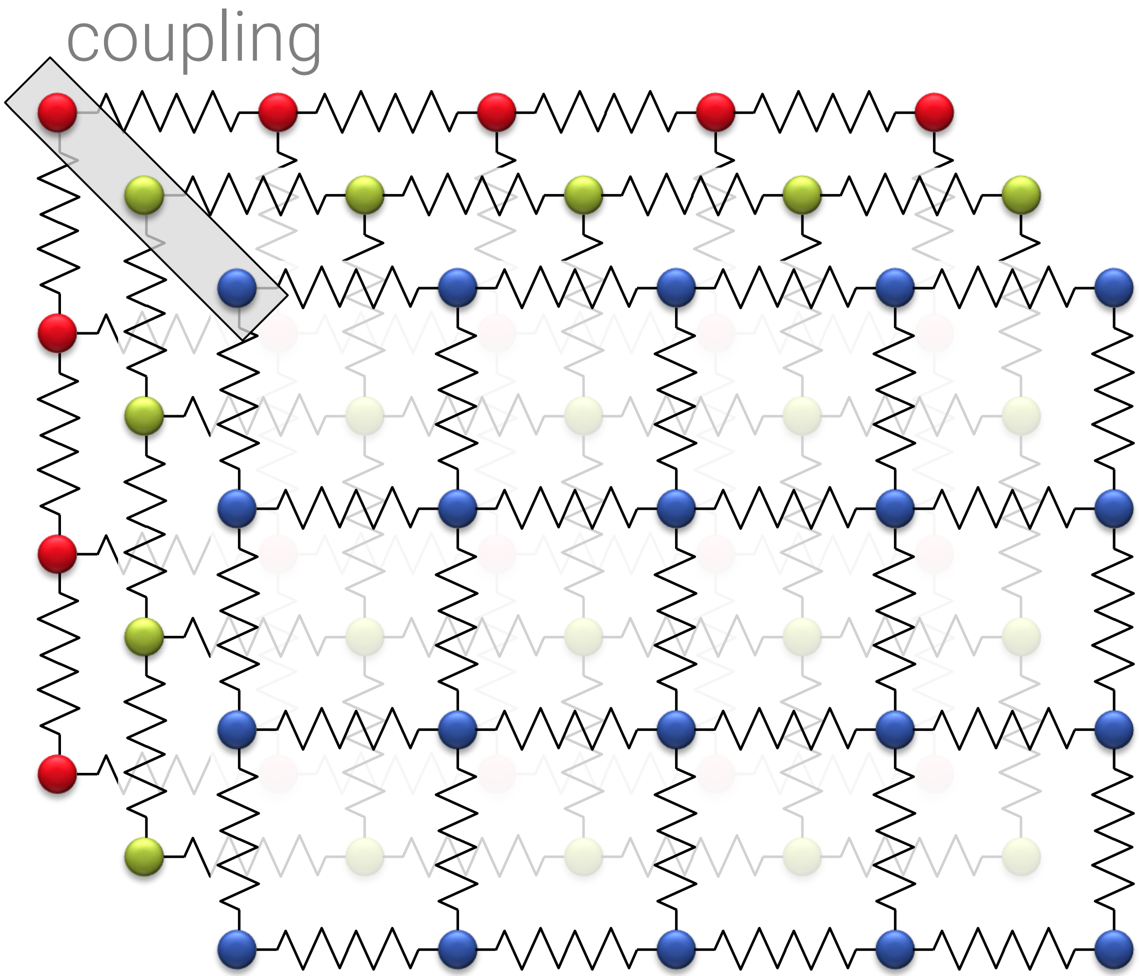
Wave Equation

$$\frac{\partial f^2}{\partial t} = -c\Delta f(\mathbf{x}, t)$$

Information

- Information transported
  - through space
  - over time
- Reversible
  - practical

Classical Wave Models



General case

- More general local interactions
- More complex dynamics
- Wave propagation still possible

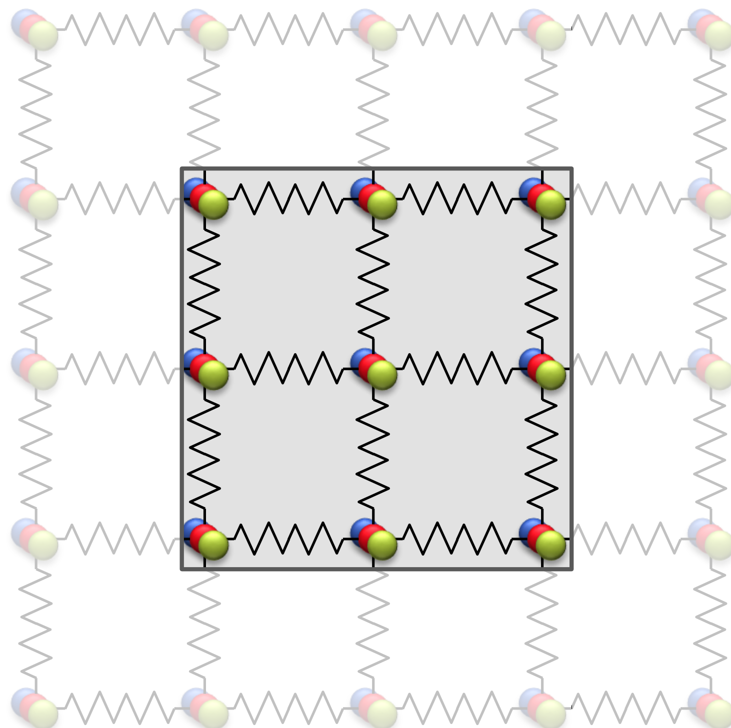
Concrete model

- Usually involves coupled fields

General Case

Local rules

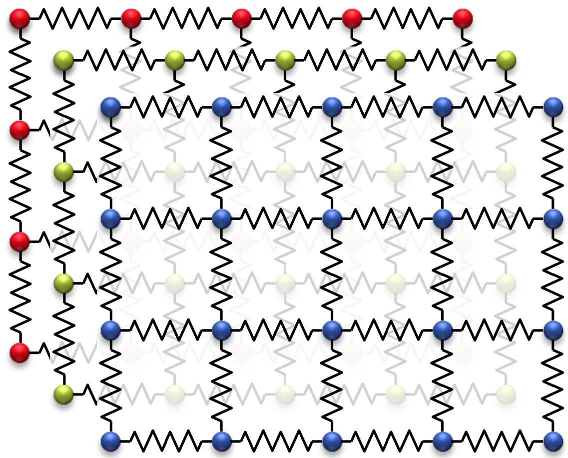
- At each time instance, affect only direct neighborhood
- Information is transported through space over time



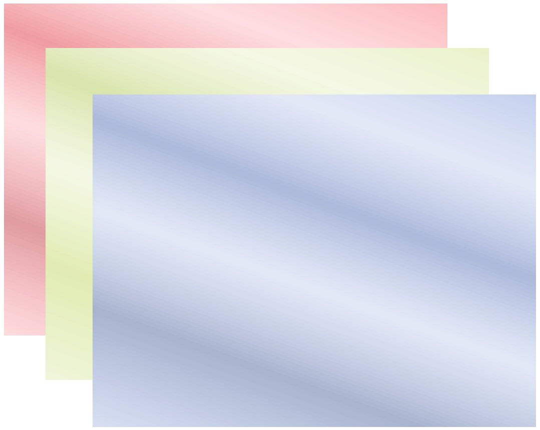
Symmetric

- Translations
- Rotations
- (Reflections)

Continuum



Local neighborhoods

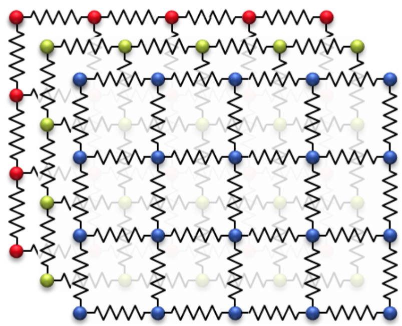


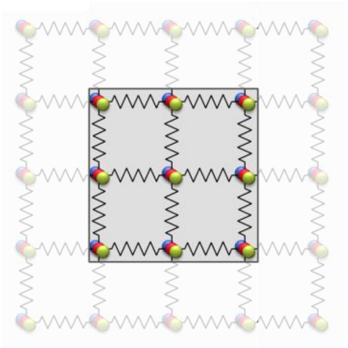
$\partial_t, \nabla_{\mathbf{x}}, \Delta_{\mathbf{x}}, \dots$

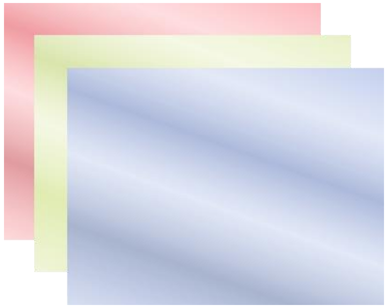
“ ”

“cellular automata”
as discrete model systems

Properties







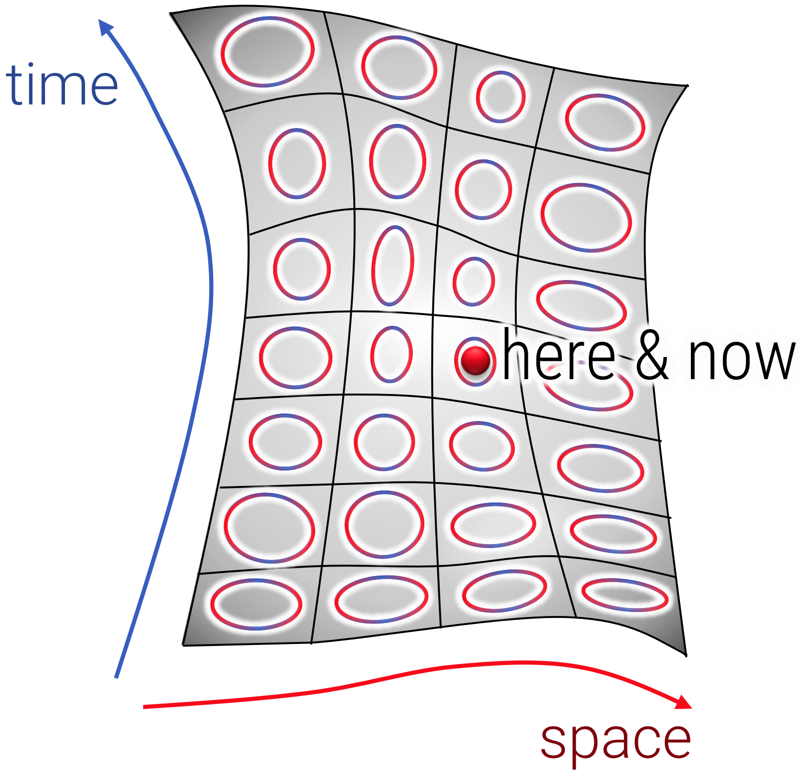
Locality

- Markovian: all memory in dynamic state
- Local interactions evolve over time

Symmetry

- Relativistic (invariant under Poincaré Group)
- Causality chains: Local interactions → global behavior

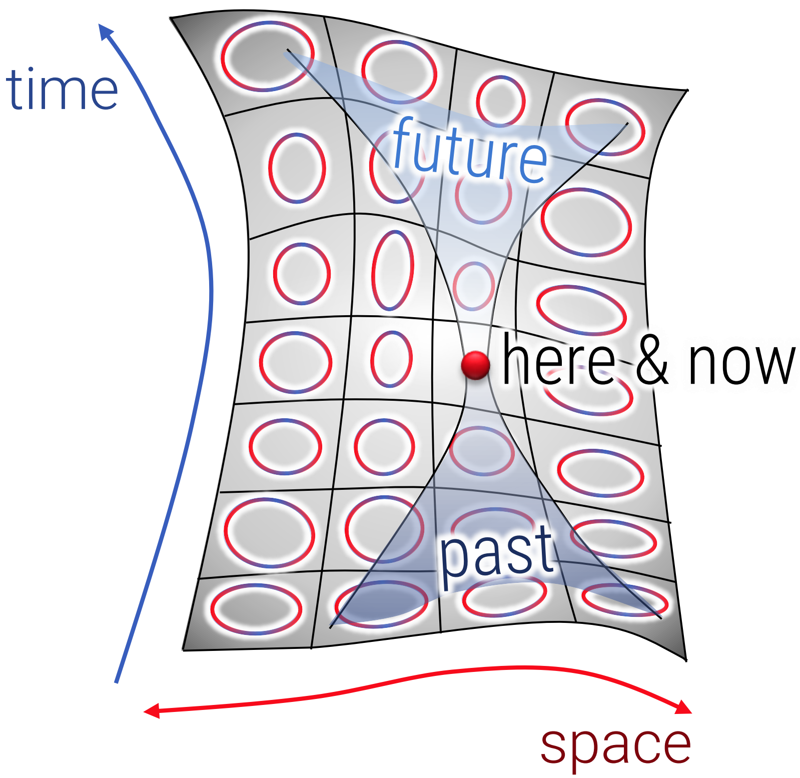
General Relativity: Locality to the max



General Relativity

- Curved spacetime
- “General covariance”
  - One could deform it arbitrarily
  - Relevant are causal chains
- Information propagates according to metric

General Relativity: Locality to the max

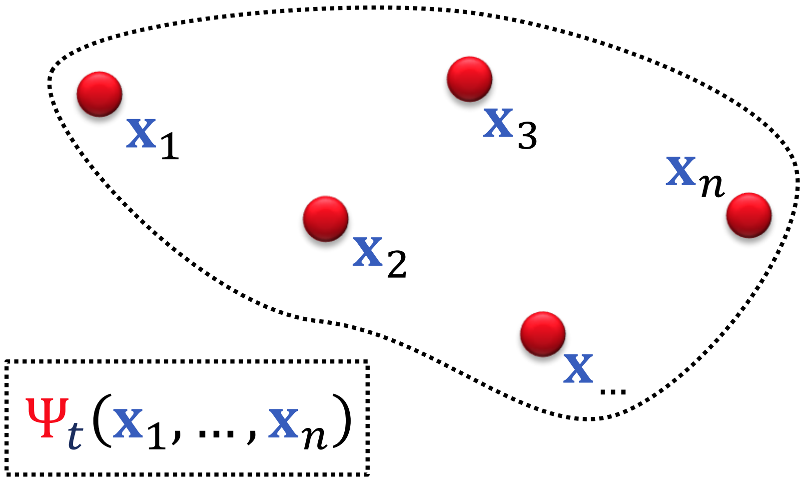


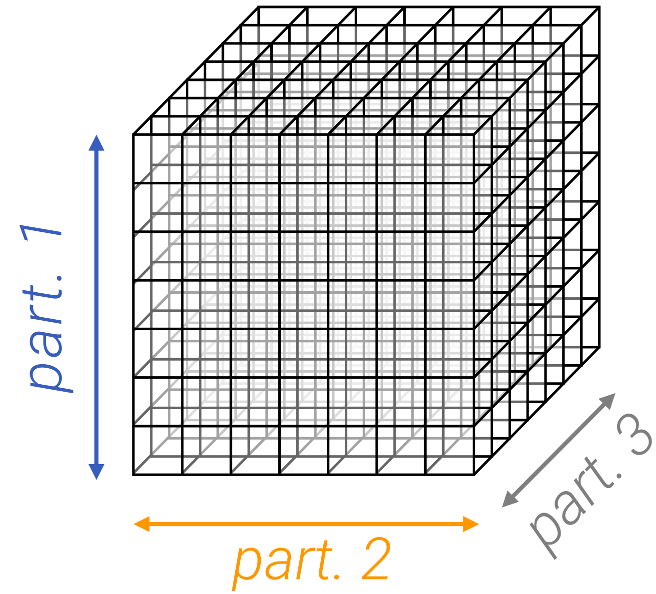
General Relativity

- Curved spacetime
- "General covariance"
  - One could deform it arbitrarily
  - Relevant are causal chains
- Information propagates according to metric

Quantum Mechanics

We Probably Like the Cat





"Schrödinger"-Style QM:

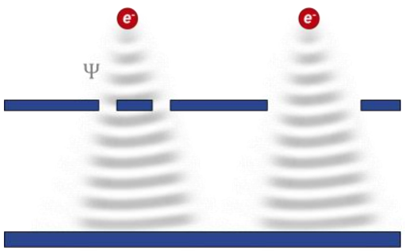
- n Particles in $\mathbb{R}^3$
- Time $t \in \mathbb{R}$
- Wave function

$$\Psi_t\big(\underbrace{\mathbf{x}_1, \dots, \mathbf{x}_n}_{=:\mathbf{X}}\big) : \mathbb{R}^{3n} \times \mathbb{R} \to \mathbb{C}$$

We Probably Like the Cat

Wave function

$$\Psi_t\Big(\underbrace{\mathbf{x}_1, \dots, \mathbf{x}_n}_{=:\mathbf{X}}\Big) : \mathbb{R}^{3n} \times \mathbb{R} \to \mathbb{C}$$



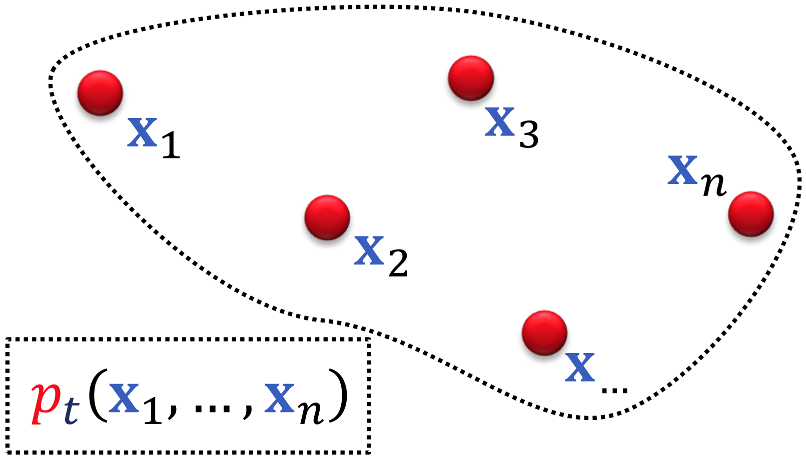
Born-Rule

$$p_t(\mathbf{X}) = |\Psi_t(\mathbf{X})|^2$$

Rules

- Dynamics $\frac{\partial}{\partial t}\Psi_t(\mathbf{X}) = \frac{1}{i\hbar}\hat{\mathbf{H}}\Psi_t(\mathbf{X})$
  - Unitarian, non-linear operator $\hat{\mathbf{H}}\Psi_t(\mathbf{X}) = \hat{\mathbf{H}}\big(\Psi_t(\mathbf{X})\big) \cdot \Psi_t(\mathbf{X})$
- General observables $\Psi_t(\mathbf{X})^{\mathrm{T}} \cdot \mathbf{M} \cdot \Psi_t(\mathbf{X})$
  - Hermitian (complex-symmetric) $\mathbf{M}$
  - Observe Eigenfunctions of $\mathbf{M}$ with p = eigenvalue

Quantum Field Theory



$$\Psi_t(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^3} z_{\mathbf{k}} e^{i\mathbf{k}\mathbf{x}}, z_{\mathbf{k}} \in \mathbb{N}$$

$$p_t(z_{\mathbf{k}_1}, z_{\mathbf{k}_2}, \dots)$$

Schrödinger's Problem

- Cannot create/remove/convert particles
- Not relativistic

Quantum Field theory

- ≈ statistics on Fourier-coefficients of fields
- Maintains symmetries of special relativity
  - Standard model: No general relativity

QM – tl;dr

We compute a Wavefunction

$$\Psi_t(\mathbf{X}): \mathbb{R}^{3n} \times \mathbb{R} \to \mathbb{C}$$

This yields a distribution

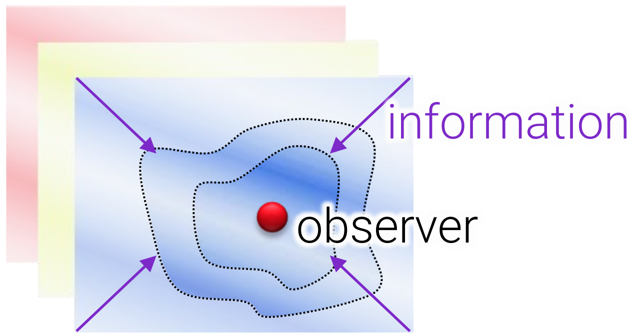
$$p_t(\mathbf{X}) = |\Psi_t(\mathbf{X})|^2$$

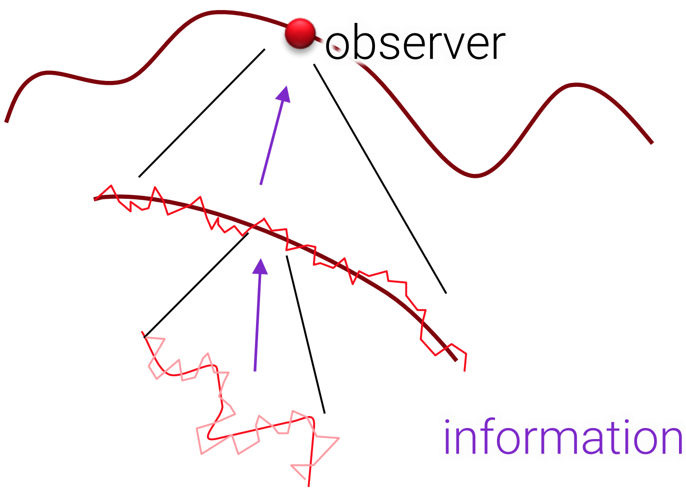
We sample once from p

- Obtain $\mathbf{X}_t \in \mathbb{R} \to \mathbb{R}^{3n}$
- This is life

Information (= Randomness) in Physics

Probabilistic Models of Physics

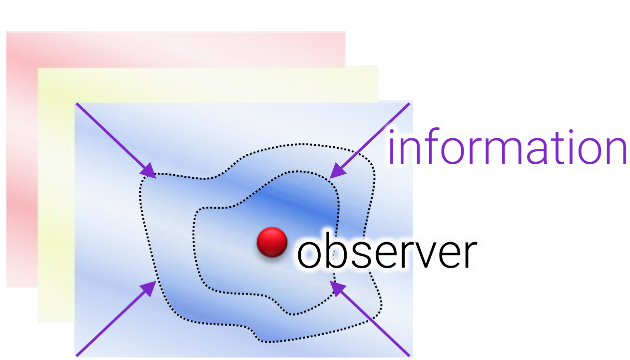


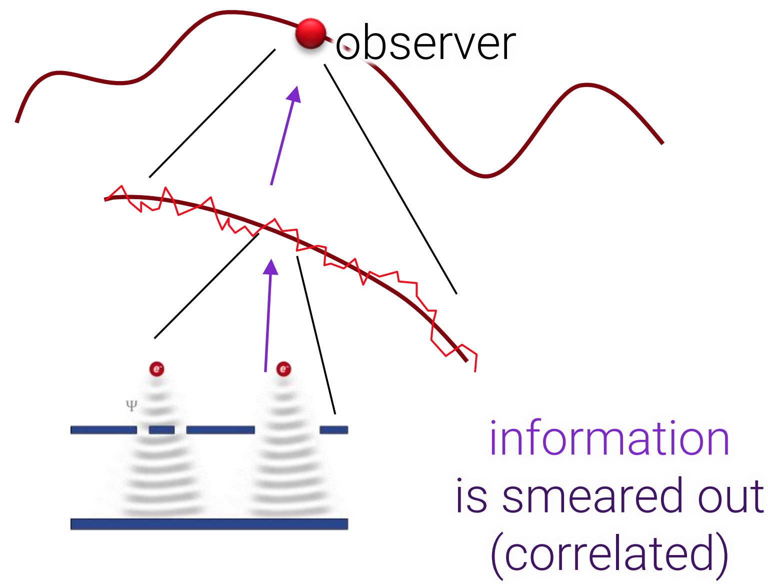


Classical Physics

- Deterministic dynamics, but only partial knowledge
- Far-away structures invisible
  - Wave equation transports information too us (e.g. light)
- Small scales invisible
  - Information transport across scales
  - Chaotic dynamics / “butterfly effect”

Probabilistic Models of Physics



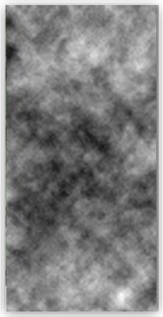


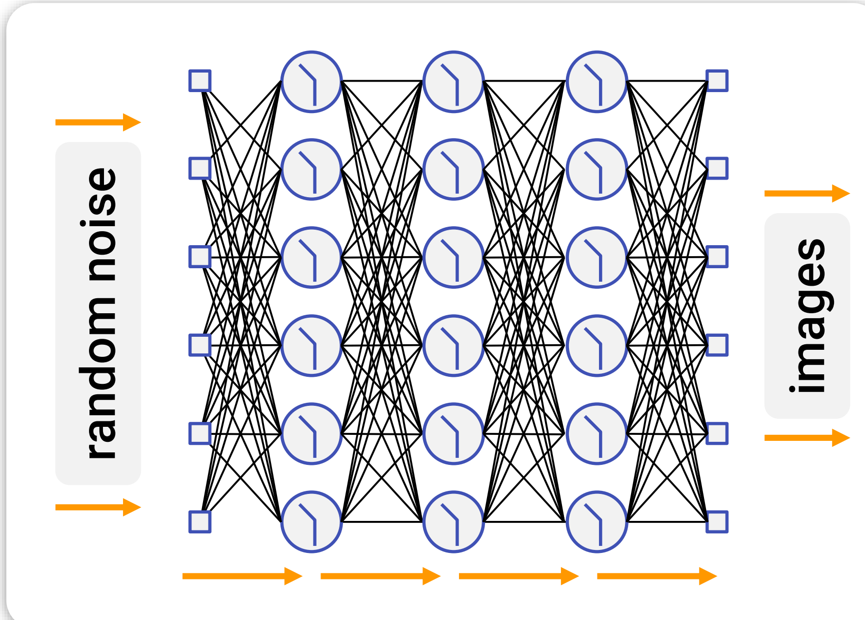
Quantum Physics

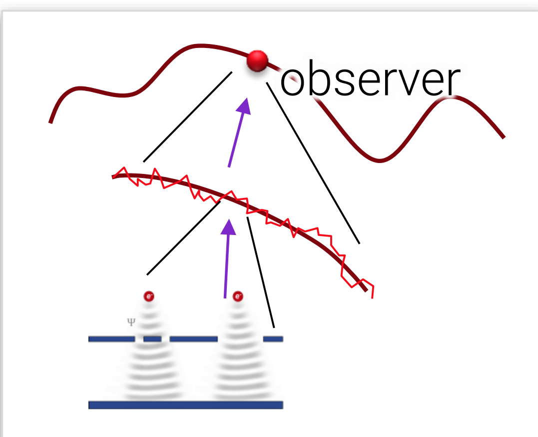
- Distribution p derived from wave Ψ
- We compute the statistical dependency structure
  - Using a non-linear wave equation
  - This is deterministic – similar to classical physics
- We sample from it: this is random

The Universe as a Generative Network

Information
(noise)
$\mathbf{z}$







There is only one wave function Ψ

- One big distribution p
- Life is correlations (stat. dependencies)

Stochastic Machines

What are we up to?

Physics as information processing

- View the dynamics of reality as computation

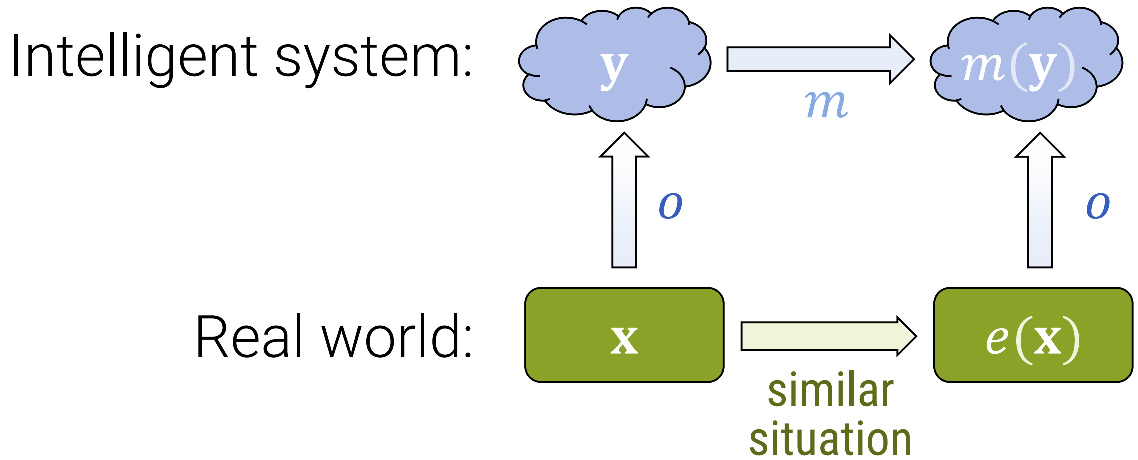
Physics as a universal machine

- We can build computers in the real world
- Rules of physics are Turing complete

Simulation

- We can simulate the rules of physics in a computer
  - Perfectly on a quantum computer
  - Approximate arbitrarily on a classical machine
  - Costs are prohibitive, of course

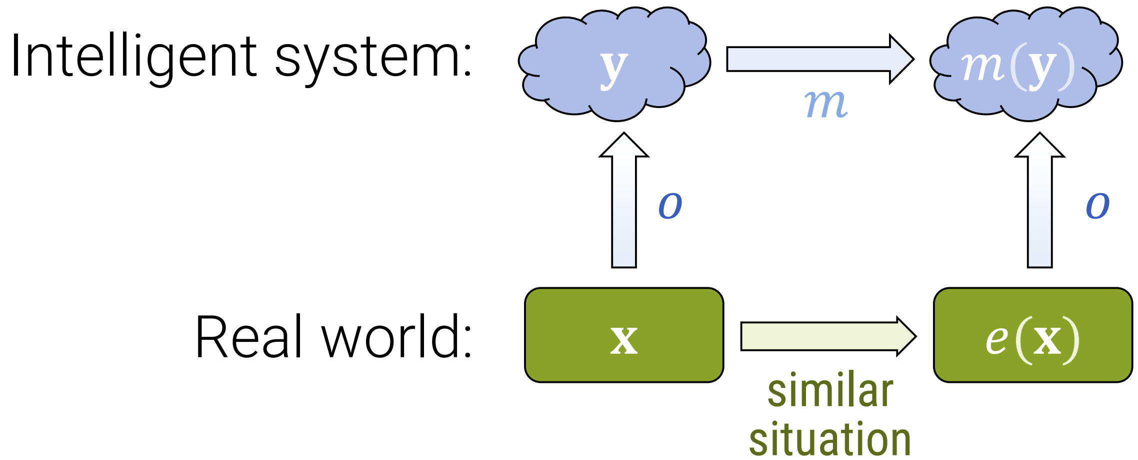
Prediction



Predictive Reasoning

- Brain tries to predict what is going to happen
  - Or other intelligent systems
- For things “we care” about
  - Macroscopic events
  - Only certain events (some “details” omitted)

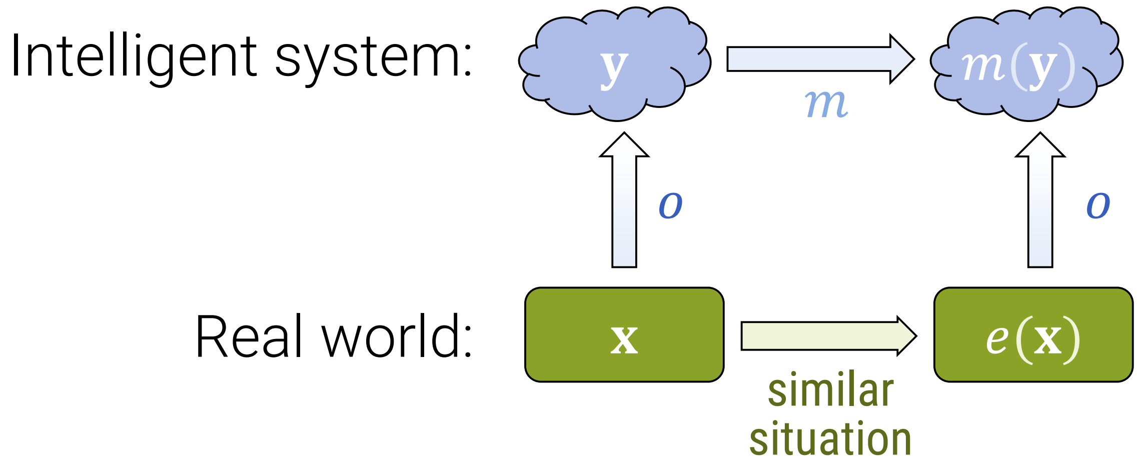
Prediction



Predictive Reasoning

- Brain / system has limits
  - Imperfect knowledge "$o(\mathbf{x})$"
  - Limited capacity of model m
  - Limited experience ("training data")
- Build "best you can"

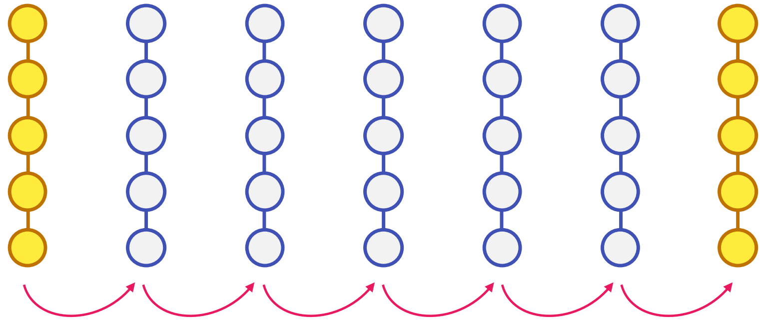
Prediction

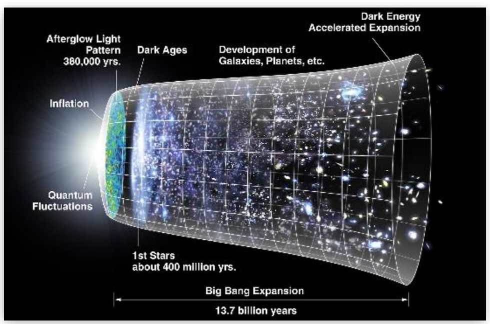


Predictive Reasoning

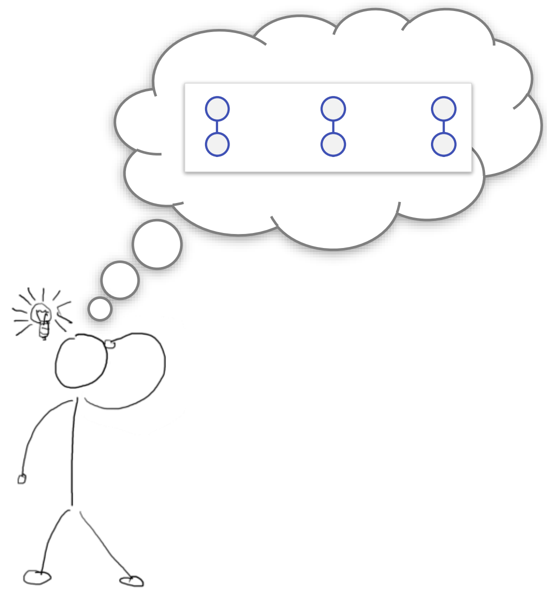
- Short-cut
  - Predict physical dynamics
  - At coarse level
- Physical computer
  - Use less time + space than original event
  - Compression (maybe evolutionary-discriminative)

Short-Cut / Coarse-Grained Reasoning



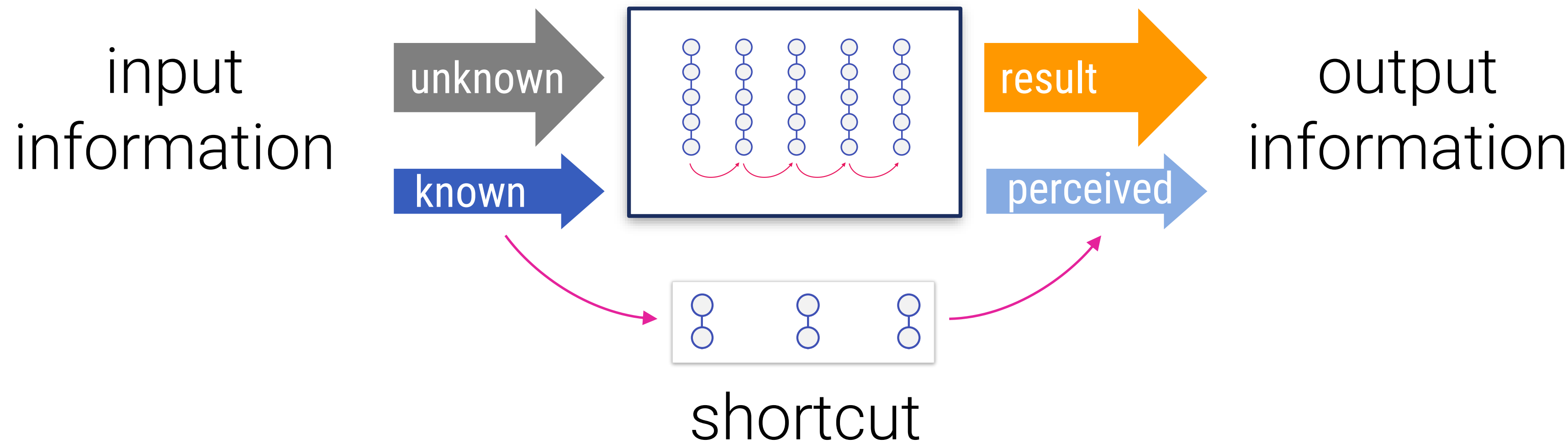


[NASA]



[https://en.wikipedia.org/wiki/Big_Bang#/media/File:CMB_Timeline300_no_WMAP.jpg]

Computability Theory



Computability (of shortcuts)

- **Infinite processes:** Undecidable
- **Limited space:** Busy-beaver (grows too fast)
- **Limited time & space:** Still a lot?

Average behavior: seems more “restrained”

Summary

Physics

A big parallel machine

- Simple rules
- Parallel computing
- Turing-capable

Properties

- Symmetric
  - Poincaré group: Relativistic models
- Strict locality (causality chains)
  - All wave-models (classic & QFT)
- Non-deterministic
  - Classical models, too, if not all state is known

Breaking the NFL-Theorem

Our learning setting

- Learning algorithm
- Pattern creation algorithm

} both implemented in physical hardware

No Free Lunch / Bias-Variance Trade-Off

- For n input bits
  - 2^n different pieces of data can be encoded
  - 2^{2^n} different binary classifications are possible
- "Storage" requirements
  - $\log_2 2^{2^n} = 2^n$ bits to encode arbitrary pattern
  - Now the generator and decoder "play the same game"

Universal Priors

Better than that?

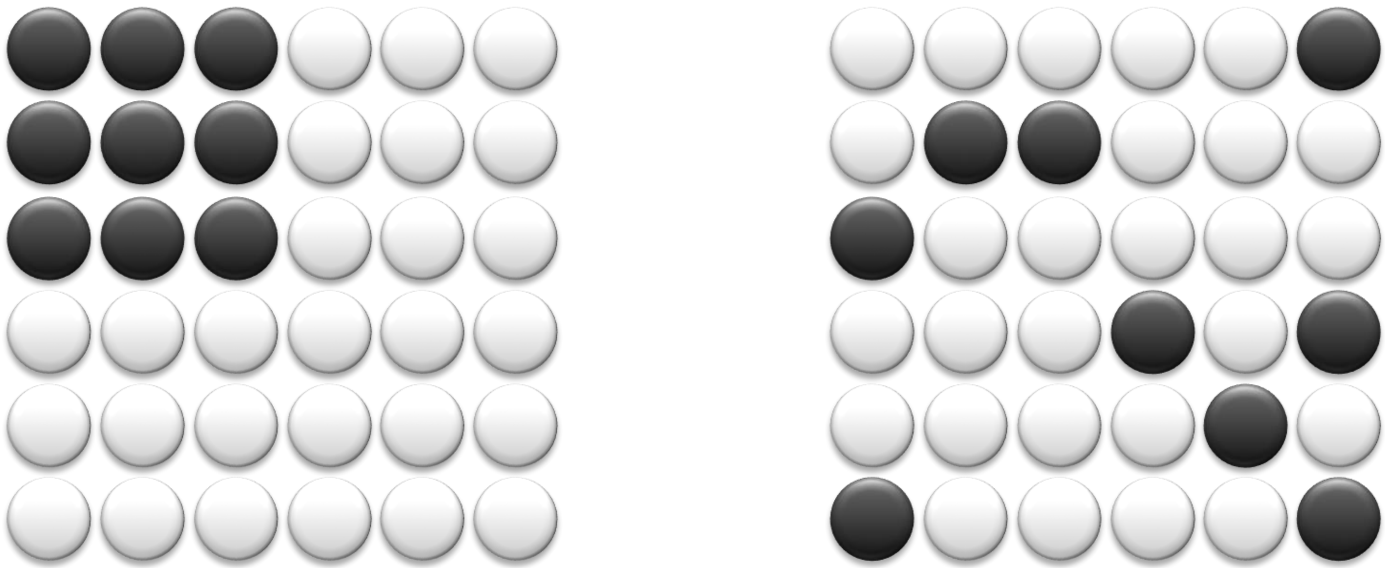
- What kind of patterns emerge naturally?
- What kind of classes / structures might we be interested in naturally?

Self-organization in physical systems

- Vast area
- “Unsolved” as far as I know
- We will take a brief glimpse in what is known

Modelling 2

STATISTICAL DATA MODELLING



Chapter 12

Physics and Self-Organization

Michael Wand · Institut für Informatik, JGU Mainz · michael.wand@uni-mainz.de

Video #12

Physics & Self-Organization

- Physics
- **Self-Organization**

Overview

Two Topics

Two self-organization scenarios

- Thermodynamic equilibrium in a gas
  - Maximum entropy for prescriped mean energy
  - Connection to log-likelihoods and “energy functions”
- Coarse-graining of processes through renormalization groups
  - Microscopical dynamical system
  - Scale symmetry
  - Understanding of macroscopical properties

Disclaimer again

- All of the above “rough sketch” from a CS person

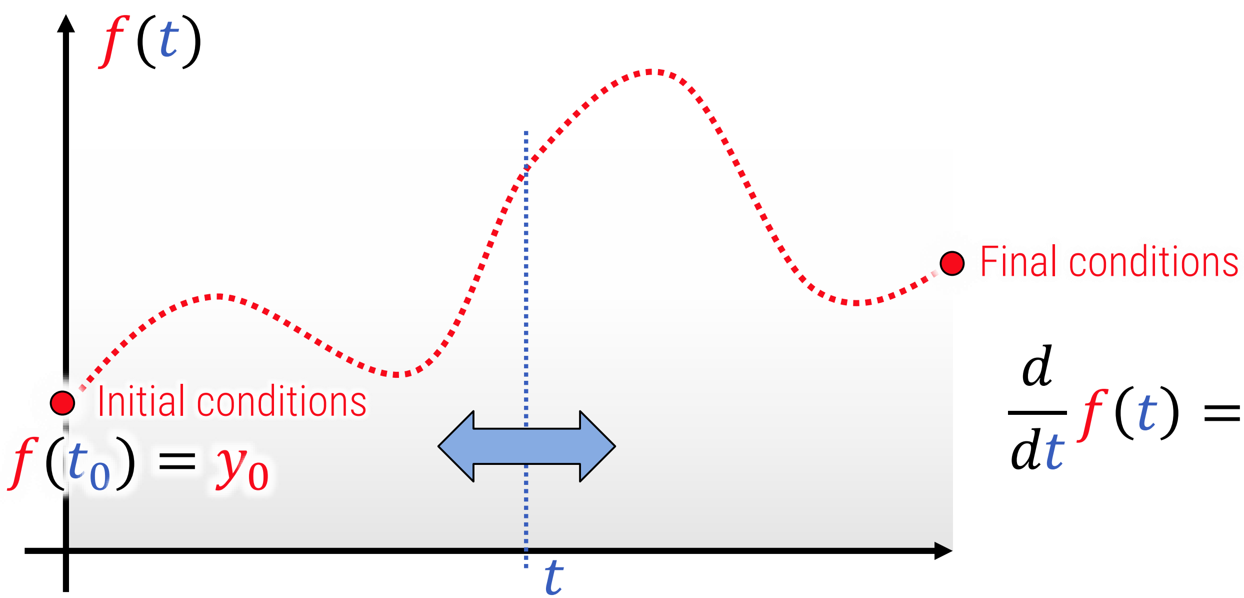
Entropy in Physical Systems

Reversibility & the Second Law

Reversible

Newtonian physics is reversible

- No information is ever lost
  - Assuming no singularities,
    such as point particles colliding
- We can play the ODE forward and backward



$$\frac{d}{dt} f(t) = F(f(t), t)$$

Reversible Dynamics

Discrete case

- F is a bijection

Continuous case

- F is a bijection (+ some regularity conditions)

Time symmetry, discrete case

- F is independent of t:
  - $f(1) = f_0, \;\; f(1) = F(f_0), \;\; f(2) = F^2(f_0), \;\; f(3) = F^3(f_0), \ldots$
  - Permutation group orbit $F^t(f_0)$ (finite cyclic group)

Where is reversibility lost?

Classical “coarse-grained” models

- Friction → thermal molecular movements
  - Btw: Variational “Hamiltonian” approach only works in the reversible case

Macroscopic view

- Abstract from small details
- Information flows into the small scales!
  - …and from the small scales – butterfly effect in fluid dyn.

Quantum physics

- Evolution of Ψ is deterministic
  - But reconstruction from observations is imperfect

Classical, Reversible Physics:

Variational Description

Quantities

Variational System Modeling

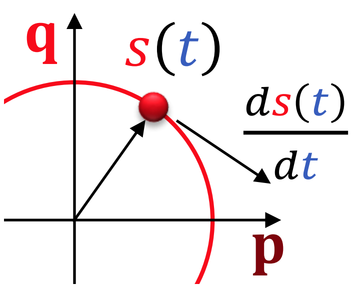
- State $s(t) \in \Omega$ (continuous)
- Energy $\mathbf{H}(s(t))$ (here $\mathbf{H}$ = "Hamiltonian", not entropy!)
- Dynamics is known when *energy function* $\mathbf{H}$ is given

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial \mathbf{H}}{\partial \mathbf{q}}, \qquad \frac{d\mathbf{q}}{dt} = +\frac{\partial \mathbf{H}}{\partial \mathbf{p}}$$

($\mathbf{q}$ = position, $\mathbf{p}$ = impulse $\mathbf{v} \cdot m$)

Reversible Dynamics

- The formulation assumes reversibility
- "No energy lost"



Example

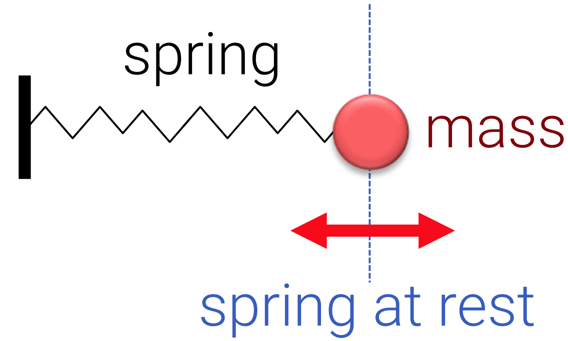
1D Mass-Spring System

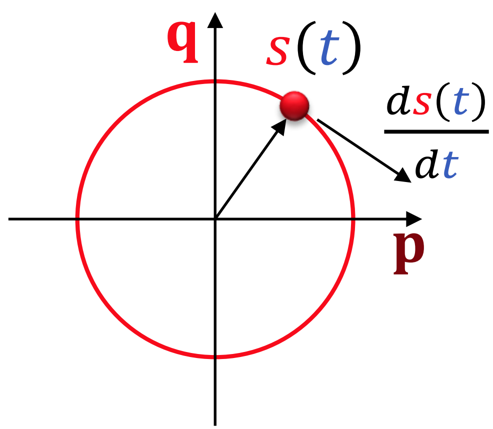
- Kinetic energy $E_{kin}(q) = \frac{1}{2} m \dot{q}^2$
  - $v = \dot{q},\ p = m\dot{q}$
- Potential energy $E_{pot}(q) = \frac{1}{2} D q^2$
  - In Hookean spring with spring constant D
- Hamiltonian

$$\mathbf{H} = \frac{1}{2} D q^2 + \frac{1}{2} m \dot{q}^2 = \frac{1}{2}\left(D q^2 + \frac{p^2}{m}\right)$$

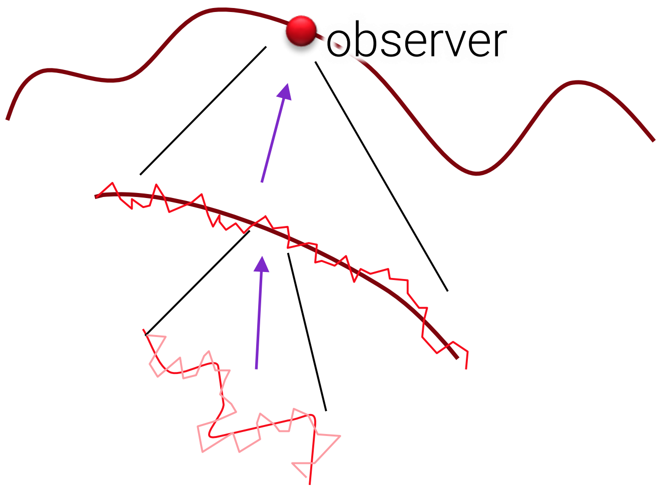
- ODE

$$\frac{dp}{dt} = -\frac{\partial \mathbf{H}}{\partial q} = -\frac{Dq}{m}, \quad \frac{dq}{dt} = \frac{\partial \mathbf{H}}{\partial p} = \frac{p}{m} = \dot{q}$$





Macro States



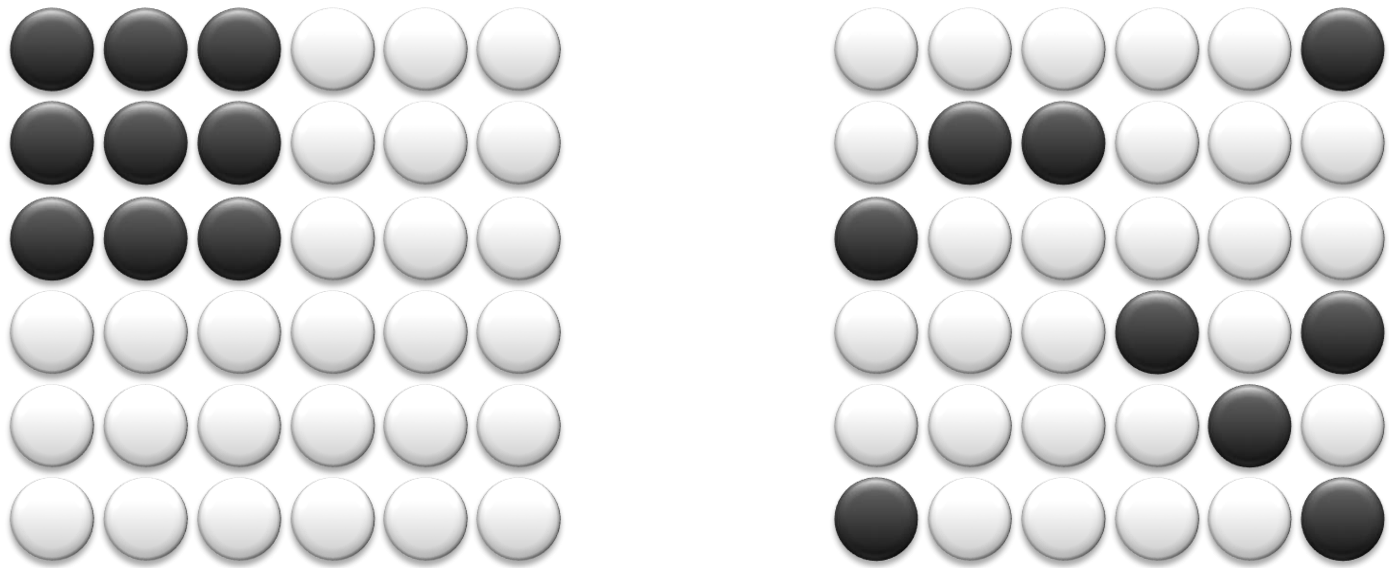
Statistical Physics

- We do not see atoms
  - "Micro states"
- Because *too small*
- We only see macroscopical phenomena
  - "Macro states"

Macro state: Descriptors for Conditions

- "Glass is half-empty" (all particles on the bottom)
- "Pressure" (force per area due to collisions)
- "Temperature" ≈ average energy per particle

Macro States



Example: “discrete” gas, 9 particles, 4×9 spots

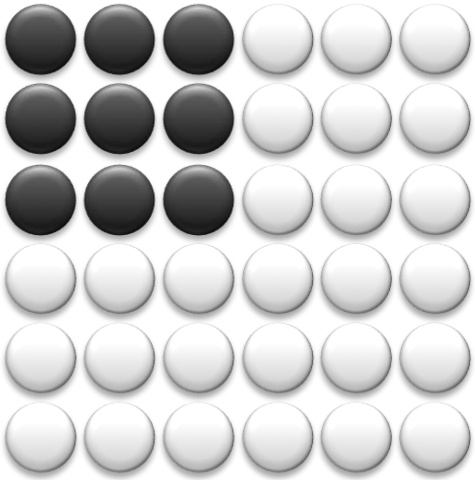
- “All particles in the upper left corner: One microstate
- “Particles can be anywhere”: $\binom{4 \times 9}{9}$ microstates (many!)

Equilibrium

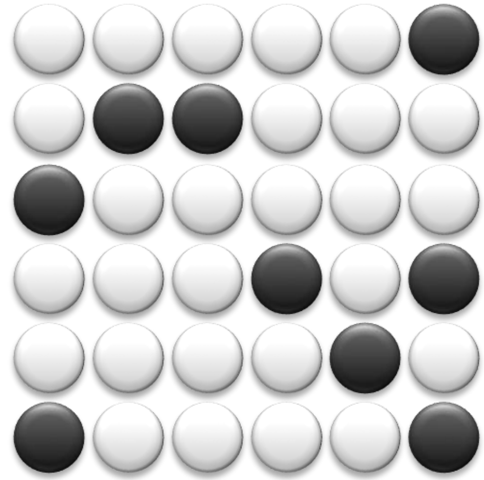
Equilibrium

- We let the system run “for ever”
  - Discrete system, for now
- Stop at a random time
  - After ages
- Sample a random configuration
  - Out of all possible states

Example



P("only upper left corner filled") = $\frac{1}{17550}$



P("any patter") = 1

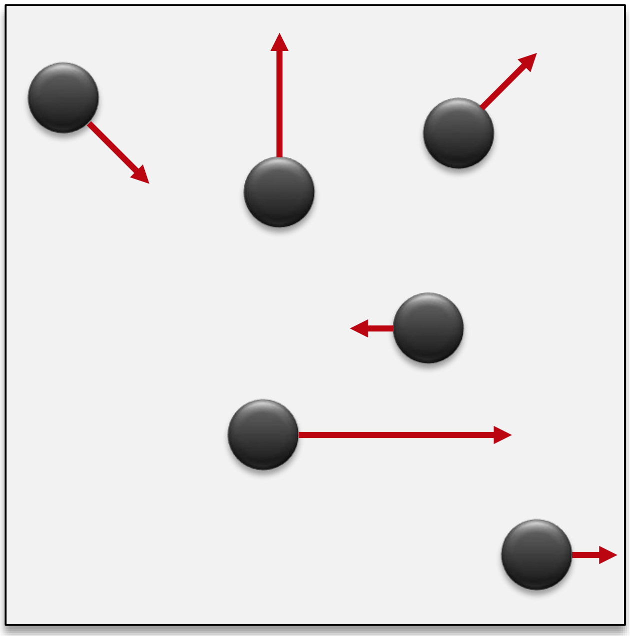
Example:

- All permutations possible
  - Implies: each visited once during each cycle
- $P(\text{Macrostate}) = \frac{\#\text{states that fit macro state}}{\#\text{all microstates}}$

“Ideal Gas”

Consider “Gas in a Box”

- Continuous
- Particles move independently
- No interaction / collisions



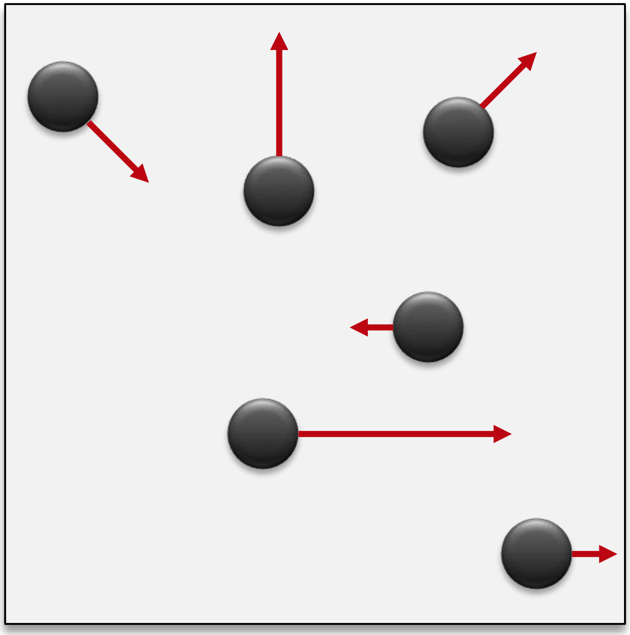
Maximum Entropy

- Each particle independent
  - Because: no interactions
- State of the particle “typical”
  - Random one out of all possible.
  - All states visited = all states similarly likely

“Ideal Gas”

Typical particle state

- Maximum Entropy principle
- Distribution has maximum uncertainty: Maximum Entropy



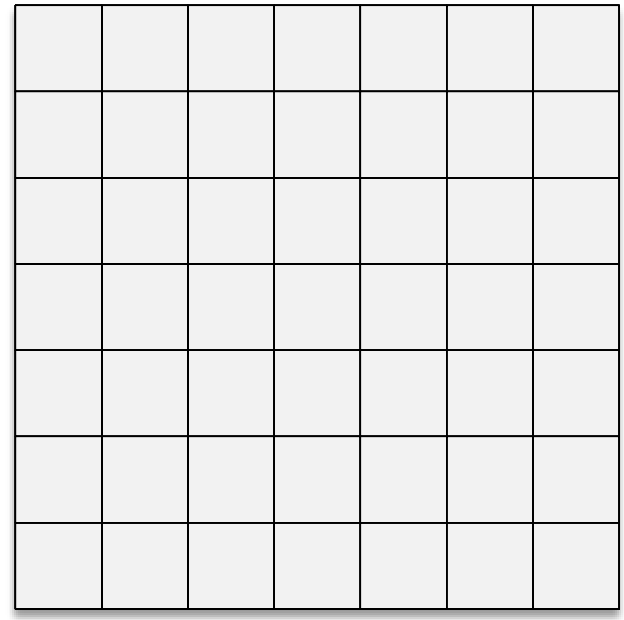
State space

- $\Omega = (\underbrace{\mathbb{R}^3}_{\text{position}} \times \underbrace{\mathbb{R}^3}_{\text{velocity}})^N$
- $P(s, v)$ chosen s.t. entropy $\mathrm{H}(P(s, v)) \to \max$
  - MaxEnt for s: Uniform distribution over box
  - MaxEnt for v: Does not make sense
    - – We need constraints!

Discretization

State Space of One Particle

- $\Omega = \left(\underbrace{[0,1]^3}_{\text{position: just the box!}} \times \underbrace{\mathbb{R}^3}_{\text{velocity: anything goes}}\right)^N$
- Discretize
  - Box is a (fine) discrete grid
  - Velocities on a (fine) discrete grid



Discrete $\mathbf{s}$: box grid
Discrete $\mathbf{v}$: infinite grid

Model assumption

- Fixed temperature
  - Fixed average energy per particle

Discretization

Model assumption

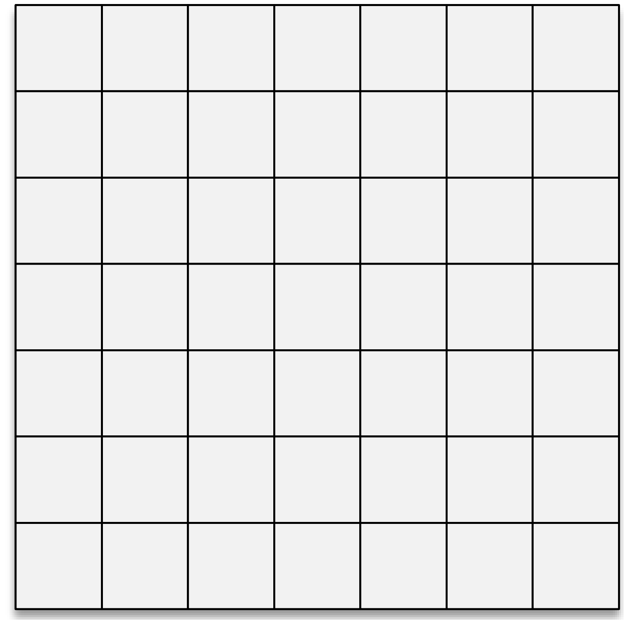
- Fixed temperature
  - Average energy per particle

Constraint

- Position must be in box
- Velocities

$$E = \frac{1}{2} m ||\mathbf{v}||^2$$

$$\frac{1}{N}\sum_{i=1}^{N} E_i \approx \mu(E) = \frac{1}{2} m \cdot \mu\left(\mathbf{v}_i^2\right)$$



Discrete $\mathbf{s}$: box grid
Discrete $\mathbf{v}$: infinite grid

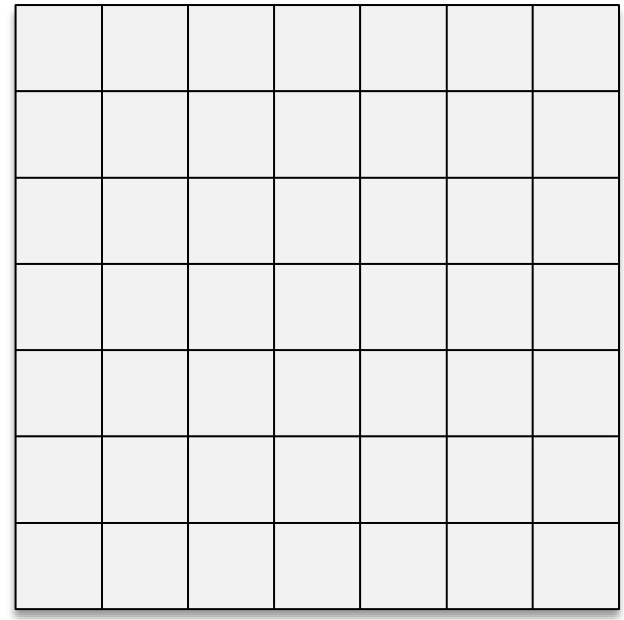
Discretization

Deriviation (Sketch)

- Mean of velocities is zero
- Normal distribution

$$\mathcal{N}_{0,\sigma^2}(\mathbf{v}) = e^{-\frac{\mathbf{v}^2}{2\sigma^2}}$$

maximizes entropy!



Discrete $\mathbf{s}$: box grid
Discrete $\mathbf{v}$: infinite grid

In general

- Gibs / Boltzmann Distribution

$$P(\text{state}) = \exp\left(-\frac{\text{energy}(\text{state})}{kT}\right)$$

maximizes entropy at temperature T (k is a constant)

Variational Model

Variational System Modeling

- State $s(t) \in \Omega$ (continuous)
- Energy $\mathbf{H}(s(t))$ (here $\mathbf{H}$ = "Hamiltonian", not entropy!)
- Dynamics is known when *energy function* $\mathbf{H}$ is given

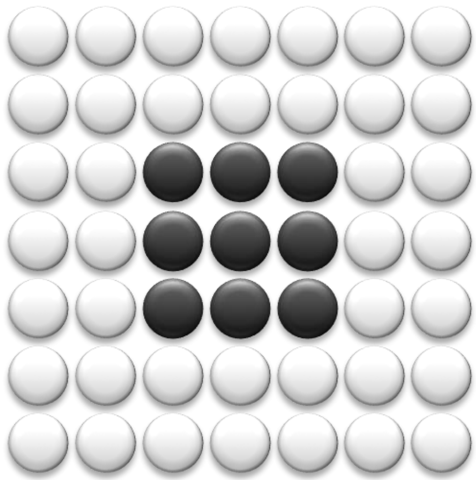
$$\frac{d\mathbf{p}}{dt} = -\frac{\partial \mathbf{H}}{\partial \mathbf{q}}, \qquad \frac{d\mathbf{q}}{dt} = +\frac{\partial \mathbf{H}}{\partial \mathbf{p}}$$

($\mathbf{q}$ = position, $\mathbf{p}$ = impulse $\mathbf{v} \cdot m$)

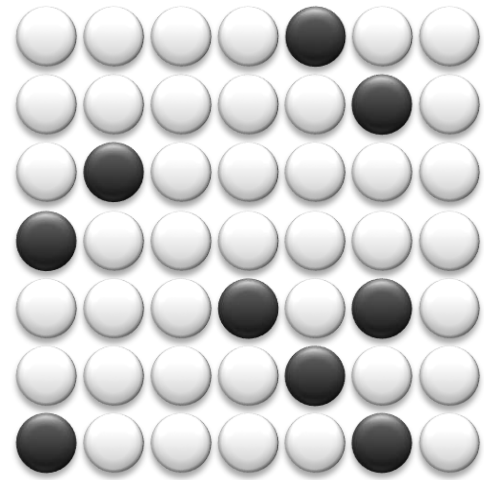
Equilibrium

- Constant temperature T: $P(s(t)) = \exp\left(-\frac{\mathbf{H}(s(t))}{kT}\right)$

The Second Law



„Big Bang“

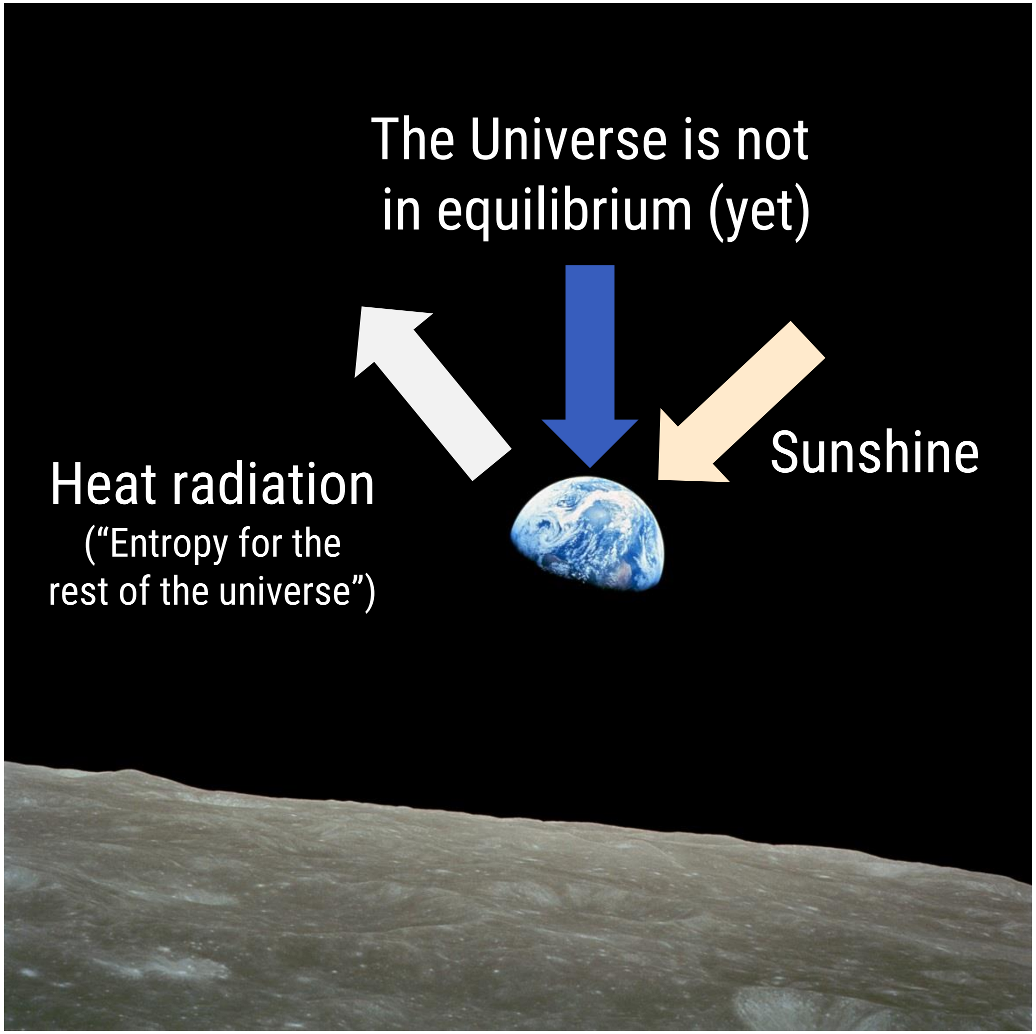


„later“

Entropy increases

- Universe starts in low-entropy state (boom)
- Reversible dynamics since then
- Entropy (of its macroscopic, observable state) increases

Statistical Doom?



"Earthrise"
Nasa/Apollo 8, Bill Anders

Renormalization Groups

Scale-Symmetry

How do systems coarse-grain?

- In general: unknown (Turing complete!)
- Special case:
  - Model family, with only changing parameters
    - For example, Hamiltonian model
  - A scale-symmetry can be established
- In this case
  - We can understand the macroscopic behavior from the microscopic

Formal tool

- “Renormalization group”

Scale-Symmetry

Consider system

- Function $s(\mathbf{x}): \mathbb{R}^d \to \mathbb{R}^k$
- Governed by (physical) law with parameters $c_1, \dots, c_n$
  - For example, a Hamiltonian

$$\mathbf{H}(s) = f_{c_1,\dots,c_n}(s)$$

  - Where f is a function with parameters $c_1, \dots, c_n$

Scale changes

- Consider system "at a different scale σ"
- New Parameters $c_1, \dots, c_n \mapsto F_\sigma(c_1, \dots, c_n)$
  - Hamiltonian: $\mathbf{H}^{(\sigma)}(s) = f_{F_\sigma(c_1,\dots,c_n)}(s)$

Examples for scale changes

Rescaling

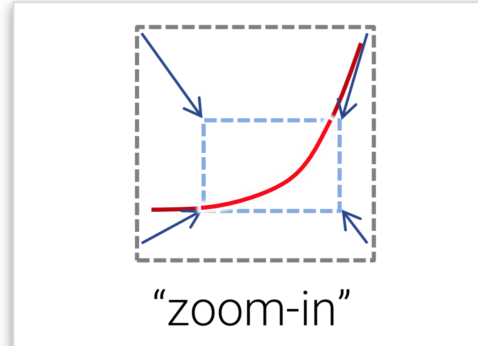
- Just replace $\mathbf{x} \mapsto \frac{1}{\sigma}\mathbf{x}$

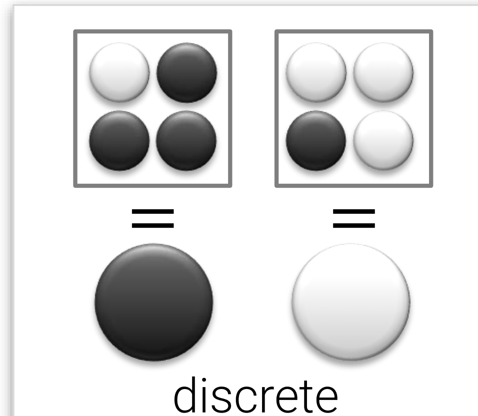
Discrete coarse-graining

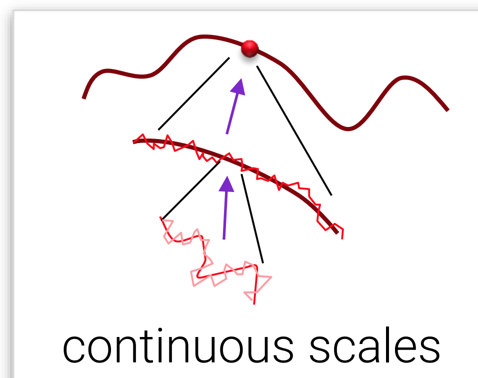
- Replace $s_1, s_2, \ldots, s_n$ by
  $m(s_1, s_2), m(s_3, s_4), \ldots, m(s_{n-1}, s_n)$
  for some averaging function m

Continuous Coarse-Graining

- Low-pass filter $\omega_\sigma(\mathbf{x})$ at scale σ
  - For example $\omega_\sigma = \mathcal{N}_{0,\sigma}$
- Coarser observables: $s^{(\sigma)} := s \otimes \omega_\sigma$







Critical Points

Examine Mapping

$$c_1, \ldots, c_n \mapsto F_\sigma(c_1, \ldots, c_n)$$

- System parameters remapped under coarse-graining
- Symmetry
  - Change scale, change parameters, then same behavior
  - Scale transforms form a "renormalization group"

Critical Points

$$c_1, \ldots, c_n = F_\sigma(c_1, \ldots, c_n)$$

- System behavior becomes scale-invariant
- This indicates fundamental changes at this point

Statistical Systems

Correlation function

- Measure correlation at distance r
  - In this example: Distance in cells
- $corr(r) = \mathbb{E}_{T=\text{const.}}(\langle s(x), s(x+r)\rangle)$

Degree of order

- Quickly dropping: unordered / random
- Slowly dropping: ordered / large-scale structure

Critical points

- Scale symmetry
- Special correlation function, such as power law r^{-2h}

Example 1: Exponential ODE

Toy Example

"Physical Law"

$$f : \mathbb{R} \to \mathbb{R}, \qquad \frac{df(t)}{dt} = cf(t), \qquad c \in \mathbb{R}$$

Solution

$$f(t) = \exp(ct)$$

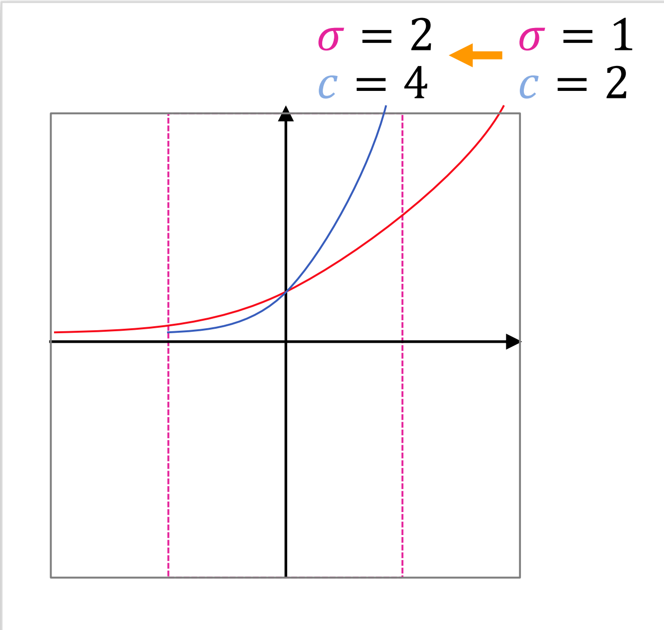
Scale Transformation

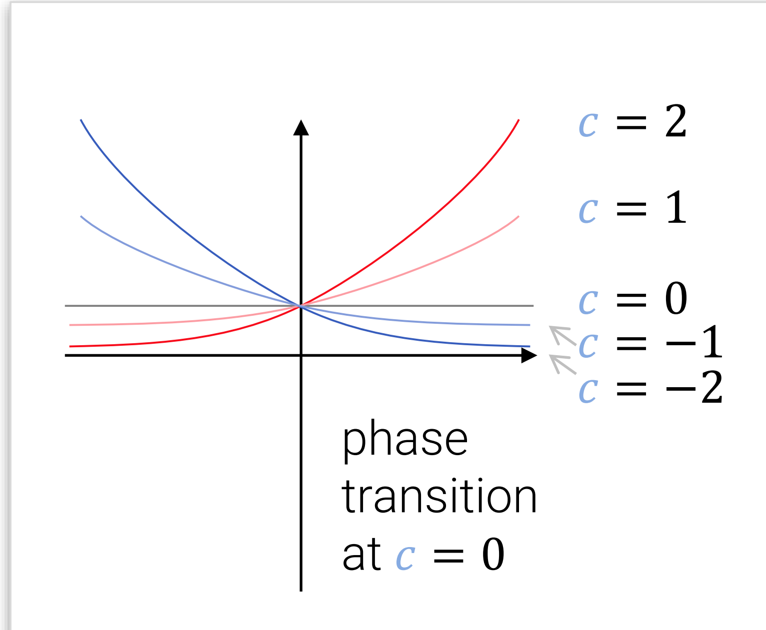
$$t = \frac{1}{\sigma} t \quad \to \quad f(t) = \exp\left(\frac{c}{\sigma} t\right) \quad \to \quad F_\sigma(c) = \sigma c$$

Fixed point

- $\forall \sigma \in \mathbb{R}: F_\sigma(c) = c$ for $c = 0$

Visualization





Scale Symmetry

- Scale symmetry at $c = 0$
  - Boring, but symmetric
- Behavior changes qualitatively at this point
  - Raising instead of shrinking

Example 2:
Fractal Brownian Motion

Toy Example

"Physical Law": Random Walk

$$f: \{1, \dots, n\} \to \mathbb{R}, \qquad f(t+1) = f(t) + \nu, \qquad \nu \sim \mathcal{N}_{0,1},$$

Solution in Fourier Space

$$f(t) = \sum_{\omega=-n}^{n} z_\omega e^{-i\omega t} \quad \text{with} \quad z_\omega \in \mathbb{C}, |z_\omega| \sim \mathcal{N}_{0,\omega^{-1}}$$

General FBM-Noise: Continuous spectrum

$$z_\omega \in \mathbb{C}, |z_\omega| \sim \mathcal{N}_{0,\omega^{-2h}}, \omega \in \mathbb{R}$$

„Fraktal exponent" h

Toy Example

Functions

$$f(t) = \int_{\mathbb{R}} \mathrm{z}_{\omega} e^{-i\omega t}\, d\omega \quad \text{with} \quad \mathrm{z}_{\omega} \in \mathbb{C}, |\mathrm{z}_{\omega}| \sim \mathcal{N}_{0,\omega^{-2h}}$$

Scale Invariance ("stochastic fractal")

$$t = \sigma t \quad \rightarrow \quad f(\sigma t) = \sigma^{h} f(t)$$

Perfect symmetry

- For $h = 1$: $f(\sigma t) = \sigma f(t)$

Example 3: Ising Model

Example System: Ising Model

State space

- Integer grid $x_{\mathbf{k}}$, $\mathbf{k} \in \Omega \subset \mathbb{Z}^d$
- Binary "spins" $s(x_{\mathbf{k}}) \in \{-1,1\}$ (→ magnetism)
  - For simplicity: enumerate as $s_1, \dots, s_n$ with $s_i = s(x_{\mathbf{k}_i})$

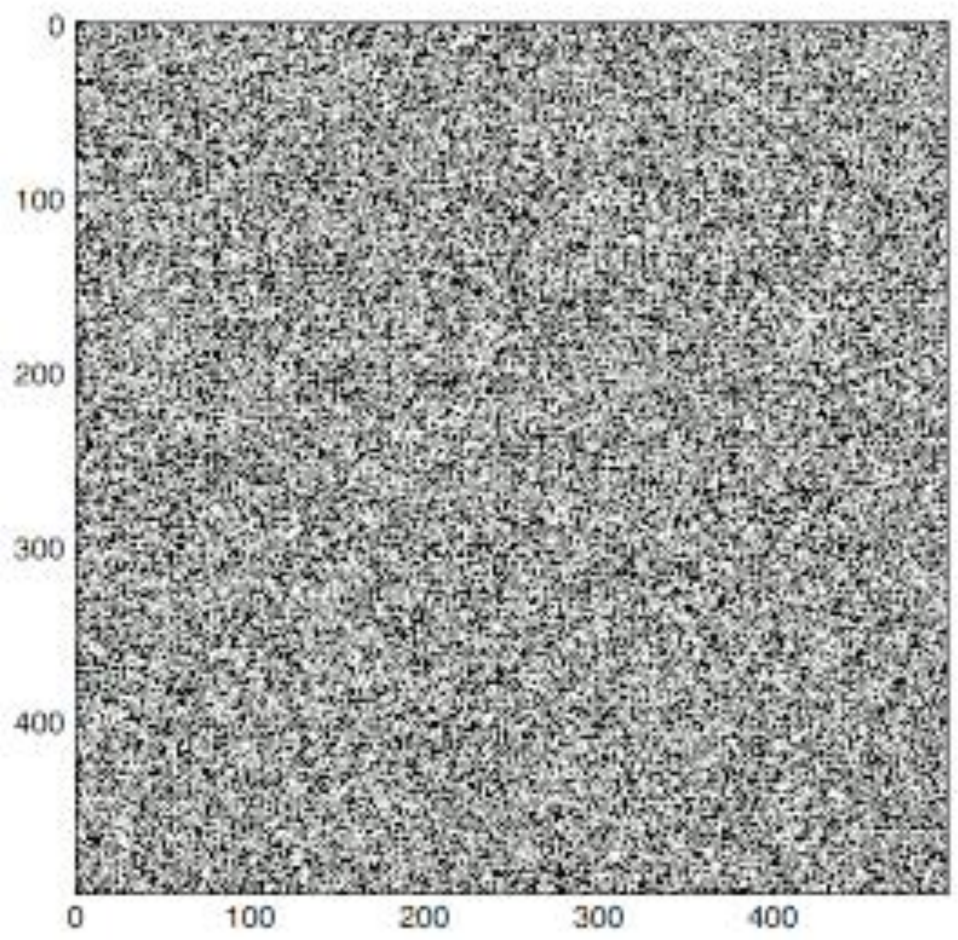
Neg-Log-Likelihood / Hamiltonian

$$\mathbf{H}(s_1, \dots, s_n) = \sum_{i=1}^{n} c_i s_i + \sum_{i=1}^{n} \sum_{j \in N(i)} J_{ij} s_i s_j$$

Symmetric, no external field

$$\mathbf{H}(s_1, \dots, s_n) = J \sum_{i=1}^{n} \sum_{j \in N(i)} s_i s_j$$

Ising Model



Equilibrium at Fixed Temperature

Probability

$$p(s_1, \ldots, s_n) = \frac{1}{Z} \exp\left(\frac{-\left(\sum_{i=1}^{n} c_i s_i + \sum_{i=1}^{n} \sum_{j \in N(i)} J_{ij} s_i s_j\right)}{kT}\right)$$

Sampling

- MCMC sampler
- Metropolis-Hastings
  - Detailed balance in equilibrium
  - $p(\mathbf{x}) \cdot p_{trans}(\mathbf{x} \to \mathbf{y}) = p(\mathbf{y}) \cdot p_{trans}(\mathbf{y} \to \mathbf{x})$
  - Random moves, accept outcome with likelihood ratio $\frac{p(new)}{p(old)}$

Scale Symmetry

Coarse-graining

- Block renormalization: 2x2 blocks with one new state
- Scale space symmetry for this system
- Hamiltonian has the same form
- Only J changes
  - Group of transformations that changes parameters with scale

Renormalized

$$\mathbf{H}(s_1, \dots, s_n) = \sum_{i=1}^{n} \sum_{j \in N(i)} J_{ij}^{(\alpha)} s_i s_j$$

Scale Symmetry

Renormalization analysis

- High temperature
  - Correlation function drops exponentially
- Low temperature
  - Correlation function drops very slowly
- Critical point: perfect scale symmetry
  - Correlation function forms a power law
  - Transition from unordered to ordered phase
  - Model for magnetism (Curie-temperature)

Deep Networks

Phase transitions

Initialization of networks

- Variance of weights
  - linear weights, bias values
- Standard initialization
  - Keeps signal variance constant
  - “critical” initialization
- Mean-field analysis [Schoenholz et al. 2017]
  - Varying weight / bias variance
  - Networks learn best close to phase transition
  - Similar observations in neuroscience (neural activity)

Ben Poole, Subhaneil Lahiri, Maithra Raghu, Jascha Sohl-Dickstein, Surya Ganguli
Exponential expressivity in deep neural networks through transient chaos. NeurIPS 2016.

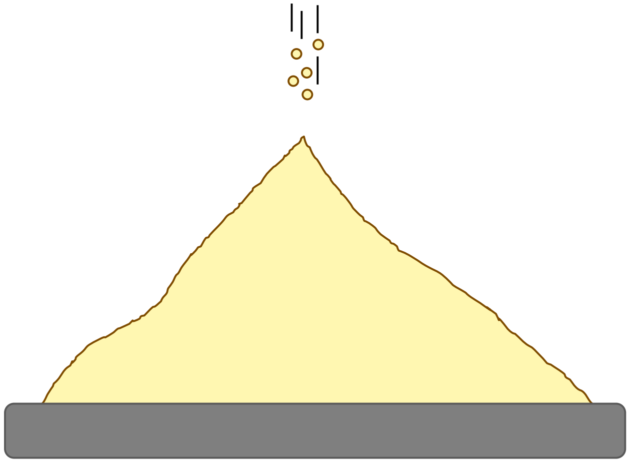
Samuel S. Schoenholz, Justin Gilmer, Surya Ganguli, Jascha Sohl-Dickstein
Deep Information Propagation. ICLR 2017.

Non-Equilibrium Self-Organization

Dynamical System View

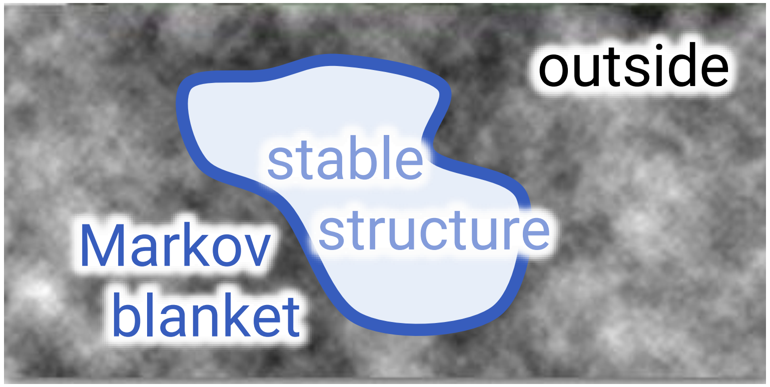
Self-organized criticality

- Many natural systems
  operate at critical point
- Self-stabilizing dynamics
- Phase transition destroy structure



Bak, Tang, Wiesenfeld: Self-organized criticality: an explanation of 1/f noise.
Physical Review Letters, 1987.

Machine Learning?



“Free Energy Principle”

- [Friston et al. 2006+]
- Hypothesis on emergence of intelligence
- Markovian systems
  - Inner & outer region
  - Interface: Markov blanket
- (Thermo-) dynamics: Outer fluctuations
  - Structure preservation implies Bayesian Inference
  - Similarities between free energy minimization and variational approximations of Bayesian inference

Karl Friston: The Free Energy Principle
https://www.youtube.com/watch?v=NIu_dJGyIQI

“Thermodynamics of Life”

Origin of life

- Why/how do complex, self-replicating structures arise from random fluctuations?
- Driven system
  - The sun shines
  - Space is cold
  - Non-maximum-entropy structure can arise

“Dissipation-driven Adaptation”

- Hypothesis by Jeremy England
- Self-replicating machines create disorder more effectively

https://www.quantamagazine.org/a-new-thermodynamics-theory-of-the-origin-of-life-20140122/

Summary

Self-Organization

Self-organizing principles

- Maximum entropy
  - As random as possible
- General dynamical systems with scale symmetry
  - Find emergent macroscopic structure through RG

Rather basic, but already very useful

More complex structures

- Wide field, beyond our lecture
- Active area of research