Thermodynamics of information processing I

Susanne Still
University of Hawaii at Manoa

Motivation

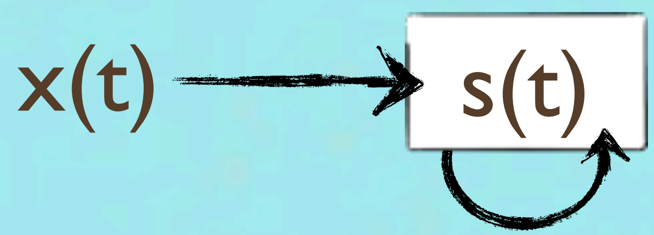
- Living systems: sensing, information processing, decision making under uncertainty, communication, learning, and adaptation are ubiquitous
  - at multiple scales (molecular to organism)
  - many different modalities (electrical, chemical, mechanical)
  - vast diversity of realizations (implementations)
- The vast diversity in biological implementations of learning and inference, just as the diversity of machine learning techniques -
  is it all "just a mess" of details, or are there *building principles?*

Motivation

- *Procedural commonalities:*
  - data is represented
  - some information is retained, some information is discarded
  - there is some utility in retaining information
- Physics of information processing should be important.
- In nature this type of information processing (learning and inference, adaptation) sets living systems apart from nonliving systems.

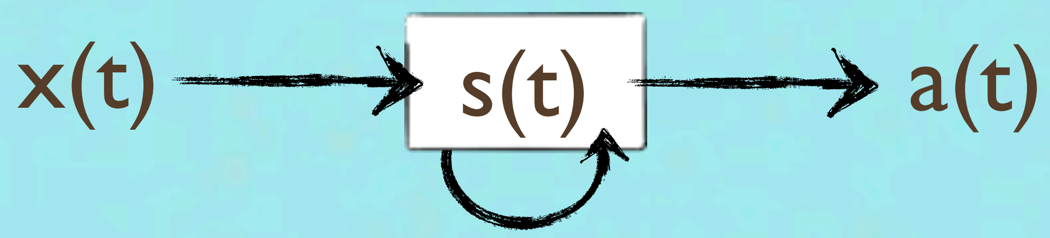
Nature as computation?

- Dynamical systems perform "computations" as part of their interactions with each other. This has long history...
- Take simplified view: distinguish "environment" - "system" "input" - "output"



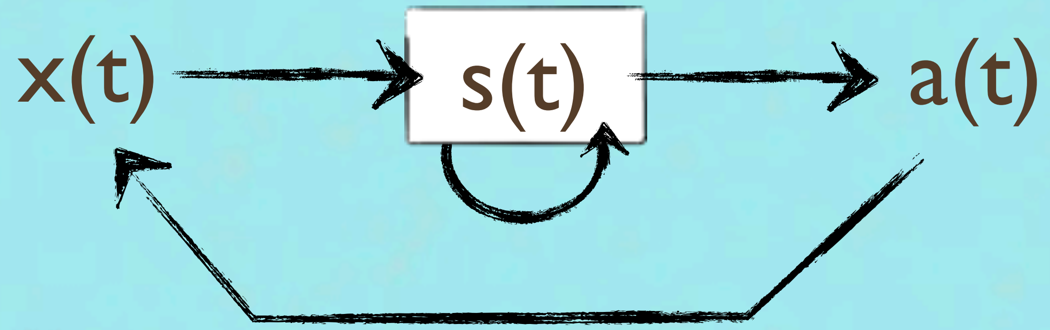
Nature as computation?

- Dynamical systems perform "computations" as part of their interactions with each other. This has long history...
- Take simplified view: distinguish "environment" - "system" "input" - "output"



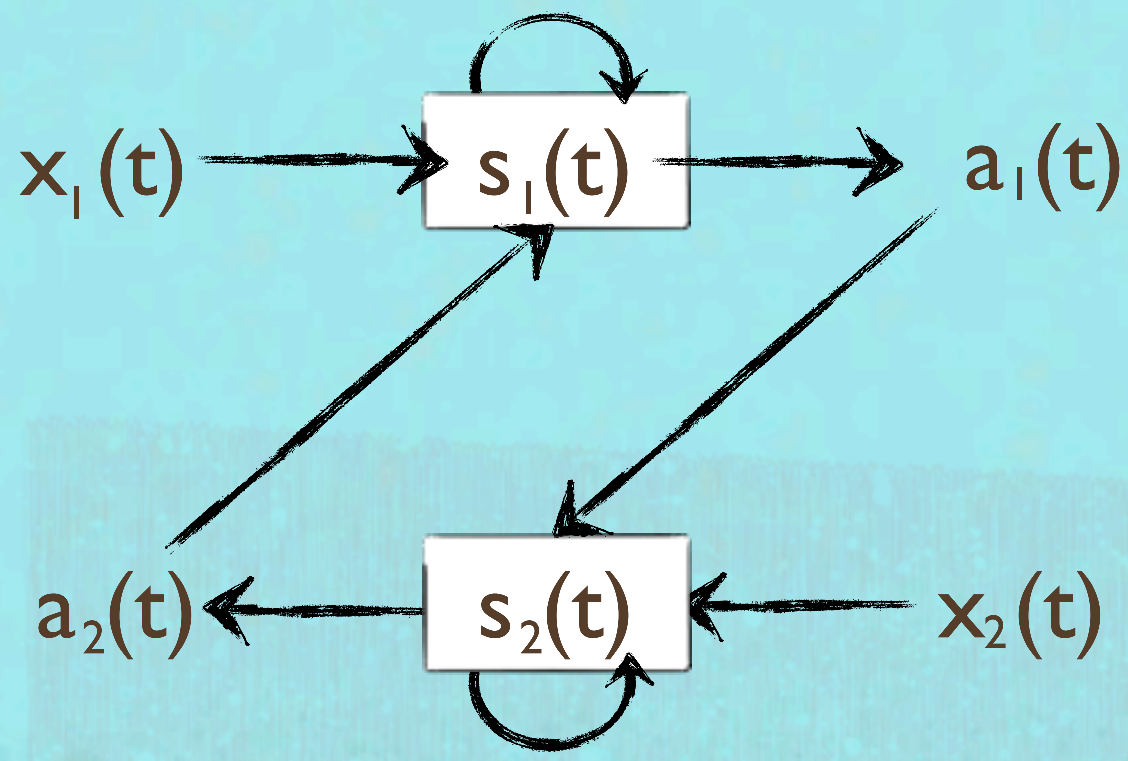
Nature as computation?

- Dynamical systems perform "computations" as part of their interactions with each other. This has long history...
- Take simplified view: distinguish "environment" - "system" "input" - "output"; define a notion of feedback



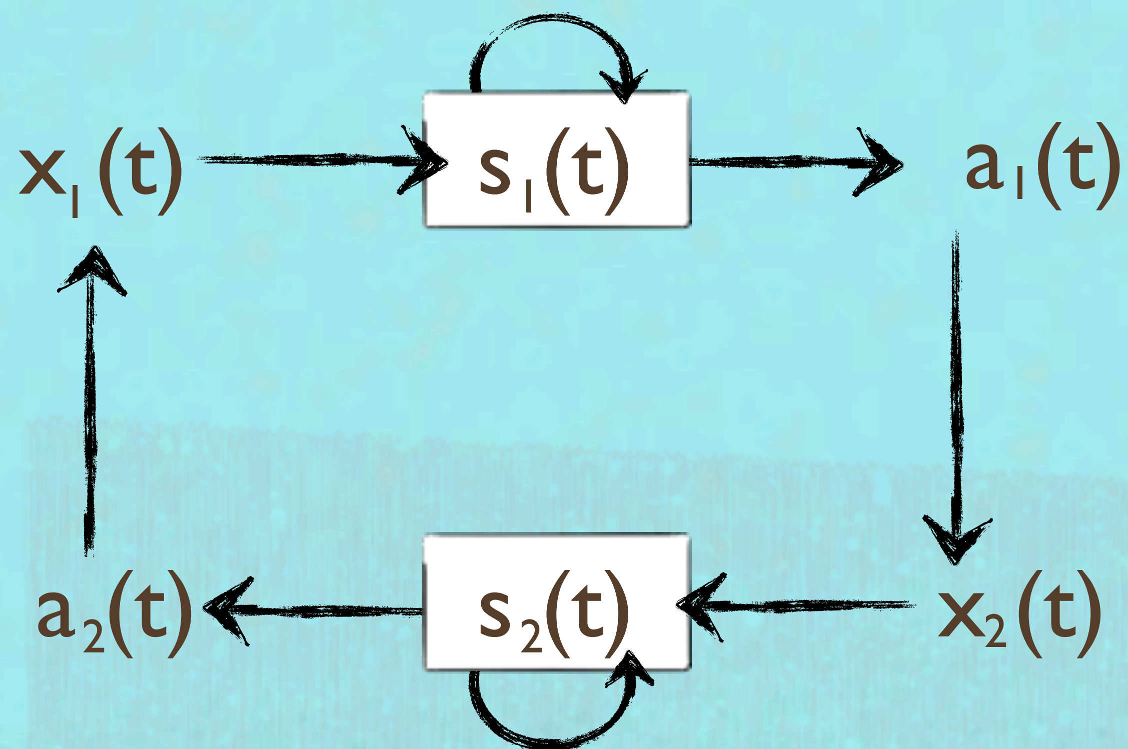
Nature as computation?

- Dynamical systems perform "computations" as part of their interactions with each other. This has long history...
- Take simplified view: distinguish "environment" - "system" "input" - "output"; define a notion of feedback and interaction.



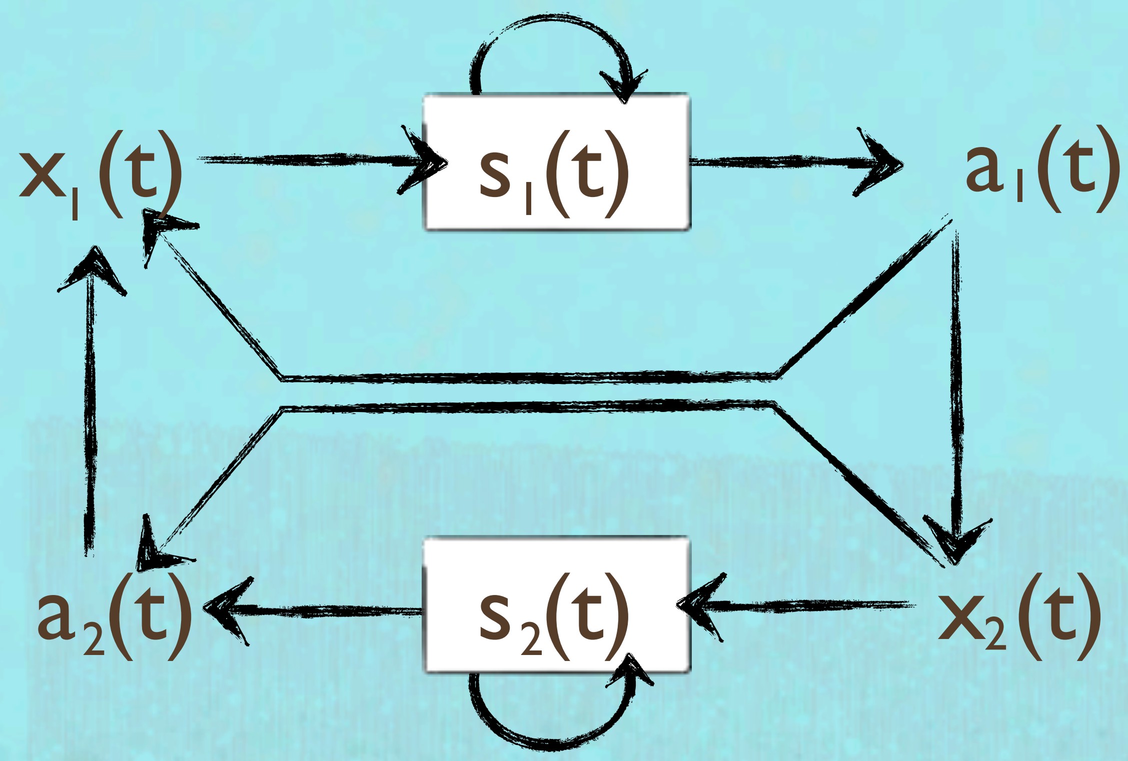
Nature as computation?

- Dynamical systems perform “computations” as part of their interactions with each other. This has long history...
- Take simplified view: distinguish “environment” - “system” “input” - “output”; define a notion of feedback and interaction.



Nature as computation?

- Dynamical systems perform "computations" as part of their interactions with each other. This has long history...
- Take simplified view: distinguish "environment" - "system" "input" - "output"; define a notion of feedback and interaction.



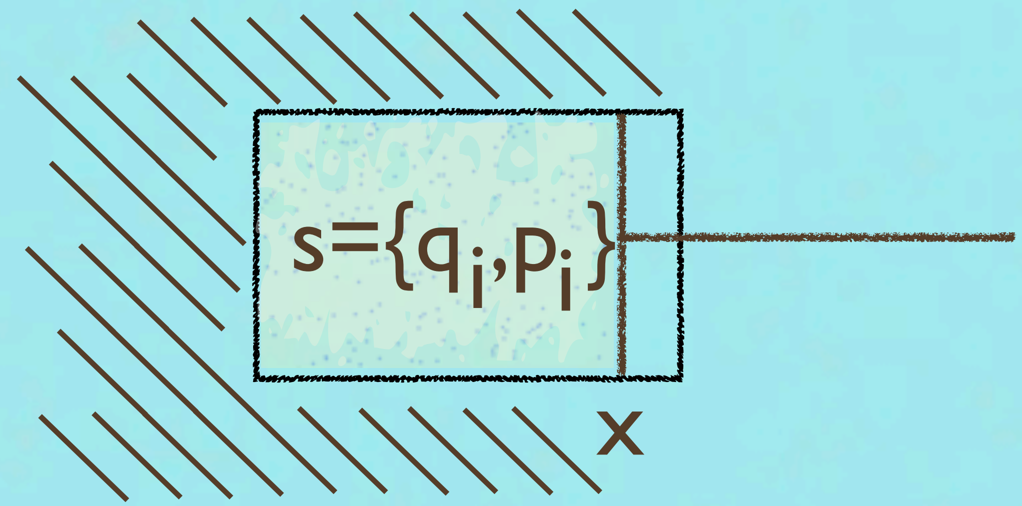
Building principles?

- **Same old?** (cybernetics and "second order" cybernetics) - requires some "goal", a notion of what ought to be controlled
- Take a more radical approach: general utility function is not given, can we get building principles from physical limitations?
- Note that biology has evolved *highly efficient* information processing hardware.
- *Perhaps we can get to some "building principles" from studying the physical limits imposed on information processing...*

Equilibrium Thermodynamics (quick review)

- The state (variable s) of a system in thermodynamic equilibrium, given an environmental parameter, x.
  Best described by the Boltzmann distribution:

$$p_{\rm eq}(s|x) = e^{-\beta[E(s,x)-F(x)]} \qquad \beta = \frac{1}{k_B T}$$



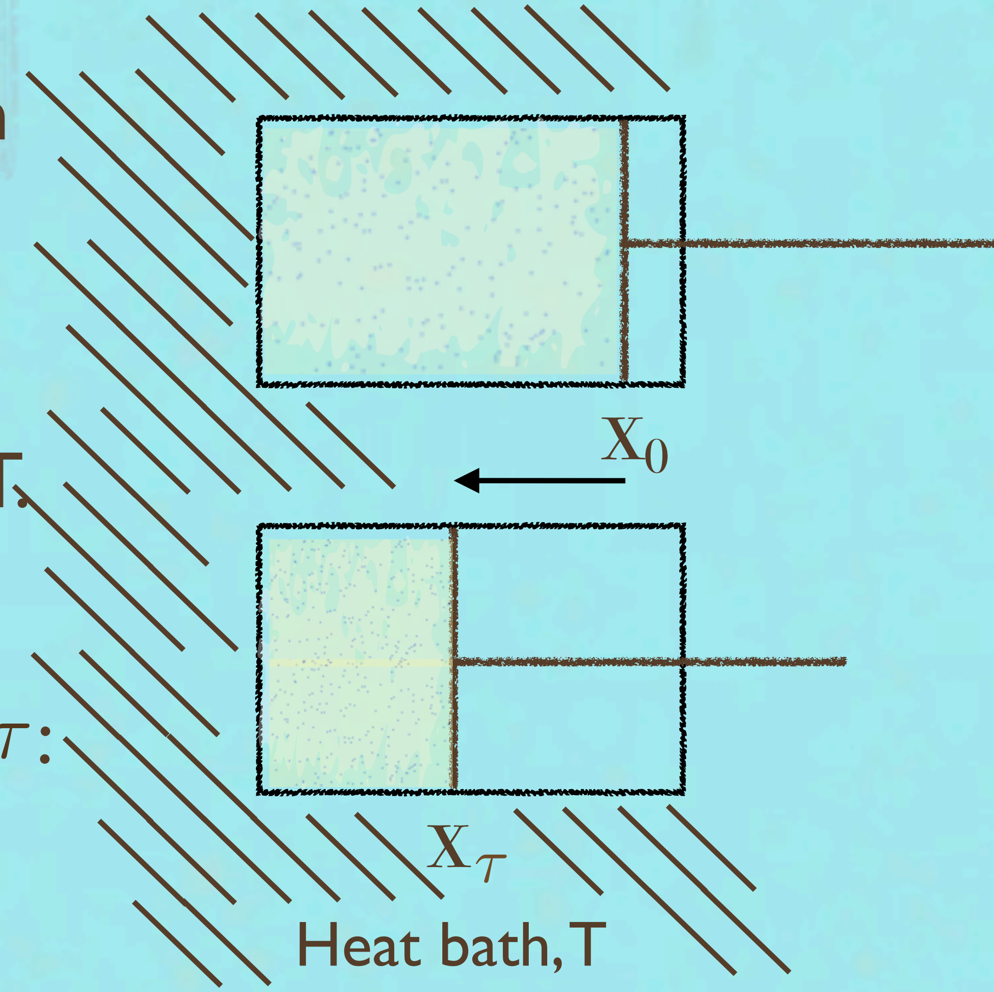
- This is the ***maximum entropy*** distribution, given the known quantities [here: the average energy $E = \langle E(s,x) \rangle$]. (Gibbs, Jaynes, ...)
- Entropy: $S = k_B H[p_{\rm eq}] = -k_B \left\langle \ln[p_{\rm eq}] \right\rangle_{p_{\rm eq}}$
- (Equilibrium) Free Energy: $F = E - TS$

Example: Gas in box with piston. x = piston position, s = positions and momenta of molecules. Temperature T.

1. Start system in thermodynamic equilibrium at t=0: $p(s_0|x_0) = e^{-\beta[E(s_0,x_0)-F(x_0)]}$

2. Drive system, performing work, W. Contact to heat reservoir at temperature T. Heat flowing into gas = Q.

3. Let system relax back to equilibrium. At t=τ:
$p(s_\tau|x_\tau) = e^{-\beta[E(s_\tau,x_\tau)-F(x_\tau)]}$
Energy is conserved: $\Delta E = W + Q$



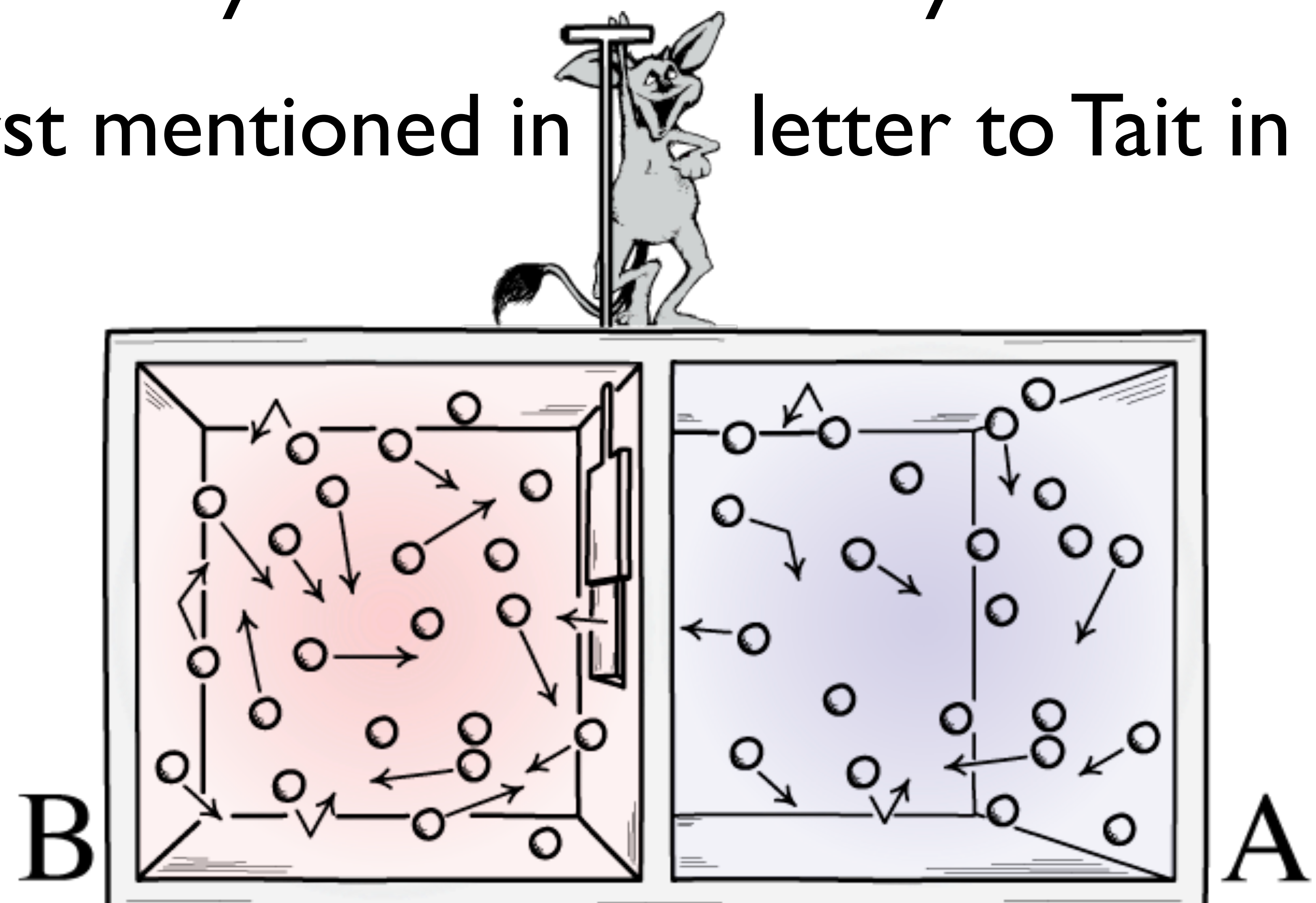
$$\Delta E := E_\tau - E_0$$
$$\Delta F := F_\tau - F_0$$
$$\Delta S := S_\tau - S_0$$

- Free energy change: $\Delta F = \Delta E - T\Delta S$
- Average work done in excess of free energy change (dissipated work):

$$\langle W_{\text{ex}} \rangle := \langle W \rangle - \Delta F = -\langle Q \rangle + T\,\Delta S \;\geq 0$$

Maxwell's demon

- Thought experiment that points out that the second law of thermodynamics can have only a statistical validity
- first mentioned in letter to Tait in 1867



Drawing: John D. Norton

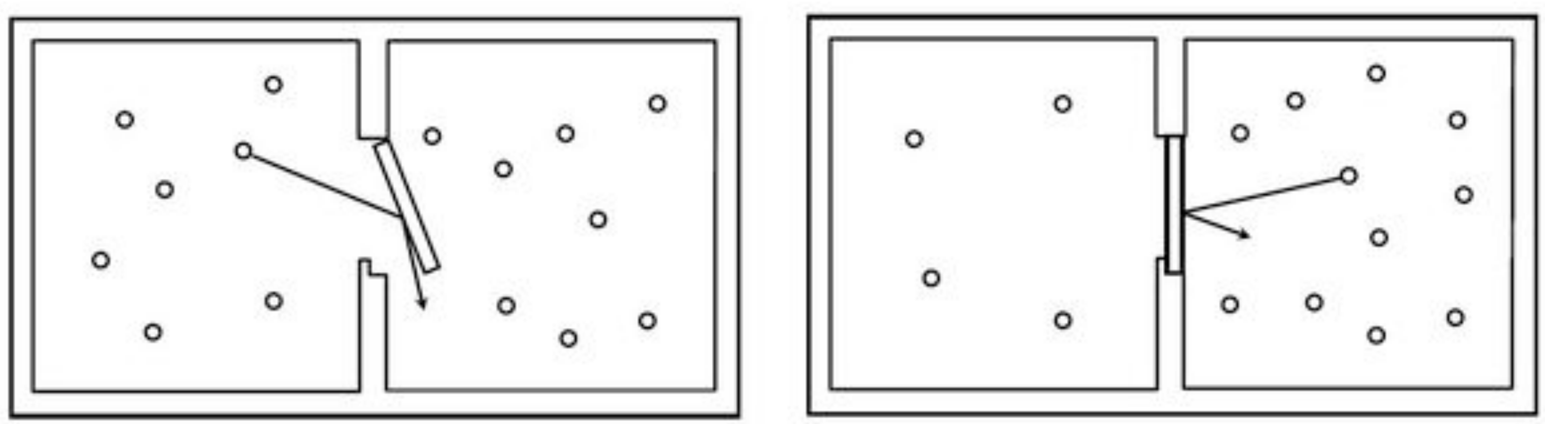
Maxwell's demon can not operate

- L. Brillouin (1951)
- Main idea: to "see" has to use light with energy exceeding the background (blackbody) radiation

In an enclosure at constant temperature, the radiation is that of a ``blackbody," and the demon cannot see the molecules. Hence, he cannot operate the trap door and is unable to violate the second principle. If we introduce a source of light, the demon can see the molecules, but the over-all balance of entropy is positive.

Trap door and spring

- Marian Smoluchowski (1914)
- replace demon by physical device
- trap door is supposed to act as a valve



- mechanism itself is subject to thermal fluctuations, hence can not work

Leo Szilard (1928)

Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen.

Von **L. Szilard** in Berlin.

Mit 1 Abbildung. (Eingegangen am 18. Januar 1928.)

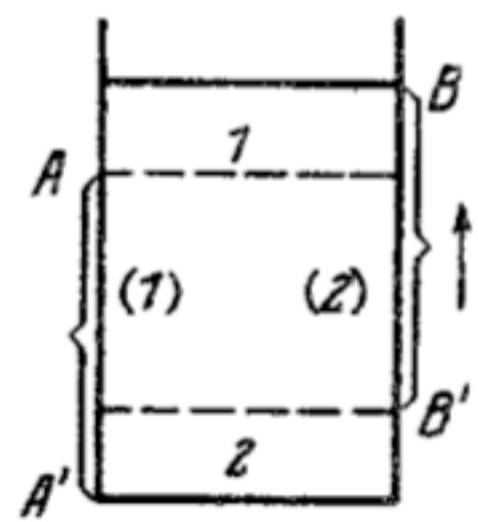
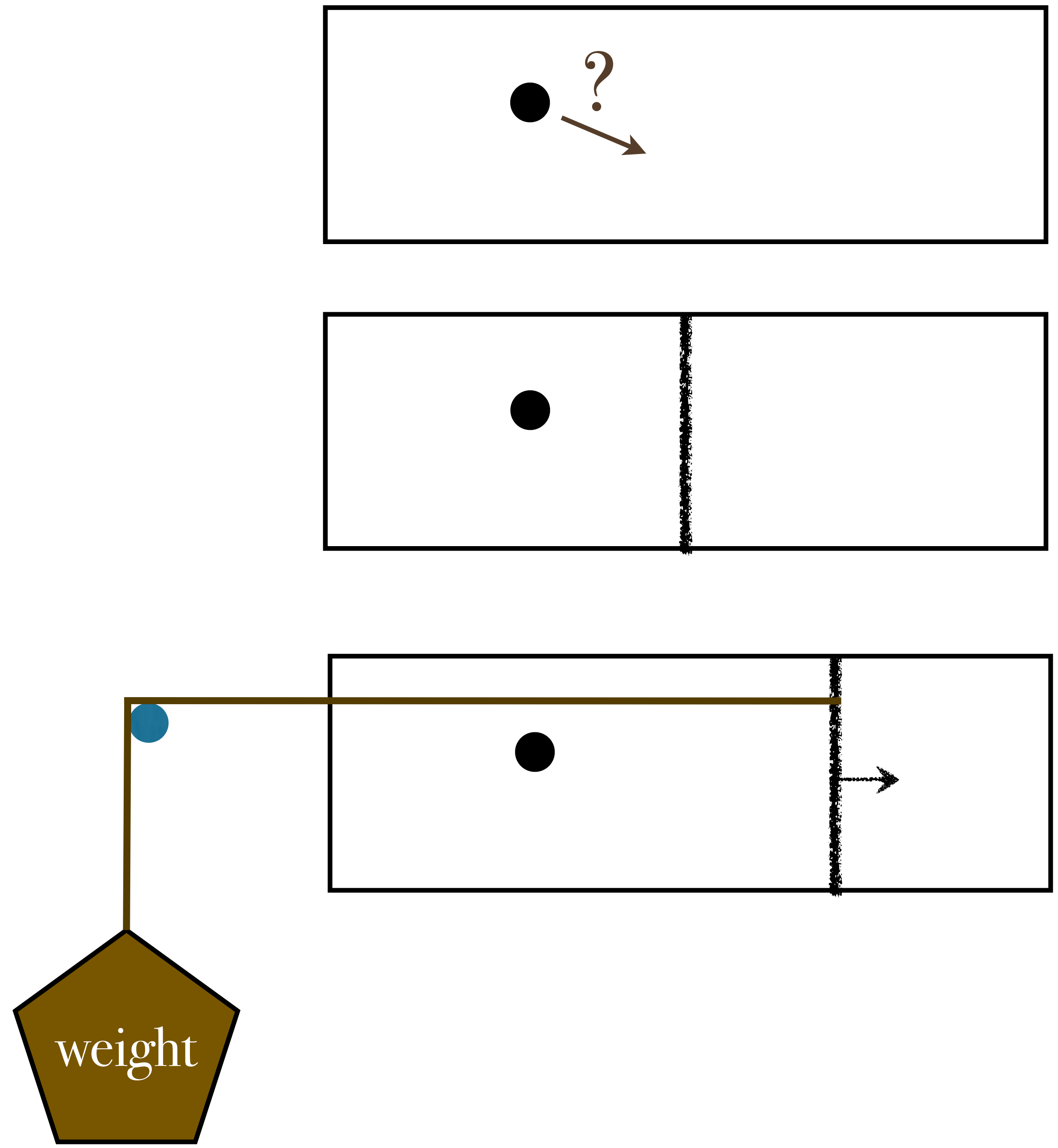


Fig. 1.

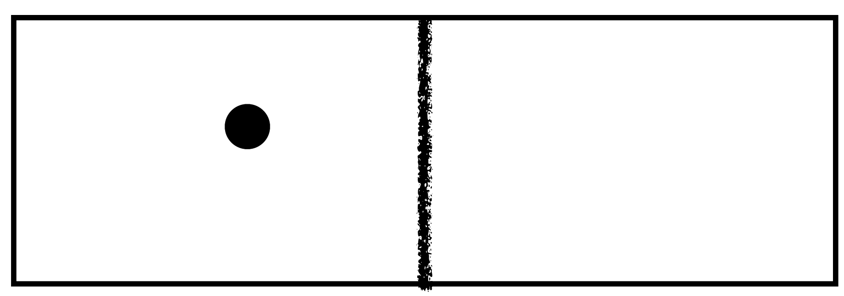
Leo Szilard (1928)

- consider a box with a single molecule
- insert partition
- measure where particle is
- move partition away from particle thus extracting work in the amount of kT ln(2)



Szilard pointed out

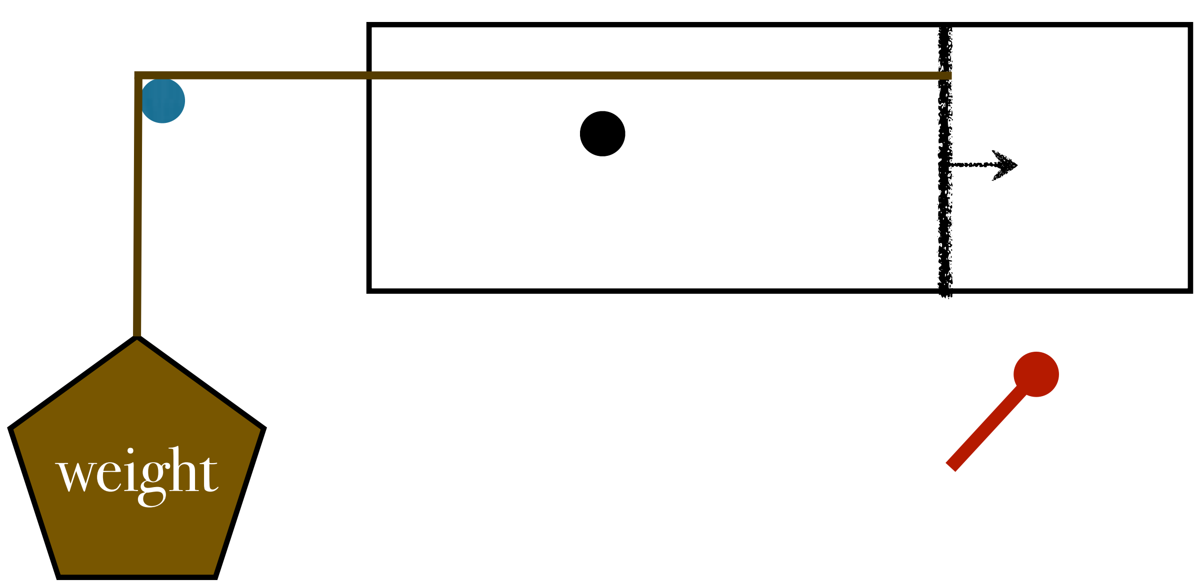
- Measurement:



- need to know which side particle is on to extract work

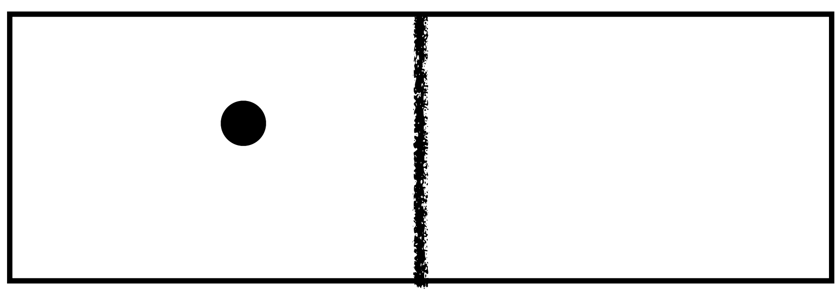
- Outcome stabilization:
  - insert partition before measuring particle location

- Memory retention:
  - need to retain the memory for the duration of work extraction scheme: turn a switch (lever).



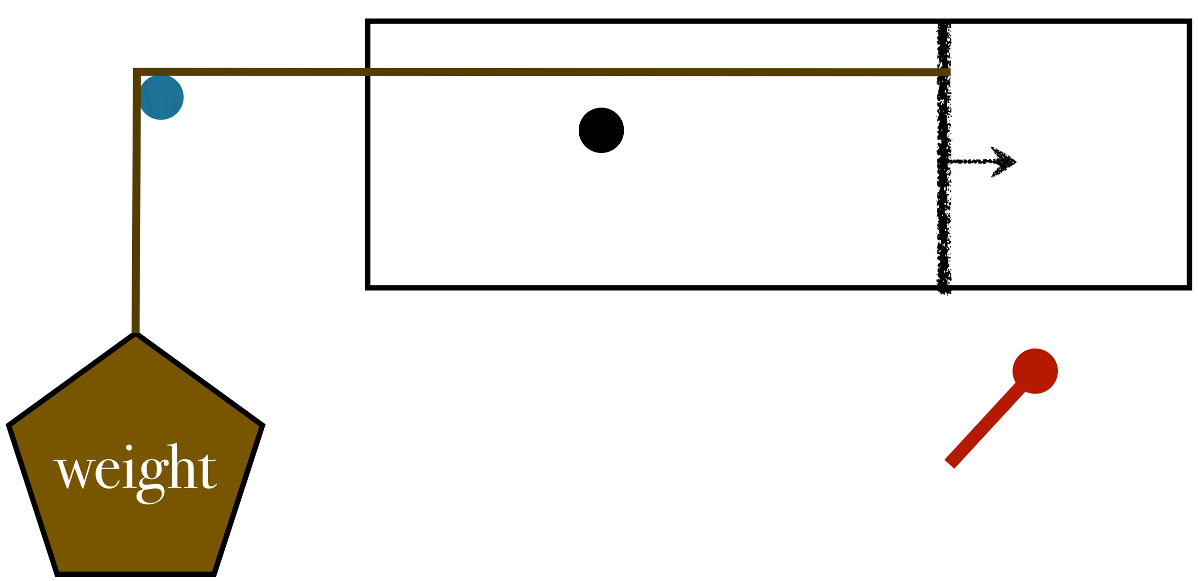
Szilard pointed out

- Measurement:
  - need to know which side particle is on to extract work



x: position of particle

- Outcome stabilization:
  - insert partition before measuring particle location
- Memory retention:
  - need to retain the memory for the duration of work extraction scheme: turn a switch (lever).



y: position of lever. Function of x *before* piston moves

Szilard's argument

- x and y *become coupled at time of measurement*, they are uncoupled before and after.
- The coupling of x and y *allows for work extraction*, but *is lost during the process*.
- The ***process of coupling*** the direction in which piston moves during work extraction (y) to the position of the particle *before* work extraction (x) ***comes at the cost of entropy increase***, in accordance with second law.

Landauer's argument (1961)

- Erasing information means:
  reset system to zero entropy state, i.e. $H_\tau = 0$

- Second law tells us

$$-\langle Q \rangle + k_B T \left(H_\tau - H_0 \right) \geq 0 \quad \Rightarrow \quad \boxed{-\langle Q \rangle \geq k_B T H_0}$$

- When one bit of information is erased, heat is produced, in the amount of *at least* kT ln(2).

- Landauer's bound is a direct consequence of the second law of thermodynamics.

Good enough?

- These arguments have discussed transformations between thermodynamic equilibrium states.
- What if there is no time to stay in equilibrium?
- Real computing machines are *out of thermodynamic equilibrium,* particularly biological ones.
- Do we need a "new thermodynamics" (S. DeDeo)?
- What is out there?
  - *Far-from-equilibrium / "stochastic" thermodynamics*
  - One-shot statistical mechanics (worry about "worst case scenarios" rather than just averages?)

Homework

Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen.

Von **L. Szilard** in Berlin.

Mit 1 Abbildung. (Eingegangen am 18. Januar 1928.)

- English translation: NASA document TT F-16723 (published 1976)
- Read and *summarize* Szillard's 1928 paper (if you can, read the original in German)
- Can you explain the figure in simple terms?
- Monday

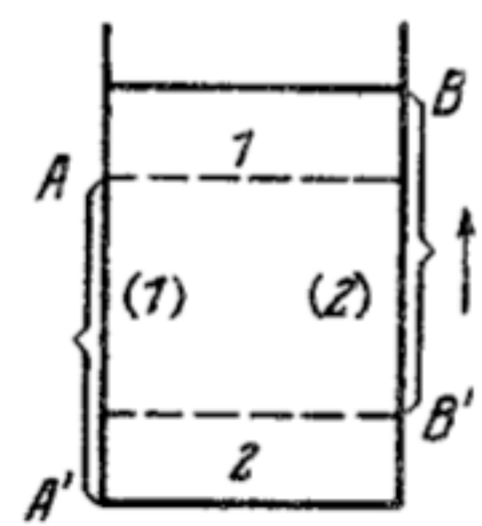


Fig. 1.

Break

- If you are really bored, you can read this (popular science journalism) article by Simon DeDeo (2015):

MATTER | PHYSICS

Nostalgia Just Became a Law of Nature

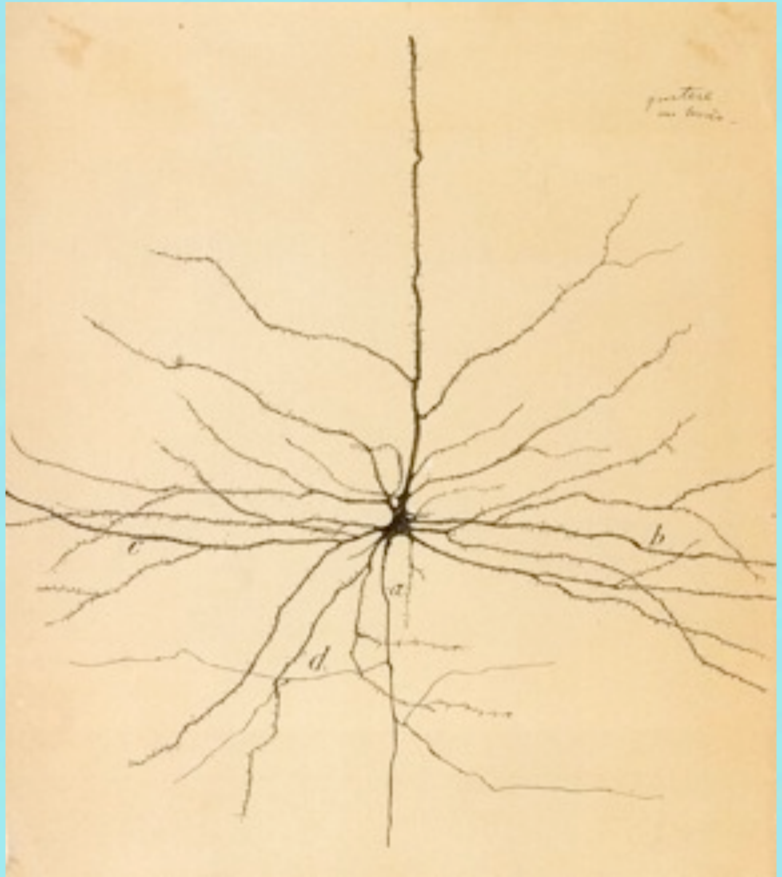
New theories have mixed perception and knowledge into the hardest of sciences.

BY SIMON DEDEO
PHOTO BY PRETTY LIFE PHOTOGRAPHY
FEBRUARY 19, 2015

http://nautil.us/issue/21/information/nostalgia-just-became-a-law-of-nature

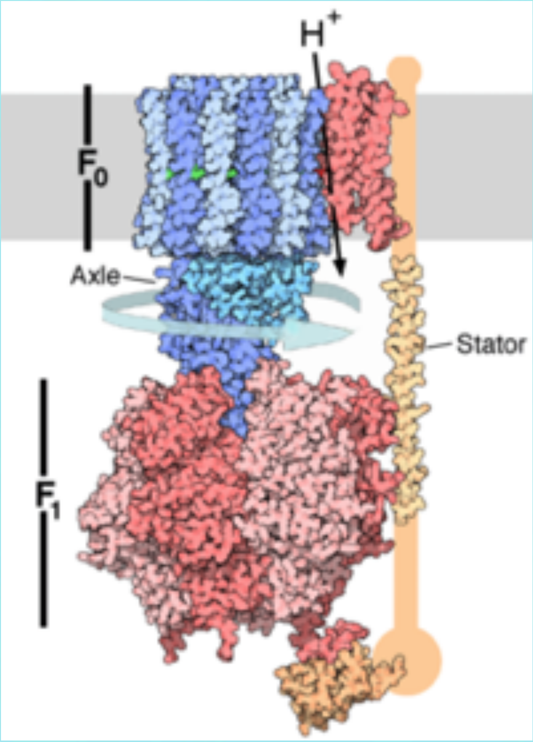
Living systems are not in equilibrium!

Neuron (nerve cell)



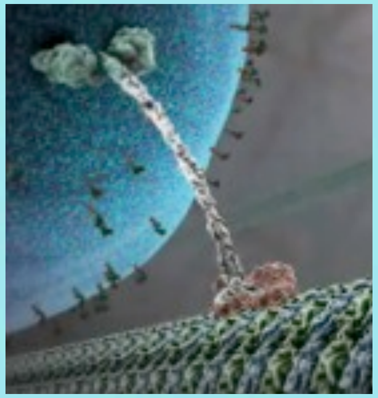
Rotary F_0-F_1 ATP synthase

(ATP - Adenosine triphosphate)



From: PDB, rcsb.org

kinesin



myosin

From: Inner Life of the Cell

Death = "the decay into thermodynamical equilibrium."
-Erwin Schrödinger, *What is Life?*

Far from thermodynamic equilibrium - what do we know?

Jarzynski's work relation (PRL 1997):

$$\boxed{\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta F}}$$

- Jensen's inequality => $\langle W \rangle \geq \Delta F$
- Write as moment generating function: $\Delta F = -\dfrac{1}{\beta} \ln \left\langle e^{-\beta W} \right\rangle$
- Expand into sum of cumulants; Gaussian noise

  => fluctuation-dissipation relation: $$\langle W \rangle - \Delta F = \frac{\sigma^2}{2 k_B T}$$

  with: $\sigma^2 = \langle W^2 \rangle - \langle W \rangle^2$

Crooks' derivation of Jarzynski's relation:

$p_{\text{eq}}(s_0|x_0)$ work step relaxation step $p(s_\tau|x_\tau)$

$$\begin{array}{l} x_0 \to x_1 \to \cdots \to x_{t-1} \longrightarrow x_t \to x_{t+1} \longrightarrow \cdots \to x_\tau \\ \quad\; s_0 \to s_1 \to \cdots \longrightarrow s_{t-1} \to s_t \longrightarrow s_{t+1} \to \cdots \to s_\tau \end{array}$$

$p(s_t|s_{t-1}, x_t)$

$p(s_0|x_0) \longleftarrow$ reverse time $p_{\text{eq}}(s_\tau|x_\tau)$

1) define:

$$W = \sum_{t=0}^{\tau-1} (E(s_t, x_{t+1}) - E(s_t, x_t))$$

$$Q = \sum_{t=1}^{\tau} (E(s_t, x_t) - E(s_{t-1}, x_t))$$

$$W + Q = E(s_\tau, x_\tau) - E(s_0, x_0) = \Delta E$$

2) consider forward and reverse time protocol

=> Crooks equation:

$$\frac{P_F(\vec{s}|\vec{x})}{P_R(\vec{s}|\vec{x})} = \frac{p_{\text{eq}}(s_0|x_0)}{p_{\text{eq}}(s_\tau|x_\tau)} \prod_t \frac{p(s_t|s_{t-1}, x_t)}{p(s_{t-1}|s_t, x_t)}$$

assume detailed balance

$$= e^{\beta(\Delta E - \Delta F)} \prod_t \frac{p_{\text{eq}}(s_t|x_t)}{p_{\text{eq}}(s_{t-1}|x_t)}$$

$$= e^{\beta(Q + W - \Delta F)} e^{-\beta \sum_t (E(s_t, x_t) - E(s_{t-1}, x_t))}$$

$$= e^{\beta(Q + W - \Delta F)} e^{-\beta Q}$$

$$= e^{\beta(W - \Delta F)}$$

3) easy proof:

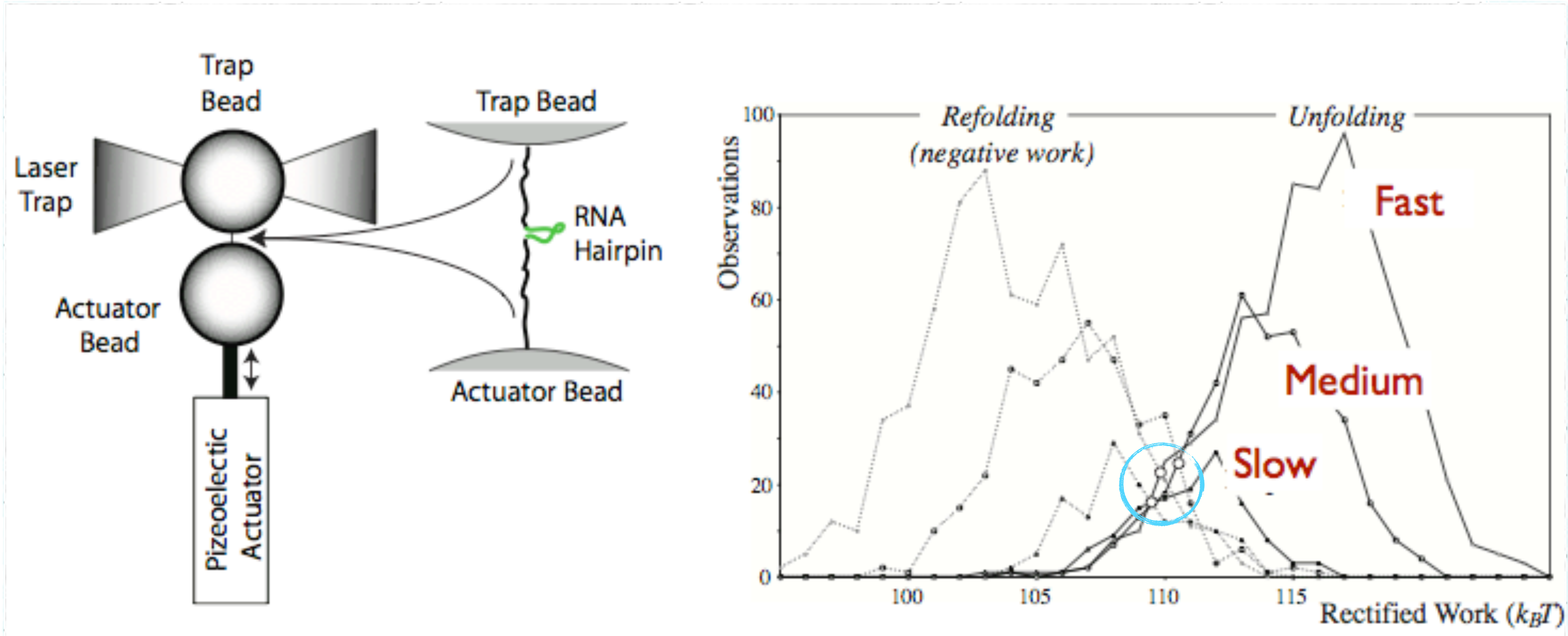
$$\langle e^{-\beta W} \rangle_{P_F} = \langle e^{-\beta W} \frac{P_F}{P_R} \rangle_{P_R}$$

$$= e^{-\beta \Delta F}$$

G. E. Crooks, J.Stat.Phys (1998)

=> Crooks' fluctuation theorem:

$$\frac{P_F(+W)}{P_R(-W)} = e^{\beta(W-\Delta F)}$$

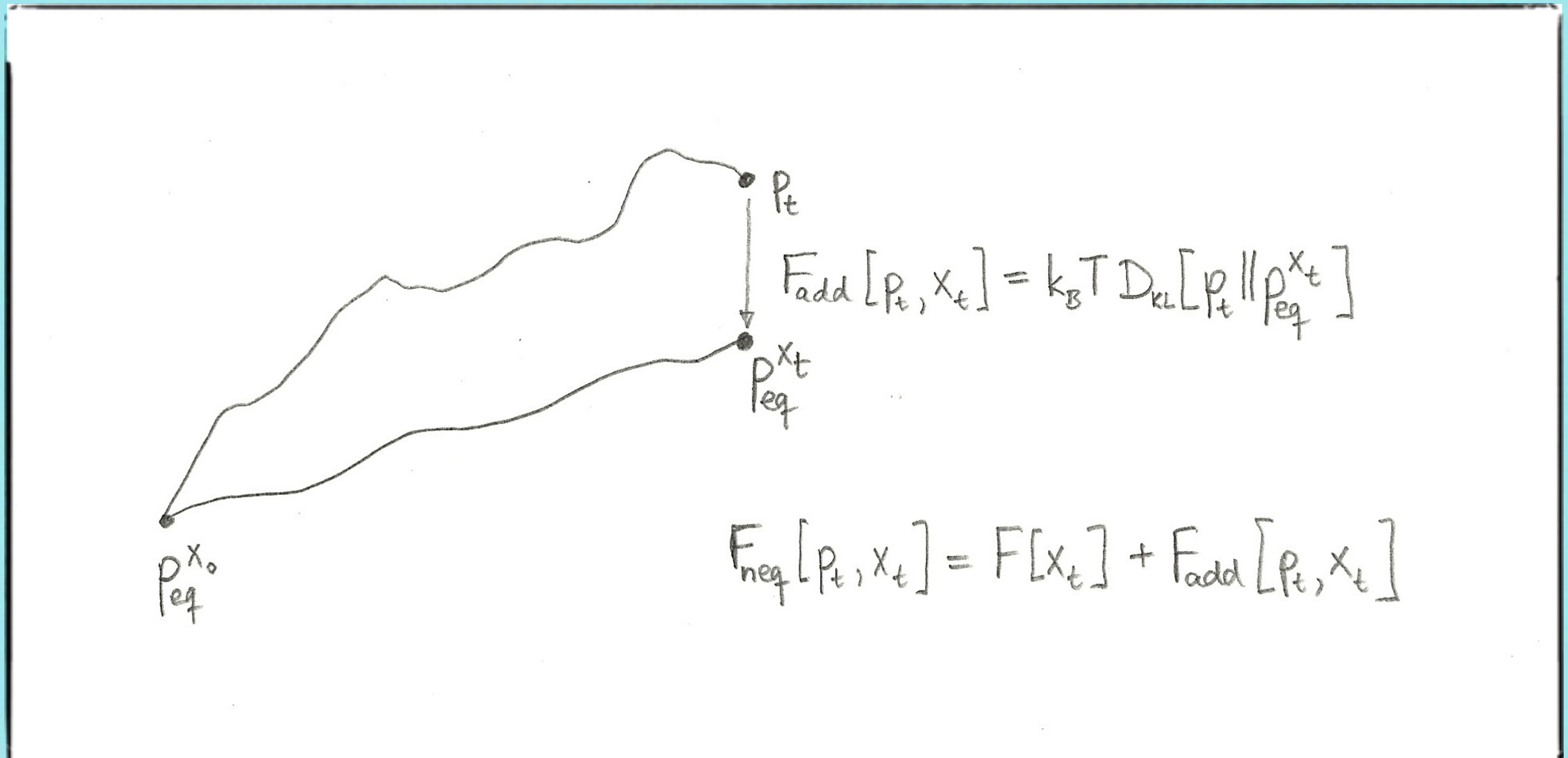


D. Collin, F. Ritort, C. Jarzynski, S.B. Smith, I.Tinoco, C. Bustamante (2005)

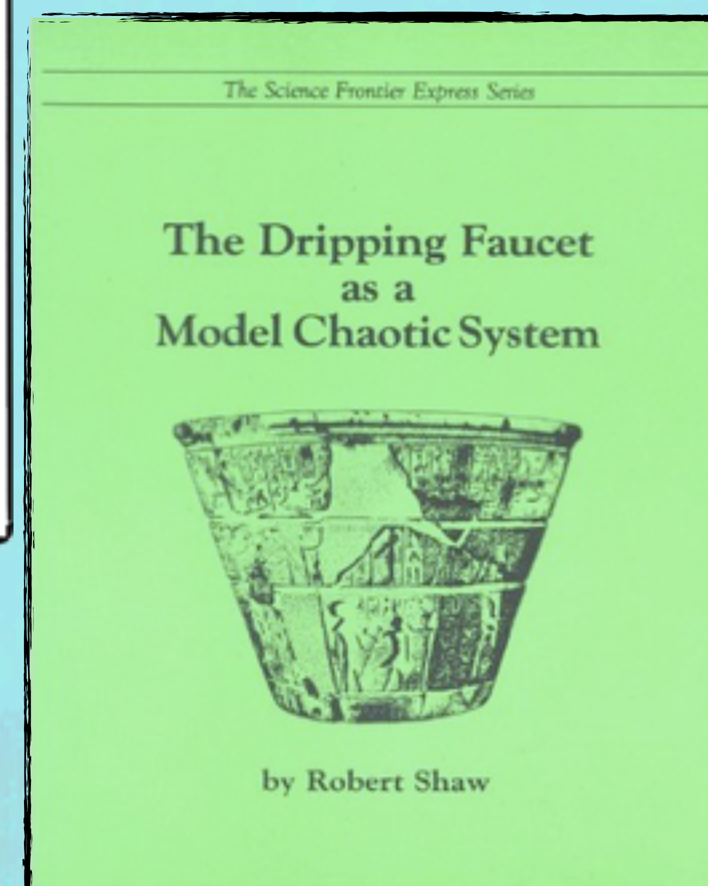
- **allows measurement of equilibrium free energy change in system that is far from equilibrium during the experiment,** e.g. optical trap RNA hairpin unfolding (& refolding).

What if there is no relaxation back to equilibrium?

* Additional free energy compared to equilibrium



Shaw (1981)



* Dissipation is the work *irretrievably* lost:

$$W_{\mathrm{diss}} = W - \Delta F_{\mathrm{neq}}$$

Esposito & Van den Broeck, EPL (2011); Still et al., PRL (2012)

Relation to work lost in excess of equilibrium free energy change

- Dissipation: $W_{\rm diss} = W - \Delta F_{\rm neq}$
  $= W_{\rm ex} - \Delta F_{\rm add}$
- Excess work:
  $W_{\rm ex} := W - \Delta F_{\rm eq}$

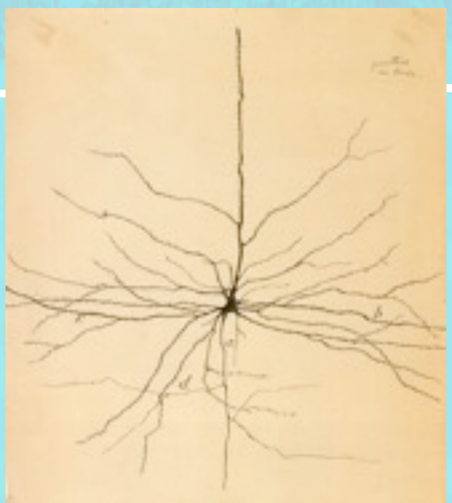
$W_{\rm diss}$ is non-negative ("generalized" second law)

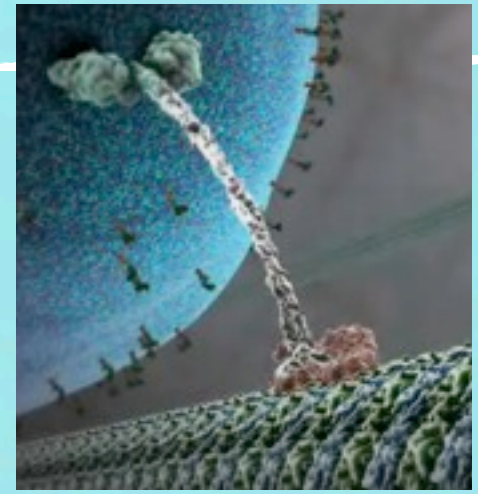
Therefore: $W_{\rm ex} \geq \Delta F_{\rm add}$

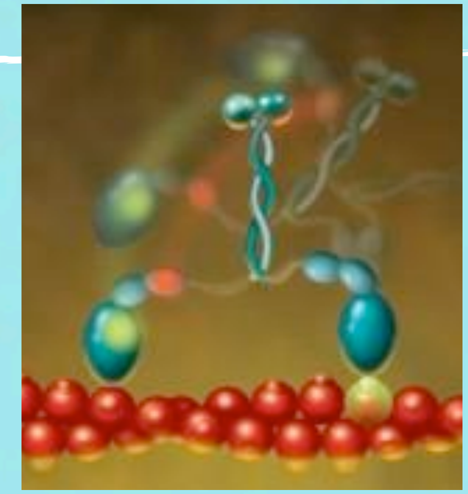
- this motivates control using measurement of the system; information could be used to lower dissipation

Esposito&VandenBroeck, EPL (2011); Sagawa&Ueda,PRL (2012)

Living systems are embedded in *stochastic* environments







- Dynamics of a stochastic environment governed by $p(x_0, \ldots, x_\tau)$
  - p is not necessarily "known"/given
  - system could be adapted to a certain type of environment

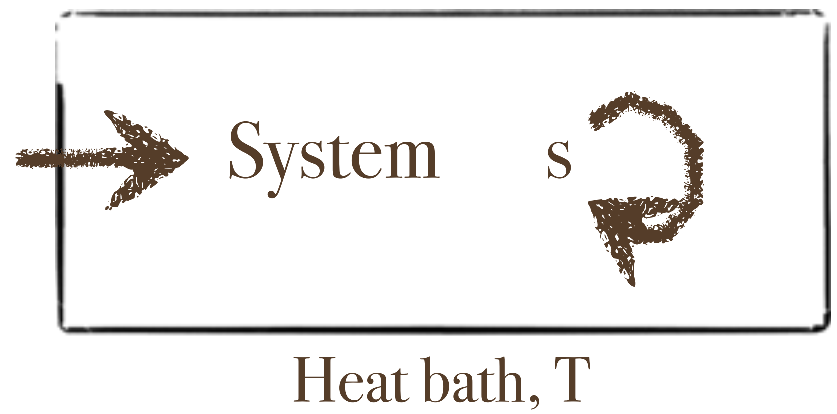
Setup:

Environmental variables x(t) $\xrightarrow{\text{drive}}$ System [state s(t)]

Heat bath, T

Physical systems process information

Environment -> external drive/data x



$$\begin{tikzcd}[column sep=small]
\text{Environment:} & x_0 \arrow[r] \arrow[d] & x_1 \arrow[r] \arrow[d] & \cdots \arrow[r] & x_{t-1} \arrow[r] \arrow[d] & x_t \arrow[r] \arrow[d, red] & x_{t+1} \arrow[r] \arrow[d] & \cdots \arrow[r] & x_\tau \arrow[d] \\
\text{(Learning) System:} & s_0 \arrow[r] & s_1 \arrow[r] & \cdots \arrow[r] & s_{t-1} \arrow[r, red] & s_t \arrow[r] & s_{t+1} \arrow[r] & \cdots \arrow[r] & s_\tau
\end{tikzcd}$$

$$p(s_t|s_{t-1}, x_t)$$

System computes by changing its state in response to changes in the environment.

Dynamics of the system encode a **model of environment**.

Still *EPL* (2009); Still et al. *PRL* (2012); Still *Entropy* (2014)

Stochastic thermodynamics of learning machines

$$
\begin{array}{l}
x_0 \to x_1 \to \cdots \to x_{t-1} \longrightarrow x_t \to x_{t+1} \longrightarrow \cdots \to x_\tau \\
\quad s_0 \to s_1 \to \cdots \longrightarrow s_{t-1} \to s_t \longrightarrow s_{t+1} \to \cdots \to s_\tau
\end{array}
$$

Dissipation due to instantaneous change in environment is proportional to instantaneous nonpredictive information:

$$I[S_t; X_t] - I[S_t; X_{t+1}] = H[S_t|X_{t+1}] - H[S_t|X_t]$$

$$= \beta \Big[\underbrace{\langle E(s_t, x_{t+1}) \rangle - \langle E(s_t, x_t) \rangle}_{\text{average work}} - \underbrace{\left(\langle F_{\text{neq}}[p_t(s|x_{t+1})] \rangle - \langle F_{\text{neq}}[p_t(s|x_t)] \rangle \right)}_{\text{av. non-eq. free energy change}} \Big]$$

$$\boxed{I[S_t; X_t] - I[S_t; X_{t+1}] = \beta \langle W_{\text{diss}}(x_t \to x_{t+1}) \rangle}$$

Still et.al. PRL (2012)

- **New bound on excess work:** $\boxed{\beta \langle W_{\text{ex}} \rangle \geq I_{\text{nonpred}}}$

(I_{nonpred} is the *total* instataneous nonpredictive information, summed over time: $I_{\text{nonpred}} = \sum_{t=0}^{\tau} I[S_t; X_t] - I[S_t; X_{t+1}]$)

- Therefore: $\boxed{-\beta Q \geq +\mathcal{I}_{\text{erasure}} + I_{\text{nonpred}}}$

Landauer refined

Landauer's bound augmented by nonpredictive information (which is a signature of the *dynamics* of the driven system)

Still et.al. PRL (2012)

Quantum systems

- **Result carries over** (Grimsmo, PRA 2013)
- Provides a new interpretation of quantum discord as: "the thermodynamic inefficiency of the most energetically efficient classical approximation of a quantum memory".
- Discord "quantifies the contribution to the lost work coming from quantum correlations".
- Possible quantum advantage in terms of lost work (when environmental signal is non-classical).

Now we have our first hint…

- Thermodynamic efficiency of a learning machine is limited by the instantaneous *non-predictive information (= "nostalgia")* retained in the state of the system

- If we want to minimize dissipation in (learning) machines that work at finite rates and are driven out of equilibrium, then we may have to minimize nostalgia.

- Trivial minimum: keep no memory (and predict nothing)
  - Obviously not useful for a learning machine
  - This optimization makes sense only with a constraint, e.g. on memory

Punchline

Dissipation of energy is fundamentally related to information processing inefficiency via nonpredictive information

Provocative statements:

- Some (parts of) living systems might predict their environment in order to operate at maximal thermodynamic efficiency.
- Predictive inference--not only a higher cognitive function? Perhaps implemented on small scales, such as bio-molecular machines?
- Is prediction/predictive inference a signature of life?