Free Energy Transduction II: detailed balance, fluxes, cycles, entropy production.[1]

Pankaj Mehta

April 2, 2021

In these notes, we talk about how cell tranduce free energy by exploiting non-equilibrium steady-states.

[1] Optional Reading: Terrence Hill's magnificent cheap and short books *Free Energy Transduction in Biology*. We will especially focus on Chapter 4. We will also use the review by Hong Qian *Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological Functions* from Annu. Rev. Phys. Chemistry 2007 58:113-42. Also discussed relevant Nelson sections scattered through out the *Biological Physics* book.

One of the central things that cells must do is be able to transduce Free energy and extract it for useful purposes. The way that the cell usually does this is by maintaining a pool of high energy "cofactors" such ATP. For example, ATP can undergo hydrolosis into ADP and release the chemical energy that has been stored in the high-energy phosphoanhydride bonds. This process releases about $\Delta G \approx -20 k_B T$ to $-30 k_B T$ or $-35 kJ/mol$ to $-50 kJ/mol$ of energy depending on the cells exact conditions. The cell essentially maintains ATP at a high non-equilibrium concentration by continuously generating ATP and then uses this energy by coupling ATP hydrolosis to less thermodynamically favorable reactions. To get a better understanding the numbers involved in the please read the discussion and look at the charts at `http://book.bionumbers.org/how-much-energy-is-released-in-atp-hydrolysis/` in detail! I will not reproduce it in these notes.

Open Systems- A simple example

How can we understand the thermodynamics of such a system?[2]. This is very far from the kind of things we usually learn about in most of our thermodynamics and statistical mechanics classes which tend to focus almost entirely on equilibrium processes. So how can we think about these kind of processes. We can actually start building on the foundation relating kinetics to thermodynamics that we have already been discussing. Let us start by reviewing and reformulating equilibrium thermodynamics.

[2] This follows Qian *Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological Functions* from Annu. Rev. Phys. Chemistry 2007 58:113-42 and this very nice discussion by Daniel Zuckerman `http://www.physicallensonthecell.org/chemical-potential`.

Reviewing some equilibrium thermodynamics

We will start by considering an ideal, non-interacting gas X in some volume V with N particles. We know that the partition function is proportional to

$$Z_X(N, V, T) \propto \lambda_{deBroglie}^{-3N} V^N / N! \tag{1}$$

Hence, we know that to leading order in N, the free energy goes like

$$F_X(N,V,T) \propto -k_B T N \log \frac{N}{V/\lambda^3_{deBroglie}} - N k_B T, \tag{2}$$

where we have uses Stirling's approximation $logN! \sim N \log N - N$. From, this we conclude that the chemical potential can be written as

$$\begin{aligned} \mu_X = \frac{\partial G_X}{\partial N} \simeq \frac{\partial F_X}{\partial N} &= k_B T N \log \frac{N}{V/\lambda^3_{deBroglie}} \\ &= \mu^0_X + RT \log [X], \end{aligned} \tag{3}$$

and in the last line we have introduced a standard concentration $1M$ and hence introduced constant $R = k_B N_A$ where $N_A$is Avagadro's number.

A simple reaction

Consider a simple reaction of the form

$$A + D \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B + E \tag{4}$$

Let us think about the Gibbs free energy associated with this reaction when it is in equilibrium. From our considerations last time, we know that we can think of this as an effective "two-state" system with one state being the "products" and the other state being the "reactants". At equilibrium, we have detailed balance so that

$$-RT \log \frac{[B]_{eq}[E]_{eq}}{[A]_{eq}[D]_{eq}} = -\log K_{eq}, \tag{5}$$

where we have defined the equilibrium constant

$$K_{eq} = \frac{k_+}{k_-}. \tag{6}$$

Now, we can make use of our formulas for the chemical potential (3) to derive a useful identity that will allow us to generalize many things to nonequilibrium open systems. Notice that at equilibrium, by definition the chemical potentials of the reactions and potentials must be equal (this is the analogues of the statement that temperatures of systems at equilibrium must be equal). Thus, we have that

$$\begin{aligned} &\mu^{eq}_B + \mu^{eq}_E = \mu^{eq}_A + \mu^{eq}_D \\ &\mu^0_B + RT \log [B] + \mu^0_E + RT \log [E] = \mu^0_A + RT \log [A] + \mu^0_D + RT \log [D] \\ &\mu^0_B + \mu^0_E - \mu^0_A - \mu^0_D = -RT \log \frac{[B]_{eq}[E]_{eq}}{[A]_{eq}[D]_{eq}} \\ &\mu^0_B + \mu^0_E - \mu^0_A - \mu^0_D = \Delta G^{eq} = -RT \log K_{eq} \end{aligned} \tag{7}$$

Generalizing to Non-equilibrium Steady-states and open systems

Revisiting our simple reaction

Now consider the same reaction out of equilibrium so that the system is not in detailed balance. We can ask how much will the Gibbs free energy change every time the reaction proceeds in the forward direction. If we denote the number of molecules of type X by N_X then by definition

$$\begin{aligned}\Delta G &= G(N_A - 1, N_D - 1, N_B + 1, N_E + 1) - G(N_A, N_D, N_B, N_E) \\ &= -\frac{\partial G}{\partial N_A} - \frac{\partial G}{\partial N_D} + \frac{\partial G}{\partial N_B} + \frac{\partial G}{\partial N_E} \\ &= \mu_B + \mu_E - \mu_A - \mu_D \end{aligned} \tag{8}$$

Thus, as expected, the change in free energy per molecular conversion is just the difference of the chemical potentials. Now we, can make use of Eq. 3 to get

$$\Delta G = RT \log \frac{[B][E]}{[D][A]} - \mu_B^0 + \mu_E^0 - \mu_A^0 - \mu_D^0. \tag{9}$$

Now we can substitute Eq. 7 to get a series of interesting, equivalent expressions for the non-equilibrium change in Free Energy:

$$\begin{aligned}\Delta G &= \mu_B + \mu_E - \mu_A - \mu_D \\ &= RT \log \frac{[B][E]}{[D][A]} - RT \log K_{eq} && (10) \\ &= RT \log \frac{k_-[B][E]}{k_+[D][A]} && (11) \\ &= RT \log \frac{J_+}{J_-} && (12)\end{aligned}$$

It is worth emphasizing these expressions measure the change in free energy in an *open, nonequilibrium* system. Furthermore, notice as expected, at equilibrium ΔG is zero. Furthermore, the more out-of-equilibrium (i.e. bigger ΔG), the more one drives the reaction in the forward direction, overcoming thermal fluctuations that want to drive the system in the opposite direction (Le Chatelier's principle). We will see examples of this in various settings in the next few lectures from Molecular machines.

Relating to Biology

In biology, this is exactly the ΔG we exploit when we use high-energy cofactors such as ATP. By maintaining the ATP to ADP concentrations artificially high, we can now extract an amount of ΔG equal to

the expression above. Furthermore, my coupling these high-energy reactions to other reactions we can do things like perform work or process information. In particular, the key is to create cyclic processes (just like engines in thermodynamics) and each time we perform a cycle, we couple to these high-energy co-factors. This allows us to drive the cycles in one direction, In fact this is how things like molecular machines work, as well as all information processing tasks that involve erasing and writing memories (Landauer's principle). These can all be thought of as cyclic processes that consume energy.

General Expressions and analogy to Circuits

Note that given a reversible chemical reaction r, we can always associate a chemical potential or equivalently change in Gibbs energy with that reaction in terms of the forward to the backward flux.

$$\Delta G_r = \Delta \mu_r = RT \log J_+^r / J_-^r \tag{13}$$

This follows from a straight forward (nearly trivial) generalization of the calculation done above [3]. In terms of electric circuits, this is exactly the analogue of the "voltage drop" across each reaction. In other words, if the particles were charged, this is how much voltage (potential) would be lost across the reaction. We can also associate a net flux with this reaction

[3] See HW Problem 1

$$J^r = J_+^r - J_-^r. \tag{14}$$

This is the analogue of the current of the electrical circuit.

Finally, we know that the power dissipated in an electric circuit just given by $P = IV$. In our chemical reactions, the power dissipation then takes the form

$$P = J^r \Delta \mu_r = RT J_r \log J_+^r / J_-^r = J_r \Delta G_r. \tag{15}$$

This makes sense, since in our derivation ΔG_r is exactly the change in the free energy when the reactants are converted into a product a single time and J_r is how many times this happens per unit time.

Just as for a very complicated electric circuit, we can just use this formula in very complicated reaction networks. Most of the times we will be concerned with cyclic processes in which case we would like to sum it over the links of a loop. We did this in class and you will revisit this for your HW problems.

Examples

Let us now apply these ideas to some simple systems.

Kinetic Proof Reading

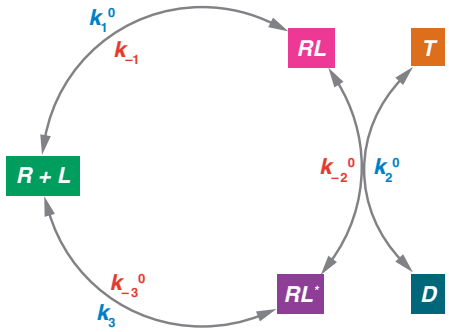

Figure 1: Kinetic Proofreading as a futile cycle from Qian Annu. Rev. Phys. Chemistry 2007 58:113-42.

Figure 8

A three-state kinetic model of receptor-ligand binding coupled with a hydrolysis reaction. The biochemical literature often refers to RL^* as an activated complex with crucial biological activity. Rate constants with positive and negative subscripts are for clockwise and counterclockwise directions, respectively. k_1^o, k_2^o, k_{-2}^o, and k_{-3}^o are second-order rate constants. $k_1 = k_1^o[L]$, $k_2 = k_2^o[T]$, $k_{-2} = k_{-2}^o[D]$, and $k_{-3} = k_{-3}^o[L]$ are pseudo-first-order rate constants. Because the concentrations of T and D are not at equilibrium in living cells, $k_1k_2k_3/(k_{-1}k_{-2}k_{-3}) = \gamma > 1$.

We previously discussed kinetic proofreading and showed that the error function is set by the equilibrium free energy difference ΔG_{eq} between the correct and incorrect substrate. In particular, the error fraction at equilibrium is set by

$$f = e^{-k_B T \Delta G_{eq}}. \tag{16}$$

The KPR scheme gets around this bound by consuming energy. However, by consuming energy we can of course increase the specificity.

Consider the two-step KPR cascade shown in Fig. 1. If we denote the rate for the correct substrate by primes (e.g. k_1') and incorrect substrates without primes (e.g. k_1). The error fraction at equilibrium is

$$f_0 = \frac{k_{-1}'}{k_{-1}}. \tag{17}$$

Now, we can characterize how out of equilibrium the process is by

$$\gamma = \frac{k_1k_2k_3}{k_{-1}k_{-2}k_{-3}} \tag{18}$$

since the nonequilibrium Free energy per cycle is

$$\Delta G_{noneq} = RT \log \gamma. \tag{19}$$

One can show through tedious calculation that in the usual Hopfield limit ($k_{-1} \ll k_2$, $k_1 \gg k_{-3}$, $k_1k_2 \gg fk_{-1}k_{-3}$, $k_3 \gg k_{-2}$) that

$$f(\gamma) \geq f_{min}(\gamma) = f_0 \left(\frac{1 + \sqrt{\gamma f_0}}{\sqrt{\gamma} + \sqrt{f_0}} \right)^2. \tag{20}$$

Notice that at equilibrium, $\gamma = 1$ so that

$$f_{min}(\gamma = 1) = f_0 \tag{21}$$

However, really far out of equilibrium when $\gamma \to \infty$ we get

$$f_{min}(\gamma = 1) = f_0^2. \tag{22}$$

Push-Pull amplifier and Gain/Bandwidth

We can also now consider a simple push-pull biochemical amplifier This is a simple biochemical circuit that many eukaryotic cells use to amplify signals. Below, I copy discussion from a relatively recent review we authored.[4]

[4] This section is copied from our review Landauer in the age of synthetic biology.

Biochemical networks can also consume energy to amplify upstream input signals. Signal amplification is extremely important in many eukaryotic pathways designed to detect small changes in input such as the phototransduction pathway in the retina [5] or the T cell receptor signaling pathway in immunology. In these pathways, a small change in the steady-state number of input messenger molecules, dI, leads to a large change in the steady-state number of output molecules, dO. The ratio of these changes is the number gain, often just called the gain,

[5] Detwiler, Peter B., et al. "Engineering aspects of enzymatic signal transduction: photoreceptors in the retina." Biophysical Journal 79.6 (2000): 2801-2817.

$$g_0 = \frac{dO}{dI} \tag{23}$$

with $g_0 > 1$ implying the ratio of output to input molecules is necessarily greater than 1.

Before proceeding further, it is worth making the distinction between the number gain, which clearly measures changes in absolute number, with another commonly employed quantity used to describe biochemical pathways called logarithmic sensitivity [6]. The logarithmic sensitivity, $\frac{d\log[O]}{d\log[I]}$, measures the logarithmic change in the concentration of an output signal as a function of the logarithmic change in the input concentration and is a measure of the fractional or relative gain. Though logarithmic sensitivity and gain are often used interchangeably in the systems biology literature, the two measures are very different. To see this, consider a simple signaling element where a ligand, L binds to a protein X and changes its conformation to X^*. The input in this case is L and the output is X^*. To have $g_0 > 1$, a small change in the number of ligands, dL must produce a large change in the number of activated X^*. Notice that by definition, in equilibrium, $\frac{dX^*}{dL} < 1$ since each ligand can bind only one receptor. If instead n ligands bind cooperatively to each X, then one would have $\frac{dX^*}{dL} < 1/n$. Thus, cooperatively in fact reduces the number gain. In contrast, the logarithmic sensitivity increases dramatically, $\frac{d\log[X]}{d\log[L]} = n$. An important consequence of this is that amplification

[6]

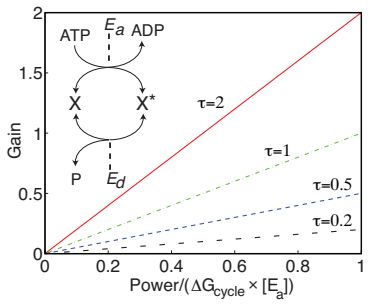

Figure 2: **Amplifying signals in a push-pull amplifier by consuming energy.** Schematic illustrates a simple push-pull amplifier where a kinase, E_a, modifies a protein from X to X^* and a phosphatase, E_d, catalyzing the reverse reaction. The plot illustrates that larger gain can be accomplished at the expense of a slower response time τ.

of input signals (as measured by number gain) necessarily requires a non-equilibrium mechanism that consumes energy.

The fact that energy consumption should be naturally related to the number gain and not logarithmic gain can be seen using both biological and physical arguments. The fundamental unit of energy is an ATP molecule. Since energy consumption is just a function of total number of ATP molecules hydrolyzed, it is natural to measure gain using changes in the absolute numbers and not concentrations. From the viewpoints of physics, this is simply the statement that energy is an extensive quantity and hence depends on the actual number of molecules.

In biochemical networks, this signal amplification is accomplished through enzymatic cascades, where the input signal couples to an enzyme that can catalytically modify (e.g. phosphorylate) a substrate. Such basic enzymatic “push-pull" amplifiers are the basic building block of many eukaryotic biochemical pathways, and are a canonical example of how energy consumption can be used to amplify input signals (see Figure 2). A push-pull amplifier consists of an activating enzyme E_a and a deactivating enzyme E_d that interconvert a substrate between two forms, X and X^*. Importantly, the post-translational modification of X is coupled to a futile cycle such as ATP hydrolysis. The basic equations governing a push-pull amplifier are

$$\frac{dX^*}{dt} = \Gamma_a(E_a)X - \Gamma_d(E_d)X^*, \tag{24}$$

where $\Gamma_a(E_a)$ is the rate at which enzyme E_a converts X to X^* and $\Gamma_d(E_d)$ is the rate at which enzyme E_d converts X^* back to X. This rate equation must be supplemented by the conservation equation on the total number of X molecules,

$$X + X^* = X_{\text{tot}}. \tag{25}$$

In the linear-response regime where the enzymes work far from saturation, one can approximate the rates in (24) as $\Gamma_a(E_a) \approx k_a[E_a]$ and $\Gamma_d(E_d) \approx k_d[E_d]$, with $k_a = k_a^{\text{cat}}/K_a$ and $k_d = k_d^{\text{cat}}/K_d$ the ratios of

the catalytic activity, k_{cat}, to the Michaelis-Menten constant, K_M, for the two enzymes. It is straightforward to show that the steady-state concentration of activated proteins is

$$\bar{X}^* = \frac{X_{\text{tot}} k_a [E_a]}{k_a [E_a] + k_d [E_d]} \tag{26}$$

Furthermore, one can define a "response time", τ, for the enzymatic amplifier to be the rate at which a small perturbation from steady-state $\delta X^* = X^* - \bar{X}^*$ decays. This yields (see Dewtiler for details)

$$\tau = (k_a [E_a] + k_d [E_d])^{-1}. \tag{27}$$

As discussed above, a key element of this enzymatic amplifier is that it works out of equilibrium. Each activation/deactivation event where the substrate cycles between the states $X \mapsto X^* \mapsto X$ is coupled to a futile cycle (e.g. ATP hydrolysis) and hence dissipates an energy ΔG_{cycle}. At steady-state, the power consumption of the enzymatic amplifier is

$$P = k_a [E_a] \bar{X} \Delta G_{\text{cycle}} = k_d [E_d] \bar{X}^* \Delta G_{\text{cycle}}. \tag{28}$$

The input of the enzymatic amplifier is the number of activating enzymes E_a and the output of the amplifier is the steady-state number of active substrate molecules X^*. This is natural in many eukaryotic signaling pathways where E_a is often a receptor that becomes enzymatically active upon binding an external ligand. Using (28), one can calculate the static gain and find

$$g_0 = (P/[E_a]) \tau (\Delta G_{\text{cycle}})^{-1}. \tag{29}$$

This expression shows that the gain of an enzymatic cascade is directly proportional to the power consumed per enzyme measured in the natural units of power that characterize the amplifier: $\Delta G_{\text{cycle}}/\tau$. This is shown in Figure 2 where we plot the gain as a function of power consumption for different response times.

Notice that the gain can be increased in two ways, by either increasing the power consumption or increasing the response time. Thus, at a fixed power consumption, increasing gain comes at the cost of a slower response. This is an example of a general engineering principle that is likely to be important for many applications in synthetic biology: the *gain-bandwidth tradeoff*. In general, a gain in signal comes at the expense of a reduced range of response frequencies (bandwidth). If one assumes that there is a maximum response frequency (ie a minimal time required for a response, a natural assumption in any practical engineering system), the gain-bandwidth tradeoff is equivalent to tradeoff between gain and response time.

For this reason, energy consumption is likely to be an important consideration for synthetic circuits such as biosensors that must respond quickly to small changes in an external input. More generally, the gain-bandwidth tradeoff highlights the general tension between signal amplification, energy consumption, and signaling dynamics.

A simple molecular machine

Homework

Problem 1 In the the main text, we considered a simple reaction of the form

$$A + D \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B + E \tag{30}$$

where one molecule of A and D are consumed to produce one molecule of B and E. Now consider a more general reaction of the form

$$n_A A + n_D D \underset{k_-}{\overset{k_+}{\rightleftharpoons}} n_B B + n_E E \tag{31}$$

where n_A molecules of A and n_D molecules of D are consumed to produce n_B molecules of B and n_E molecules of E. Derive the equivalent of Eq. 7 and Eq. 12 for this case. Can you write down an expression for the free energy change of a general chemical reaction?

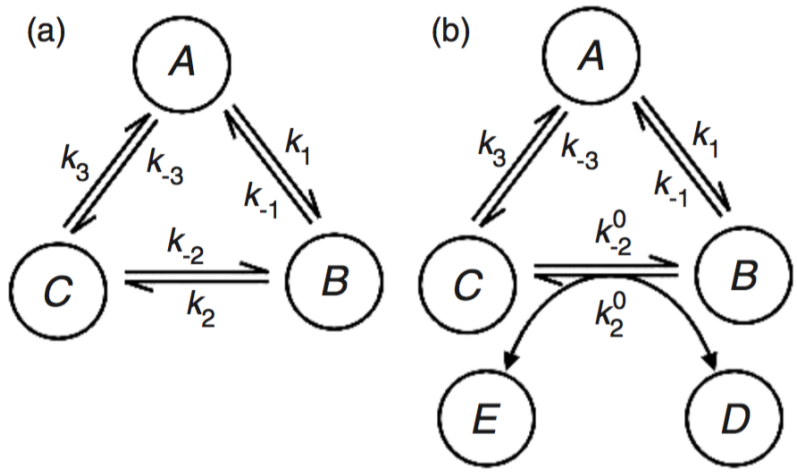

Figure 3: Figure for HW problem 2.

Problem 2 Thermodynamics of a simple cycle. Here we will work through some thermodynamics for a simple cycle. Consider the two cycles in Fig. 3. Let us think about these from a thermodynamic perspective. In the cycle in (b), we know that the full rates $k_{\pm 2}$ can be written as the product of a bare rate $k^0_{\pm 2}$ times a concentration (e.g.

$k_2 = k_2^0[D]$, $k_{-2} = k_{-2}^0[E]$). We can define for both cycles the quantity

$$\gamma \equiv \frac{k_1 k_2 k_3}{k_{-1} k_{-2} k_{-3}} \tag{32}$$

- Show that in the closed cycle in (a), detailed balance implies that $\gamma = 1$.
- We can associate a Gibbs free energy for each reaction

$$\begin{aligned} \Delta G_{AB} &= k_B T \log \frac{k_{-1}[B]}{k_1[A]} \\ \Delta G_{BC} &= k_B T \log \frac{k_{-2}[C]}{k_2[B]} \\ \Delta G_{CA} &= k_B T \log \frac{k_{-3}[A]}{k_3[C]}. \end{aligned}$$

  Show that detailed balance is equivalent to the statement $\Delta G_{AB} + \Delta G_{BC} + \Delta G_{CA} = 0$
- Show that in the cycle in (b) that we have that

$$\Delta G_{AB} + \Delta G_{BC} + \Delta G_{CA} = -k_B T \log \gamma = \Delta G_{DE}. \tag{33}$$

  where $\Delta G_{DE} = k_B T \log \frac{[E]}{[E]_{eq}} - k_B T \log \frac{[D]}{[D]_{eq}}$
- Show that with non-zero flux $\gamma \neq 1$, the current reaches a steady-state with non-zero flux equal to

$$J^{ss} = \frac{k_1 k_2 k_3 - k_{-1} k_{-2} k_{-3}}{k_2 k_3 + k_{-1} k_{-2} + k_{-1} k_3 + k_3 k_1 + k_{-2} k_{-3} + k_{-2} k_1 + k_1 k_2 + k_{-3} k_{-1} + k_{-3} k_2}$$

- Set $k_{\pm 1} = k_{\pm 2}^0 = k_{\pm 3} = 1$. Plot the steady-state current, the thermodynamic driving force (i.e. $-k_B T \log \gamma$), as a function of the ratio $\theta = [E]/[D]$.

Problem 3 We are now going to think about this in terms of energy conservation. We can define the total Gibbs Free energy of the open system. We can associate a Gibbs free energy G_α with each state $\alpha = A, B, C$. Notice that by definition we have that (*check signs*)

$$G_C - G_A = \Delta G_{CA} \tag{34}$$

$$G_B - G_A = \Delta G_{BA} \tag{35}$$

$$G_C - G_B = \Delta G_{BC} - G_{DE}. \tag{36}$$

The last of this equalities tells us that the difference in Gibbs free energy of state C and B also depends on the chemical potentials of D and E.

We can define a Gibbs Free energy associated with the protein by

$$G = p_A G_A + p_B G_B + p_c G_C \tag{37}$$

where p_α is the probability of being in state $\alpha = A, B, C$.

These probabilities can be written in terms of the fluxes as

$$\begin{aligned} \frac{dp_A}{dt} &= -J_1 + J_3 \\ \frac{dp_B}{dt} &= -J_2 + J_1 \\ \frac{dp_C}{dt} &= -J_3 + J_2, \end{aligned}$$

where $J_1 = k_1 p_A - k_{-1} p_B$, $J_2 = k_2 p_B - k_{-2} p_c$, and $J_3 = k_3 p_C - k_{-3} p_A$.

- Show that

$$\frac{dG}{dt} = (J_1 \Delta G_{AB} + J_2 \Delta G_{BC} + J_3 \Delta G_{CA}) - J_2 \Delta G_{DE}. \tag{38}$$

  The term of the left is the *rate of increase of energy*. The term in the parenthesis is the *energy dissipation rate* (i.e. $P = IV$). The last term is the *power of energy input*.

- Show that in a nonequibrium steady-state $J^{ss} = J_1 = J_2 = J_3$ and we recover Eq. 33 from problem 1.

- Set $k_{\pm 1} = k^0_{\pm 2} = k_{\pm 3} = 1$. Plot the entropy production rate as a function of θ. Argue that these must always be positive.