Characterization of Learning Cost in Neural Networks through Stochastic Thermodynamics

M. Mahdi Mostafaei

Computational Cosmology Group Of Shahid Beheshti University

COMPLEXLAB SHAHID BEHESHTI UNIVERSITY

Main Questions:

What characteristics of the system is related to the maximum thermodynamical cost of learning?

Is this limit inviolable?

What is the contribution of synaptic learning algorithms to this cost?

Stochastic Thermodynamics

Thermodynamics

\+

Stochastic Processes

»

Stochastic Thermodynamics

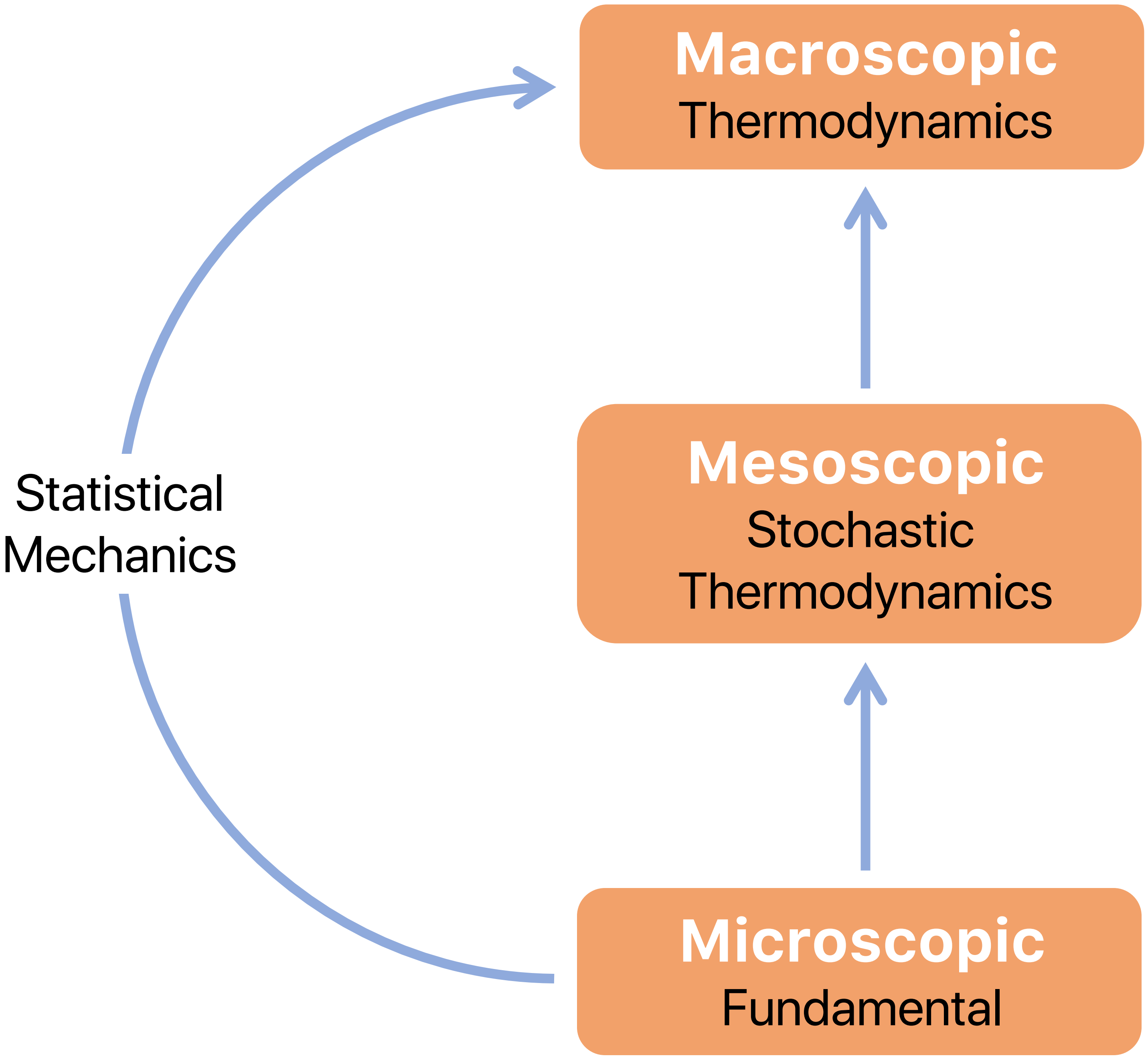
What is stochastic thermodynamics about?

It is a thermodynamic theory for physical systems which are:

mesoscopic $\Delta E \sim K_B T$ like colloidal particles, macromolecules, etc.

non-equilibrium

interacting with equilibrium reservoirs



Macroscopic
Thermodynamics
Mesoscopic
Stochastic
Thermodynamics
Microscopic
Fundamental
Statistical
Mechanics

Stochastic Thermodynamics	Classical Thermodynamics
Mesoscopic	Macroscopic
Large fluctuations	No or negligible fluctuations
Non-equilibrium	Equilibrium or quasi-equilibrium
Work, heat and entropy are defined in individual trajectories	Work, heat and entropy are averaged ensembles of observables
Probabilistic	Deterministic

Why use stochastic thermodynamics?

It's simpler.

In finding the thermodynamics quantities of mesoscopic systems, it is pretty much simpler than the general problem in non-equilibrium statistical mechanics.

Less problems!

several fundamental problems that arise in non-equilibrium statistical physics do not even appear in stochastic thermodynamics.

One example is understanding how the irreversible nature of macroscopic thermodynamics systems emerges from the microscopic dynamics. This conceptual issue is known as **Loschmidt's paradox.**

Ways to perturb a system out of equilibrium:

1

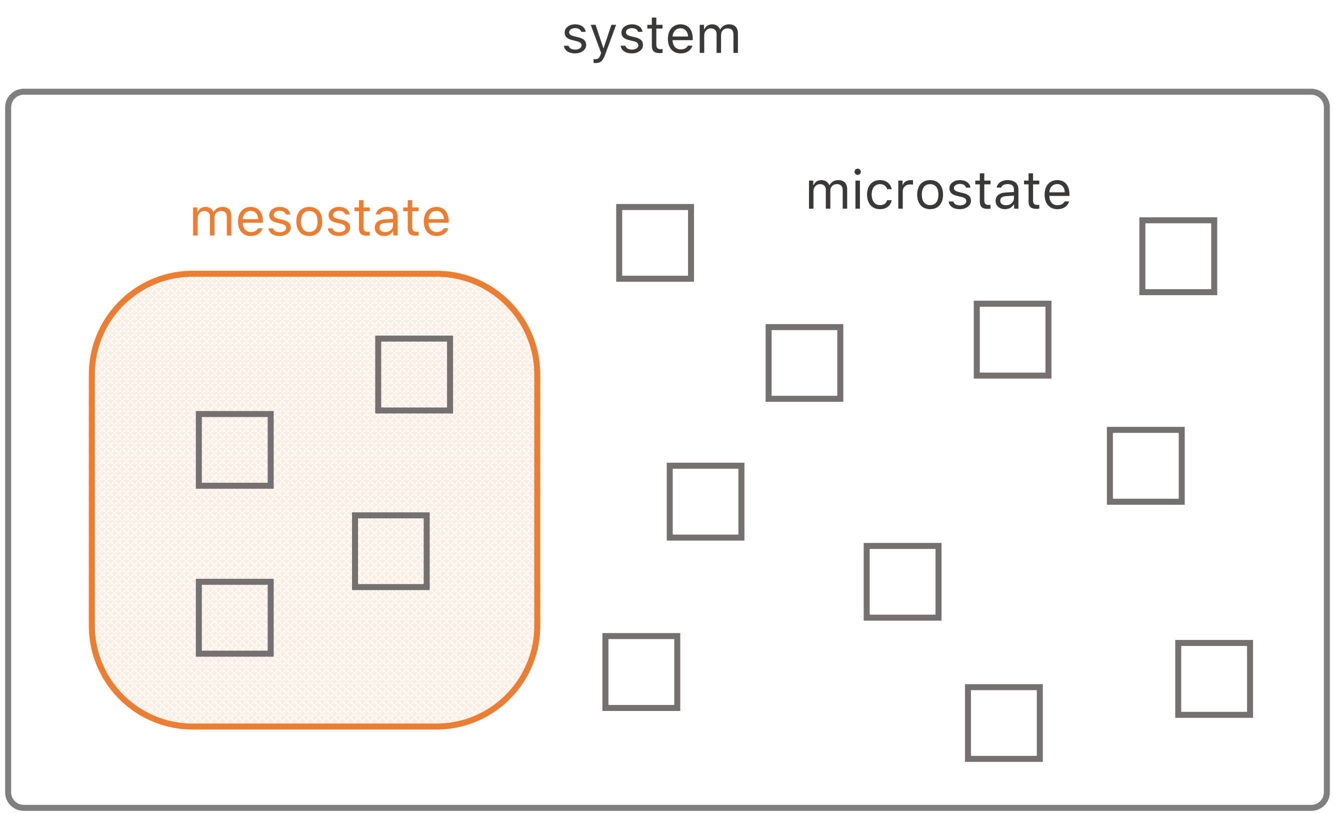
Manipulation

An external, time-dependent control of the system. In a manipulation, the energies of states depend on an external control parameter $\epsilon_x = \epsilon_x(\lambda)$ which is made to change with time $\lambda = \lambda(t)$.

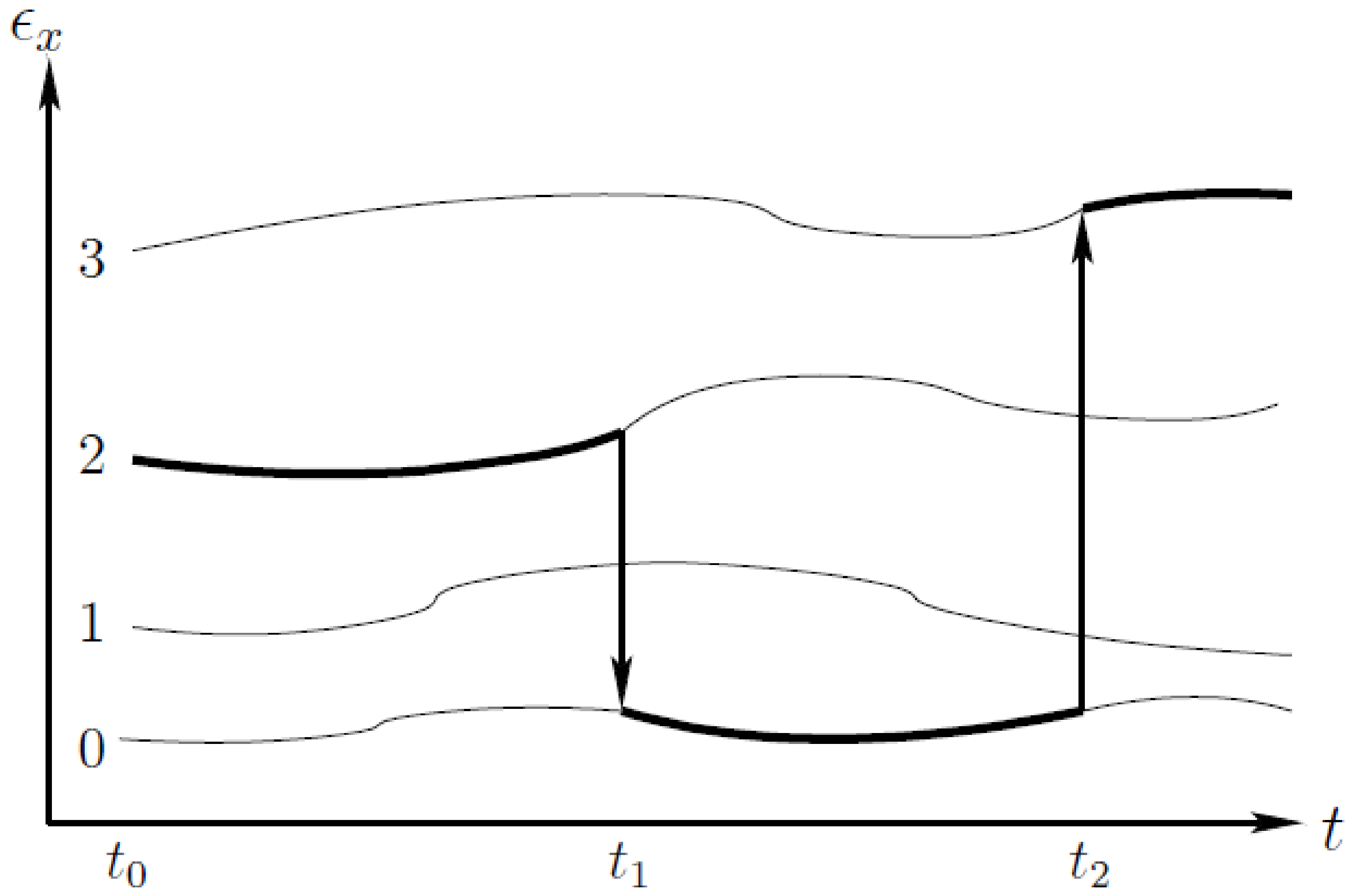
2

Driving

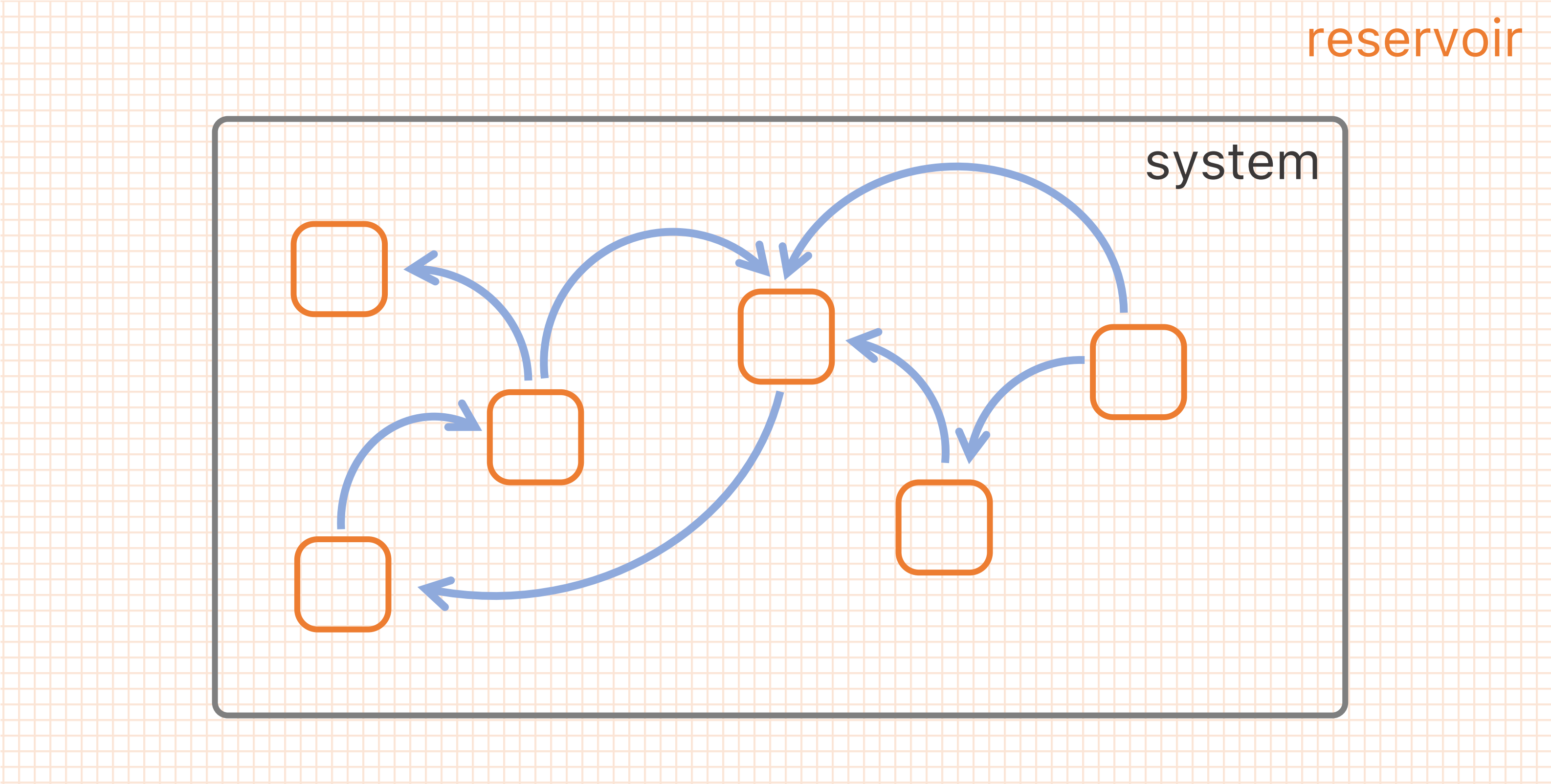
A coupling of the system to an external agent that exchanges with it a certain amount of energy when the system performs specific jumps.

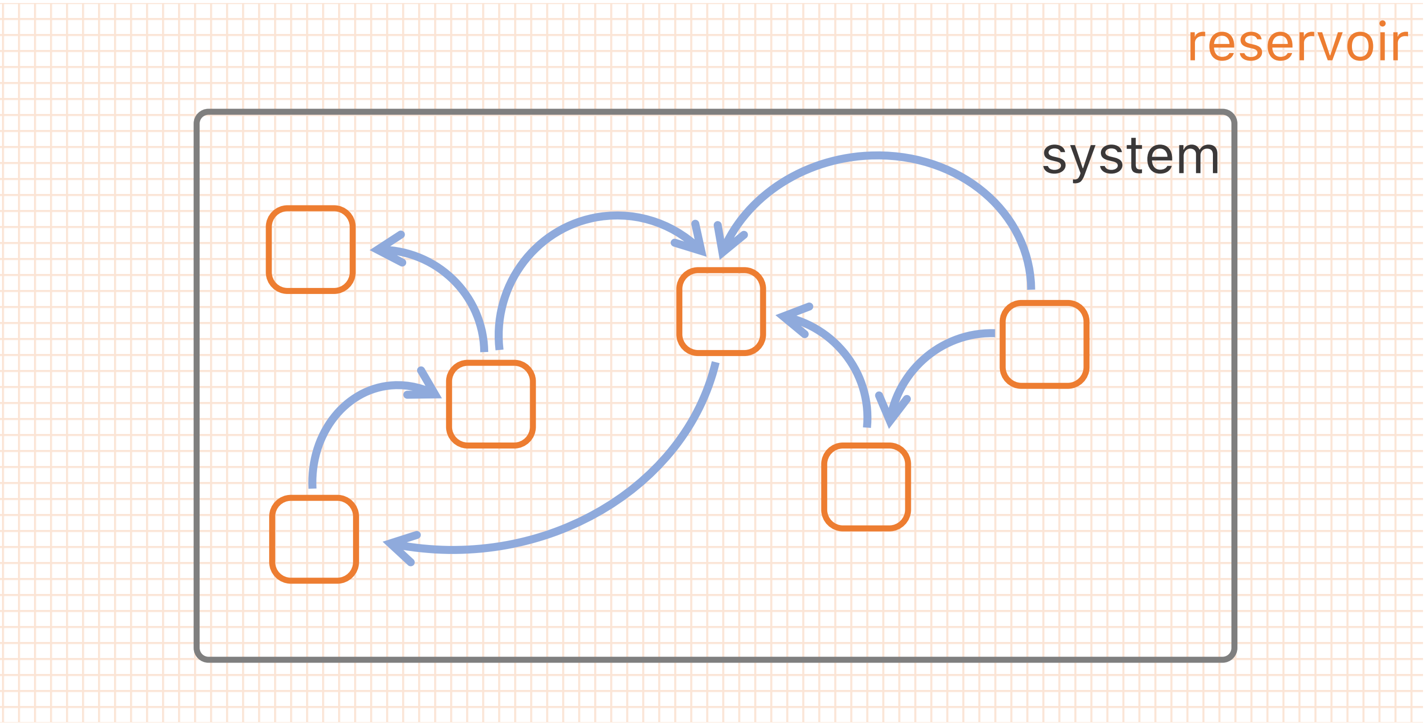

system
mesostate
microstate

Trajectory of a manipulated system.



Due to uncontrolled interactions of mesoscopic system with reservoir,
the dynamic of the system is intrinsically stochastic.





Master equation

$$\frac{d}{dt}p_x(t) = \sum_{x'(\neq x)} \left[k_{xx'}p_{x'}(t) - k_{x'x}p_x(t)\right]$$

Stationary distribution

$$\lim_{t \to \infty} p_x(t) = p_x^{st}, \qquad \forall x$$

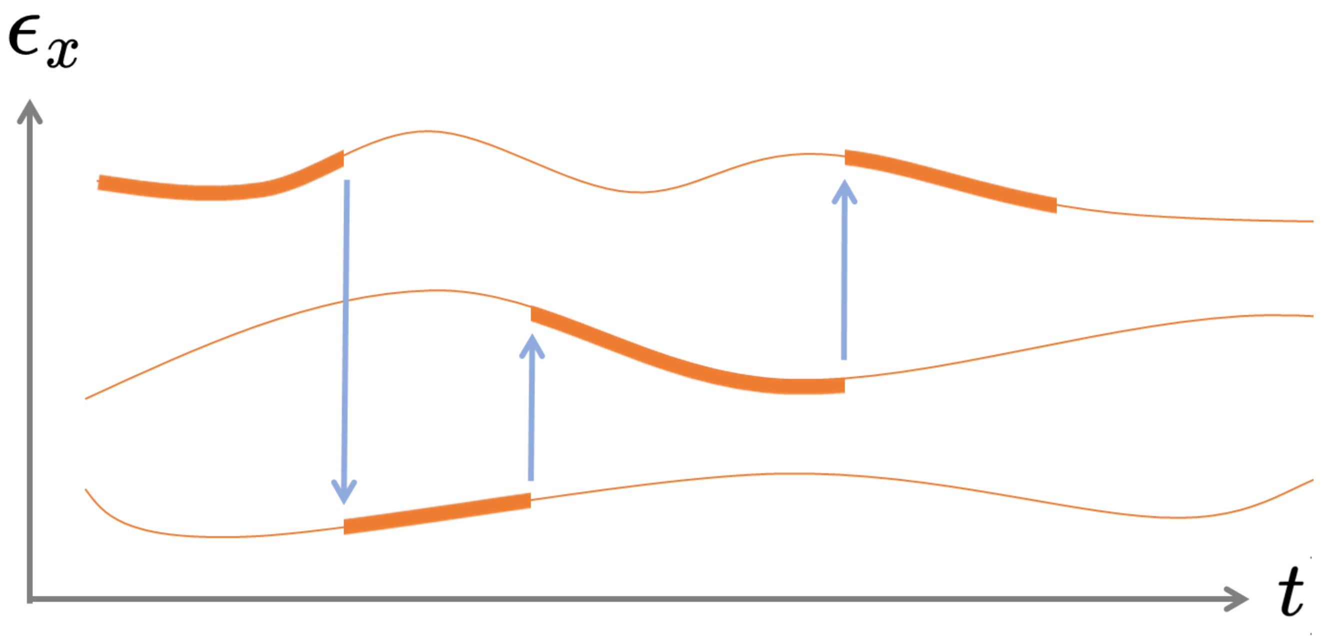
Detailed balance condition

$$k_{xx'} p_{x'}(t) = k_{x'x} p_x(t), \qquad \forall x \neq x'$$

If it is satisfied, the stationary distribution is called the **equilibrium distribution** and we denote it by p^{eq}.

In thermodynamic systems, the condition of detailed balance is often associated with **thermodynamic equilibrium**, and the stationary distribution p^{eq} is the Boltzmann distribution.

$$\frac{k_{xx'}}{k_{x'x}} = \frac{p_x^{eq}}{p_{x'}^{eq}} = e^{-(\epsilon_x - \epsilon_{x'})k_B T}$$



$$q_{xx'} = \epsilon_{x'}(\lambda) - \epsilon_x(\lambda) + \delta_{xx'}$$

$\epsilon_{x'}(\lambda) - \epsilon_x(\lambda)$ Energy lost by the system

$\delta_{xx'}$ the energy provided by the external agent during a jump from x_0 to x.

Generalized detailed balance condition

$$\frac{k_{xx'}}{k_{x'x}} = exp(\frac{q_{xx'}}{k_B T}) = exp(\frac{\epsilon_{x'}(\lambda) - \epsilon_x(\lambda) + \delta_{xx'}}{k_B T})$$

It implies in particular that the jump $x \to x'$ is allowed if and only if the reverse $x' \to x$ is also allowed, i.e., that the system exhibits **microscopic reversibility**.

Next...

- Giving a long talk on ST on Tuesday of next week.
- Finishing main topic Stochastic thermodynamics by the end of **Azar**.
- Then, starting applied ST in neural networks and learning.

سال		سال ۱۴۰۲												سال ۱۴۰۳					
ردیف	عنوان فعالیت	فروردین	اردیبهشت	خرداد	تیر	مرداد	شهریور	مهر	آبان	آذر	دی	بهمن	اسفند	فروردین	اردیبهشت	خرداد	تیر	مرداد	شهریور
۱	مطالعه‌ی نسبتاً جامع و تخصصی از مفاهیم ترمودینامیک تصادفی	■	■	■	■	■	■	■											
۲	شبکه‌های عصبی					■	■	■	■	■									
۳	الگوریتم‌های یادگیری در شبکه‌های عصبی						■	■	■	■	■								
۴	شبیه‌سازی								■	■	■	■	■	■	■	■			
۶	اثر بازخورد (Feedback) توزیع وزن سیناپسی در قالب مدل‌هایی نظیر Hebbian and real								■	■	■	■	■	■	■	■			
۵	تحلیل و بررَسی نتایج به دست آمده و ارزیابی اثرات کمیت‌های فیزیک بر روی فرآیند یادگیری در شبکه‌های عصبی										■	■	■	■	■	■	■		
۷	ارائه گزارش و تهیه مقاله و نگارش پایان‌نامه													■	■	■	■	■	■
۸	دفاع از پایان‌نامه																		■

Thank you.