Notes

Stochastic Thermodynamics

Sabarno Saha

Email: sabarnohere@gmail.com

Date: 28 January 2026

A word

This note serves the purpose of familiarizing me with stochastic thermodynamics and its related concepts. This note has been divided into several sections.

Contents

Chapter 1: Stochastic Processes

1. Langevin Equations

A langevin equation is a type of stochastic differential equation that describes the evolution of a system under the influence of both deterministic and random forces. **TODO 1: Need to include projection methods to find langevin equations from microscopic mechanics. See Sekimoto**

An example of a langevin equation is the inertial motion of a particle in a fluid, where the particle experiences both viscous drag and random collisions with fluid molecules. This is called the **underdamped langevin equation**.

Definition 1.1 (Underdamped Langevin Equation): The underdamped langevin equation for a particle of mass m moving in a potential $U(x)$, subject to viscous drag with coefficient γ and random thermal forces $\hat{\xi}(t)$, is given by:

$$\frac{\mathrm{d}\hat{p}}{\mathrm{d}t} = -\frac{\gamma}{m}\hat{p}(x,t) + F(\hat{x},t) + \sqrt{\frac{2\gamma k_B T}{m}}\hat{\xi}(t)$$
$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = \frac{\hat{p}(x,t)}{m} \tag{1.1}$$

where $\hat{\xi}(t)$ is a Gaussian white noise term with zero mean and correlation function:

$$\left\langle \hat{\xi}(t)(t) \right\rangle = 0 \quad ; \quad \left\langle \hat{\xi}(t)(t')\hat{\xi}(t') \right\rangle = \delta(t - t') \tag{1.2}$$

The diffusion coefficient $D = k_B T/\gamma$ is related to the temperature T and the drag coefficient γ via the fluctuation-dissipation theorem. It also can be seen, that D must take this form to ensure that the system reaches thermal equilibrium, Boltzmann distribution, at long times. One can, in theory , solve the above equations to find the time evolution of the position and velocity of the particle, by finding the moments of the probability distribution function $p(x,\hat{p},t)$. A much nicer and neater way to do this is to use the Fokker-Planck equation, which we will see later.

In, cases where the inertial effects are negligible compared to the viscous drag, especially for small particles in a fluid, we can simplify the underdamped langevin equation by neglecting the acceleration term, leading to the overdamped langevin equation.

Definition 1.2 (Overdamped Langevin Equation): In many cases, the inertial effects are negligible compared to the viscous drag, especially for small particles in a fluid. In such cases, we can simplify the underdamped langevin equation by neglecting the inertial term, leading to the overdamped langevin equation:

$$\gamma\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} = F(\hat{x},t) + \sqrt{2\gamma k_B T}\hat{\xi}(t) \tag{1.3}$$

---

2. Fokker-Planck Equation

The Fokker-Planck equation is a partial differential equation that describes the time evolution of the probability density function of a stochastic process.

2.1. Single Variable Fokker-Planck Equation

The Fokker-Planck equation for a single variable x is given by:

$$\frac{\partial p(x;t)}{\partial t} = \mathcal{L}p(x;t) = -\frac{\partial x}{\partial t}A(x)p(x;t) + \frac{1}{2}\frac{\partial^2 x}{\partial x^2}B(x)p(x;t) \tag{2.1}$$

Chapter 2: Thermodynamics

3. Stochastic Entropy

4. Fluctuation Dissipation Relation

In the regime of linear response theoery, we consider systems, which are driven slightly out of equilibrium by a time-dependent manipulation. We assume that the observables deviate from their equilibrium values with linear dependence on the manipulation.

TODO 2: Need to read the derivation of the Fluctuation Dissipation Relation from the generalized Langevin equation.

Following $< PP >$, this derivation is done in the context of Stochastic thermodynamics. Suppose that the energy of a state x can be manipulated in r different ways. We denote the manipulation parameters by $\lambda = (\lambda_\alpha), \alpha = 1, ..., r$. Again, under the assumption of linear response, the energies deviate linearly from their unperturbed values,

$$\epsilon_x(\lambda) = \epsilon_x - \sum_{\alpha=1}^{r} \lambda_\alpha X_{\alpha,x} \tag{4.1}$$

5. Coarse graining

We need coarse graining to create seperations between the slow timescales of mesoscopic dynamics and the fast timescales of the heat reservoirs. This helps in achieving a simple closed description without worrying about the fast micrscoppic dynamics.

Generally there are quite a few techniques used to eliminate fast dynamics and achieve an effective description. They are,

- **Coarse Graining**: These technique group states together in such a way that the internal dynamics among states in a group are much faster than the dynamics among different groups.
- **Decimation Procedures**: These seek to eliminate states that have fast dynamics.

5.1. Adiabatic Elimination

Let's understand via an example. Consider three states $1, 2$ and f. "f" is a high energy state that stands for fast and is an intermediate step from 1 to 2. Since the f states is a high energy state, the system is quite unstable in that state. As a result, the dwelling time on the f state is very short. We perform **adiabatic elimination**. We are in such a time-scale of measurement that the dwells on "f" seem very short. So we could, somehow, create an effective theory by eliminating the "f" state and modifying the jump rates.

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = k_{1f}p_f - k_{f1}p_1$$

$$\frac{\mathrm{d}p_f}{\mathrm{d}t} = k_{f1}p_1 + k_{f2}p_2 - (k_{1f} + k_{2f})p_f \tag{5.1}$$

$$\frac{\mathrm{d}p_2}{\mathrm{d}t} = k_{2f}p_f - k_{f2}p_2$$

Since the dynamics of the f state is much faster compared to the slower dynamics of the other states, it always stays close to a steady state. Basically over the time dt required to see resolvable change in the states $1, 2$, p_f would already tend to a stationary-esque state. Thus, we would get,

$$\frac{\mathrm{d}p_f}{\mathrm{d}t} = 0$$

$$p_f = \frac{k_{f1}p_1 + k_{f2}p_2}{k_{1f} + k_{2f}}$$

(5.2)

Plugging these in the master equation results in,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} -k_{21}^e & k_{12}^e \\ k_{21}^e & -k_{12}^e \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

(5.3)

The effective jump rates in the matrix are given by

$$k_{21}^e = \frac{k_{2f}k_{f1}}{k_{1f} + k_{2f}}$$

$$k_{12}^e = \frac{k_{1f}k_{f2}}{k_{1f} + k_{2f}}$$

(5.4)

What does this mean physically? The first of the effective rates can be seem as jumps of the form,

$$1 \longrightarrow 2 = \frac{(1 \longrightarrow f)(f \longrightarrow 2)}{(f \longrightarrow 1) + (f \longrightarrow 2)}$$

(5.5)

The system goes from $1 \longrightarrow 2$, by traversing the intermediate step $1 \longrightarrow f \longrightarrow 2$. This results in the form of our numerator in the effective jump rates. However, the rates need to be scaled down. Due to microscopic reversibility, the system might come back to state 1, $f \longrightarrow 1$. On the other hand, if the system jumps to state 2, it might not have come from 1, it might have just gone from $2 \longrightarrow f$ back to 2, $f \longrightarrow 2$. This same idea applies to the second of the jump rates. This is an example of a Decimation procedure. Suppose that before coarse graining, the system satisfied detailed balance with distributions $p_1^{\mathrm{ss}}, p_f^{\mathrm{ss}}, p_2^{\mathrm{ss}}$.

$$k_{f1}p_1^{\mathrm{ss}} = k_{1f}p_f^{\mathrm{ss}}$$

$$k_{f2}p_2^{\mathrm{ss}} = k_{2f}p_f^{\mathrm{ss}}$$

(5.6)

Using these, now note that,

$$\frac{k_{1f}k_{2f}}{k_{1f} + k_{2f}}p_f^{\mathrm{ss}} = \frac{k_{1f}k_{2f}}{k_{1f} + k_{2f}}p_f^{\mathrm{ss}}$$

$$\implies \frac{k_{1f}}{k_{1f} + k_{2f}}k_{f2}p_2^{\mathrm{ss}} = \frac{k_{2f}}{k_{1f} + k_{2f}}k_{f1}p_1^{\mathrm{ss}}$$

$$\implies k_{21}^e p_1^{\mathrm{ss}} = k_{12}^e p_2^{\mathrm{ss}}$$

(5.7)

Thus the system satisfies detailed balance even after the decimation procedure.

5.2. Free energy over different coarse-graining levels

The coarse graining also creates an important distinction in the boltzmann probabilities. The energies of a coarse-grained mesostate have a free energy structure. Suppose the mesostate x is composed of microstates $\xi \in x$. Then the boltzmann weight of x is given by the sum of boltzmann weights of the inner structures.

$$\exp(-\beta\epsilon_x) \sim \sum_{\xi \in x} \exp(-\beta\epsilon_\xi)$$

$$\implies \epsilon_x = -k_B T \ln\left(\sum_{\xi \in x} \exp(-\beta\epsilon_\xi)\right)$$

(5.8)

Thus the coarse graining causes the energies to assume a free energy like structure. The nice thing about this structure is that it remains consistent over different coarse graining levels. Suppose a coarse grained mesostate x comprised of mesoscates $y \in x$ which are composed of microstates $\xi \in y$. Then,

$$
\begin{aligned}
\epsilon_x &= -k_B T \ln\left(\sum_{y \in x} \exp(-\beta \epsilon_y) \right) \\
\epsilon_y &= -k_B T \ln\left(\sum_{\xi \in y} \exp(-\beta \epsilon_\xi) \right)
\end{aligned}
\tag{5.9}
$$

It is then obvious to see, if we did not have an intermediate coarse grained layer $y \in x$, the energies of x would still be the same. We could group the microstates into the same group as before,

$$
\begin{aligned}
\epsilon_x &= -k_B T \ln\left(\sum_{\xi \in x} \exp(-\beta \epsilon_\xi) \right) = -k_B T \ln\left(\sum_{y \in x} \sum_{\xi \in y} \exp(-\beta \epsilon_\xi) \right) \\
&= -k_B T \ln\left(\sum_{y \in x} \exp(-\beta \epsilon_y) \right)
\end{aligned}
\tag{5.10}
$$

The last equality follows from our definitions of coarse grained energy. In doing this consistency check, we have not assumed how the states are assigned to different groups.

5.3. Work done over different mesoscopic scales

This coarse graining also effects the calculation of the work done. We compare the definitions of work done in the two coarse graining levels x and y. Consider a trajectory $\boldsymbol{x}$ with N jumps. Consider manipulation with the parameter $\lambda = \lambda(t)$.**Q1: Need to think about why temps do not change with λ.**

$$
w^{(y)} = \sum_{k=0}^{N} \int_{t_k}^{t_{k+1}} \frac{d\lambda}{dt} \frac{\partial \epsilon_{y_k}}{\partial \lambda}
\tag{5.11}
$$

$$
\begin{aligned}
w^{(x)} &= \sum_{k=0}^{N} \int_{t_k}^{t_{k+1}} \frac{d\lambda}{dt} \frac{\partial \epsilon_{x_k}}{\partial \lambda} = \sum_{k=0}^{N} \int_{t_k}^{t_{k+1}} \frac{d\lambda}{dt} \frac{\partial}{\partial \lambda}\left(-k_B T \ln\left(\sum_{y \in x_k} \exp(-\beta \epsilon_y) \right) \right) \\
&= \sum_{k=0}^{N} \int_{t_k}^{t_{k+1}} \frac{d\lambda}{dt} \frac{\partial}{\partial \lambda}\left(-k_B T \ln\left(\sum_{y \in x_k} \exp(-\beta \epsilon_y) \right) \right) \\
&= \sum_{k=0}^{N} \int_{t_k}^{t_{k+1}} \frac{d\lambda}{dt} \left\langle \frac{\partial \epsilon_{y_k}}{\partial \lambda} \right\rangle_{x,\text{eq}}
\end{aligned}
\tag{5.12}
$$

The two definitions are equivalent if the states $y \in x$ are at equilibrium with each other. This generally happens when the dynamics inside the mesostate x is quite fast and the states $y \in x$ quickly settle to equilibrium. **Q2: Do we require equilbrium or is steady state fine?** This is a coarse graining procedure, as opposed to adiabatic elimination which is a decimation procedure. We can see that the descriptions of work are equivalent provided we seperate the timescales of the dynamics.
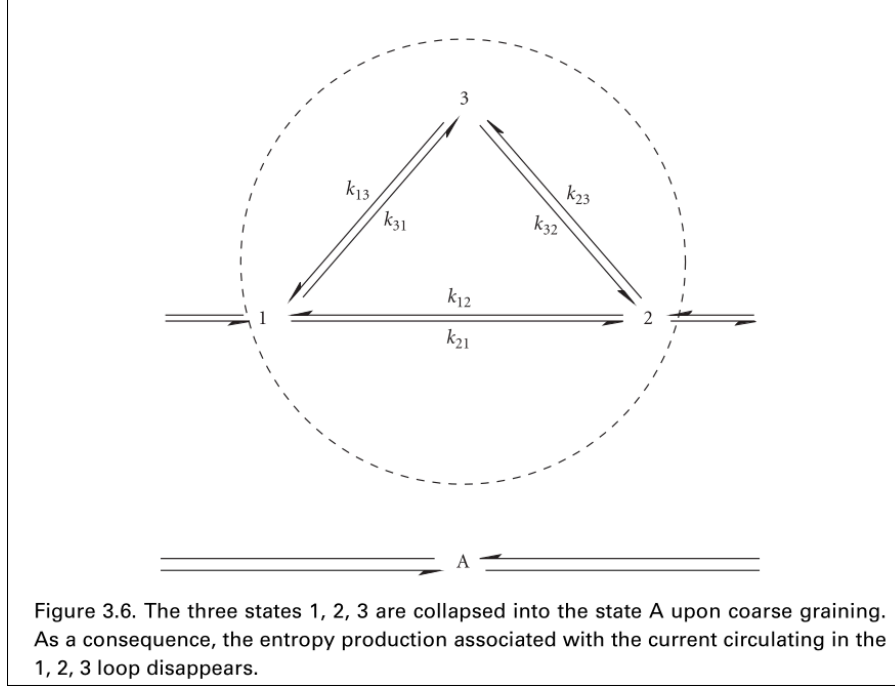
5.4. Coarse grained Entropy production

The Schnackenberg formula gives us the form of entropy production in discrete systems,

$$
\dot{S}_{\text{tot}} = \frac{k_B}{2} \sum_{x,x'} J_{xx'} \ln\left(\frac{k_{xx'} p_x^{\text{ss}}}{k_{x'x} p_{x'}^{\text{ss}}} \right)
\tag{5.13}
$$

However, this is a lower bound for the actual entropy production, as we shall see now. To see why this happens, we can break our coarse-grained systems into two cases,

1. The states y in a coarse grained group x are in equilibrium with themselves and the reservoir. When this happens, there is no entropy production within the mesostate x.
2. If the states y are out of equilibrium with each other, coarse-graining described the microscopic pathways between the states y, reducing the entropy production rate.

Figure 3.6. The three states 1, 2, 3 are collapsed into the state A upon coarse graining. As a consequence, the entropy production associated with the current circulating in the 1, 2, 3 loop disappears.

Then the jumps say between x and x' can happen via multiple pathways, r in number, with the effective rate $k_{xx'} = \sum_r k_{xx'}^{(r)}$. The actual total entropy production is $\dot{S}_{\text{tot,act}}$,

$$\dot{S}_{\text{tot,act}} = \frac{k_B}{2} \sum_{xx'r} J_{xx'}^{(r)} \ln\left(\frac{k_{xx'}^{(r)} p_{x'}^{\text{ss}}}{k_{x'x}^{(r)} p_x^{\text{ss}}}\right) \tag{5.14}$$

where $J_{xx'}^{(r)}$ is,

$$J_{xx'}^{(r)} = k_{xx'}^{(r)} p_{x'}^{\text{ss}} - k_{x'x}^{(r)} p_x^{\text{ss}} \quad \forall x \neq x, \forall r \tag{5.15}$$

The effecitve current $J_{xx'}$ is,

$$J_{xx'} = \sum_r J_{xx'}^{(r)} \tag{5.16}$$

The effective entropy production or the Schnackenberg formula, with the effective current and effective rates. The actual entropy production can be related to the schnakenberg formula,

$$\dot{S}_{\text{tot,act}} = \frac{k_B}{2} \sum_{xx'r} J_{xx'}^{(r)} \ln\left(\frac{k_{xx'}^{(r)} p_{x'}^{\text{ss}}}{k_{x'x}^{(r)} p_x^{\text{ss}}}\right)$$

$$= k_B \sum_{xx'} \sum (r) k_{xx'}^{(r)} p_{x'}^{\text{ss}} \ln\left(\frac{k_{xx'}^{(r)} p_{x'}^{\text{ss}}}{k_{x'x}^{(r)} p_x^{\text{ss}}}\right) \tag{5.17}$$

$$\geq k_B \sum (xx') k_{xx'} p_{x'}^{\text{ss}} \ln\left(\frac{k_{xx'} p_{x'}^{\text{ss}}}{k_{x'x} p_x^{\text{ss}}}\right) = \dot{S}_{\text{tot}}$$

To obtain to the inequality of entropy productions we have used the logsum inequality.

6. Entropy production in continuous systems

TODO 3: Need to talk about continuous systems and stochastic integral conventions.

Let's see the entropy production in continuous systems. The overdamped langevin equation is given by,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu_p \mathcal{F}(\hat{x}, t) + \sqrt{2D}\hat{\xi}(t) \tag{6.1}$$

where $\mathcal{F}(\hat{x}, t)$ is some generalized force, which can be broken into two parts, A term $-\partial u(\hat{x}, \lambda)/\partial x$ representing the force originating from a conservative force field $u(\hat{x}, \lambda)$ and $f(\hat{x}, t)$ a non conservative force, which represents non-equilibrium driving. In common literature, one often encounters this with the *friction coeeficient* γ. Note $\mu_p = \gamma^{-1}$, where μ_p is the *mobility* of the particle. The corresponding Fokker planck is given by,

$$\frac{\partial p(x; t)}{\partial t} = \frac{\partial}{\partial x}\left[-\mu_p \frac{\partial \mathcal{F}(\hat{x}, t)}{\partial x} p(x; t) + D\frac{\partial p(x; t)}{\partial x}\right] \tag{6.2}$$

This can be written in the form of a continuity equation,

$$\frac{\partial p(x; t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x} \tag{6.3}$$

where, the probability current is given by,

$$J(x, t) = \mu_p \mathcal{F}(\hat{x}, t)p(x; t) - D\frac{\partial p(x; t)}{\partial x} \tag{6.4}$$

In the absence of driving $f(\hat{x}, t)$ and λ is held constant, we obtain the equilibrium state $J = 0$,

$$p^{\mathrm{eq}}(x) = e^{F(\lambda)}e^{-\beta u(x, \lambda)} \tag{6.5}$$

where the Free energy $F(\lambda)$ is given by the normalization condition,

$$F(\lambda) = -k_B T \ln\left(\int \mathrm{d}x \exp(-\beta u(x, \lambda))\right) \tag{6.6}$$

The **canonical distribution** is attained provided the fluctuation dissipation relation is given as,

$$D = k_B T \mu_p \tag{6.7}$$

Q3: With driving and suitable boundary conditions, the solution is a NESS, where work done by driving equals heat to reservoir. Read further

The energy change can be seen as

$$\mathrm{d}u = \frac{\partial u(\hat{x}, \lambda)}{\partial x} \circ \mathrm{d}\hat{x} + \frac{\partial u(\hat{x}, \lambda)}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t}\,\mathrm{d}t \tag{6.8}$$

TODO 4: Why the stratonovich convention. Read KS and NS. The $\circ$ signifies that the term should be interpreted in the stratonovich sense.

Recall the discrete definition of work **TODO 5: insert relevant equation.**. In continuous systems, the infinitesimal work done is given by,

$$\mathrm{d}w = \frac{\partial u(\hat{x}, \lambda)}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t}\,\mathrm{d}t + f(\hat{x}, t) \circ \mathrm{d}\hat{x} \tag{6.9}$$

Combining Equation (6.8) and Equation (6.9), we define the heat given to the reservoir,

$$\mathrm{d}q = \mathrm{d}w - \mathrm{d}u = \left[-\frac{\partial u(\hat{x}, \lambda)}{\partial x} + f(\hat{x}, t)\right] \circ \mathrm{d}\hat{x} = \mathcal{F}(\hat{x}, t) \circ \mathrm{d}\hat{x} \tag{6.10}$$

The system entropy is given using the Shannon-Entropy formula,

$$s^{\mathrm{sys}}(\hat{x}, t) = -k_B \ln p(x; t) \tag{6.11}$$

where $p(x; t)$ is the probability that the random variable $\hat{x}$ takes the value x at time t. One can regard this probability distribution as a function of the random variable itself. **Q4: I think one requires strato convention in order for these to be equivalent.** We also use the fact that the stratonovich

convention preserves the chain rule. Using Equation (6.4) and Equation (6.7), the rate of increase of system entropy is given by,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} s^{\mathrm{sys}}(\hat{x}(t), t) &= -k_B \ln p(x(t), t) \\
&= -\frac{k_B}{p(x;t)} \left[\frac{\partial p(x;t)}{\partial x} \circ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} + \frac{\partial p(x;t)}{\partial t} \right] \\
&= k_B \left[\frac{J(x,t)}{Dp(x;t)} \circ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} - \frac{\mathcal{F}(\hat{x},t)}{k_B T} \circ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} - \frac{1}{p(x;t)} \frac{\partial p(x;t)}{\partial t} \right]
\end{aligned}
\tag{6.12}
$$

Since the reservoir, has fast relaxation times and is held at the same temperature, the reservoir entropy is $s^{\mathrm{res}} = \frac{q}{T}$. The total stochastic entropy production,

$$
\begin{aligned}
\dot{s}^{\mathrm{tot}} &= \frac{\mathrm{d}s^{\mathrm{sys}}}{\mathrm{d}t} + s^{\dot{\mathrm{res}}} \\
&= k_B \left[\frac{J(x,t)}{Dp(x;t)} \circ \frac{\mathrm{d}\hat{x}}{\mathrm{d}t} - \frac{1}{p(x;t)} \frac{\partial p(x;t)}{\partial t} \right]
\end{aligned}
\tag{6.13}
$$

A small note on notation here. Strictly speaking $\langle f \rangle$ are used when the function f is a random variable. However, here sometimes, we need to average out functions, which in the truest sense are not random variables.

The biggest culprit is the common notation of the Shannon entropy $S[p(x)]$ as the "average" of stochastic entropy $\sigma(p(x)) = \ln(p(x))$.

$$
S[p(x)] = -\sum_x p_x \ln p_x = -\sum_x p_x \sigma(x) = \langle \sigma_x \rangle
\tag{6.14}
$$

Many books use the last definition. However, in the strictest sense, that is an abuse of notation. We have just introduced a measure for addition or integration, which just so happens to be similar to the integral forms of conventional expectations. So whenever, the integral or additive measure is $p(x;t)$, but we are not averaging over random variables, we introduce a new notation,

$$
\langle\!\langle f(x,t) \rangle\!\rangle = \int \mathrm{d}x\, p(x;t) f(x,t)
\tag{6.15}
$$

Using this, we can calculate the average entropy production rate,

$$
\dot{S}^{\mathrm{tot}} = \int \mathrm{d}x\, \dot{s}^{\mathrm{tot}}(\boldsymbol{x}) = \langle \dot{s}^{\mathrm{tot}} \rangle
\tag{6.16}
$$

Note that the last term vanishes due to conservation of probability

$$
\int \mathrm{d}x\, \frac{1}{p(x;t)} \frac{\partial p(x;t)}{\partial t} p(x;t) = \frac{\partial}{\partial t} \int \mathrm{d}x\, p(x;t) = 0
\tag{6.17}
$$

The other terms can be easily found in the Ito convention, so we shall switch to that,

$$
\begin{aligned}
\langle s^{\mathrm{tot}} \rangle &= k_B \left\langle\!\!\left\langle \frac{J(x,t)}{Dp(x;t)} \circ \mathrm{d}\hat{x} \right\rangle\!\!\right\rangle \\
&= k_B \left\langle\!\!\left\langle \frac{J(x,t)\mu_p \mathcal{F}(\hat{x},t)}{Dp(x;t)} + \sqrt{\frac{2}{D}} \frac{J(x,t)}{p(x;t)} \circ \mathrm{d}W \right\rangle\!\!\right\rangle \\
&= k_B \left\langle\!\!\left\langle \frac{J(x,t)\mu_p \mathcal{F}(\hat{x},t)}{Dp(x;t)} + \underbrace{\sqrt{\frac{2}{D}} \frac{J(x,t)}{p(x;t)} \cdot \mathrm{d}W}_{2} + \underbrace{\frac{1}{p(x;t)} \frac{\partial J(x,t)}{\partial x} \mathrm{d}t - \frac{J(x,t)}{(p(x;t))^2} \frac{\partial p(x;t)}{\partial x} \mathrm{d}t}_{3} \right\rangle\!\!\right\rangle
\end{aligned}
\tag{6.18}
$$

The nonanticipation of the Ito integral gives

$$
\exp\left(\int f(\hat{x}) \cdot \mathrm{d}W \right) = \int \exp(f(\hat{x})) \cdot \exp(\mathrm{d}W) = 0
\tag{6.19}
$$

Q5: Different conventions for nonanticipation of ito in KS, NS, and PP. Why? Thus, on an average over different time intervals term **2** equals 0. Also term **3** can be converted to a surface term, which vanishes under proper boundary conditions. We are left with,

$$
\begin{aligned}
\dot{S}^{\text{tot}} &= k_B \left\langle\!\!\left\langle \frac{J(x,t)\mu_p \mathcal{F}(\hat{x},t)}{Dp(x;t)} - \frac{J(x,t)}{(p(x;t))^2} \frac{\partial p(x;t)}{\partial x} \, \mathrm{d}t \right\rangle\!\!\right\rangle \\
&= k_B \left\langle\!\!\left\langle \frac{J(x,t)}{p(x;t)} \left(\frac{\mu_p \mathcal{F}(\hat{x},t)}{D} - \frac{1}{p(x;t)} \frac{\partial p(x;t)}{\partial x} \, \mathrm{d}t \right) \right\rangle\!\!\right\rangle \\
&= k_B \left\langle\!\!\left\langle \frac{J^2(x,t)}{Dp^2(x;t)} \right\rangle\!\!\right\rangle = k_B \int \mathrm{d}x \, \frac{J^2(x,t)}{Dp(x;t)}
\end{aligned}
\tag{6.20}
$$

Thus we have obtained a formula for the total entropy production rate for continuous systems.

Result 1 (Entropy Production Rate for Continuous Systems):
The total entropy production for a system in both continuous space and time is

$$
\dot{S}^{\text{tot}} = k_B \left\langle\!\!\left\langle \frac{J^2(x,t)}{Dp^2(x;t)} \right\rangle\!\!\right\rangle = k_B \int \mathrm{d}x \, \frac{J^2(x,t)}{Dp(x;t)}
\tag{6.21}
$$

References