Sequence Complexity and Work Extraction

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

ITA 2016, La Jolla, CA, February 2016

Information Thermodynamics – Some Background

Common theme: the presence of information "violates" the second law.

Systems with measurement and feedback control:

Systems controlled by parameters that depend on past (noisy) measurements.

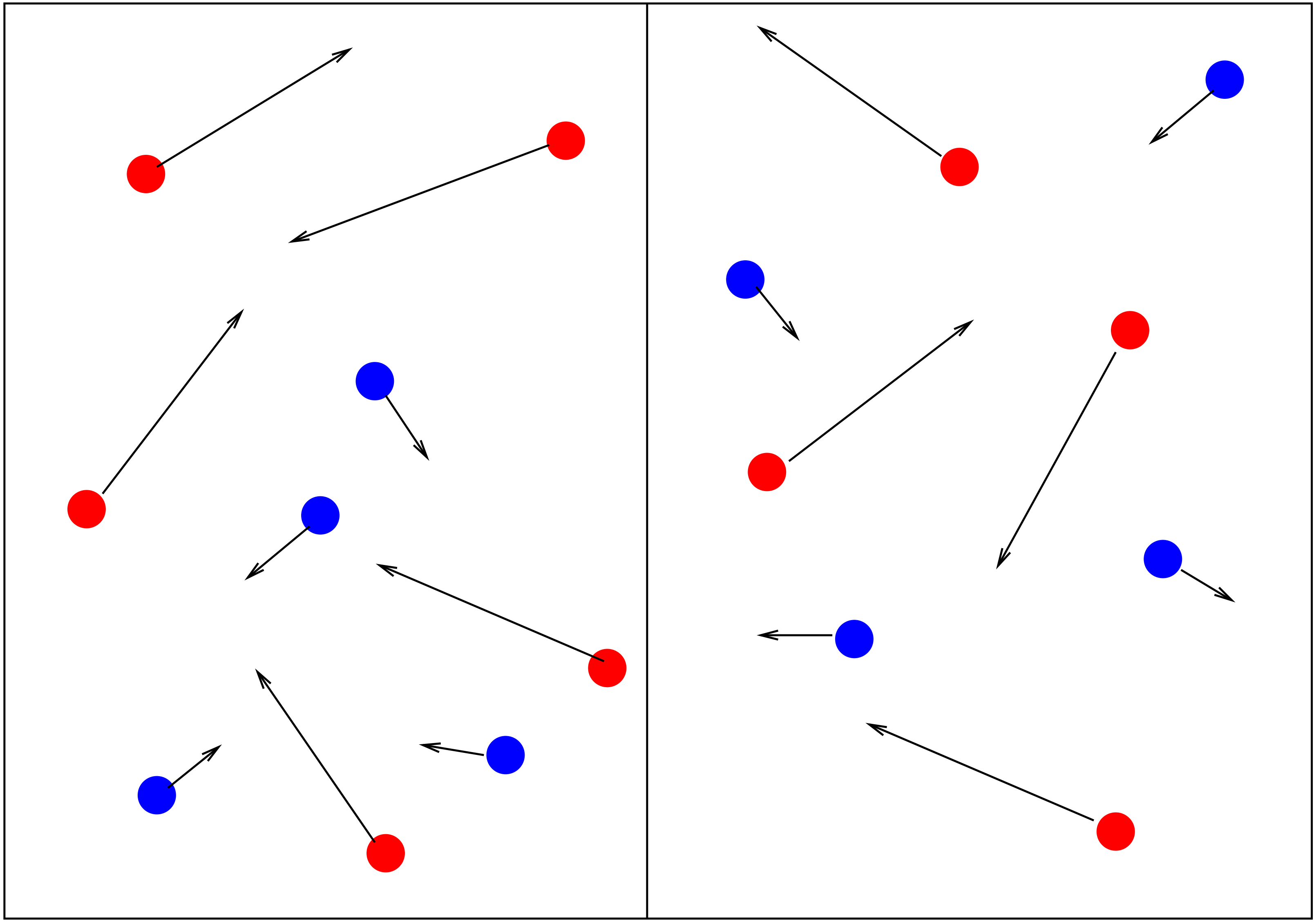
Barato & Seifert (2013, 2014); Deffner (2013); Deffner & Jarzynski (2013); Hoppenau & Angel (2014); Sagawa & Ueda (2011, 2012, 2013); Parrondo, Horowitz & Sagawa (2015); ...

Systems with information reservoirs:

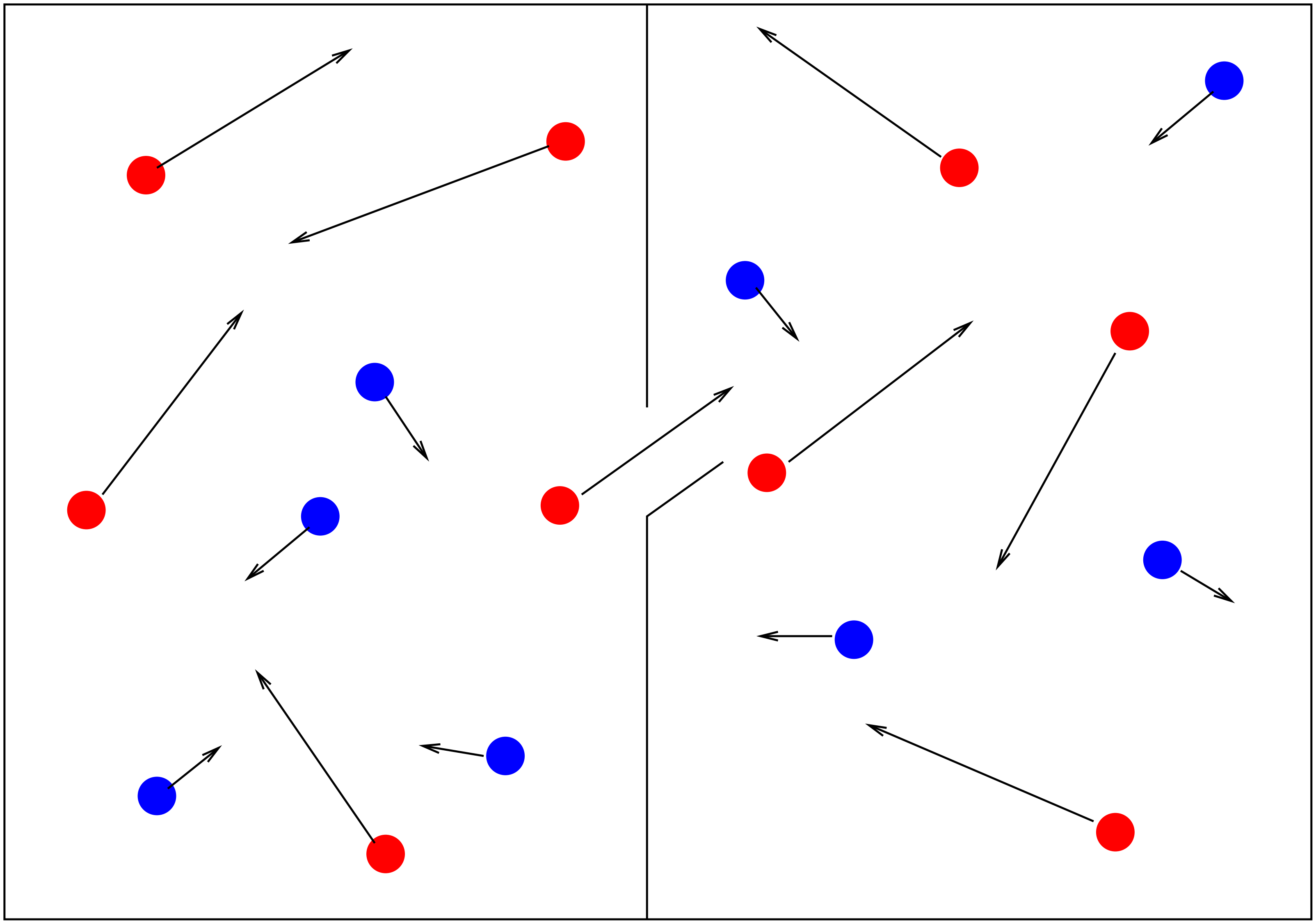
Systems that interact with an informational device (memory, digital tape, etc.).

Barato & Seifert (2014); Mandal & Jarzynski (2012, 2014); Boyd, Mandal & Crutchfield (2015); Mandal, Quan & Jarzynski (2013); Horowitz & Esposito (2014); ...

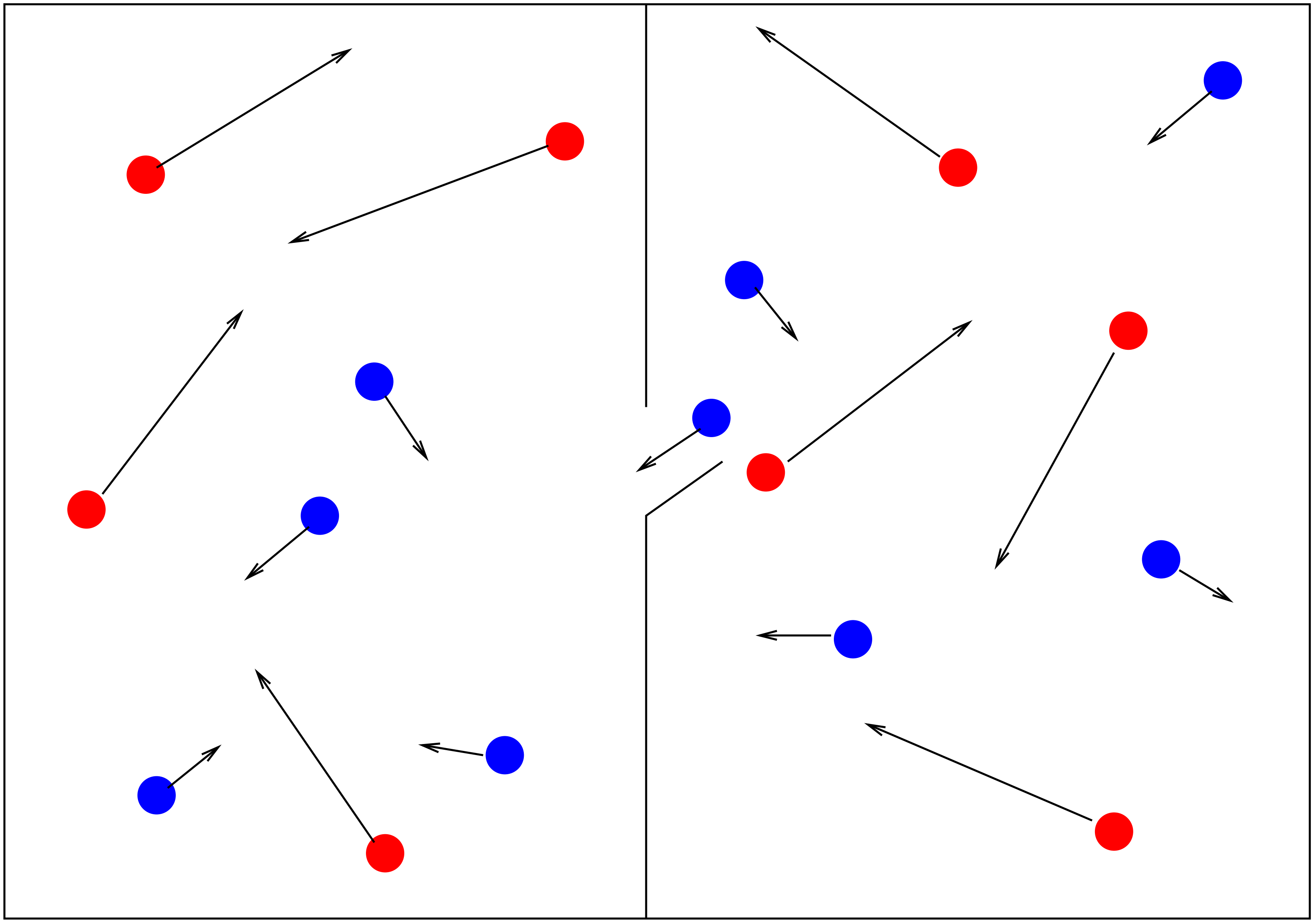
The Maxwell Demon (1867)



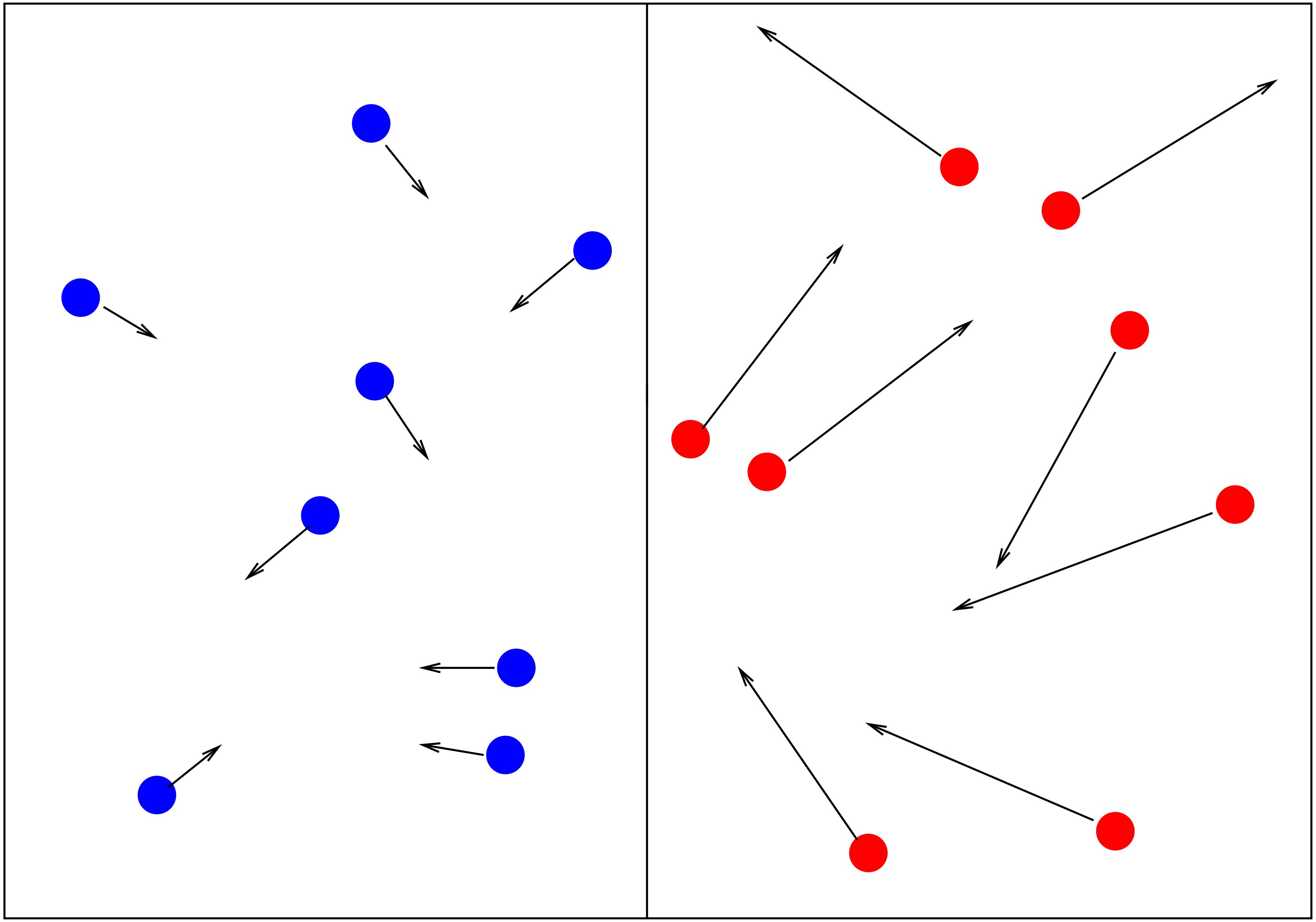
The Maxwell Demon (1867)



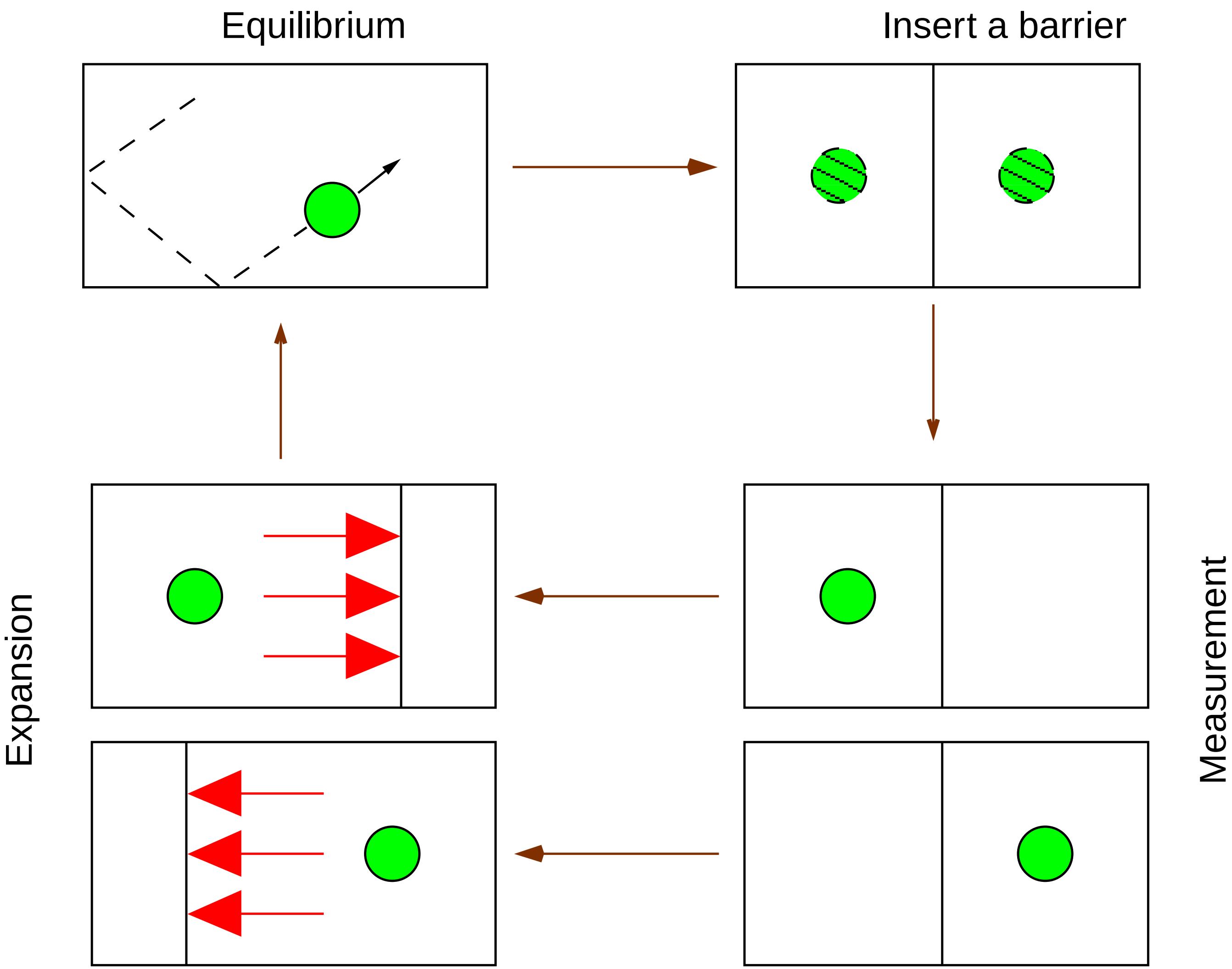
The Maxwell Demon (1867)



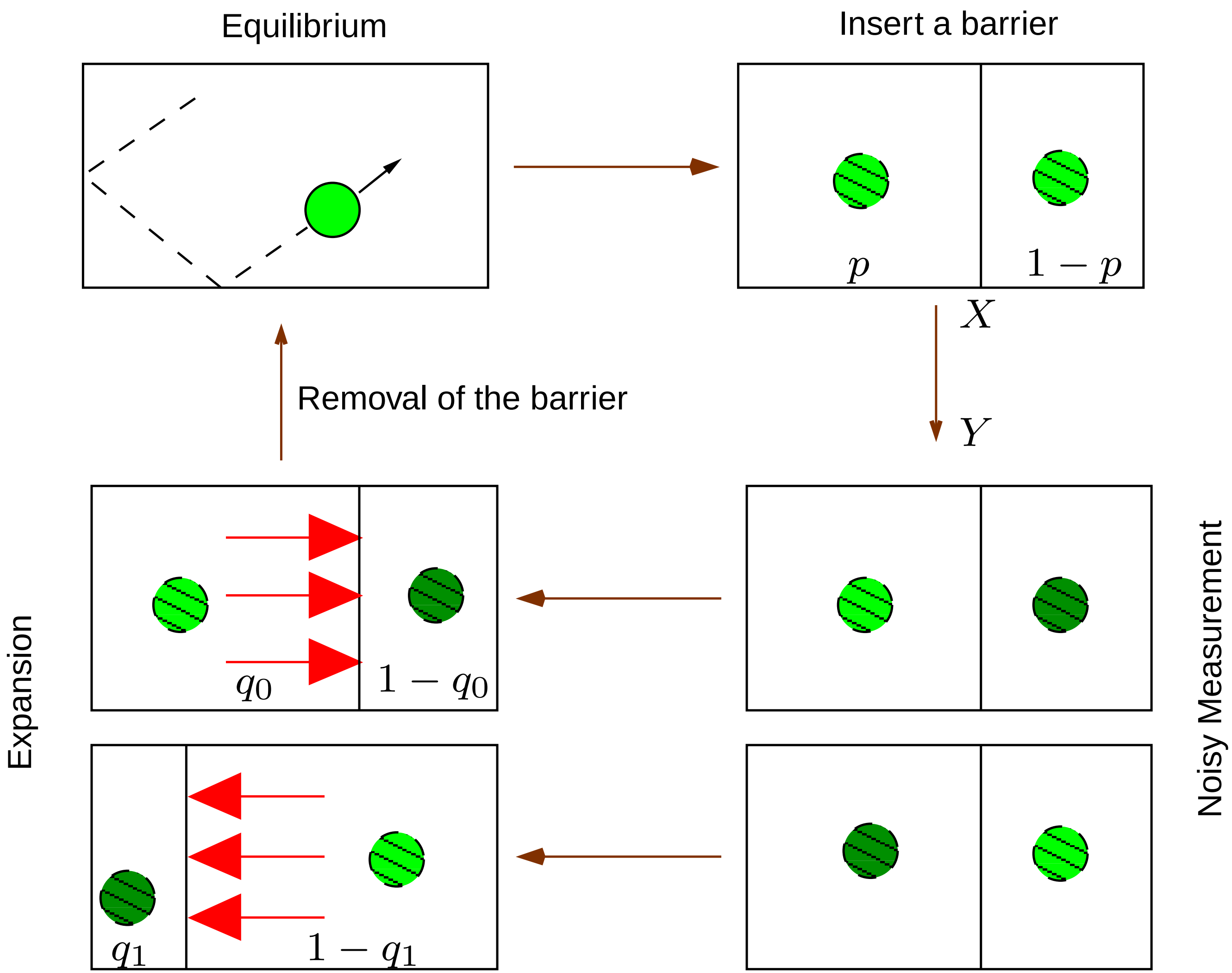
The Maxwell Demon (1867)



The Szilard Engine (1929)



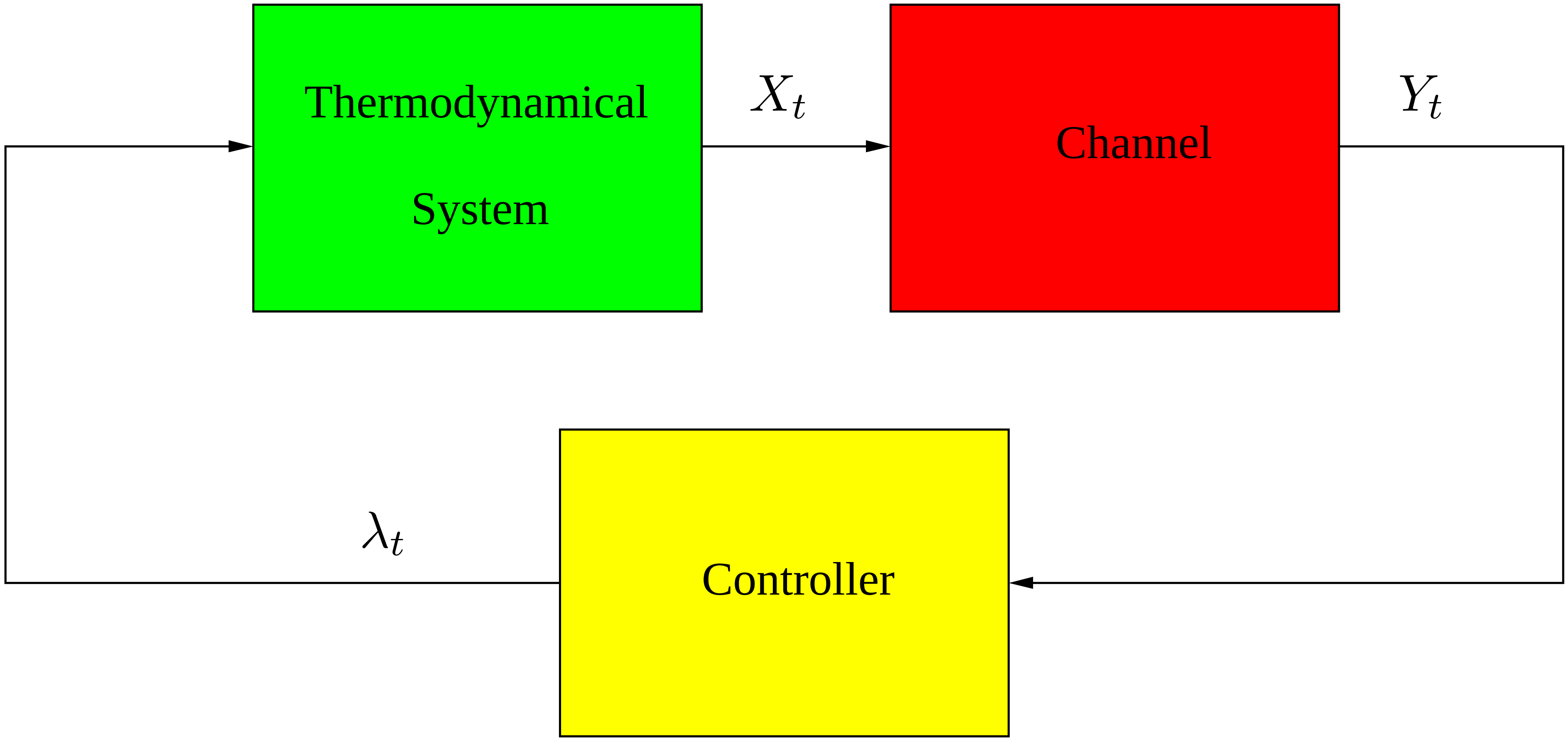
Generalized Szilard Engine (Sagawa & Ueda, 2011)



$\max_{q_0, q_1} W = kT \cdot I(X;Y)$

Vinkler, Permuter and Merhav (2014): relation to gambling.

System with Measurement and Feedback Control



Sagawa and Ueda (2008–2011) : Extracted work $\leq -\Delta F + kT \cdot nI(X;Y)$

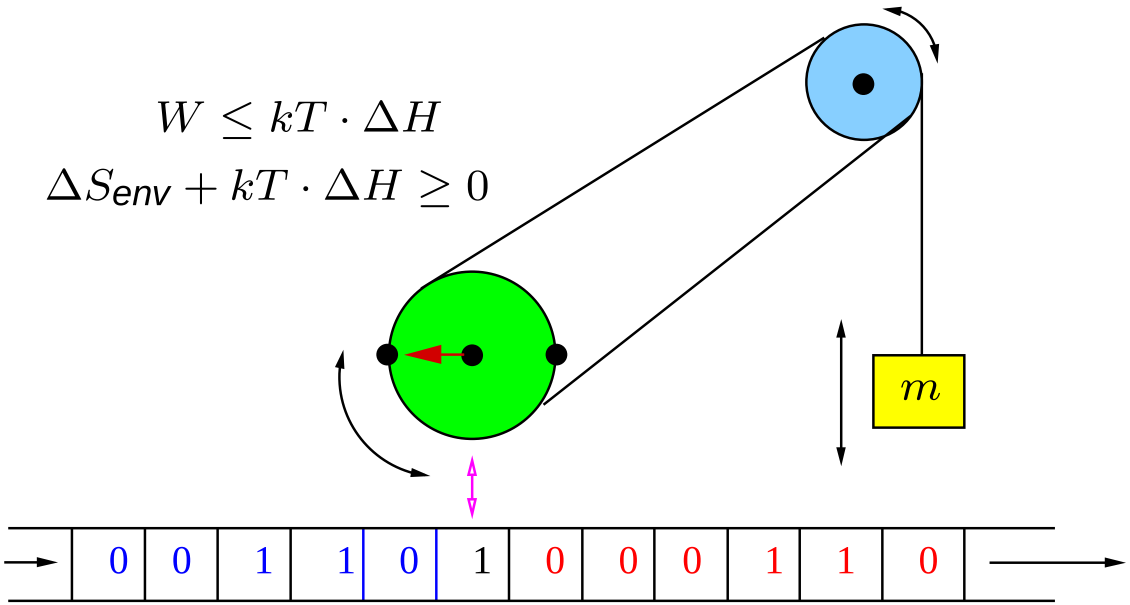
More generally, $nI(X;Y) \to I(X^n \to Y^n) =$ capacity of channel w. feedback.

Physical Systems with an Informational Device:

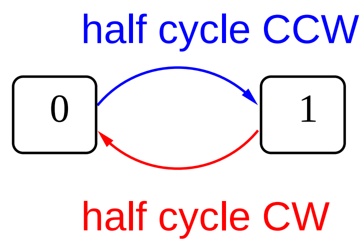
Mandal & Jarzynski (2012): system converting thermal fluctuations to work while writing info.

$$W \le kT \cdot \Delta H$$

$$\Delta S_{\text{env}} + kT \cdot \Delta H \ge 0$$



$$\frac{\lambda_{0\to1}}{\lambda_{1\to0}} = e^{-mg\Delta h/kT}$$

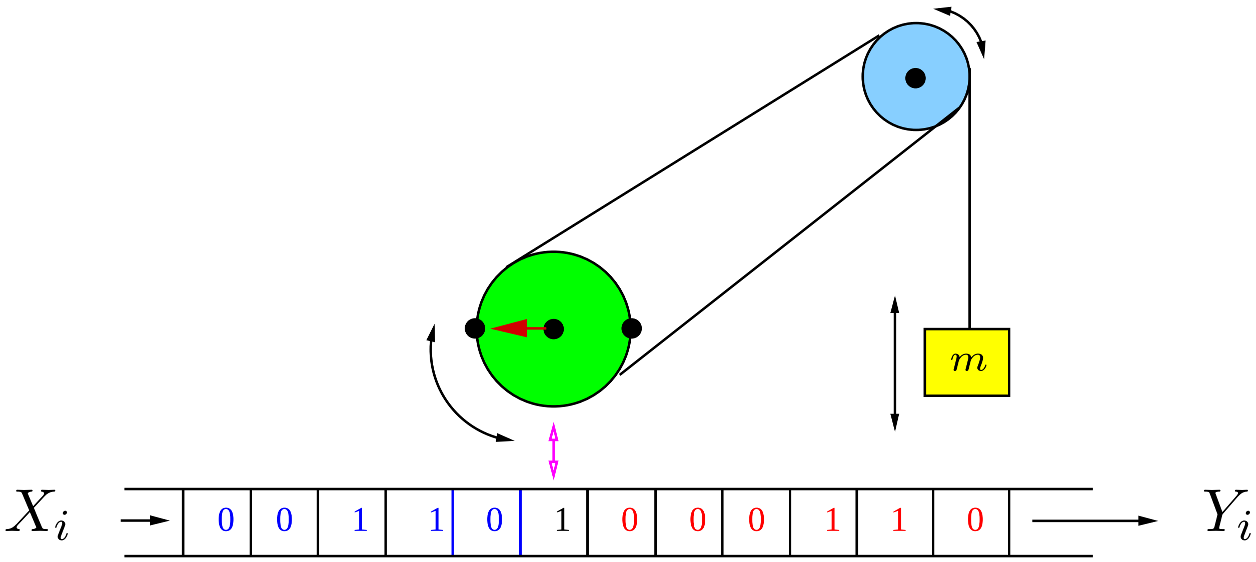


More generally, Deffner and Jarzynski (2013):

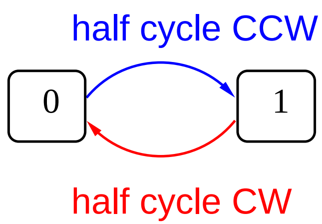
- A system (device) + a heat bath (heat reservoir),
- An information reservoir, e.g., a memory device with N bits (2^N states).

An extended 2nd law: $\Delta S_{\text{dev}} + \Delta S_{\text{heat-res}} + \Delta S_{\text{info-res}} \ge 0.$

More Details on the Simplified MJ Model



$$\frac{\lambda_{0\to 1}}{\lambda_{1\to 0}} = e^{-mg\Delta h/kT}$$



- $\{X_i\}$ binary, i.i.d. Each bit interacts for τ seconds.
- state =binary Markov jump process $s(i\tau) = X_i$; $s((i+1)\tau-) = Y_i$.
- $0 \to 1 \iff$ CCW; $1 \to 0 \iff$ CW. $\Delta W_i = mg\Delta h \cdot (Y_i - X_i)$.
- Main result in M&J (2012): $\Delta W_i \leq kT \cdot [H(Y_i) - H(X_i)]$.

This Work is About Several Extensions

- Non–binary sequences.
- Correlated input bits.
- Bounding ΔW in terms of generalized entropies.
- Individual (deterministic) input sequences.

Non–Binary Sequences

- In each interaction interval τ: Markov jump process with K states.
- Each state s is associated with energy $E(s) = mg\Delta(s)$ relative to some s_0.
- Transition $s \to s'$: $\Delta W = mg\Delta(s') - mg\Delta(s)$.

Using the inequality

$$D(P_\tau \| P_{\text{eq}}) \leq D(P_0 \| P_{\text{eq}}),$$

and

$$\log \frac{1}{P_{\text{eq}}(s)} \leq \ln Z + \frac{mg\Delta(s)}{kT},$$

we readily obtain

$$\langle \Delta W_i \rangle \leq kT[H(Y_i) - H(X_i)].$$

Correlated Inputs

Mandal & Jarzynski (2012): In a single cycle

$$\langle \Delta W_i \rangle \leq kT \cdot [H(Y_i) - H(X_i)].$$

In n cycles,

$$\langle W_n \rangle = \sum_{i=1}^{n} \langle \Delta W_i \rangle \leq kT \cdot \sum_{i=1}^{n} [H(Y_i) - H(X_i)].$$

Necessary (and sufficient) condition for work extraction is entropy increase.

When the inputs (and hence also the outputs) are correlated:

$$\langle W_n \rangle \leq kT \cdot [H(Y_1, \ldots, Y_n) - H(X_1, \ldots, X_n)].$$

- Extended second law: $\Delta S_{\text{total}} = k \cdot [H(Y^n) - H(X^n)] - \langle W_n \rangle / T \geq 0$.
- Entropy production is minimized if the inputs $\{X_i\}$ are independent.

Individual Input Sequences

- How would the “new” 2nd law look like if $x_1, x_2, \ldots$ is deterministic?
- What would replace the entropies $H(X^n)$ and $H(Y^n)$?
- In information theory, $H(X^n)$ measures complexity (data compressibility).
- Analogue for indiv. sequences: Ziv & Lempel (‘78)– LZ compressibility.
- We show that LZ compressibility plays a role in the 2nd law.

The LZ Compressibility

For a sequence of bits $x^n = (x_1, x_2, \ldots, x_n)$, let

$$c(x^n) = \text{\# phrases obtained by incremental parsing}$$

Incremental parsing: parse x^n sequentialy into distinct phrases, s.t. each phrase = shortest string not seen earlier as a phrase.

Example: 1000110111010001O is parsed as 1,0,00,11,01,110,10,001,0.

The LZ complexity is defined as

$$\rho(x^n) = \frac{c(x^n) \log c(x^n)}{n}.$$

Ziv & Lempel ('78): $\rho(x^n) =$ best compression with finite memory.

Compressing a phrase: pointer to previous phrase + new symbol uncoded.

The MJ Model for Individual Input Sequences

Let

$$P(y|x) = \text{Pr}\{Y_i = y|X_i = x\}, \qquad x, y \in \{0, 1\}$$

and define

$$U(z) = \max\{H(Y) : \ H(X) \geq z\}.$$

$U(z)$ is concave and monotonically deceasing.

Our main result is:

$$\langle W_n \rangle \leq kT \cdot \{U[\rho(x^n)] - \rho(x^n) + o(n)\}$$

Discussion

$$\langle W_n \rangle \leq kT \cdot \{U[\rho(x^n)] - \rho(x^n) + o(n)\}$$

- The function $U(z)$ has a simple closed–form formula.
- The more x^n is LZ–compressible, the more work one can hope for.
- The bound is tight in the sense that for some x^n it is saturated.
- LZ complexity: entropic meaning, not only in IT, but also in physics.
- Zurek (late 80s) used the Kolmogorov complexity.
- LZ complexity is both more computable and yields a tighter bound.
- Other measures of complexity, e.g., finite–memory predictability.