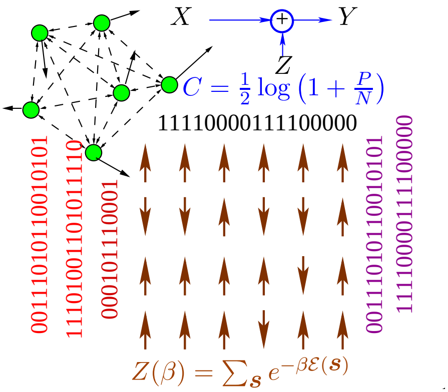
Information and Physics – bettertogether



Neri Merhav, Technion

ITW 2015, Jerusalem, April–May 2015

Some Meeting Points Between the Fields

- The Maxwell demon; the Szilárd engine.
- The maximum entropy principle: Jaynes, Burg, Shore & Johnson,...
- Physics of information: Landauer, Bennet, Maroney,...
- Phys. systems with informational resources: Ueda, Sagawa, Jarzynski,...
- Large deviations theory: Ellis, Oono, McAllester,...
- Random matrix theory: Wigner, Balian, Foschini, Telatar, Verdú,...
- Spin glasses and coding: Sourlas, Kabashima, Mézard, Montanari,...
- Replica theory: same and many others..

Talk Outline

Analogies and relations:

- Conceptual level
  - Information measures $\leftrightarrow$ physical quantities.
  - Coding theorems $\leftrightarrow$ physical laws.
  - Physical systems with informational ingredients.
  - Phase transitions.
- Technical level
  - Mapping physical models to communication system models.
  - The replica method and other analysis tools.
  - Random energy model (REM) and error exponents.
  - MSE analysis using statistical physics.

Very Brief Background in Statistical Physics

Consider a system with $N \gg 1$ particles which can lie in various microstates, $\{\boldsymbol{x} = (x_1, \ldots, x_N)\}$, e.g., a combination of locations, momenta, angular momenta, spins, ...

For every $\boldsymbol{x}$, $\exists$ energy $\mathcal{E}(\boldsymbol{x})$ – Hamiltonian.

Example: For $x_i = (\boldsymbol{p}_i, \boldsymbol{r}_i)$,

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \left(\frac{\|\boldsymbol{p}_i\|^2}{2m} + mgh_i \right).$$

Background (Cont'd)

In thermal equilibrium, $\boldsymbol{x} \sim$ Boltzmann–Gibbs distribution:

$$P_\beta(\boldsymbol{x}) = \frac{e^{-\beta \mathcal{E}(\boldsymbol{x})}}{Z(\beta)}$$

where $\beta = \frac{1}{kT}$, k – Boltzmann's constant, T – temperature, and

$$Z(\beta) = \sum_{\boldsymbol{x}} e^{-\beta \mathcal{E}(\boldsymbol{x})}, \quad \text{a normalization factor} = \text{partition function}$$

$\phi(\beta) = \ln Z(\beta) \quad \Rightarrow$ many physical quantities:

mean internal energy: $E = -\frac{d\phi}{d\beta}$;

entropy: $S = k \cdot \log \Omega(E) = \phi - \beta \frac{d\phi}{d\beta}$; $\quad \Omega(E) \stackrel{\triangle}{=} \text{Vol}\{\boldsymbol{x} : \ \mathcal{E}(\boldsymbol{x}) = E\}$.

free energy: $F = -\frac{\phi}{\beta} = E - TS$; $\Delta F =$ min. work between two equil. states.

heat capacity: $C = k\beta^2 \frac{d^2\phi}{d\beta^2}$

Part I

Relations in the Conceptual Level

Information Measures and Physical Quantities

Entropy:

$$\begin{aligned} S &\propto \log \text{Vol}\{\boldsymbol{x} : \mathcal{E}(\boldsymbol{x}) = E\} \\ &= \log \text{Vol}\left\{\boldsymbol{x} : \; -\log P_\beta(\boldsymbol{x}) = \underbrace{\beta E + \phi(\beta)}_{H}\right\} \qquad \beta \text{ 'tuned' to } E \\ &\approx H = \beta E + \phi(\beta) \qquad \text{weak typicality} \end{aligned}$$

The 'matching' β turns out to be the minimizer of $\beta E + \phi(\beta)$, i.e.,

$$H = \min_{\beta \geq 0} [\beta E + \phi(\beta)] \qquad \text{Legendre–Fenchel transform}$$

The two entropies are equivalent.

Information Measures and Physical Quantities (Cont'd)

Divergence (Bağci 2007): Let P_β be the B–G distribution and let Q be arbitrary:

$$\begin{aligned} D(Q\|P_\beta) &= -H_Q - \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) \ln P_\beta(\boldsymbol{x}) \\ &= -H_Q - \beta \sum_{\boldsymbol{x}} Q(\boldsymbol{x})[F_\beta - \mathcal{E}(\boldsymbol{x})] \qquad F_\beta \equiv F_{P_\beta} \\ &= -H_Q - \beta F_\beta + \beta \mathbf{E}_Q\{\mathcal{E}(\boldsymbol{X})\} \\ &= \beta(F_Q - F_\beta) \end{aligned}$$

or $$F_Q = F_\beta + kT \cdot D(Q\|P_\beta)$$

Divergence is proportional to the free energy difference.

In equilibrium ($\boldsymbol{X} \sim P_\beta$), F is minimum.

Information Measures and Physical Quantities (Cont'd)

Rate–distortion. Parametric representation (Legendre transform relation):

$$-R(D) = \min_{\beta \geq 0} \left[\beta D + \sum_x P(x) \underbrace{\ln \left(\sum_y Q(y) e^{-\beta d(x,y)} \right)}_{\phi_x(\beta)} \right],$$

$=$ entropy of a mixture of systems, each with $NP(x)$ particles, and Hamiltonian $\mathcal{E}_x(y) = d(x,y)$, in thermal equilibrium with total normalized energy D.

- Each x contributes normalized distortion $D_x = -\phi'_x(\beta^*)$.
- Equilibrium temperature $= 1/\beta^* =$ negative slope of $D(R)$.
- Channel capacity: same with $D = H(Y|X)$, $d(x,y) = -\ln P(y|x)$, $\beta^* = 1$.

Rate–distortion (Cont'd) – "I-MMSE" Relations

Parametric representation in terms of MMSE:

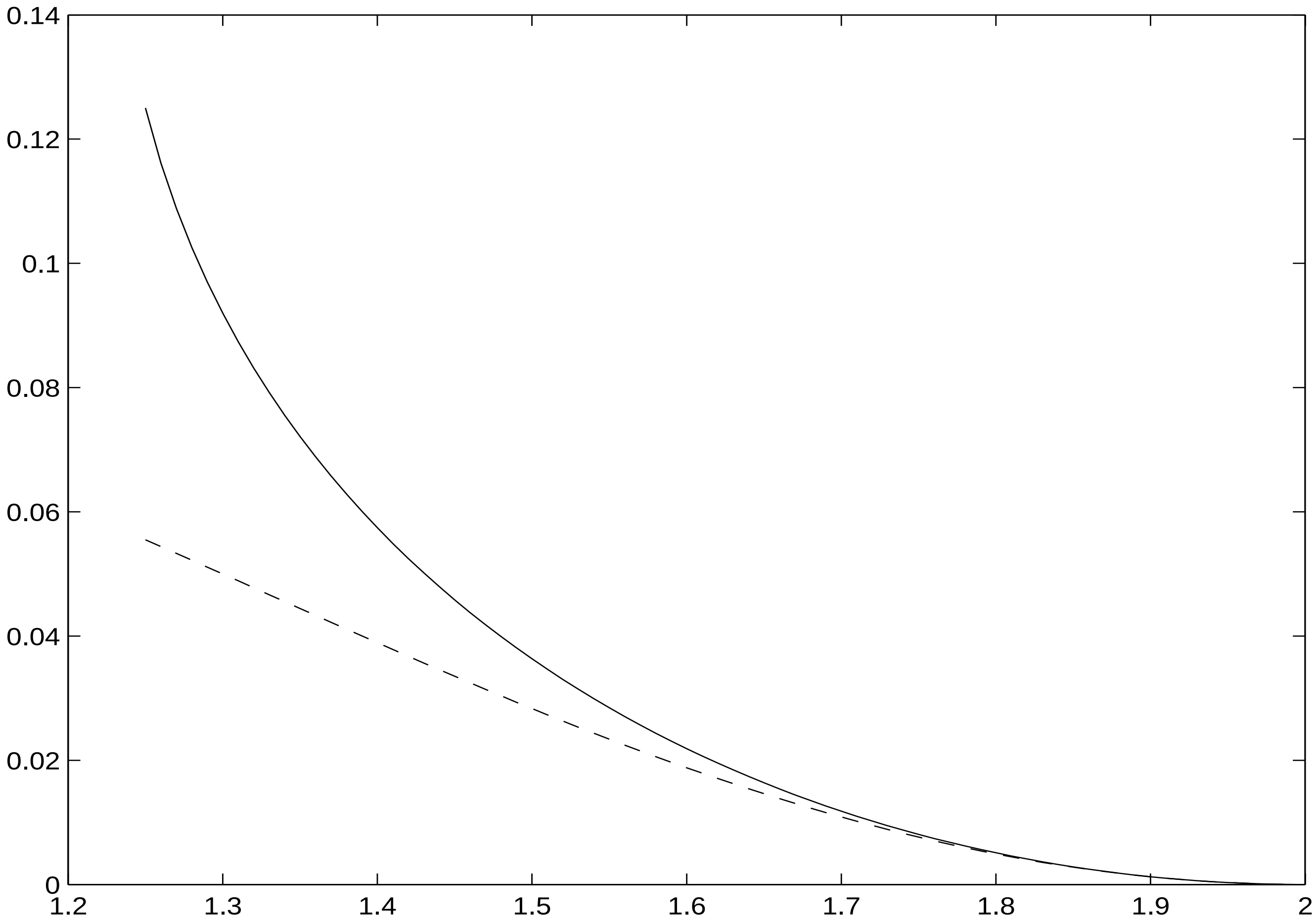
$$
\begin{aligned}
R(D) &= \int_0^{\beta} \mathrm{d}\hat{\beta} \cdot \hat{\beta} \cdot \mathrm{mmse}_{\hat{\beta}}\{d(X,Y)|X\} \\
D_{\beta} &= D_0 - \int_0^{\beta} \mathrm{d}\hat{\beta} \cdot \mathrm{mmse}_{\hat{\beta}}\{d(X,Y)|X\}
\end{aligned}
$$

where $P(x,y) \propto P(x)Q(y)e^{-\hat{\beta}d(x,y)}$.

- $\mathrm{mmse}_{\hat{\beta}}\{d(X,Y)|X\}$ related to heat capacity, $C(T)$.
- First integral – related to entropy: $S = \int \frac{C(T)\mathrm{d}T}{T}$.
- Second integral – related to heat: $E = \int C(T)\mathrm{d}T$.
- Enables derivation of bounds on $R(D)$ via bounds on MMSE.

Example: $P = \mathcal{N}(0,1)$; $Q =$ BSS $\{-1,+1\}$; $d(x,y) = (x-y)^2$.

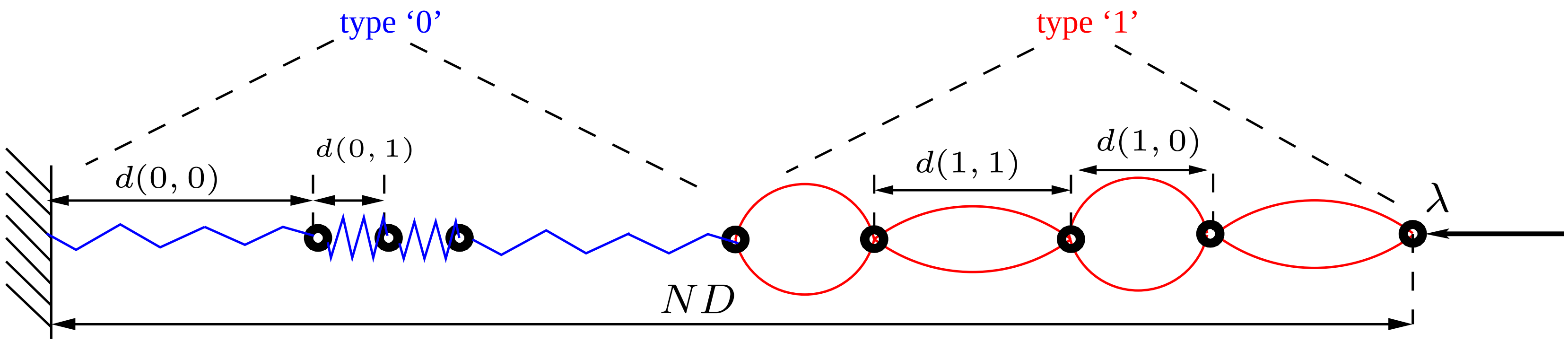
$$R(D) \geq \frac{(2-D)^2}{8} - \frac{3(2-D)^4}{64}; \quad R(D) \leq \frac{2}{3}\sin^2\left[\frac{1}{3}\sin^{-1}\left(\frac{3\sqrt{3}(2-D)}{4}\right)\right].$$



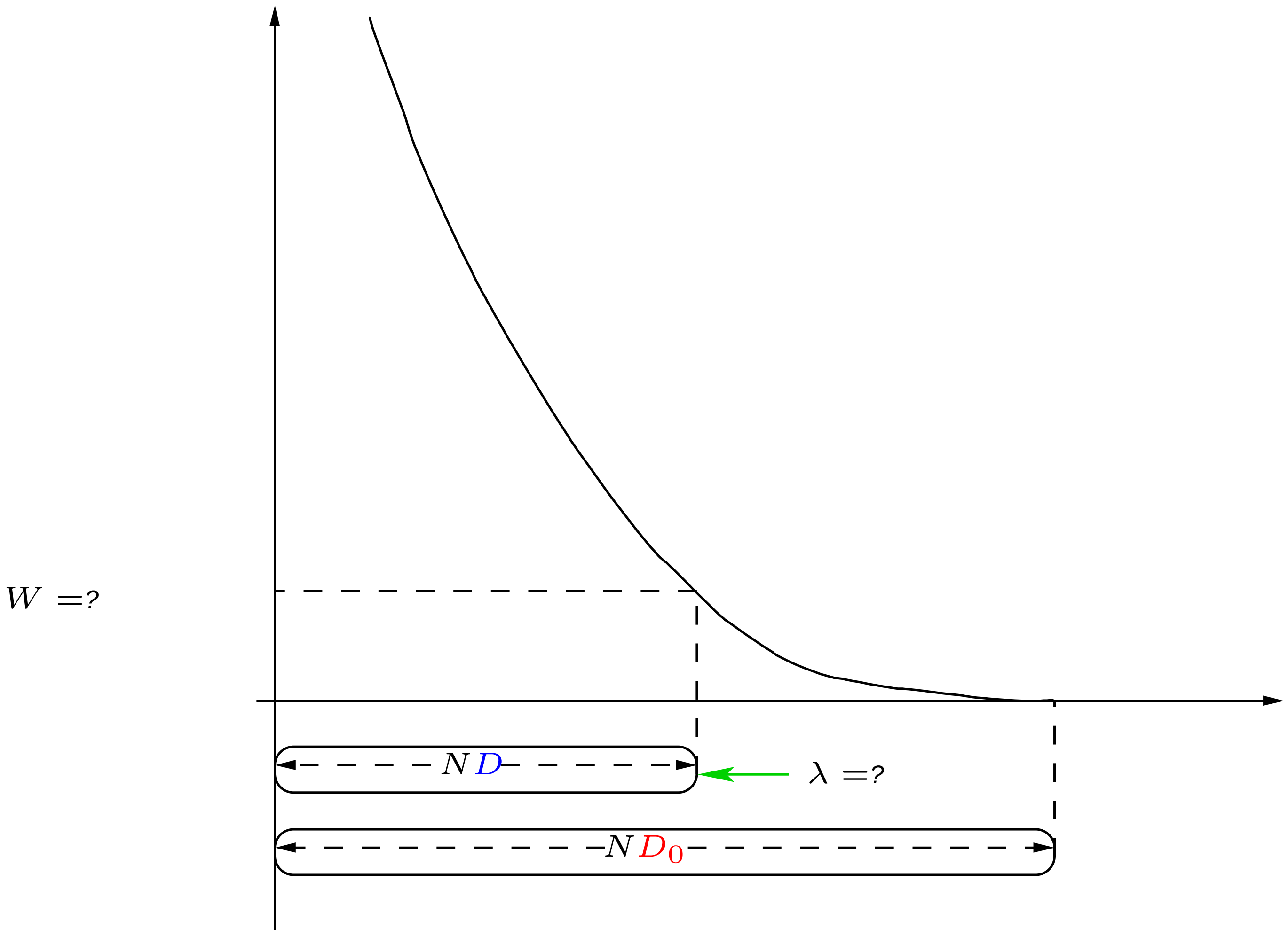
Physics of Coding Theorems

Physics of rate–distortion theory (or vice versa):

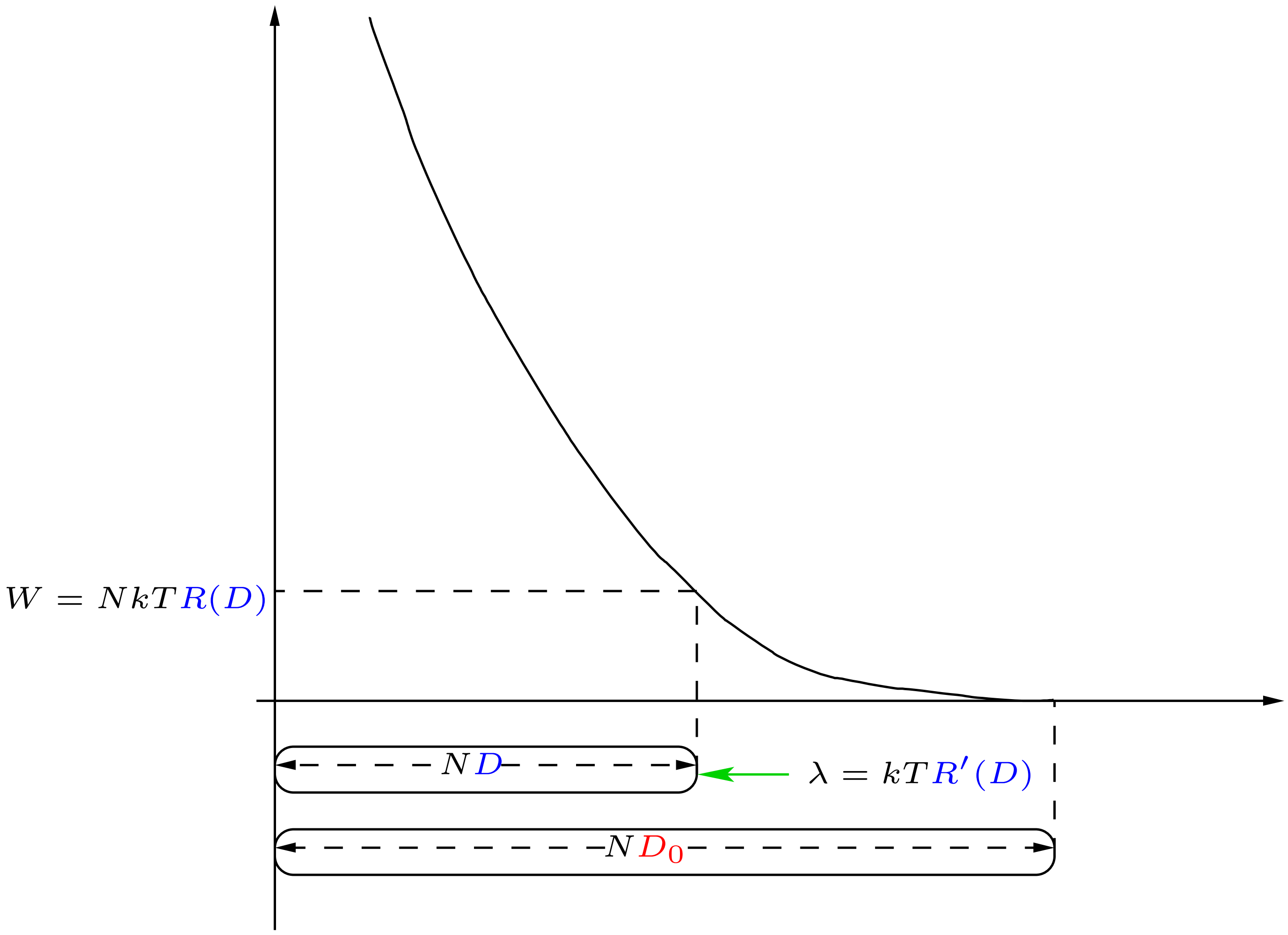
- Chain of N connected elements at temperature T.
- $|\mathcal{X}|$ types of elements, indexed by x.
- Number of elements of type $x \in \mathcal{X}$ is $NP(x)$.
- Each element – in one of $|\mathcal{Y}|$ states, labeled by y.
- Length of type–x element at state y is $d(x,y)$
- Energy of any element at state y is $\epsilon(y) = -kT \ln Q(y)$.
- What is the minimum work needed to shrink to length ND?



Minimum Required Work =?



Minimum Required Work $\propto R(D)$



Achievability: λ should grow slowly from zero to $\lambda = kTR'(D)$.

Information Loss and the 2nd Law

Essentially all fundamental limits of IT are based on the information inequality in some form (DPT, Fano's inequality, "conditioning reduces entropy," ...).

For any two distributions, P and Q, over an alphabet $\mathcal{X}$:

$$D(P\|Q) \triangleq \sum_x P(x) \log \frac{P(x)}{Q(x)} \geq 0.$$

Strict inequality is normally associated with some information loss.

An equivalent inequality – the log–sum inequality:

$$\sum_i a_i \log \frac{a_i}{b_i} \geq A \log \frac{A}{B},$$

where $A = \sum_i a_i$ and $B = \sum_i b_i$.

In physics, this is Gibbs inequality (A and B are partition functions).

The Gibbs Inequality

Let $\mathcal{E}_0(x)$ and $\mathcal{E}_1(x)$ be two Hamiltonians of a system. For a given β, let

$$P_i(x) = \frac{e^{-\beta\mathcal{E}_i(x)}}{Z_i}, \qquad Z_i = \sum_x e^{-\beta\mathcal{E}_i(x)}, \quad i = 0, 1.$$

Then,

$$0 \le D(P_0 \| P_1) = \mathbf{E}_0 \left\{ \ln \frac{e^{-\beta\mathcal{E}_0(X)}/Z_0}{e^{-\beta\mathcal{E}_1(X)}/Z_1} \right\}$$

or

$$\begin{aligned} \mathbf{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\} &\ge kT\ln Z_0 - kT\ln Z_1 \\ &= F_1 - F_0 \end{aligned}$$

Interpretation of $\mathbf{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\} \geq \Delta F$

- A system with Hamiltonian $\mathcal{E}_0(x)$ – in equilibrium $\forall\, t < 0$.
  Free energy $= -kT \ln Z_0$.
- At $t = 0$, the Hamiltonian jumps, by $W = \mathcal{E}_1(x) - \mathcal{E}_0(x)$: from $\mathcal{E}_0(x)$ to $\mathcal{E}_1(x)$ – by abruptly applying a force. Energy injected:
  $\mathbf{E}_0\{W\} = \mathbf{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\}$.
- New system, with Hamiltonian $\mathcal{E}_1$, equilibrates.
  Free energy $= -kT \ln Z_1$.

Gibbs inequality: $\mathbf{E}_0\{W\} \geq \Delta F$.

$$\mathbf{E}_0\{W\} - \Delta F = kT \cdot D(P_0 \| P_1)$$

$=$ dissipated work $= T\times$ entropy production (system $+$ environment) due to irreversibility of the abruptly applied force.

- Irreversibility of information loss $\leftrightarrow$ irreversibility of entropy production.
- Converse theorems $\leftrightarrow$ second law of thermodynamics.

Extension

Consider, more generally, $\mathcal{E}_\lambda(\boldsymbol{x}) = (1-\lambda)\mathcal{E}_0(\boldsymbol{x}) + \lambda\mathcal{E}_1(\boldsymbol{x})$, where the "force" λ varies from 0 to 1 according to

$$\lambda(t) = \sum_{i=0}^{n-1} (\lambda_{i+1} - \lambda_i) u(t - t_i), \quad \lambda_i \in (0,1)\ \forall i,\ \lambda_0 = 0,\ \lambda_n = 1.$$

$$\text{Dissipated work} = kT \cdot \sum_{i=0}^{n-1} (\lambda_{i+1} - \lambda_i) \mathbf{E}_{\lambda_i} \left\{ \ln \frac{P_0(\boldsymbol{X})}{P_1(\boldsymbol{X})} \right\} \geq 0.$$

A new information inequality inspired from physical considerations.

The ordinary divergence $D(P_0 \| P_1)$ is a special case where $n = 1$ and $t_0 = 0$.

DPT and the 2nd Law: Dynamical System Perspective

Let $\{X_t\}$ be a finite–alphabet Markov jump process:

$$\text{Pr}\{X_{t+\delta} = x' | X_t = x\} = W_{xx'}\delta + o(\delta) \qquad x' \neq x$$

Define

$$H(X_t) = -\sum_{x \in \mathcal{X}} P_t(x) \log P_t(x).$$

H–theorem: if $\{X_t\}$ obeys detailed balance $W_{xx'} = W_{x'x}$ then:

$$\frac{\text{d}H(X_t)}{\text{d}t} \geq 0.$$

Comments:

- $\{X_t\}$ corresponds to an isolated dynamical system $P_\infty(x) = 1/|\mathcal{X}|$.
- Discrete–time analogue – holds too.
- Similar to the 2nd law of thermo, but not precisely equivalent.

DPT and the Second Law (Cont'd)

What if $P_\infty(x)$ is not uniform?
Cover & Thomas (2006):

$$D(P_t \| P_\infty) = \sum_{x \in \mathcal{X}} P_t(x) \log \frac{P_t(x)}{P_\infty(x)} \quad \searrow$$

Indeed, for P_∞ uniform

$$D(P_t \| P_\infty) = \log |\mathcal{X}| - H(X_t).$$

- Detailed balance is not needed.
- Maximum entropy $\rightarrow$ minimum free energy.
- Characterizes monotonic convergence $P_t \rightarrow P_\infty$ in the divergence sense.

More generally, for P_t and P_t', two time–varying state distributions pertaining a given Markov process, $D(P_t \| P_t') \quad \searrow$ Cover & Thomas (2006).

DPT and the Second Law (Cont'd)

Kelly (1979): If Q is convex and P_∞ is a steady–state distribution

$$D_Q(P_\infty \| P_t) = \sum_{x \in \mathcal{X}} P_\infty(x) Q\left(\frac{P_t(x)}{P_\infty(x)}\right) \quad \searrow$$

for whatever P_t that evolves according to the Markov process. This covers both $D(P_\infty \| P_t)$ and $D(P_t \| P_\infty)$, but not $D(P_t \| P_t')$. To be handled soon...

Define $P_t(x, x') = P(X_0 = x, X_t = x')$ and $P_t'(x, x') = P(X_0 = x)P(X_t = x')$ then

$$D_Q(P_t \| P_t') = \sum_{x, x'} P_t(x, x') Q\left(\frac{P_t'(x, x')}{P_t(x, x')}\right) \quad \searrow$$

because here $D_Q(P_t \| P_t') = I_Q(X_0; X_t)$, and the above is the Ziv–Zakai DPT (1973) for the Markov chain $X_0 \to X_t \to X_{t+1}$.

DPT and the Second Law (Cont'd)

This monotonicity thm does not cover the entire picture. Can we put everything under one umbrella?

Yes, we can! including the 1975 Ziv–Zakai information measure.

Theorem: Let $\mu_t^0, \mu_t^1, \ldots, \mu_t^k$ be arbitrary measures that obey the Markov recursion

$$\mu_{t+1}^i(x) = \sum_{x'} \mu_t^i(x')P(x|x').$$

Then,

$$V_t \triangleq \sum_x \mu_t^0(x) Q\left(\frac{\mu_t^1(x)}{\mu_t^0(x)}, \ldots, \frac{\mu_t^k(x)}{\mu_t^0(x)}\right) \quad \searrow$$

The 1975 ZZ DPT for the Markov chain $X_0 \to X_t \to X_{t+1}$ is obtained for $\mu_t^0(x, x') = P(X_0 = x, X_t = x')$.

- Temporal monotonicity of gen. info measures for Markov processes $\leftrightarrow$ gen. DPT for systems with a spatial Markov structure.
- Both are special cases of the above theorem.

A New Perspective on the 1973 Ziv–Zakai DPT

While the 1973 Ziv–Zakai info measure is

$$I_Q(X;Y) = \sum_{x,y} P(x,y) Q\left(\frac{P(x)P(y)}{P(x,y)}\right),$$

one can use any μ_0 and μ_1 (satisfying the Markov relations) and define

$$I_Q(X;Y) = \sum_{x,y} \mu_0(x,y) Q\left(\frac{\mu_1(x,y)}{\mu_0(x,y)}\right),$$

Consider μ's of the form: $\mu(x,y) = s_0 P(x,y) + \sum_{x_i \in \mathcal{X}} s_i P(x) P(y|x = x_i), \;\; s_0, \; s_i \geq 0$

For example, $I_Q(X;Y) = \sum_{x,y} [P(x,y) + sP(x)P(y)] \cdot Q\left(\frac{P(x)P(y)}{P(x,y) + sP(x)P(y)}\right)$

satisfies a DPT for every $s \geq 0$. $s = 0 \;\; \rightarrow$ 1973 Ziv–Zakai info measure.

Even for the 1973 ZZ DPT (univariate Q), we have added a degree of freedom. Important since only few functions Q, are easy to work with.

Example

Source U and the reconstruction V are uniform over $\{0, 1, \ldots, K-1\}$.

$$d(u, v) = \begin{cases} 0 & v = u \\ 1 & v = (u+1) \text{ mod } K \\ \infty & \text{elsewhere} \end{cases}$$

Channel: clean L–ary channel.

For $Q(z) = -\sqrt{z}$, we obtain

$$I_Q(U; V) = -\sum_{u,v} P(u)P(v)\sqrt{s + \frac{P(v|u)}{P(v)}}.$$

Example (Cont'd)

Applying the DPT $R_Q(d) \leq C_Q$ (for a given s), we obtain the lower bound

$$d \geq d_s.$$

For $s = 0$ (ZZ '73), we have:

$$d_0 = \frac{1}{2} - \frac{1}{2}\sqrt{2\theta - \theta^2},$$

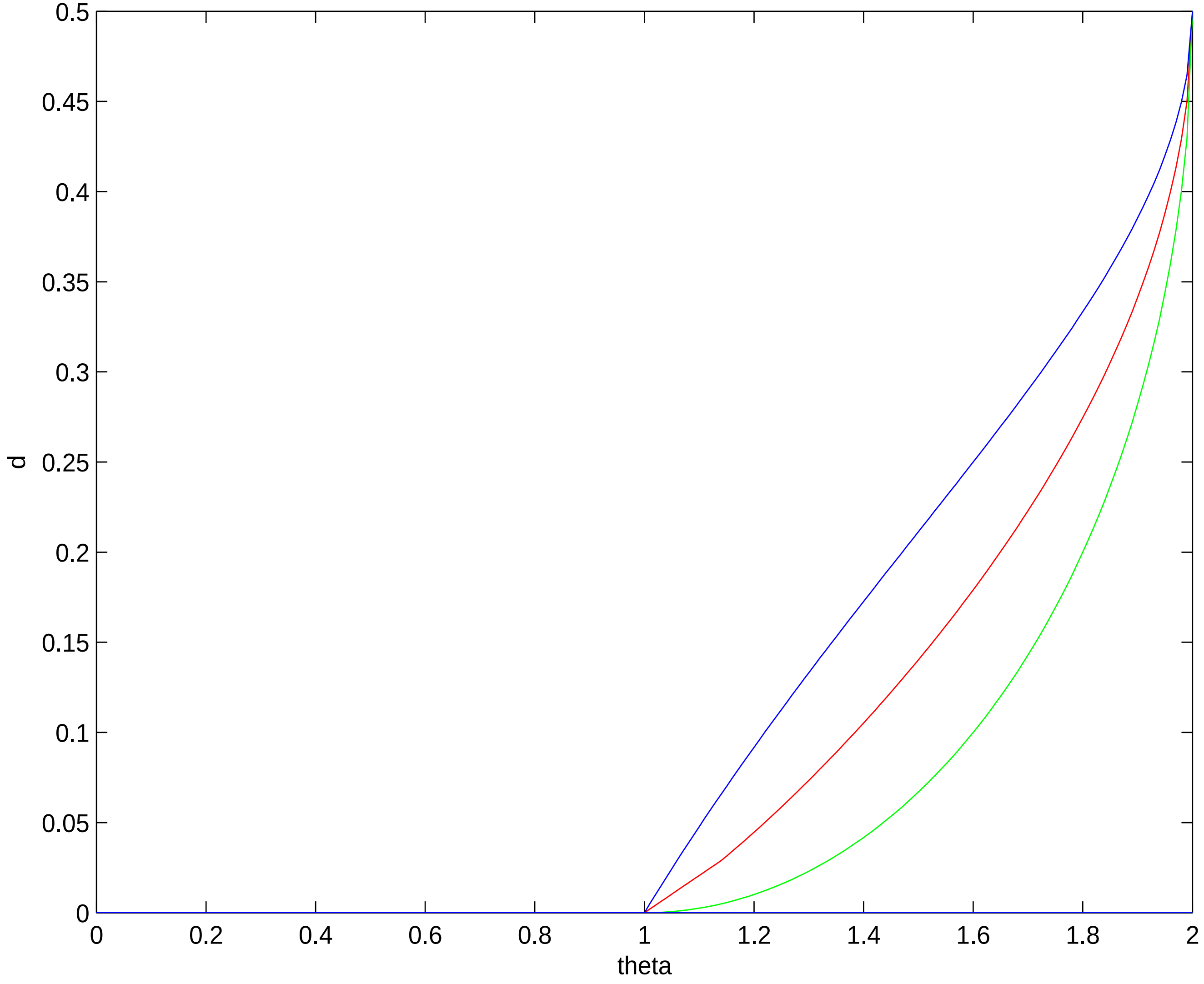
where $\theta \triangleq K/L$.

For $s \to \infty$,

$$d_\infty = \frac{1}{2} - \frac{1}{2\theta}\sqrt{2\theta - \theta^2},$$

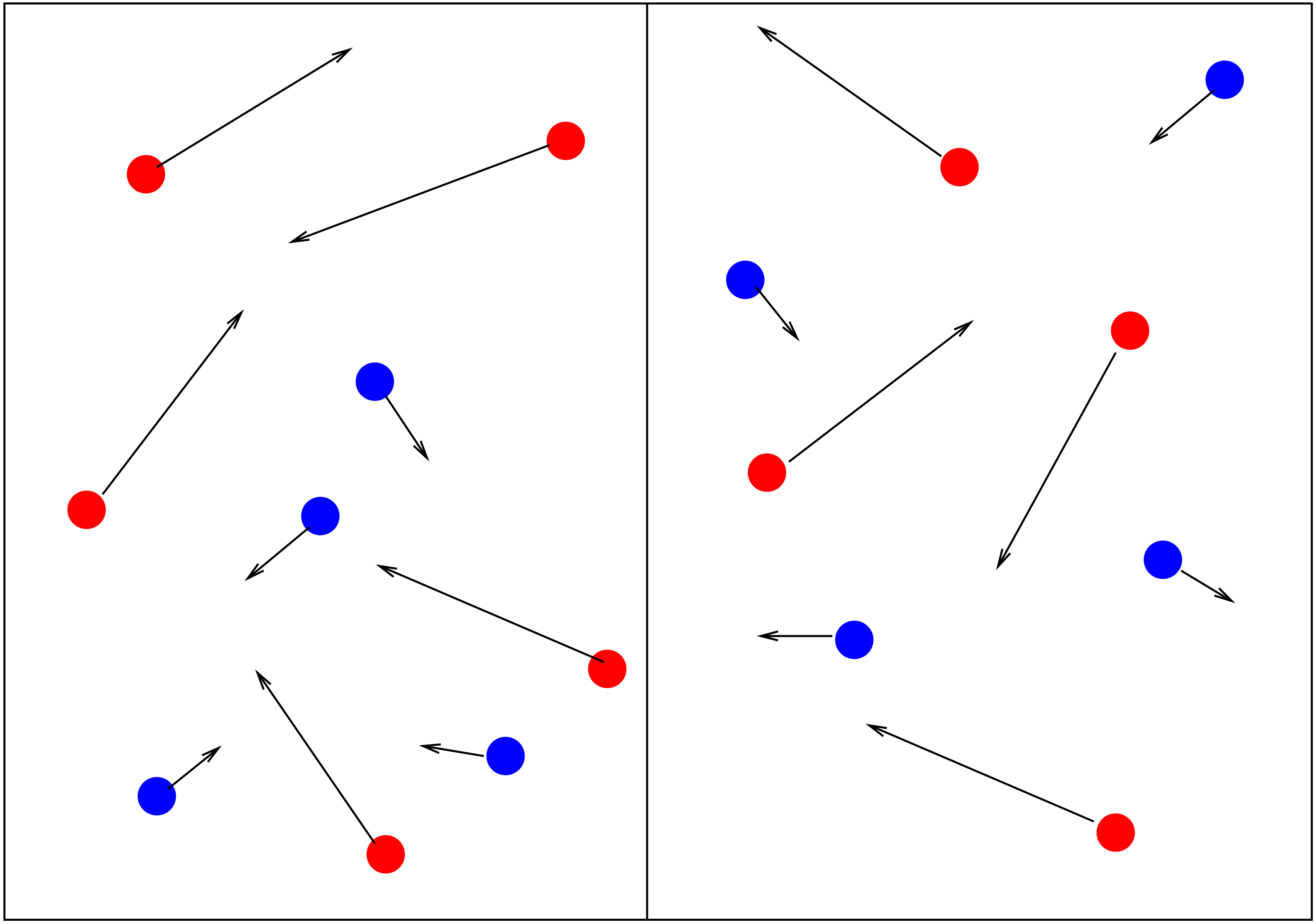
which is larger than d_0 for all $1 < \theta < 2$.

The Shannon bound: $d_{\text{Shannon}} = h^{-1}(\log \theta)$ is in between.

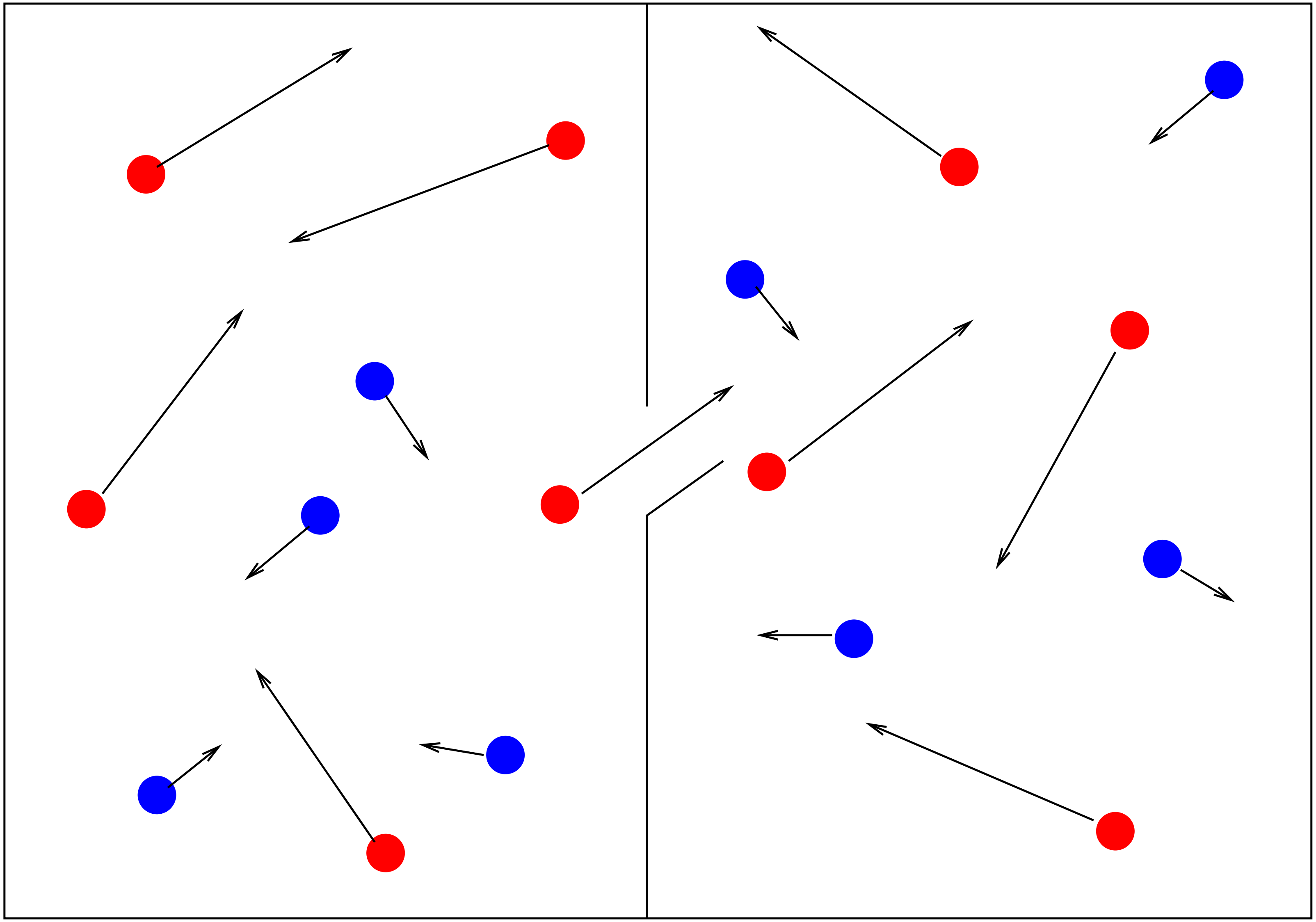


$s = 0$, $s = \infty$, Shannon bound

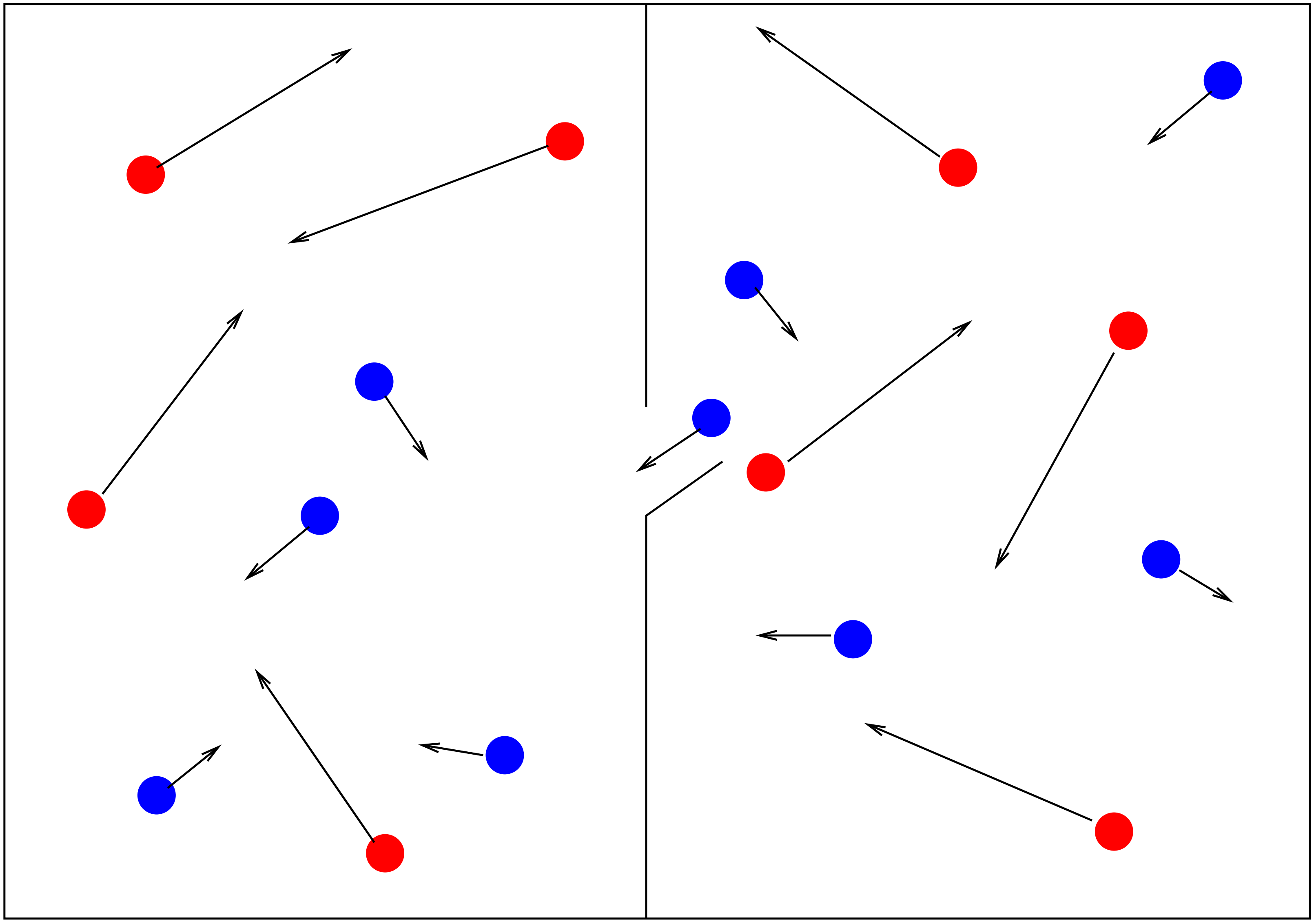
The Maxwell Demon



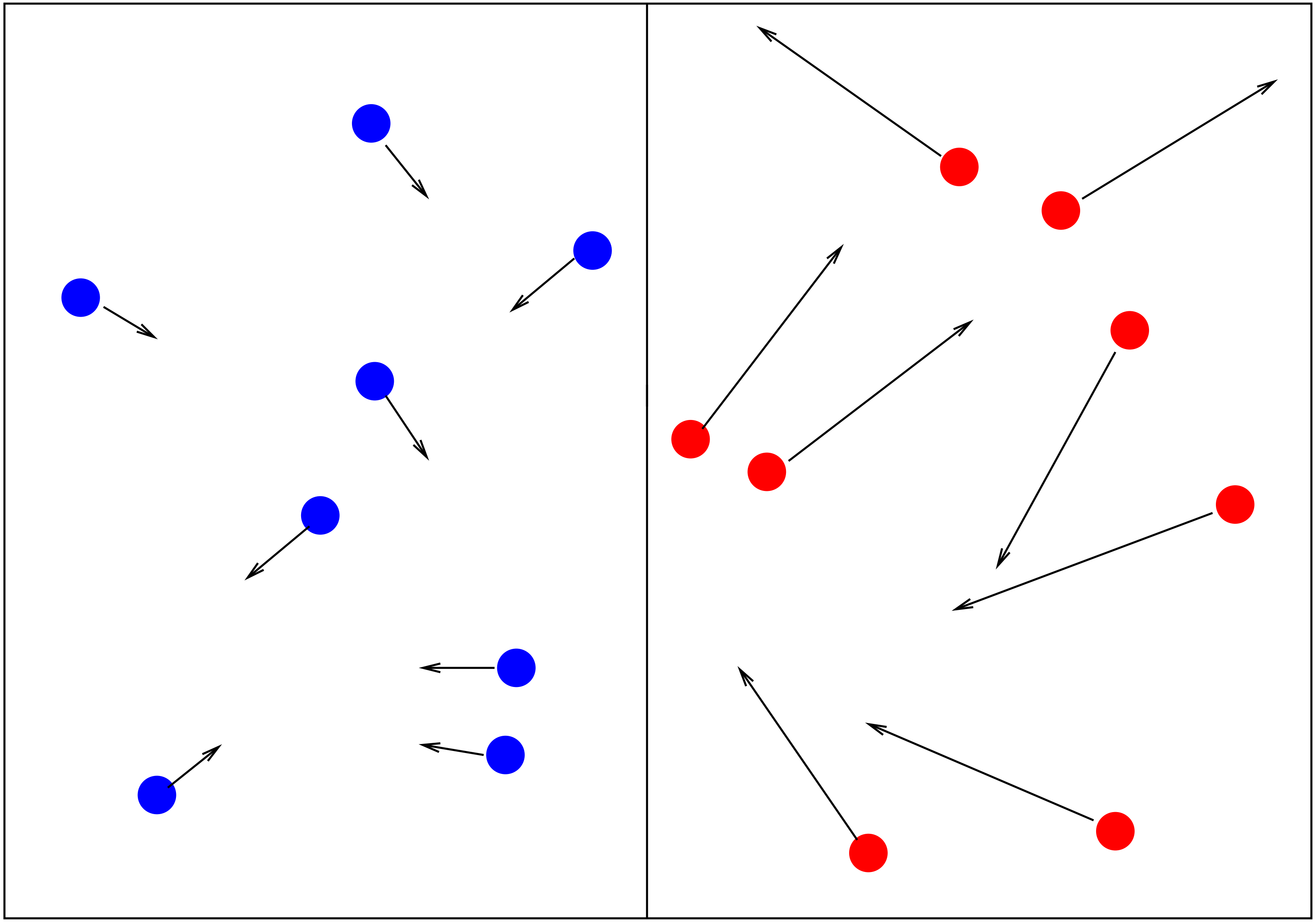
The Maxwell Demon



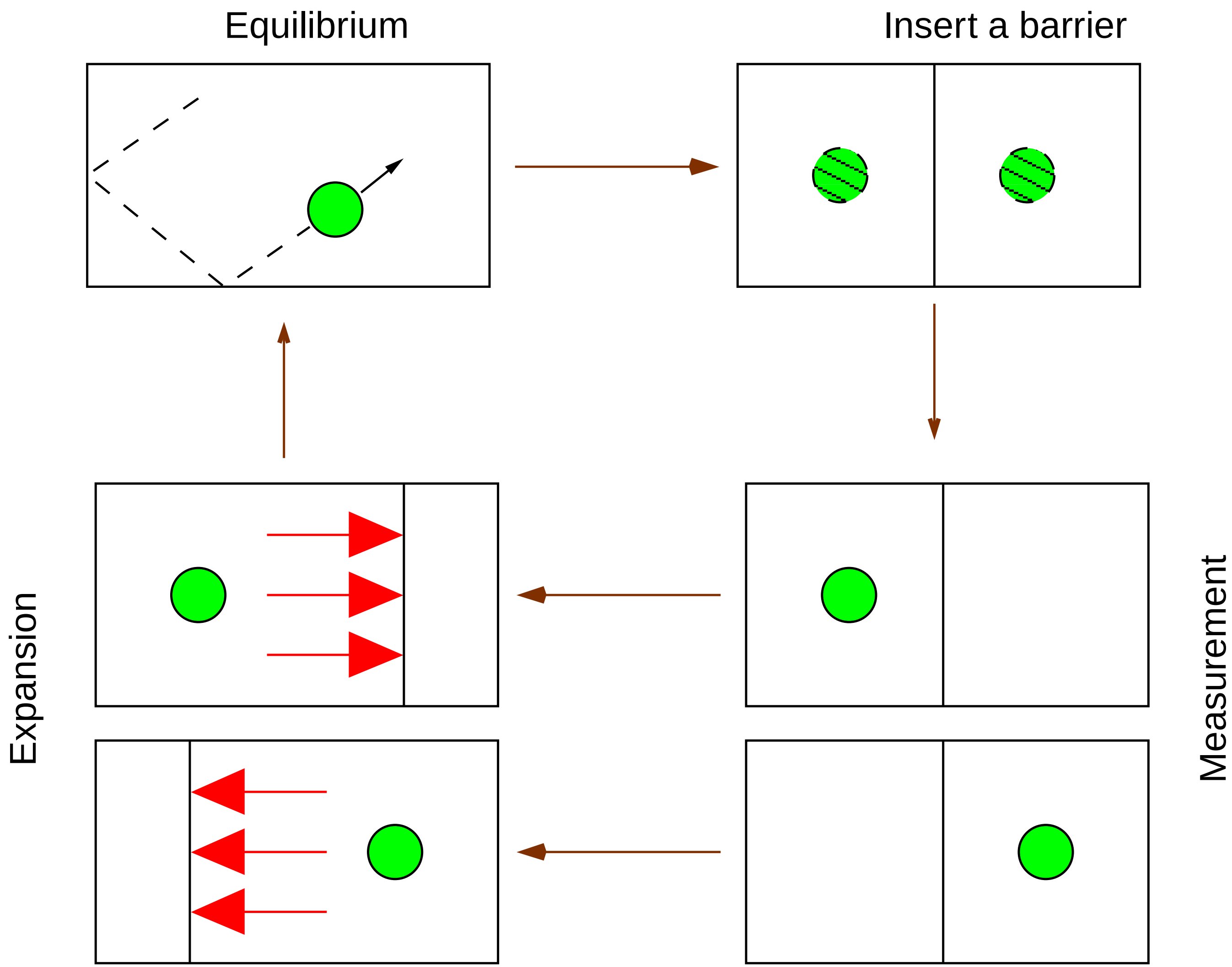
The Maxwell Demon



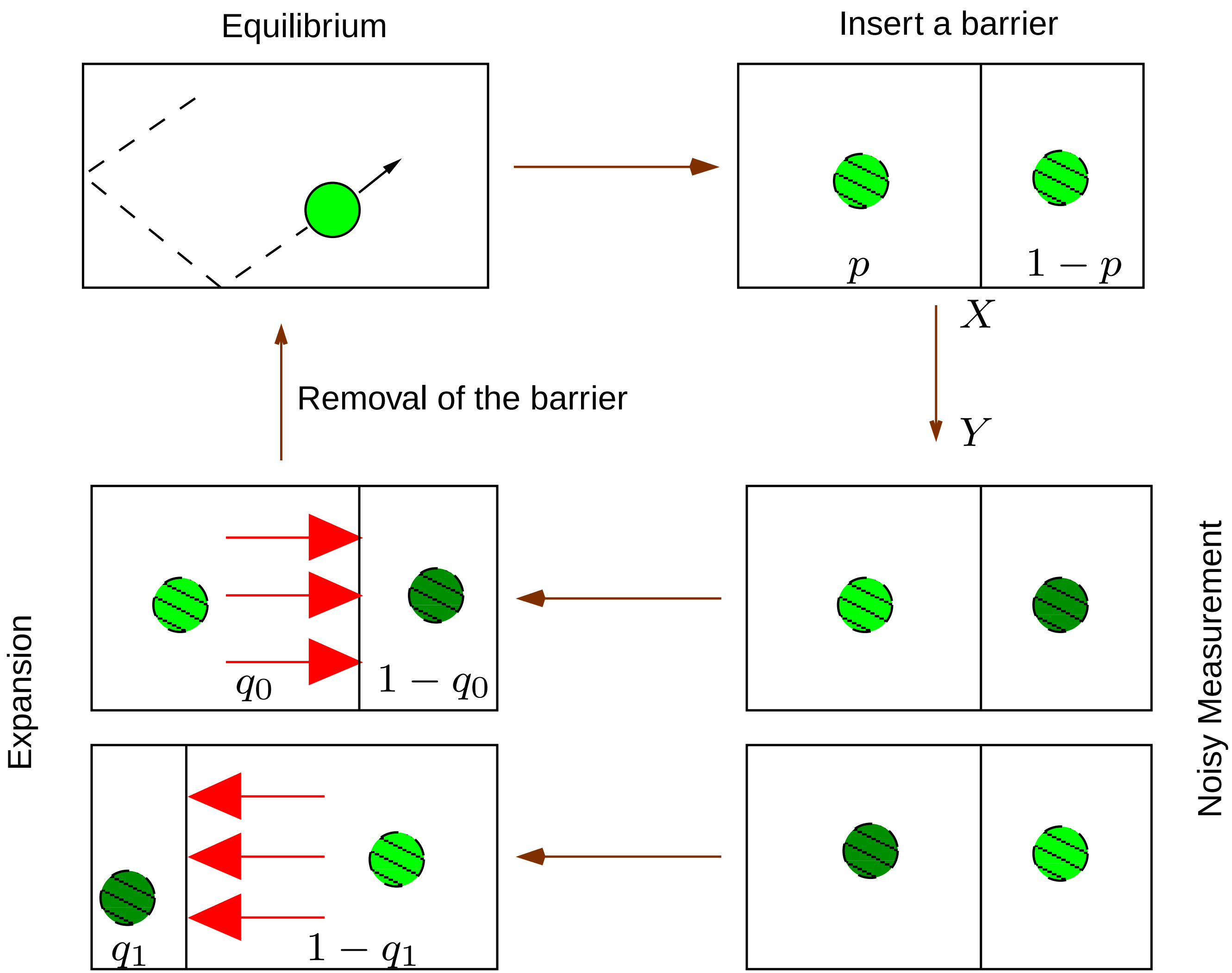
The Maxwell Demon



The Szilard Engine



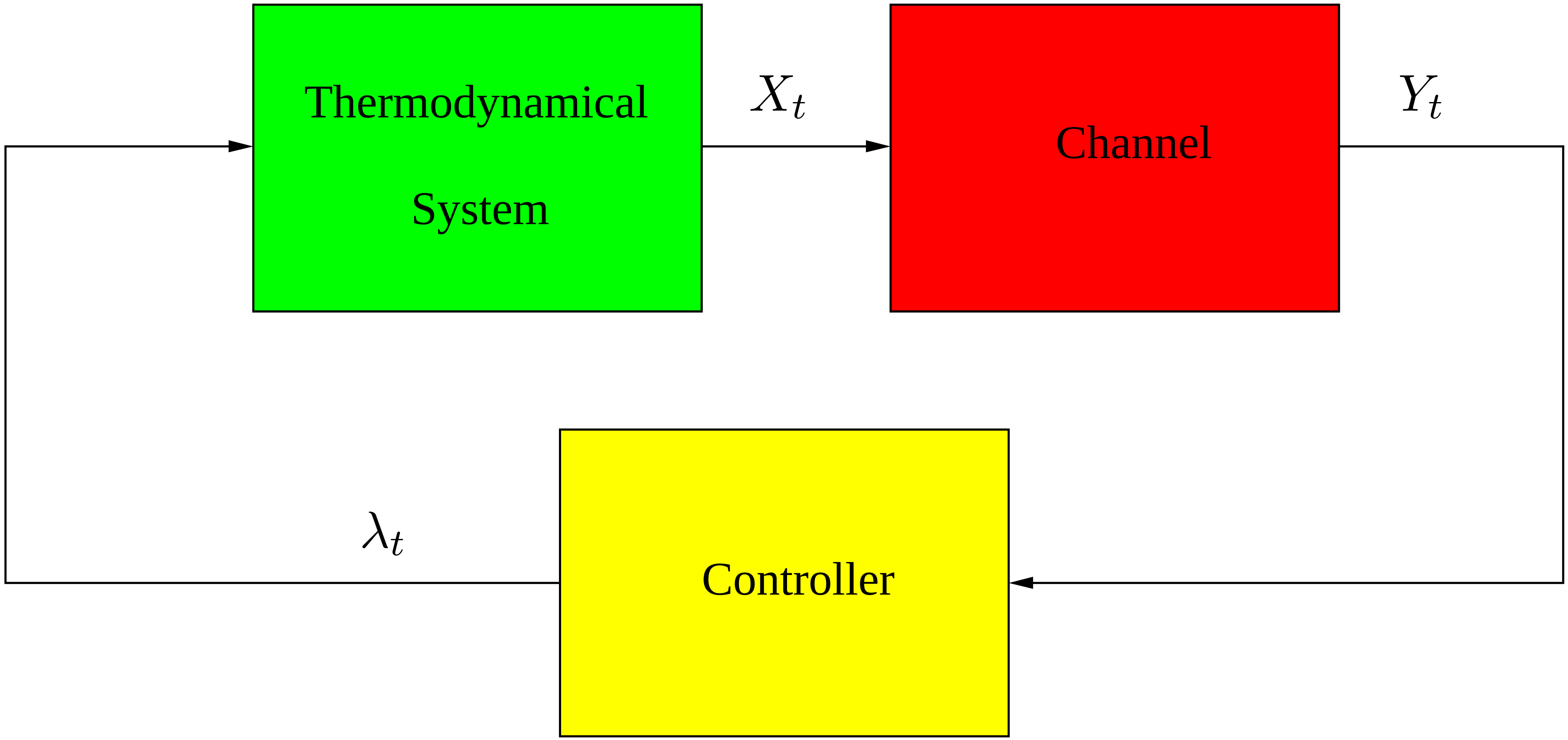
Generalized Szilard Engine (Sagawa & Ueda, 2011)



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

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

System with Feedback Control – Directed Information



Sagawa and Ueda (2008–2011) : Extracted work $\leq -\Delta F + kT \cdot I(X^n \to Y^n)$.

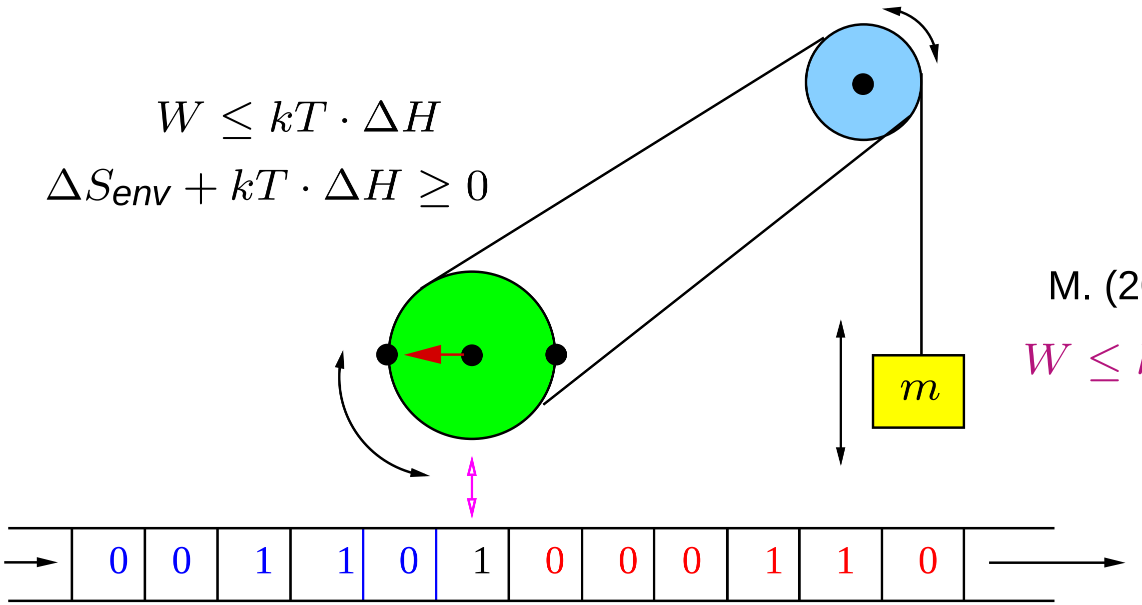
Feedback is linked to directed information in physics too!

Physical Systems with an Informational Ingredient

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

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

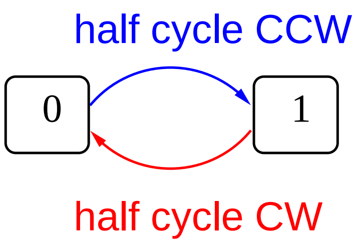
$$\Delta S_{\text{env}} + kT \cdot \Delta H \ge 0$$



M. (2015):

$$W \le kT \cdot F[\textit{LZ complexity}]$$

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



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: $\Delta S_{\text{dev}} + \Delta S_{\text{heat-res}} + \Delta S_{\text{info-res}} \ge 0$.

Phase Transitions

In physics: an abrupt change in the “behavior” of a physical system upon a small change in a parameter (temperature, magnetic field,...).

- Occurs with “strong” interactions: $\mathcal{E}(\boldsymbol{x}) = \sum_i \mathcal{E}(x_i) + \sum_{i,j} \mathcal{E}(x_i, x_j)$.
- Mathematically: a discontinuity in a derivative of $\phi(\beta)$.
- An asymptotic concept – in the thermodynamic limit $N \to \infty$.
- Xmpls: magnetization below T_c; glassy ϕ–transition; freezing/boiling.

In communication systems: an abrupt change in the behavior/performance ... upon a small change in a parameter (SNR, bandwidth, ...).

- Occurs in coded coded systems and in non–linear modulation.
- A discontinuity in performance, e.g., $P_e \approx 0 \ \leftrightarrow \ P_e \approx 1$.
- An asymptotic concept - in the limit of long blocks $N \to \infty$.
- Xmpls: un/reliable comm. as $R \uparrow C$; est. threshold effects, comp. sensing.

Part II

Relations in the Technical Level

Mapping Models Between Physics and IT

Customary model of disordered magnetic materials (spin arrays)

$$\mathcal{E}(\boldsymbol{s}) = -\sum_i B_i s_i - \sum_{i,j} J_{ij} s_i s_j \qquad \boldsymbol{s} \in \{-1,+1\}^N$$

B_i – (random) local magnetic fields; J_{ij} – (random) coupling coefficients.
Let $\mathcal{C} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$ be a code for $P(\boldsymbol{y}|\boldsymbol{x})$ and let

$$\begin{aligned} P(\boldsymbol{x}|\boldsymbol{y}) &= \frac{\frac{1}{M} P(\boldsymbol{y}|\boldsymbol{x})}{\frac{1}{M}\sum_{\boldsymbol{x}'\in\mathcal{C}} P(\boldsymbol{y}|\boldsymbol{x}')} \qquad \boldsymbol{x} \in \mathcal{C} \\ &= \frac{\exp\{-\overbrace{\ln[1/P(\boldsymbol{y}|\boldsymbol{x})]}^{\mathcal{E}_y(\boldsymbol{x})}\}}{\sum_{\boldsymbol{x}'\in\mathcal{C}} \exp\{-\ln[1/P(\boldsymbol{y}|\boldsymbol{x}')]\}} \end{aligned}$$

Or more generally, $P_\beta(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{\exp\{-\beta\mathcal{E}_y(\boldsymbol{x})\}}{Z_y(\beta)}$

Example: for a BPSK AWGNC: $\mathcal{E}_y(\boldsymbol{x}) \propto \|\boldsymbol{y} - \boldsymbol{x}\|^2 \sim -c\sum_i x_i y_i$.
x_i – spin; y_i – local magnetic field.
For $\mathcal{E}_y(\boldsymbol{x}) \propto \|\boldsymbol{y} - A\boldsymbol{x}\|^2 \sim -c\sum_i x_i \tilde{y}_i + \sum_{i,j} J_{ij} x_i x_j$. Sourlas (1989, 1994).

Performance Analysis

Let $\mathcal{C}$ be a random code and consider the calculation of

$$\mathbf{E}\{I(\boldsymbol{X};\boldsymbol{Y})\} = \underbrace{\mathbf{E}\{H(\boldsymbol{Y})\}}_{\text{difficult}} - \underbrace{\mathbf{E}\{H(\boldsymbol{Y}|\boldsymbol{X})\}}_{\text{easy for additive channels}}$$

$$\begin{aligned}
\mathbf{E}\{H(\boldsymbol{Y})\} &= -\mathbf{E}\{\log P(\boldsymbol{Y})\} = -\mathbf{E}\left\{\log\left[\frac{1}{M}\sum_{m=0}^{M-1} P(\boldsymbol{y}|\boldsymbol{X}_m)\right]\right\} \\
&= \log M - \mathbf{E}\log \underbrace{\sum_{m=0}^{M-1} \exp[-\mathcal{E}_y(\boldsymbol{X}_m)]}_{\text{partition function}} \qquad (1)
\end{aligned}$$

$\{\mathcal{E}_y(\boldsymbol{X}_m)\}_{m=0}^{M-1}$ are i.i.d. RV's.

Analogous to the Random Energy Model (Montanari 2001).

Also, many (Gallager–style) performance bounds include expressions like

$$\sum_m P^\beta(\boldsymbol{y}|\boldsymbol{X}_m) = Z_y(\beta) \quad \text{for some } \beta$$

The Random Energy Model (REM)

The REM (Derrida 1980s) is a toy model obtained as limit of strong disorder, where

$$\mathcal{E}(\boldsymbol{s}) \sim \mathcal{N}(0, NJ^2/2) \quad \text{i.i.d.}$$

It exhibits a ϕ-transition since $\Omega(E) = |\{\boldsymbol{s} : \ \mathcal{E}(\boldsymbol{s}) \approx N\epsilon\}|$ is a binomial RV with $2^N = e^{N \ln 2}$ trials and probability of 'success' $\sim e^{-N\epsilon^2/J^2}$:

$$\text{Typically: } \Omega(E) \doteq \begin{cases} \exp\{N[\ln 2 - \epsilon^2/J^2]\} & |\epsilon| \le J\sqrt{\ln 2} \\ 0 & |\epsilon| > J\sqrt{\ln 2} \end{cases}$$

The entropy jumps to $-\infty$ for high energies!

Thus, $\phi(\beta)$ is non–smooth – a phase transition: below a certain temperature the system freezes, i.e., dominated by a sub–exponential number of ground–state $\{\boldsymbol{s}\}$ – glassy phase.

A Few Words on Useful Analysis Methods

While the REM can be analyzed rigorously, the calculation of $\mathbf{E} \log Z(\beta)$ for random $Z(\beta)$ is difficult in general. A very popular (but non–rigorous) technique: the replica method – based on the identity

$$\mathbf{E} \log Z = \lim_{n \to 0} \frac{\mathbf{E} Z^n - 1}{n}.$$

Works in many cases, but not always..

Another useful tool (in general): saddle point integration:

$$\int_{\mathcal{P}} g(z) e^{N f(z)} \mathrm{d}z \sim e^{i\theta} \sqrt{\frac{2\pi}{N |f''(z_0)|}} g(z_0) e^{N f(z_0)}$$

where z_0 is a saddle point ($f'(z_0) = 0$) and $\theta = (\pi - \arg\{f''(z_0)\})/2$.
If the integrand includes a non-analytic function (e.g., the Dirac delta function or the unit step function) then a common trick is to present it as the inverse transform of a 'nice' function

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega x} \mathrm{d}\omega$$ and switch the order of the integrals.

REM–Based Analysis of Correct Decoding Exponent

For the BSC(p)

$$P(\boldsymbol{y}|\boldsymbol{x}) = (1-p)^N e^{-Bd(\boldsymbol{x},\boldsymbol{y})}, \qquad B = \ln\frac{1-p}{p}$$

$$\mathcal{E}_y(\boldsymbol{x}) = d(\boldsymbol{x},\boldsymbol{y})$$

$$\Omega_y(N\delta) = |\{\boldsymbol{x}:\ \boldsymbol{x}\in\mathcal{C},\ \ d(\boldsymbol{x},\boldsymbol{y}) = N\delta\}|$$

Binomial RV with e^{NR} trials and probability of success $\exp\{-N[\ln 2 - h_2(\delta)]\}$.

$$\begin{aligned}
\overline{P_c} &= \mathbf{E}\left\{\frac{1}{M}\sum_{\boldsymbol{y}}\max_m P(\boldsymbol{y}|\boldsymbol{X}_m)\right\} \\
&= \mathbf{E}\left\{\frac{1}{M}\sum_{\boldsymbol{y}}\lim_{\beta\to\infty}\left[\sum_{m=0}^{M-1}P^\beta(\boldsymbol{y}|\boldsymbol{X}_m)\right]^{1/\beta}\right\} \\
&= \frac{1}{M}\sum_{\boldsymbol{y}}\lim_{\beta\to\infty}\mathbf{E}\left\{Z(\beta|\boldsymbol{y})^{1/\beta}\right\}
\end{aligned}$$

The Correct Decoding Exponent (Cont'd)

$$\begin{aligned}\mathbf{E}\{Z(\beta|\boldsymbol{y})^{1/\beta}\} &= \mathbf{E}\left\{\left[(1-p)^{N\beta}\sum_\delta \Omega_{\boldsymbol{y}}(N\delta)e^{-\beta BN\delta}\right]^{1/\beta}\right\} \\ &\doteq (1-p)^N\mathbf{E}\left\{\sum_\delta \Omega_{\boldsymbol{y}}^{1/\beta}(N\delta)e^{-BN\delta}\right\} \\ &= (1-p)^N\sum_\delta \mathbf{E}\left\{\Omega_{\boldsymbol{y}}^{1/\beta}(N\delta)\right\}\cdot e^{-BN\delta}\end{aligned}$$

$$\mathbf{E}\left\{\Omega_{\boldsymbol{y}}^{1/\beta}(N\delta)\right\} = \begin{cases} \exp\{N[R+h(\delta)-\ln 2]\} & \delta \le \delta_{GV}(R) \text{ or } \delta \ge 1-\delta_{GV}(R) \\ \exp\{N[R+h(\delta)-\ln 2]/\beta\} & \delta_{GV}(R) < \delta < 1-\delta_{GV}(R)\end{cases}$$

This gives the correct exponential behavior, unlike the traditional use of Jensen's inequality:

$$\mathbf{E}Z^{1/\beta} \le (\mathbf{E}Z)^{1/\beta}.$$

Tighter bounds on random coding and expurgated error exponents!

Statistical Physics of Signal Estimation

In a joint work with Guo and Shamai (2010), stat–mech methods were applied on the I–MMSE relation:

$$\frac{\mathrm{d}I(\boldsymbol{X};\sqrt{snr}\boldsymbol{X}+\boldsymbol{N})}{\mathrm{d}\, snr}=\frac{1}{2}\mathrm{mmse}(\boldsymbol{X}|\sqrt{snr}\boldsymbol{X}+\boldsymbol{N}),\quad \boldsymbol{N}\sim\mathcal{N}(0,I)$$

to compute MMSE and to relate threshold effects (in estimation) to phase transitions (in physics).

An alternative approach: For the purpose of evaluating the MMSE (using stat-mech methods), more direct relations can be used: Given $P(\boldsymbol{x},\boldsymbol{y})$, $\boldsymbol{x}\in\mathbb{R}^N$, $\boldsymbol{y}\in\mathbb{R}^K$:

$$Z(\boldsymbol{y},\boldsymbol{\lambda})=\sum_{\boldsymbol{x}}\exp\{\boldsymbol{\lambda}^T\boldsymbol{x}\}P(\boldsymbol{x},\boldsymbol{y})$$

$$\hat{\boldsymbol{x}}=\mathbf{E}\{\boldsymbol{X}|\boldsymbol{y}\}=\nabla_{\boldsymbol{\lambda}}\ln Z(\boldsymbol{y},\boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0};\quad \mathrm{Cov}\{(\boldsymbol{X}-\hat{\boldsymbol{X}})\}=\mathbf{E}\left\{\nabla^2_{\boldsymbol{\lambda}}\ln Z(\boldsymbol{Y},\boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0}\right\}$$

Example: Codeword Sent Over an AWGN Channel

Channel input: $M = e^{NR}$; $\mathcal{C} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{M-1}\}$;
$\boldsymbol{x}_i \sim$ Surf{sphere of radius $\sqrt{NP}$}.

AWGN channel:

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{W}; \quad \boldsymbol{W} \sim \mathcal{N}(0, \sigma^2 I)$$

Partition function:

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) = \sum_{\boldsymbol{x} \in \mathcal{C}} e^{-NR} \cdot \exp\{-\|\boldsymbol{y} - \boldsymbol{x}\|^2/(2\sigma^2) + \boldsymbol{\lambda}^T \boldsymbol{x}\}.$$

Analyzable using the above REM technique.
The MMSE undergoes a phase transition:

$$\text{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = \begin{cases} 0 & R < C \\ \frac{P\sigma^2}{P+\sigma^2} & R > C \end{cases}$$

Huleihel and M. (2014): extension to $\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{W}$ (H – filter) + mismatch.
Huleihel and M. (2014): $\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{W}$, where $\boldsymbol{X}$ is sparse and H is random.

Other Applications of This and Related Techniques

- Ordinary random coding exponents (2009).
- Source coding w. hierarchical ensembles (2009); tree codes (2010).
- JSC coding – source–channel equilibrium (2009).
- IFC (w. Etkin and Ordentlich – 2009); revisited (w. Huleihel – 2015).
- Estimation of codewords in AWGN (w. Guo and Shamai – 2010).
- Broadcast channels (w. Kaspi – 2011); improved (2014).
- Erasure/list decoding (w. Somekh–Baruch – 2011);
- Implications of the above on info rates (w. Huleihel and Shamai – 2014).
- Expurgated exponents (w. Scarlett *et al.* – 2014).
- Ordinary and erasure/list decoding for S–W codes (2014).
- Codeword or noise? (w. Weinberger – 2014).
- The wiretap channel with optimal decoding (2014).
- List decoding: random coding and expurgated bounds.
- Universal erasure/list decoding (w. Weinberger and Huleihel – 2014).
- Statistical physics of random binning (2015).

Summary

- Relations in the conceptual level:
  - Info measures $\leftrightarrow$ phys. quantities
  - Coding theorems $\leftrightarrow$ physical laws.
  - Information plays a role in physics.
  - Phase transitions
- Relations in the technical level:
  - Mapping models.
  - Replica mtd + more (+ cavity mtd, fin.–size scaling, interpolation,...).
  - Rigorous REM analysis and error exponents.
  - MMSE via I–MMSE relations and gradients of partition functions.
- Bounds inspired by stat. phys. ($R(D)$, DPT, err. exp., new info ineq).
- Stat. physics has much more to offer us than just the replica method!
- We have just scratched the surface in this talk ...